DISCOVERY AND CHARACTERIZATION OF LONG NON-CODING RNAs IN PROSTATE CANCER

by

John R. Prensner

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular and Cellular Pathology)
in the University of Michigan
2012

Doctoral Committee:

Professor Arul M. Chinnaiyan, Chair
Professor David Beer
Professor Kathleen Cho
Professor David Ginsburg
Assistant Professor Yali Dou

ACKNOWLEDGEMENTS

This work represents the combined efforts of many talented individuals. First, my mentor, Arul Chinnaiyan, deserves much credit for guiding me through this experience and pushing me to produce the best science that I could. My thesis committee—Kathy Cho, David Beer, David Ginsburg, and Yali Dou—was an integral part of my development and I am truly grateful for their participation in my education.

I am also indebted to the many collaborators who have contributed to this work. Foremost, Matthew Iyer has been responsible for much of my progress and has been an invaluable colleague. Saravana M. Dhanasekaran mentored me and taught me much of what I know. Wei Chen has spent countless hours working on these projects. I have also been fortunate to collaborate with Felix Feng, Qi Cao, Alejandro Balbin, Sameek Roychowdhury, Brendan Veeneman, Lalit Patel, Anirban Sahu, Chad Brenner, Irfan Asangani, and Xuhong Cao who have helped to make this thesis possible.

I would like to extend my appreciation to my wonderful support network at MCTP and the MSTP office: Karen Giles, Christine Betts, Dianna Banka, Ron Koenig, Ellen Elkin, and Hilkka Ketola, who have been immensely supportive and helpful during my training. I would also like to thank my funding sources: the University of Michigan Cancer Biology Training Grant, the Department of Defense Predoctoral Grant, and the Department of Defense Idea Award supported me financially.

Finally, my life has depended on my friends and family. My wife, Sarah Barbrow, has been a source of enduring kindness that has lifted my spirits on innumerable occasions. I also would like to thank my parents for supporting my decision to remain a student in perpetuity.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

ABSTRACT

Discovery and characterization of long non-coding RNAs in prostate cancer

by

John R. Prensner

Chair: Arul M. Chinnaiyan

The comprehensive delineation of cancer-causing genes is an essential step in understanding the molecular basis of cancer. While much research has interrogated the role of protein-coding genes in cancer, recent discoveries demonstrate that the human genome may additionally contain thousands of non-protein-coding genes, termed non-coding RNAs (ncRNAs). However, the identity and function of these genes is largely unknown.

Here, I describe the systematic discovery and functional characterization of long non-coding RNAs (lncRNAs) in prostate cancer. We used transcriptome sequencing (RNA-Seq) of human prostate cancer samples to identify over 1,800 unannotated, intergenic lncRNAs, of which 121 *P*rostate *C*ancer *A*ssociated *T*ranscripts (PCATs) demonstrated aberrant expression patterns between benign prostate samples, localized cancers, and metastatic cancers.

To study novel lncRNAs in prostate cancer, we focused on two: *PCAT-1* and *SChLAP-1* (*S*econd *C*hromosome *L*ocus *A*ssociated with *P*rostate-*1*, also referred to as *PCAT-114)*, which are overexpressed in subsets of prostate cancer. We found that upregulation of *PCAT-1* mediates increased cellular proliferation *in vitro* and *in vivo* through the regulation of genes involved in DNA maintenance, including *BRCA2*, a tumor suppressor gene essential for DNA break repair by homologous recombination (HR). Mechanistically, *PCAT-1* expression repressed *BRCA2* in a microRNA-like manner via the *BRCA2* 3' untranslated region. *BRCA2* repression resulted in defective HR in *PCAT-1*-expressing prostate cells, leading to increased cell sensitivity to PARP1 inhibitors, which engender synthetic lethality in the context of impaired HR.

By contrast, our investigation of *SChLAP-1* revealed a nuclear lncRNA that is involved in prostate cell invasiveness and metastasis *in vitro* and *in vivo*. Mechanistically, *SChLAP-1* antagonizes the SWI/SNF chromatin remodeling complex, a tumor suppressor complex inactivated in cancer, by directly binding SWI/SNF proteins and impairing their ability to regulate gene expression. Clinically, *SChLAP-1* expression defined a subset of prostate cancers associated with aggressive phenotypes and poor outcome.

Taken together, this thesis work represents the first comprehensive assessment of lncRNAs in a major cancer type and describes novel oncogenic lncRNAs in prostate cancer that may further serve as predictive (*PCAT-1*) or prognostic (*SChLAP-1*) biomarkers. Broadly, this work suggests that uncharacterized lncRNAs may play critical roles in the pathogenesis of other diseases.

CHAPTER 1

Introduction: Long non-coding RNAs in biology and disease

THE CENTRAL DOGMA OF MOLECULAR BIOLOGY

In the 50 years following the discovery of DNA in the 1950s, molecular biology—though ever more complex with each passing year—was grounded by a unifying principle: the central dogma. Part science, part philosophy, the central dogma was formulated by Francis Crick in 1958, only 5 years after he and James Watson reported the basic structure of DNA (*1, 2*). In this short time, Crick had catapulted his thinking beyond nucleic acids and toward proteins. What kind of relationship existed between the two? He posited his thesis thus:

> The Central Dogma. This states that once 'information' has passed into protein it cannot get out again. In more detail, the transfer of information from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible. Information means here the precise determination of sequence, either of bases in the nucleic acid or of amino-acid residues in the protein. (*1*)

Soon, Crick and others expanded this initial model. In addition to DNA and protein, RNA, the enzymatic synthesis of which earned Severo Ochoa the 1959 Nobel Prize in Medicine, was quickly added (*2, 3*). Originally, transfer of genetic information was possible from DNA to RNA, RNA to protein, RNA to DNA, and DNA to protein,

although Crick believed that transfer of genetic information from RNA to DNA or DNA directly to protein was unlikely (*1, 2, 4*). Thus was born the central dogma of biology. Another generation of researchers would codify these concepts in a simple dictum: DNA is transcribed into RNA, and RNA is translated proteins (**Figure 1.1**).

A genome full of genes?

According to the central dogma, the regions of DNA that ultimately give rise to RNA and proteins are the complement of genes in an organism. A "gene" is therefore defined somewhat empirically: if a certain part of DNA is essential for the generation of RNA and protein, it is then part of a gene. Thus, the question of which regions of the human genome constitute its genes has long been a central topic in biology. While early cloning-based methods revealed more than 7000 human genes in the 1970s and 1980s (*5*), large-scale analyses of expressed sequence tags (ESTs) in the 1990s suggested that the estimated number of human genes lay in the 35,000–100,000 range (*6*). The completion of the Human Genome Project narrowed the focus considerably by highlighting the surprisingly small number of protein-coding genes, which is now conventionally cited as less than 25,000.

While the number of protein-coding genes (20,000–25,000) has maintained broad consensus, recent studies of the human transcriptome have revealed an astounding number of non-coding RNAs (ncRNAs). These transcribed elements, which lack the capacity to code for a protein, are bafflingly abundant in all organisms studied to date, from yeast to humans (*7-9*), and they represent a break with the central dogma of biology because they do not engage in protein synthesis. Yet, over the past decade, numerous

studies have demonstrated that ncRNAs have distinct biological functions and operate through defined mechanisms. Still, their sheer abundance—some reports estimate that up to 70% of the human genome is transcribed into RNA (*7, 10, 11*)—has sparked debates as to whether ncRNA transcription reflects true biology or byproducts of a leaky transcriptional system. Encompassed within these studies are the broad questions of what constitutes a human gene, what distinguishes a gene from a region that is simply transcribed, and how we interpret the biological meaning of transcription.

These developments have been matched by equally insightful discoveries analyzing the role of ncRNAs in human diseases, especially cancer, lending support to the importance of their cellular functions (*12, 13*). Initial evidence suggests that ncRNAs, particularly long ncRNAs (lncRNAs), have essential roles in tumorigenesis (*12*), and that lncRNA-mediated biology occupies a central place in cancer progression (*14*). With the number of well-characterized cancer-associated lncRNAs growing, the study of lncRNAs in cancer is now generating new hypotheses about the biology of cancer cells. Here, I review the current understanding of ncRNAs in cancer, with particular focus on lncRNAs as novel drivers of tumorigenesis.

NON-CODING RNAs: A NEW KIND OF GENE

ncRNAs are RNA transcripts that do not encode for a protein. In the past decade, a great diversity of ncRNAs have been observed. Depending on the type of ncRNA, transcription can occur by any of the three RNA polymerases (RNA Pol I, RNA Pol II, or RNA Pol III). General conventions divide ncRNAs into two main categories: small ncRNAs less than 200 bp and long ncRNAs greater than 200 bps (*15*). Within these two

categories, there are also many individual classes of ncRNAs (**Table 1.1**), although the degree of biological and experimental support for each class ranges substantially and should be evaluated individually.

Small ncRNAs

The diversity of small ncRNAs has perhaps grown the most, where several dozen classes of small ncRNAs have been proposed (*15, 16*). These include well-characterized housekeeping ncRNAs (transfer RNA (tRNA) and some ribosomal RNA (rRNA)) essential for fundamental aspects of cell biology, splicing RNAs (small nuclear RNAs (snRNAs)), and a variety of recently-observed RNAs associated with protein-coding gene transcription, such as tiny transcription-initiation RNAs, promoter-associated short RNAs, termini-associated short RNAs, 3'UTR-derived RNAs, and antisense termini-associated short RNAs (*15*).

To date, the most extensively studied small RNAs in cancer are microRNAs (miRNAs). Elegant studies over the past 15 years have defined an intricate mechanistic basis for miRNA-mediated silencing of target gene expression through the RNA-induced silencing complex (RISC), which employs Argonaute family proteins (such as *AGO2*) to cleave target mRNA transcripts or to inhibit the translation of that mRNA (**Figure 1.2A**) (*17*). Aberrant expression patterns of miRNAs in cancer has been well documented in most tumor types (**Figure 1.2B**), and detailed work from many labs have shown that numerous miRNAs, including miR-10b, let-7, miR-101, and the miR-15a-16-1 cluster, possess oncogenic or tumor suppressive functions (**Figure 1.2C**) (*17*).

Long ncRNAs

Recent observations of novel long ncRNA species has led to a complex set of terms and terminologies used to describe a given long ncRNA. These include antisense RNAs, which are transcribed on the opposite strand from a protein-coding gene and frequently overlap that gene (*18*), transcribed ultraconserved regions (T-UCRs), which originate in regions of the genome showing remarkable conservation across species, and ncRNAs derived from intronic transcription.

Although many RNA species are >200 bp in length, such as repeat or pseudogene-derived transcripts (*19*), the abbreviated term lncRNA (also referred to as lincRNA, for long intergenic ncRNA) does not uniformly apply to all of these. While the nomenclature is still evolving, lncRNA typically refers to a polyadenylated long ncRNA that is transcribed by RNA polymerase II and associated with epigenetic signatures common to protein-coding genes, such as trimethylation of histone 3 lysine 4 (H3K4me3) at the transcriptional start site (TSS) and trimethylation of histone 3 lysine 36 (H3K36me3) throughout the gene body (*20*). This description similarly suits many T-UCRs and some antisense RNAs, and the overlap between these categories may be substantial. lncRNAs also commonly exhibit splicing of multiple exons into a mature transcript, as do many antisense RNAs but not RNAs transcribed from gene enhancers (eRNAs) or T-UCRs (*21-23*). Transcription of lncRNAs occurs from an independent gene promoter and is not coupled to the transcription of a nearby or associated parental gene, as with some classes of ncRNAs (promoter/termini-associated RNAs, intronic ncRNAs) (*15*). In this thesis, I will use the term lncRNA in this manner. When the data

is supportive, I include specific T-UCRs and antisense RNAs under the lncRNA umbrella term, and I distinguish other long ncRNAs, such as eRNAs, where appropriate.

Identification of long ncRNAs

Many initial lncRNAs, such as *XIST* and *H19,* were discovered in the 1980s and 1990s by searching cDNA libraries for clones of interest (*24, 25*). In these studies, the intention was generally to identify new genes important in a particular biological process—X chromosome inactivation in the example of *XIST*—by studying their expression patterns. At the time, most genes uncovered were protein-coding, and this tended to be the assumption, with a handful exceptions, such as *XIST*, which were subsequently determined to be noncoding as a secondary observation (*24*).

In the past decade, however, large-scale analyses have focused on identifying ncRNA species in a comprehensive fashion. This paradigm shift was mediated by dramatic advances in high-throughput technologies, including DNA tiling arrays and next generation RNA sequencing (RNA-Seq) (*10, 11, 14, 26, 27*). These platforms provide systems with which RNA transcription can be observed in an unbiased manner, and have thereby highlighted the pervasive transcription of ncRNAs in cell biology. Moreover, whereas conventional cDNA microarrays detected only the transcripts represented by probes on the array, the introduction and popularization of RNA-Seq as a standard tool in transcriptome studies has removed many barriers to detecting all forms of RNA transcripts (*14, 28*). RNA-Seq studies now suggest that several thousand uncharacterized lncRNAs are present in any given cell type (*14, 20*). Observations that many lncRNAs demonstrate tissue-specific expression therefore enables speculations that the human

genome may harbor nearly as many lncRNAs as protein-coding genes (perhaps ~15,000 lncRNAs), though only a fraction are expressed in a given cell type.

LONG NON-CODING RNAs IN CANCER

Emerging evidence suggests that lncRNAs constitute an important component of tumor biology (**Table 1.2**). Dysregulated expression of lncRNAs in cancer marks the spectrum of disease progression (*14*) and may serve as an independent predictor for patient outcomes (*29*). Mechanistically, most well-characterized lncRNAs to date show a functional role in gene expression regulation, typically transcriptional rather than post-transcriptional regulation. This can occur by targeting either genomically local (*cis*-regulation) or genomically distant (*trans*-regulation) genes. Recently, a new type of long ncRNAs at gene enhancers, termed eRNAs, have also been implicated in transcriptional regulation (*30*).

cis-Regulatory lncRNAs

cis-regulation by lncRNAs contributes to local control of gene expression by recruiting histone modification complexes to specific areas of the genome (**Figure 1.3**). This effect can either be highly specific to a particular gene, such as the regulation of *IGF2* by lncRNAs (*31*); or, it can encompass a wide chromosomal region, such as X-chromosome inactivation in women through *XIST*. Historically, *cis*-regulation through lncRNAs was studied earlier than *trans*-regulation, as several *cis*-regulatory lncRNAs, including *H19*, *AIR*, *KCNQ1OT1*, and *XIST* were earlier discoveries (*24, 25, 32*). Several

cis-regulatory lncRNAs, including *H19*, *AIR* and *KCNQ1OT1*, are also functionally related through their involvement in epigenetic imprinting regions.

Imprinting lncRNAs

lncRNA involvement in imprinted regions of the genome is critical for maintaining parent-of-origin-specific gene expression. In particular, an imprinted region of human chromosome 11 (orthologous to mouse chromosome 7) has been extensively studied for the role of lncRNAs. In humans, most well-known are the *H19* and *KCNQ1OT1* lncRNAs (*25*, *32*), which are expressed on the maternal and paternal alleles, respectively, and maintain silencing of the *IGF2* and *KCNQ1* genes on those alleles (**Figure 1.3A**) (*33*).

Of the imprinting-associated ncRNAs, *H19* has been most extensively studied in cancer. Aberrant expression of *H19* is observed in numerous solid tumors, including hepatocellular and bladder cancer (*31, 34*). The functional data on *H19* point in several directions, and it has been linked to both oncogenic and tumor suppressive qualities (*35*). For example, there is evidence for its direct activation by cMYC (*36*) as well as its downregulation by p53 and during prolonged cell proliferation (*37*). In model systems, siRNA-mediated knockdown of *H19* expression impaired cell growth and clonogenicity in lung cancer cell lines *in vitro* (*36*) and decreased xenograft tumor growth of Hep3B hepatocellular carcinoma cells *in vivo* (*31*). Together, these data support a general role for *H19* in cancer, although its precise biological contributions are still unclear.

Other imprinting-associated lncRNAs are only tangentially associated with cancer. While loss of imprinting is observed in many tumors, the role for lncRNAs in

this process is not well defined. For example, Beckwith-Wiedemann syndrome (BWS), a disorder of abnormal development with an increased risk for cancer, displays aberrant imprinting patterns of *KCNQ1OT1* (*33, 38*); but a direct association or causal role for *KCNQ1OT1* in cancer is not described (*38*). Conversely, aberrant *H19* methylation in BWS appears to predispose to cancer development more strongly (*38*).

XIST

XIST, perhaps the most well studied lncRNA, is transcribed from the inactivated X chromosome, in order to facilitate that chromosome's inactivation, and manifests as multiple isoforms (*39, 40*). On the active X allele, *XIST* is repressed by its antisense partner ncRNA, *TSIX* (*40*). *XIST* contains a double-hairpin RNA motif in the RepA domain, located in the first exon, which is crucial for its ability to bind Polycomb Repressive Complex 2 (PRC2) and propagate epigenetic silencing of an individual X chromosome (**Figure 1.3B**) (*41*).

Despite the body of research on *XIST*, a precise role for *XIST* in cancer has remained elusive (*42*). Some evidence initially suggested a role for *XIST* in hereditary *BRCA1*-deficient breast cancers (*43, 44*), where data indicated that *BRCA1* was not required for *XIST* function in these cells (*45*). Others have reasoned that *XIST* may be implicated in the X chromosome abnormalities observed in some breast cancers. There have also been surprising accounts of aberrant *XIST* regulation in other cancers, including lymphoma and male testicular germ-cell tumors, where *XIST* hypomethylation is, unexpectedly, a biomarker (*46*). Yet, it remains unclear whether these observations

reflect a passenger or driver status for *XIST*, as a well-defined function for *XIST* in cancer has yet to attain a broad consensus.

ANRIL

Located on Chr9p21 in the INK4A/ARF tumor suppressor locus, *ANRIL* was initially described by examining the deletion of this region in hereditary neural system tumors, which predispose for hereditary cutaneous malignant melanoma (*47*). *ANRIL* was subsequently defined as a polyadenylated lncRNA antisense to the *CDKN2A* and *CDKN2B* genes. *In vitro* data have suggested that *ANRIL* functions to repress the INK4A/INK4B isoforms (*48*), but not ARF. This repression is mediated through direct binding to CBX7 (*48*), a member of Polycomb Repressive Complex 1 (PRC1), and SUZ12 (*49*), a member of PRC2, which apply repressive histone modifications to the locus. However, these studies were performed in different cell types and it is not known whether *ANRIL* binds both complexes simultaneously.

ANRIL also displays a highly complicated splicing pattern, with numerous variants, including circular RNA isoforms (*50*). Currently it is unclear whether these isoforms have tissue-specific expression patterns or unique functions, which may suggest a biological basis for this variation. Through genome wide association studies (GWAS), *ANRIL* has also been identified by single nucleotide polymorphisms (SNPs) correlated with a higher risk of atherosclerosis and coronary artery disease (*51*), and *ANRIL* expression has been noted in many tissues. The function and isoform-level expression of *ANRIL* in these tissue types is not yet elucidated, but may shed light onto its role in diverse disease processes.

HOTTIP and *HOTAIRM1*

An intriguing theme emerging in developmental biology is the regulation of HOX gene expression by lncRNAs. Highly conserved among metazoan species, HOX genes are responsible for determining tissue patterning and early development, and in humans HOX genes reside in four genomic clusters. Within these clusters, HOX genes display intriguing anterior-posterior and proximal-distal expression patterns that mirror their genomic position 5' to 3' in the gene cluster.

Two recently-discovered lncRNAs, termed *HOTTIP* and *HOTAIRM1*, may help to explain this co-linear patterning of HOX gene expression. *HOTTIP* and *HOTAIRM1* are located at opposite ends of the HoxA cluster, and each helps to enhance gene expression of the neighboring HoxA genes (*52, 53*). *HOTAIRM1*, located at the 3' end, coordinates *HOXA1* expression and has tissue-specific expression patterns identical to *HOXA1* (*52*). *HOTTIP*, by contrast, is at the 5' end of the cluster and similarly enhances expression of the 5' HoxA genes, most prominently *HOXA13* (*53*). Mechanistic studies of *HOTTIP* suggest that it binds WDR5 and recruits the MLL H3K4 histone methyltransferase complex to the HoxA cluster to support active chromatin confirmation (*53*). These observations distinguish *HOTTIP* and *HOTAIRM1*, as most lncRNAs to date facilitate gene repression.

While *HOTAIRM1* and *HOTTIP* have not been extensively studied in cancer, expression of these may have important roles in the differentiation status of cancer cells. For example, differentiation of myeloid cancer cell lines, such as K562 and NB4, by treatment with retinoic acid led to an increase in *HOTAIRM1* expression, implicating it in

myeloid differentiation (*52*). Moreover, HoxA genes are broadly known to be important for many cancers, particularly *HOXA9*, which is essential for oncogenesis in leukemias harboring MLL rearrangements. Thus, *HOTAIRM1* and *HOTTIP* also suggest a potential role for lncRNAs in MLL-rearranged leukemias.

trans-Regulatory lncRNAs

Like most *cis*-acting lncRNAs, *trans*-acting lncRNAs typically facilitate epigenetic regulation of gene expression. However, because *trans*-acting lncRNAs may operate at geographically distant locations of the genome, it is generally thought that the mature lncRNA transcript is the primary actor in these cases, as opposed to *cis*-regulating lncRNAs, like *H19*, *AIR* and *KCNQ1OT1,* which may function through the act of transcription itself (*35, 54, 55*).

HOTAIR

trans-Regulatory lncRNAs were brought to widespread attention by the characterization of *HOTAIR*. First described in fibroblasts, *HOTAIR* is located in the HoxC cluster; but unlike *HOTTIP* and *HOTAIRM1*, *HOTAIR* was found to regulate HoxD cluster genes in a *trans*-regulatory mechanism (**Figure 1.3C**) (*56*). These observations raise the question of whether all Hox clusters are regulated by lncRNAs, either by a *cis*-regulatory or a *trans*-regulatory mechanism.

In cancer, *HOTAIR* is upregulated in breast and hepatocellular carcinomas (*15*), and in breast cancer overexpression of *HOTAIR* is an independent predictor of overall survival and progression-free survival (*29*). Work by Howard Chang and colleagues has

further defined a compelling mechanistic basis for *HOTAIR* in cancer. *HOTAIR* has two main functional domains, a PRC2-binding domain located at the 5' end of the RNA, and a LSD1/CoREST1-binding domain located at the 3' end of the RNA (*56, 57*). In this way, *HOTAIR* is thought to operate as a tether that links two repressive protein complexes in order to coordinate their functions. In breast cancer, *HOTAIR* overexpression facilitates aberrant PRC2 function by increasing PRC2 recruitment to the genomic positions of target genes. By doing so, *HOTAIR* mediates the epigenetic repression of PRC2 target genes, and profiling of repressive (H3K27me3) and active (H3K4me3) chromatin marks shows widespread changes in chromatin structure following *HOTAIR* knockdown (*29*).

Furthermore, HOTAIR dysregulation results in a phenotype in both *in vitro* and *in vivo* models. Ectopic overexpression of *HOTAIR* in breast cancer cell lines increases their invasiveness both *in vitro* and *in vivo*. Supporting this, in benign immortalized breast cells overexpressing a core component of PRC2 (*EZH2)*, knockdown of *HOTAIR* mitigated *EZH2*-induced invasion *in vitro* (*29*). Taken together, these data provide the most thorough picture of a lncRNA in cancer.

GAS5

GAS5, first identified in murine NIH-3T3 cells, is a mature, spliced lncRNA manifesting as multiple isoforms up to 12 exons in size (*58*). Using HeLa cells engineered to express *GAS5*, Kino et al. recently described an intriguing mechanism by which *GAS5* modulates cell survival and metabolism by antagonizing the glucocorticoid receptor (GR). The 3' end of *GAS5* both interacts with the GR DNA-binding domain

(DBD) and is sufficient to repress GR-induced genes, such as *cIAP2*, when cells are stimulated with dexamethasone. By binding to the GR, *GAS5* serves as a decoy that prevents GR binding to target DNA sequences (**Figure 1.3D**) (*58*).

In cancer, *GAS5* induces apoptosis and suppresses cell proliferation when overexpressed in breast cancer cell lines, and in human breast tumors *GAS5* expression is downregulated (*59*). Although it is unclear whether this phenotype is due to an interaction with GR, it is intriguing that *GAS5* may also be able to suppress signaling by other hormone receptors, such as androgen receptor (AR), though this effect was not seen with estrogen receptor (ER) (*58*).

Other long ncRNAs

eRNAs

eRNAs are transcribed by RNA polymerase II at active gene enhancers (*21*). But unlike lncRNAs, they are not polyadenylated and are marked by a H3K4me1 histone signature denoting enhancer regions (*21*), rather than the H3K4me3/H3K36me3 signature classically associated with lncRNAs. While research on eRNAs is still in the earliest phases, an emerging role for them in hormone signaling is already being explored. Nuclear hormone receptors, such as AR and ER, are critical regulators of numerous cell growth pathways and are important in large subsets of prostate (AR), breast (ER), and thyroid (PPARγ) cancers. To date, eRNAs have been most directly implicated in prostate cancer, where they assist in AR-driven signaling and are maintained by *FOXA1*, a transcription factor that mediates cell lineage gene expression in several cell types (*30*).

T-UCRs

Ultraconserved regions in the genome were initially described as stretches of sequence >200 bp long with 100% conservation between humans and mice but harboring no known gene (*60*). As high levels of sequence conservation are hallmarks of exonic sequences in protein-coding genes, ultraconserved regions strongly suggest the presence of either a gene or a regulatory region, such as an enhancer. Subsequently, numerous ultraconserved sequences were found to be transcriptionally active, defining a class of T-UCRs as ncRNAs (*22*). Many transcripts from T-UCRs are polyadenylated and associated with H3K4me3 at their transcriptional start sites (TSSs), indicating that many are likely lncRNAs according to our definition (*61*).

Aberrant expression of T-UCRs has been noted in several cancer types, including neuroblastoma (*61*), leukemia (*22*), and hepatocellular carcinoma (*23*). Most notably, one T-UCR gene, termed *TUC338*, has been shown to promote both cell proliferation and anchorage-independent growth in hepatocellular carcinoma cell lines (*23*), and *TUC338* transcript is localized to the nucleus, suggesting a role in regulation of expression (*23*). Calin et al. further demonstrated that T-UCRs are targets for miRNAs (*22*). While T-UCRs remain poorly characterized as a whole, additional exploration of the role and mechanism of these ncRNAs will likely elucidate novel aspects of tumor biology.

FUNCTIONS AND MECHANISMS OF LONG NON-CODING RNAs

Like protein-coding genes, there is considerable variability in the function of long ncRNAs. Yet, clear themes in the data suggest that many long ncRNAs contribute to

associated biological processes. These processes typically relate to transcriptional regulation or mRNA processing, which is reminiscent of miRNAs and may indicate a similar sequence-based mechanism akin to miRNA binding to seed sequences on target mRNAs. However, unlike miRNAs, long ncRNAs show a wide spectrum of biological contexts that demonstrate greater complexity to their functions.

Epigenetic transcriptional regulation

The most dominant function explored in lncRNA studies relates to epigenetic regulation of target genes. This typically results in transcriptional repression, and many lncRNAs were first characterized by their repressive functions, including *ANRIL*, *HOTAIR*, *H19*, *KCNQ1OT1*, and *XIST* (*15, 48, 56*). These lncRNAs achieve their repressive function by coupling with histone modifying or chromatin remodeling protein complexes.

The most common protein partners of lncRNAs are the PRC1 and PRC2 polycomb repressive complexes. These complexes transfer repressive post-translational modifications to specific amino acid positions on histone tail proteins, thereby facilitating chromatin compaction and heterochromatin formation in order to enact repression of gene transcription. PRC1 may be comprised of numerous proteins, including BMI1, RING1, RING2 and Chromobox (CBX) proteins, which act as a multi-protein complex to ubiquitinate histone H2A at lysine 119 (*62*). PRC2 is classically composed of EED, SUZ12, and EZH2, the latter of which is a histone methyltransferase enzymatic subunit that trimethylates histone 3 lysine 27 (*62*). Both EZH2 and BMI1 are upregulated in numerous common solid tumors, leading to tumor progression and aggressiveness (*62*).

Indeed, *ANRIL*, *HOTAIR*, *H19*, *KCNQ1OT1*, and *XIST* have all been linked to the PRC2 complex, and in all except *H19*, direct binding has been observed between PRC2 proteins and the ncRNA itself (*41, 49, 56, 63, 64*). Binding of lncRNAs to PRC2 proteins, however, is common and observed for ncRNAs which do not appear to function through a PRC2-mediated mechanism. It is estimated that nearly 20% of all lncRNAs may bind PRC2 (*65*), though the biological meaning of these observations remains unclear. It is possible that PRC2 promiscuously binds lncRNAs in a non-specific manner. However, if lncRNAs are functioning in a predominantly *cis*-regulatory mechanism—such as *ANRIL*, *KCNQ1OT1*, and *XIST*—then numerous lncRNAs may bind PRC2 to facilitate local gene expression control throughout the genome. Relatedly, studies of PRC2-ncRNA binding properties have been able to determine a putative PRC2-binding motif that includes a GC-rich double hairpin, indicating a structural basis for PRC2-ncRNA binding in many cases (*41*).

Similarly, PRC1 proteins, particularly CBX proteins, have been implicated in ncRNA-based biology. For example, *ANRIL* binds CBX7 in addition to PRC2 proteins, and this interaction with CBX7 recruits PRC1 to the *INK4A*/*ARF* locus to mediate transcriptional silencing (*48*). More broadly, work with mouse polycomb proteins demonstrated that treatment with RNAse abolished CBX7 binding to heterochromatin on a global level, supporting the notion that ncRNAs are critical for PRC1 genomic recruitment (*66*).

While PRC1 and PRC2 are perhaps the most notable partners of lncRNAs, numerous other epigenetic complexes are implicated in ncRNA-mediated gene regulation. For example, the 3' domain of *HOTAIR* contains a binding site for the

LSD1/CoREST, a histone deacetylase complex that facilitates gene repression by chromatin remodeling (**Figure 1.4A**) (*57*). *AIR* is similarly reported to interact with G9a, an H3K9 histone methyltransferase (*67*). *KCNQ1OT1* has been shown to interact with PRC2 (*64*), G9a (*64*), and DNMT1, which methylates CpG dinucleotides in the genome. More rarely, lncRNAs have been observed in activating epigenetic complexes. In a recent example, *HOTTIP* interacts with WDR5 to mediate recruitment of the MLL histone methyltransferase to the distal HoxA locus. MLL transfers methyl groups to H3K4me3, thereby generating open chromatin structures that promote gene transcription (*53*).

In some cases, the mere act of lncRNA transcription is critical for the recruitment of protein complexes. Studies for both *H19*, *KCNQ1OT1* and *AIR* suggest that transcriptional elongation of these genes is an important component of their function (*35, 54, 55*). By contrast, other lncRNAs, including *HOTTIP* as well as many *trans*-regulatory ones, do not show this relationship (*53*). For these lncRNAs, biological function may be centrally linked to their role as flexible scaffolds. In this model, lncRNAs serve as tethers that rope together multiple protein complexes through a loose arrangement. Supporting this model are the multiple lncRNAs found to bind multiple protein complexes, such as *ANRIL* (binding PRC1 and PRC2) and *HOTAIR* (binding PRC2 and LSD1/CoREST) (**Figure 1.4A**).

Enhancer-associated long ncRNAs

In addition to facilitating epigenetic changes that impact gene transcription, emerging evidence suggests that some ncRNAs contribute to gene regulation by

influencing the activity of gene enhancers. For example, *HOTTIP* is implicated in chromosomal looping of active enhancers to the distal HoxA locus (*53*), though knockdown and overexpression of *HOTTIP* is not sufficient to alter chromosomal confirmations (*53*). There is also a report of local enhancer-like ncRNAs that typically lack the H3K4me1 enhancer histone signature, but possess H3K4me3, and function to potentiate neighbor gene transcription in a manner independent of sequence orientation (*68*).

A major recent development has been the discovery of eRNAs, which are critical for the proper coordination of enhancer genomic loci with gene expression regulation. While the mechanism of their action is still unclear, in prostate cancer cells, induction of AR signaling increased eRNA synthesis at AR-regulated gene enhancers, suggesting that eRNAs facilitate active transcription upon induction of a signaling pathway (*30*). Using chromatin conformation assays, Wang et al. showed that eRNAs are also important for the establishment of enhancer-promoter genomic proximity by chromosomal looping. Moreover, eRNAs work in conjunction with cell lineage-specific transcription factors, such as *FOXA1* in prostate cells, thereby creating a highly specialized enhancer network to regulate transcription of genes in individual cell types (**Figure 1.4B**) (*30*). Future work in this area will likely provide insight into signaling mechanisms important in cancer.

Modulating tumor suppressor activity

The role of many lncRNAs as transcriptional repressors lends itself to inquiry as a mechanism for suppression of tumor suppressor genes. Here, one particular hotspot is

the chromosome 9p21 locus, harboring the tumor suppressor genes *CDKN2A* and *CDKN2B*, which give rise to multiple unique isoforms, such as p14, p15, and p16, and function as inhibitors of oncogenic cyclin dependent kinases. Expression of this region is impacted by several repressive ncRNAs, such as *ANRIL* (**Figure 1.4C, upper**), and the p15-antisense RNA, the latter of which also mediates heterochromatin formation through repressive histone modifications and was observed in leukemias (*48, 69*).

Moreover, several lncRNAs are implicated in the regulation of p53 tumor suppressor signaling. *MEG3*, a maternally-expressed imprinted lncRNA on Chr14q32, has been shown to activate p53 and facilitate p53 signaling, including enhancing p53 binding to target gene promoters (*70*). *MEG3* has also been linked to p53 signaling in meningioma (*71*), and *MEG3* overexpression suppresses cell proliferation in meningioma and hepatocellular carcinoma cell lines (*71, 72*). In human tumors, *MEG3* downregulation is widely noted, with frequent hypermethylation of its promoter observed in pituitary tumors (*15*) and leukemias (*73*). Taken together, these data implicate *MEG3* as a putative tumor suppressor.

A recently described murine lncRNA located near the p21 gene, termed *linc-p21*, has also emerged as a promising p53-pathway gene. In murine lung, sarcoma, and lymphoma tumors, *linc-p21* expression is induced upon activation of p53 signaling and represses p53 target genes through a physical interaction with hnRNP-K, a protein that binds the promoters of genes involved in p53 signaling (**Figure 1.4C, lower**) (*74*). *linc-p21* is further required for proper apoptotic induction (*74*). These data highlight *linc-p21* as a candidate tumor suppressor gene. However, due to sequence differences between

species, it is currently unclear whether the human homologue of *linc-p21* plays a similarly important role in human tumor development.

Regulation of mRNA processing and translation

While many lncRNAs operate by regulating gene transcription, post-transcriptional processing of mRNAs is also critical to gene expression. A primary actor in these processes is the nuclear paraspeckle, a sub-cellular compartment found in the interchromatin space within a nucleus and characterized by PSP1 protein granules (*75*). While nuclear paraspeckle functions are not fully elucidated, this structure is known to be involved in a variety of post-transcriptional activities, including splicing and RNA editing (*75*). Paraspeckles are postulated to serve as storage sites for mRNA prior to its export to the cytoplasm for translation, and one study discovered a paraspeckle-retained, polyadenylated nuclear ncRNA, termed *CTN-RNA*, that is a counterpart to the protein-coding murine *CAT2* (*mCAT2*) gene (*76*). *CTN-RNA* is longer than *mCAT2*, and under stress conditions, cleavage of *CTN-RNA* to the *mCAT2* coding transcript resulted in increased *mCAT2* protein (*76*).

In cancer, two ncRNAs involved in mRNA splicing and nuclear paraspeckle function, *MALAT1* and *NEAT1*, are overexpressed. *MALAT1* and *NEAT1* are genomic neighbors on Chr11q13, and both are thought to contribute to gene expression by regulating mRNA splicing, editing, and export (**Figure 1.4D**) (*77*, *78*). *MALAT1* may further serve as a precursor to a small, 61-base-pair ncRNAs that is generated by RNase P cleavage of the primary *MALAT1* transcript and exported into the cytoplasm (*79*).

Although a unique role for *MALAT1* in cancer is not yet known, its overexpression in lung cancer predicts for aggressive, metastatic disease (*80*).

Regulatory RNA-RNA interactions

Recent work on mechanisms of RNA regulation has highlighted a novel role for RNA-RNA interactions between ncRNAs and mRNA sequences. These interactions are conceptually akin to miRNA regulation of mRNAs, as sequence homology between the ncRNA and the mRNA is important to the regulatory process.

This sequence homology may be derived from ancestral repeat elements that contribute sequence to either the untranslated sequences of a protein-coding gene, or, less frequently, the coding region itself. For example, STAU1-mediated mRNA decay involves the binding STAU1, a RNA degradation protein, to protein-coding mRNAs that interact with lncRNAs containing ancestral Alu repeats. In this model, sequence repeats—typically Alus—in lncRNAs and mRNAs partially hybridize, forming double-stranded RNA complexes that then recruit STAU1 to implement RNA degradation (**Figure 1.4E**) (*81*). A related concept is found with *XIST*, which contains a conserved repeat, termed RepA, in its first exon. RepA is essential for *XIST* function and the RepA sequence is necessary to recruit PRC2 proteins for X-chromosome inactivation (*41*).

Another model for mRNA regulation was recently posited by Pandolfi and colleagues, who suggested that transcribed pseudogenes serve as decoy for miRNAs that target the protein-coding mRNA transcripts of their cognate genes (*82*). Sequestration of miRNAs by the pseudogene then regulates the gene expression level of the protein-coding mRNA indirectly (**Figure 1.4F**). Beyond just pseudogenes, this model more

broadly suggests that all long ncRNAs, as well as other protein-coding mRNAs, may function as molecular "sponges" that bind and sequester miRNAs in order to control gene expression indirectly. In their study, Pandolfi and colleagues demonstrate that pseudogenes of two cancer genes, *PTEN* and *KRAS*, may be biologically active, and that *PTENP1*, a pseudogene of *PTEN* that competes for miRNA binding sites with *PTEN*, itself functions as a tumor suppressor in *in vitro* assays and may be genomically lost in cancer (*82*). This intriguing hypothesis may shed new light onto the functions or ncRNAs, pseudogenes, and even the UTRs of a protein-coding gene.

METHODS TO DISCOVER LONG NON-CODING RNAs

ncRNAs have historically been more difficult to detect than protein-coding genes. This is largely because ncRNAs tend to be more tissue-specific in their expression (*83*), exhibit lower overall levels of expression (*26, 27, 84*), and cannot be predicted by computational algorithms scanning for open reading frames (ORFs) in the human genome. This last point is particularly germane, as the Human Genome Project employed hidden Markov Model-based methods to predict protein coding genes from DNA sequence, arriving at estimates between 20,000 – 25,000 (*85, 86*).

However, the advance of high-throughput RNA profiling methods has enabled more precise and accurate cataloguing of ncRNAs. While the Human Genome Project emphasized only protein-coding genes in their computational analysis of DNA, groups investigating RNA had long observed a great number of unannotated transcripts (*5, 6*). Many of these transcripts were discovered using cloning and sequencing of Expressed Sequence Tags (ESTs), and the detection of a spliced EST suggested the presence of two

exons being post-transcriptionally processed (*87*). Prior to the completion of the Human Genome Project, analyses of ESTs routinely generated estimates of 30,000 – 120,000 human genes (*6, 88*).

The first studies to systematically catalogue ncRNAs used DNA tiling arrays, which utilized probes designed to densely cover a specific genomic region. While these analyses were not genome-wide, they did cover several chromosomes in full (*10, 11*). The major drawback of this methodology is that there is no reliable way to exclude all of the degraded portions of gene introns, which can be detected as independent fragments of RNA but do not represent unique ncRNAs themselves. Thus, these methods predicted a genome that was nearly universally transcribed into RNA (*7, 10, 11*).

A major advance came when Guttman *et al.* combined the methodology of DNA tiling arrays with the logic of epigenetics (*20*). Here, the authors reasoned that ncRNAs could resemble protein-coding genes in their fundamental structure and epigenetic characteristics. Thus, ncRNAs could be polyadenylated and spliced, and, like protein-coding genes, they could have a gene promoter marked by H3K4me3 and a gene body marked by H3K36me3. By using ChIP-seq data of these epigenetic marks as well as RNA polymerase II, the authors observed several thousand regions of unannotated transcription by DNA tiling arrays in the same locations as these epigenetic marks (*20*). These were defined as lncRNAs based on their length and characteristics.

The introduction of RNA-seq subsequently enabled the direct detection of all RNA species in a cell. RNA-seq-based methods have been adapted to selectively measure small RNAs (*89*), long RNAs (*26, 27, 84*), RNA stability (*90*), the act of RNA transcription (*30, 91, 92*), and RNA species bound to protein complexes such as PRC2

(*93*). RNA-seq is now the gold standard method to discover lncRNAs, but a major challenge with these data is their interpretation. Sequence reads commonly harbor multi-mapping potential, especially for lncRNAs whose DNA sequence is overall less conserved and harbors a greater degree of repetitive elements. Thus, stringent filtration and rigorous analysis is required to eliminate spurious transcripts.

PROSTATE CANCER

This year, over one million prostate biopsies will be performed (*94*), over 200,000 new diagnoses of prostate cancer are expected (*95*), and over 30,000 deaths will result from prostate cancer—the second leading cause of death from cancer (after lung cancer) in American men (*95*). While autopsy-based analysis reveals prostate cancer as one of the most common male diseases—afflicting an estimated 65% of 70 year old men (*96*)—there is only a 1 in 6 lifetime risk of developing clinically-detected disease (*95*), with African Americans having a increased risk and Asians having a lower risk of disease compared to Caucasians (*97*). Furthermore, the incidence of lethal prostate cancer is approximately 1 in 35 (*98*, *99*). Given this gap in prevalence and clinical incidence, deciphering the molecular events that both cause and distinguish high-risk prostate cancer lesions from low risk ones continues to be an area of significant research.

In the normal adult male, the prostate gland is a walnut-sized exocrine gland that contributes clear, slightly alkaline fluid secretions to a male's semen as the sperm course down the reproductive tract. Grossly, it is divided into five lobes: the anterior, median, and posterior lobes, as well as two lateral lobes. Pathologically, however, the prostate is sectioned into four zones: the peripheral, central, transition, and anterior fibro-muscular

zones. Whereas benign prostatic hyperplasia (BPH), an enlargement of the prostate commonly seen in aging men, is generally a product of hyperplasia of the transition zone, prostate cancer typically arises in the peripheral zone (*100, 101*).

The molecular basis of prostate cancer

While the development and progression of prostate cancer is linked to many factors, including dietary and lifestyle changes, advancements in genomics in the past 20 years have placed a major emphasis on the causative role of somatic aberrations in tumorigenesis. Aided by RNA and DNA sequencing technologies, the ability to study the underlying genomic alterations that distinguish cancers from normal tissue has led to a widespread paradigm shift in the way in which both cancer tumorigenesis and cancer therapeutics are conceptualized. In 2000, Hanahan and Weinberg delineated the currently accepted mechanistic steps involved in cancer development: (1) self-sufficiency in growth signals; (2) insensitivity to anti-growth signals; (3) evading apoptosis; (4) sustained angiogenesis; (5) limitless replicative potential; (6) tissue invasion and metastasis (*102*). For prostate cancer, these criteria have been investigated as multiple somatic events that occur as normal or benign tissues transform into malignant disease.

Genetic aberrations in prostate cancer

Prostate cancer is characterized by a low mutational rate compared to other cancers, such as lung, melanoma and colon (*103*). As a result, the landscape of prostate cancer lacks the preponderance of recurrent mutations observed in other cancers. For example, mutations in canonical oncogenes such as *KRAS*, *IDH1/2*, *BRAF*, *EGFR*, and

PIK3CA are absent or relatively uncommon in prostate cancer (*103*). A recent study using exome sequencing to comprehensively profile all known coding exons found low levels of recurrent mutations in *BRAF*, *RET*, and multiple androgen receptor cofactor genes, including *FOXA1*, *MLL2*, *MLL*, and *ASHXL* (*104*). The ubiquitin ligase *SPOP*, a novel cancer gene discovered through whole-genome sequencing of prostate cancer (*105*), is also mutated in ~5% of all prostate cancers (~10% of ETS-negative cancers), though it is unclear how this gene is functioning in the disease. Several tumor suppressor genes, most prominently p53, are also mutated in prostate cancer (*104*).

Although prostate cancer is not dominated by mutations, gene rearrangements and copy-number changes characterize much of the genomic landscape of this disease (*104, 105*). Several well-characterized tumor suppressor genes are commonly deleted, including *PTEN* and *RB1* (**Figure 1.5**) *(106, 107)*. Deletion of *PTEN*, in particular, has been widely studied as a prognostic indicator of poor outcomes in approximately 1/3rd of patients (*108-111*). Loss of *PTEN* activates the PI3K pathway (*112*), which results in resistance to androgen-based therapies in transgenic mouse models of prostate cancer (*113*), suggesting that patient cancers with a *PTEN* deletion may also be predisposed to resistance to androgen deprivation therapy. Interestingly, a recent report suggests that deletion of *MAGI2*, a PTEN-interacting gene, may also occur a smaller subset of patients lacking *PTEN* deletion (*105*).

The NK3 homeobox 1, *NKX3.1*, encodes a prostate-specific transcription factor that is positively regulated by androgen signaling (*114-116*). A putative tumor suppressor gene, *NKX3.1* is located on chromosome 8p21.2, a region that exhibits loss of heterozygosity in 50% to 85% of prostate cancers (*115, 117-120*) (**Figure 1.5**).

Decreased levels of *NKX3.1* protein are also observed in ~75% of PINs, though only ~12% of PIN lesions harbor genomic loss of *NKX3.1*, indicating that loss of the protein generally precedes loss of the gene itself (*117*). While the exact functional role of *NKX3.1* loss in prostate cancer in unclear, *NKX3.1* appears both to regulate cell cycle and cell death pathways through stabilization of TP53 and to inhibit AKT1 activation (*121*). Further research has linked *NKX3.1* with protection against damage from reactive oxygen species (*122*).

Amplifications of several oncogenes also helps to drive prostate cancer pathogenesis. The *cMYC* oncogene is frequently amplified as part of an arm-level copy number gain of Chr8q, leading to its transcriptional upregulation in cancer (**Figure 1.5**) (*123*). In mouse models, *cMYC* contributes to cellular proliferation angiogenesis, and apoptosis (*124*), and amplification of chromosome 8 occurs commonly in PIN, carcinomas, and metastases, identifying it as an early lesion in prostate cancer (*125*). In addition, *PIK3CA*, though rarely mutated, resides in a common Chr3q amplicon (*104*).

Androgen receptor alterations in prostate cancer

Among the most well-studied genomic alterations in prostate cancer are those relating to the androgen receptor (*AR*) itself. Androgen signaling is a major component of prostate cancer development, and androgen ablation therapy has been a mainstay in cancer therapy since Huggins first described the method in the early 1940s (*126, 127*). Though androgen ablation therapy can be an effective means for tumor regression, it frequently fails to cure the disease, leading to an aggressively metastatic, often fatal, castrate-resistant prostate cancer (CRPC) instead. Whereas *AR* mutations are very rare in

androgen-naïve cancers, CRPC is characterized by a number of genetic alterations, with 30 - 60% of CRPC tumors harboring an amplified *AR* gene (*104, 128-131*) and perhaps 20 - 30% of metastatic CRPC tumors possessing mutations of *AR* (**Figure 1.5**) (*130-135*).

Many of these alterations result in hypersensitivity to androgen signaling. Increased *AR* expression (*128*), amplified co-activators, increased *AR* stability, and enhanced nuclear localization all contribute to abnormal *AR* sensitivity (*131, 136, 137*). Variants in the 5-alpha reductase gene, which converts testosterone into the more potent dihydrotestosterone (DHT), can also result in enhanced enzymatic activity and increased DHT levels (*138-140*). Supporting this, there is some evidence that, after androgen ablation therapy, DHT levels in the prostate tissue decrease by only 60% though serum testosterone drops by 95%, indicating that androgen levels may remain relatively high (*141*).

Another major mechanism in the alteration of androgen signaling is through increased *AR* promiscuity. Many mutations of the *AR*, most occurring within the ligand binding domain, have been correlated with the ability for the *AR* to bind nonclassical ligands (*130, 131, 136*). Androgen receptors harboring the T877A and L701H mutations, in particular, are able to bind and respond to corticosteroids, which ordinarily function as *AR* antagonists (*130, 133, 136, 142*). Overexpression of *AR* co-regulators, such as TIF2, ARA70, SRC1, and oncostatin M, an IL-6-related cytokine, have also been shown to contribute to *AR* promiscuity with nonclassical ligands (*136, 137, 143, 144*).

Gene fusions in prostate cancer

In 2005, our group described the presence of recurrent fusions between the ETS family transcription factors, *ERG* and *ETV1*, and the androgen-regulated transmembrane serine protease, *TMPRSS2* (*145*) (**Figure 1.5**). Subsequently, multiple other 5' fusion partners have been described for *ERG* and *ETV1*, as well as other members of the ETS family (**Table 1.3**) (*145, 146*). All told, ETS fusions are found in roughly 50% of prostate cancers, of which 90% are *TMPRSS2-ERG* fusions (*145, 147-149*). Because *ERG* and *TMPRSS2* are separated by only ~ 3 kb on the long arm of chromosome 21, two mechanisms are implicated in their creation: an intrachromosomal deletion within chromosome 21, and an inversion of that same region (*145, 150*). For gene fusion generation, it is thought that *AR* plays a causative role in bringing *TMPRSS2* and *ERG* into spatial proximity, which allows for faulty DNA repair following genotoxic stress (*151, 152*). While these mechanisms may create TMPRSS2-ETS gene fusions, *TMPRSS2* typically contributes only untranslated sequences to the final transcript, thereby functioning only to upregulate ETS transcription factor expression (*145, 153*).

Following this discovery, the *TMPRSS2-ERG* fusions have emerged as a major factor in prostate tumorigenesis, contributing to cellular invasiveness *in vitro* (*146-149*). Moreover, *TMPRSS2-ERG* fusions are detectable in pre-malignant prostate lesions such as high-grade prostatic intraepithelial neoplasia (HGPIN) if these lesions are proximal to, or continguous with, regions of cancer (*154-156*), suggesting that these fusions may be important in early tumorigenesis.

In addition, the detection of *TMPRSS2-ERG* RNA in patient urine became widely investigated. Since the presence of these transcripts is diagnostic for prostate cancer, a

positive *TMPRSS2-ERG* urine assay may indicate cancer even if prostate needle biopsy results failed to detect cancer. Because of this remarkable attribute, clinical *TMPRSS2-ERG* urine assays significantly improve upon serum PSA for the delineation of prostate cancer in PSA-prescreened cohorts. Our lab recently published a large study of more than 1300 men demonstrating that combined measurement of *PCA3*, a prostate cancer biomarker, and *TMPRSS2-ERG* in urine outperforms serum PSA for prostate cancer diagnosis (AUC = 0.71 – 0.77 for *TMPRSS2-ERG* + *PCA3*; AUC = 0.61 for PSA) and adds to available clinical in the Prostate Cancer Prevention Trial (PCPT) risk estimates for predicting cancer (*157*).

There has been some debate as to whether the presence of a *TMPRSS2-ERG* fusion is itself a prognostic biomarker when detected in tissues. Several groups have reported an association between *TMPRSS2-ERG* and aggressive prostate cancer (*157-159*). However, others have not observed this association (*160, 161*). One complication to these studies has been heterogeneity in the patient populations studied and the clinical outcomes evaluated. Interestingly, quantitative levels of urine *TMPRSS2-ERG*, however, appear to be associated with clinically significant prostate cancer based on Epstein criteria, which stratifies disease aggressiveness using PSA density and characteristics of the patient's biopsy (Gleason score, the percent tumor observed, and number of cores with tumor) (*157*).

Additional efforts to discover gene fusions using RNA-seq have been employed to interrogate prostate cancers lacking an ETS gene fusion. Here, our group defined functionally-recurrent gene fusions of the *BRAF* and *RAF* oncogenes in a subset of ETS-negative prostate cancers (*162*). These fusions promoted cell transformation and growth,

and, of note, these fusions were found to retain sensitivity to BRAF inhibitors *in vitro*, suggesting that they may be appropriate as a predictive biomarker in patients. Similar efforts also described an activating *KRAS* gene fusion in the Du145 prostate cancer cell line, and rare genomic disruption of the *KRAS* locus in prostate cancer may contribute to a subset of cancers (*163*).

lncRNAs in prostate cancer

To date, the most prominent lncRNA investigated in prostate cancer has been *PCA3*. *PCA3* is a prostate-specific gene elevated in >90% of prostate cancer tissues, but not in normal or BPH tissues (*164, 165*). However, the biological function of *PCA3* remains unknown, and thus the high sensitivity and specificity of *PCA3* in tissues have led to its evaluation as a non-invasive biomarker for prostate cancer, where numerous assays have been developed to detect it in patient urine samples, which contain cells shed from the prostate during urination (**Figure 1.6**). Over the past decade, several iterations of *PCA3* urine tests have also emerged (*166*) and, currently, a clinical-grade assay based on transcription-mediated amplification is available (*166*).

Urine *PCA3* measurements add to the diagnostic information obtained from the PSA test, with higher AUC values of 0.66 to 0.72, compared to 0.54 to 0.63 for serum PSA alone (*167*). Unlike PSA, *PCA3* levels are independent of prostate size (*168*). Sensitivities for urine *PCA3* levels range from 47 to 69%, with most between 58 and 69%, although it is difficult to directly compare the studies because of different analysis platforms, different criteria for enrolling patients (for example, serum PSA concentrations), and relatively small patient cohorts of several hundred men rather than

thousands (*167*). In addition, combining a serum PSA value with a urine *PCA3* analysis improves both measures, with the combination AUC of 0.71 to 0.75 (*169*). In 2012, *PCA3* was FDA-approved as a diagnostic for prostate cancer in the setting of a prior negative biopsy (http://www.accessdata.fda.gov/cdrh_docs/pdf10/p100033a.pdf).

Another lncRNA, *PCGEM1*, has been reported to be elevated in prostate cancer (*170*). *PCGEM1* expression increased cell proliferation and transformation *in vitro* (*171*). Interestingly, *PCGEM1* expression was associated with patients of African-American ethnicity, a patient population with worse disease outcomes, though it was not associated with disease grade or stage (*171*). Overexpression of *PCGEM1* decreased cell apoptosis following genotoxic stress (*172*), suggesting that this gene may be involved in apoptotic processes, though the mechanism remains unknown.

CONCLUSION

lncRNAs represent a promising avenue for uncovering previously unknown mechanisms of cancer biology. However, to date only a handful of lncRNAs have been well-characterized in human disease, partly because systematic discovery of all lncRNA species has been technically challenging to accomplish. The introduction of RNA-seq promises to facilitate the genome-wide characterization of thousands of unannotated lncRNAs, revealing a new complement of genes to investigate. In cancer, several lncRNAs have been described, such as *HOTAIR* and *ANRIL*, and most of these appear to coordinate epigenetic regulation of gene expression by physically partnering with major histone-modifying complexes. In prostate cancer, however, little is known about the role of lncRNAs. The most prominent lncRNA in prostate cancer, *PCA3*, remains

functionally uncharacterized, and only anecdotal studies have suggested a role for lncRNAs in prostate cancer pathogenesis.

FIGURES

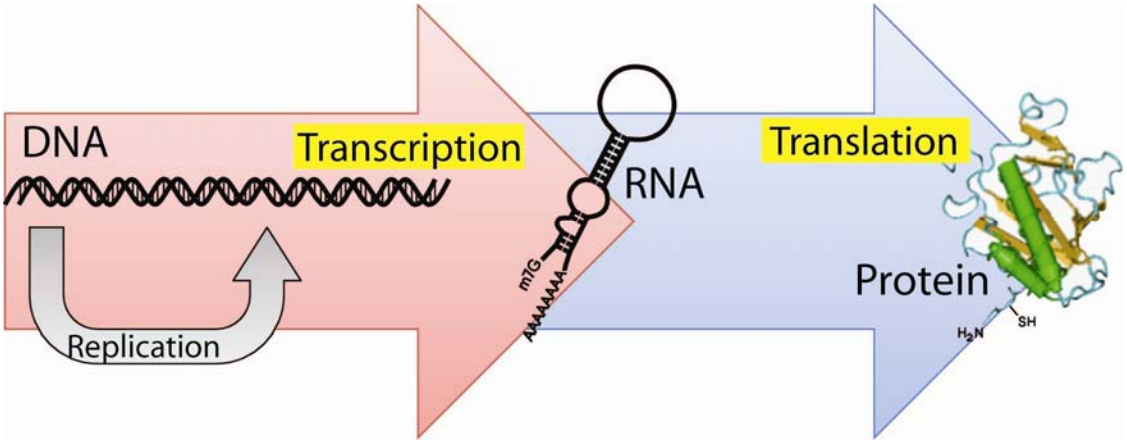


Figure 1.1: A schematic of the central dogma of molecular biology, wherein DNA is transcribed into RNA, and RNA is translated into proteins.

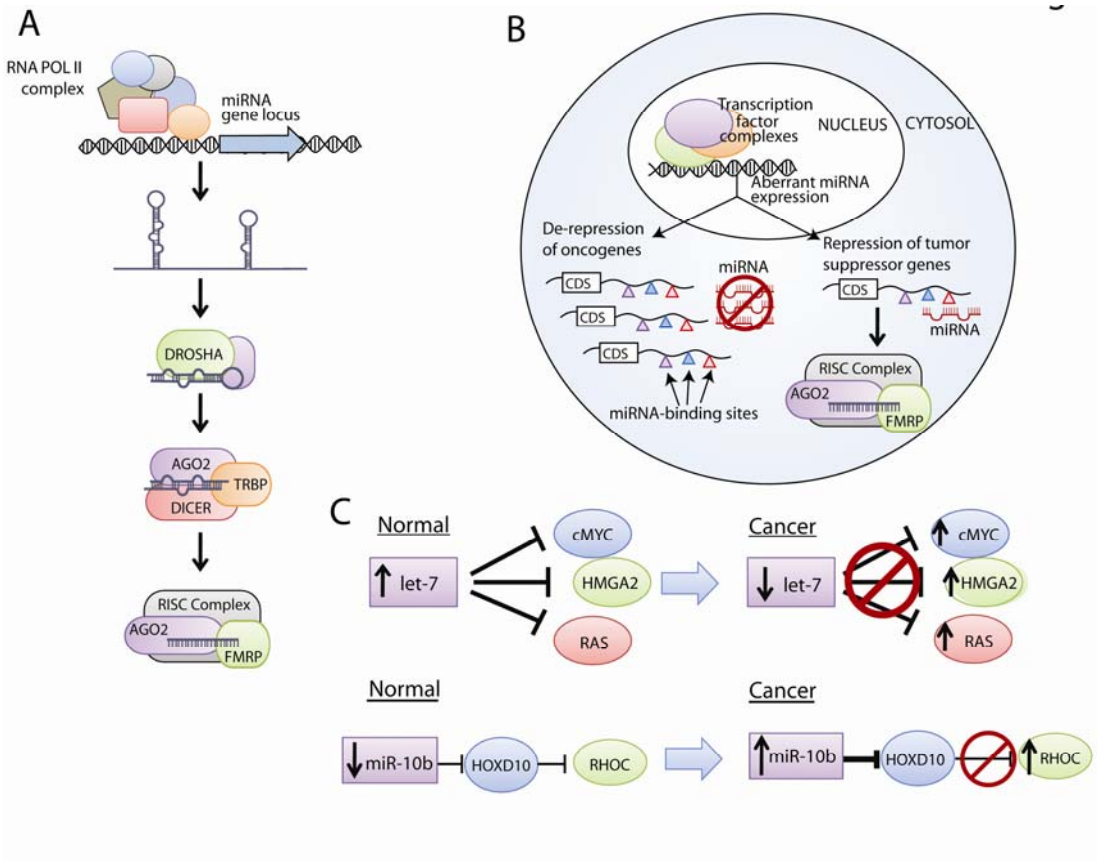


Figure 1.2. MicroRNA-mediated pathways in cancer. **(A)** MicroRNA (miRNA) transcription usually occurs by RNA Polymerase II, generating a primary pri-miRNA transcript. The pri-miRNA is processed by *DROSHA* and cleaved by *DICER* to generate a mature miRNA, which then associates with Argonaute family proteins in the RNA-induced silencing complex (RISC) to achieve gene expression control. (**B**) In cancer, aberrant miRNA expression levels can lead either to the repression of tumor suppressor (typically when miRNA levels are upregulated) or de-repression of oncogenes (typically when miRNA levels are downregulated). The colored triangles indicate different miRNA binding sites in the 3' untranslated region (UTR) of a protein-coding mRNA. Abbreviation: CDS, coding sequence. (**C**) Two examples of aberrant miRNA signaling in cancer are let-7, which is downregulated in cancer and regulates oncogenes such as *cMYC*, and miR-10b, which is upregulated in cancer metastases and indirectly upregulates *RHOC*.

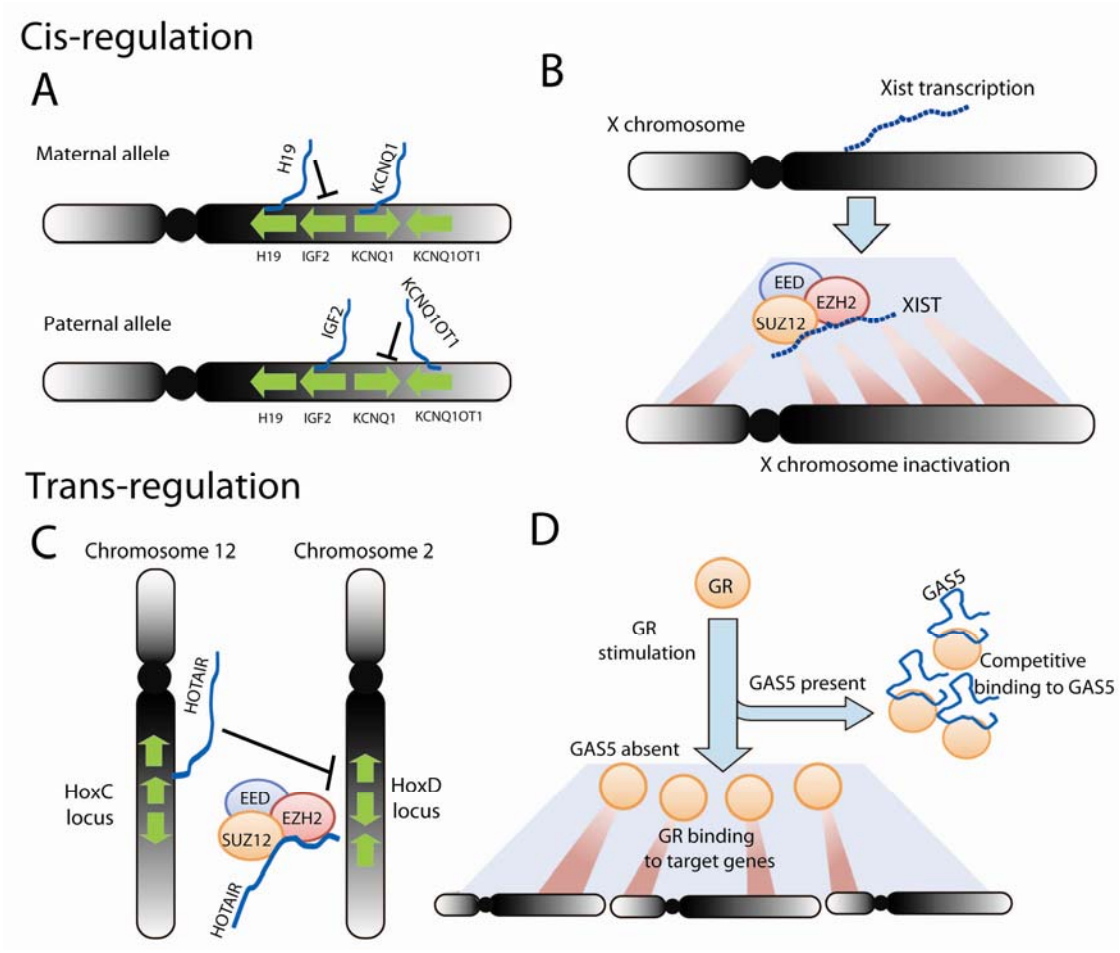


Figure 1.3: Gene expression regulation by lncRNAs. (**A**) and (**B**): *cis*-regulation of gene expression results in local control of genes neighboring, or on the same chromosome as, lncRNA transcription. (**A**) *H19* and *KCNQ1OT1* are imprinted lncRNAs on chromosome 11 associated with allele-specific expression of *IGF2* and *KCNQ1*. (**B**) *XIST* transcription facilitates inactivation of an individual X chromosome in women by recruiting the Polycomb Repressive Complex 2 (PRC2). (**C**) and (**D**) *trans*-regulation of gene expression results in control of genomically-distant genes. (**C**) *HOTAIR* is transcribed from the HoxC cluster on chromosome 12 but represses the HoxD locus via PRC2-mediated epigenetic modifications. (**D**) The *GAS5* lncRNA binds Glucocorticoid Receptor (GR) and sequesters it, preventing upregulation of GR-target genes across the genome. The blue line represents the *GAS5* transcript.

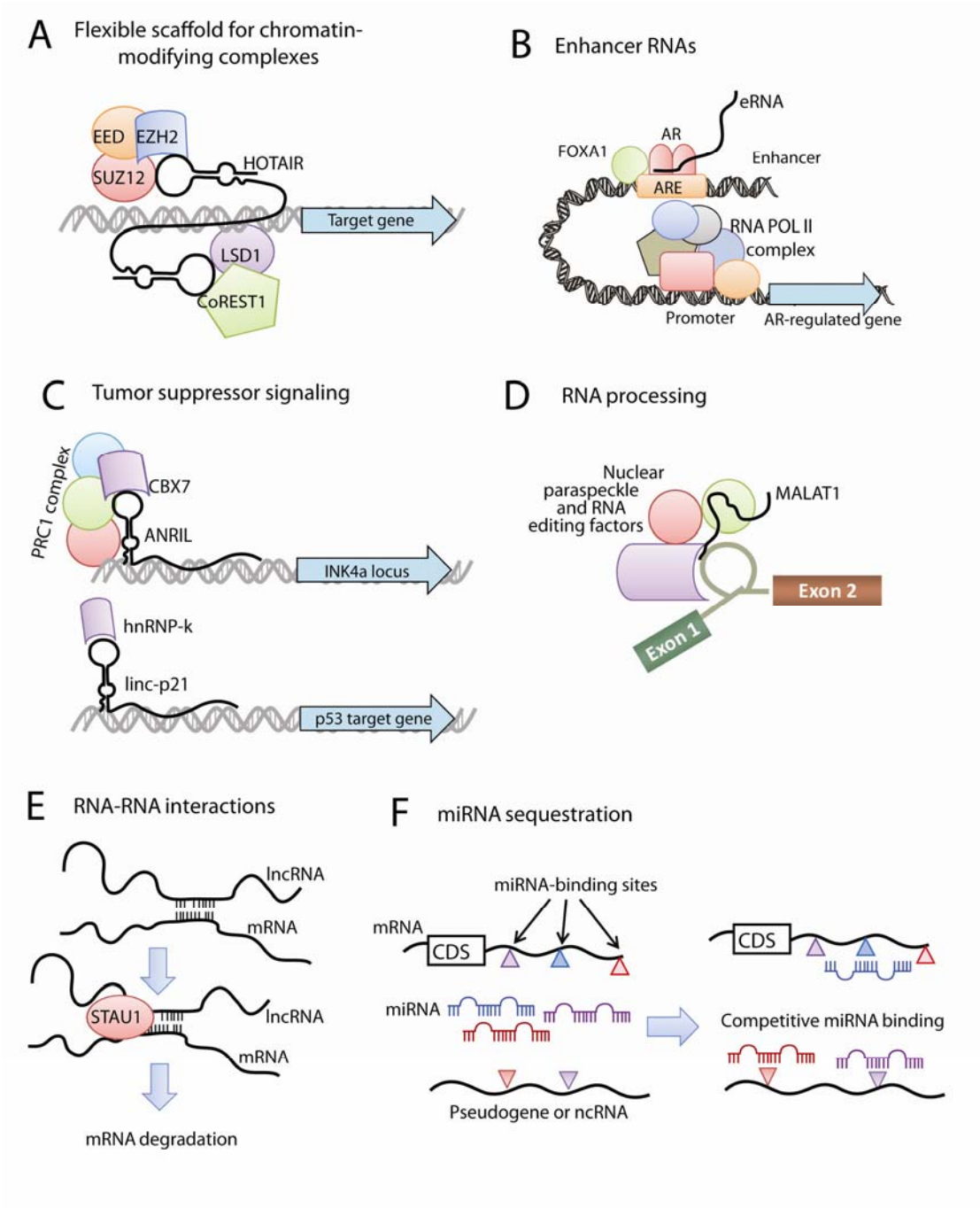


Figure 1.4: Mechanisms of lncRNA function. **(A)** lncRNAs, such as *HOTAIR*, may serve as a scaffolding base for the coordination of epigenetic or histone-modifying complexes, including Polycomb repressive complexes and LSD1/CoREST. (**B**) Enhancer RNAs (eRNAs) transcribed from gene enhancers may facilitate hormone signaling by cooperating with lineage-specific complexes such as FOXA1 and Androgen Receptor. (**C**) lncRNAs may directly impact tumor suppressor signaling either by

transcriptional regulation of tumor suppressor genes through epigenetic silencing (e.g. *ANRIL*, **upper**) or by mediating activation of tumor suppressor target genes (e.g. *linc-p21*, **lower**). (**D**) *MALAT1* and *NEAT2* lncRNAs may be integral components of the nuclear paraspeckle and contribute to post-transcriptional processing of mRNAs. (**E**) Gene expression regulation may occur through direct lncRNA-mRNA interactions which arise from hybridization of homologous sequences and can serve as a signaling for STAU1-mediated degradation of the mRNA. (**F**) RNA molecules, including mRNAs, pseudogenes, and ncRNAs, can serve as molecular sponges for miRNAs. This generates an environment of competitive binding of miRNAs to achieve gene expression control based upon the degree of miRNA binding to each transcript. The colored triangles represent different miRNA binding sites in a transcript. Abbreviation: CDS, coding sequence.

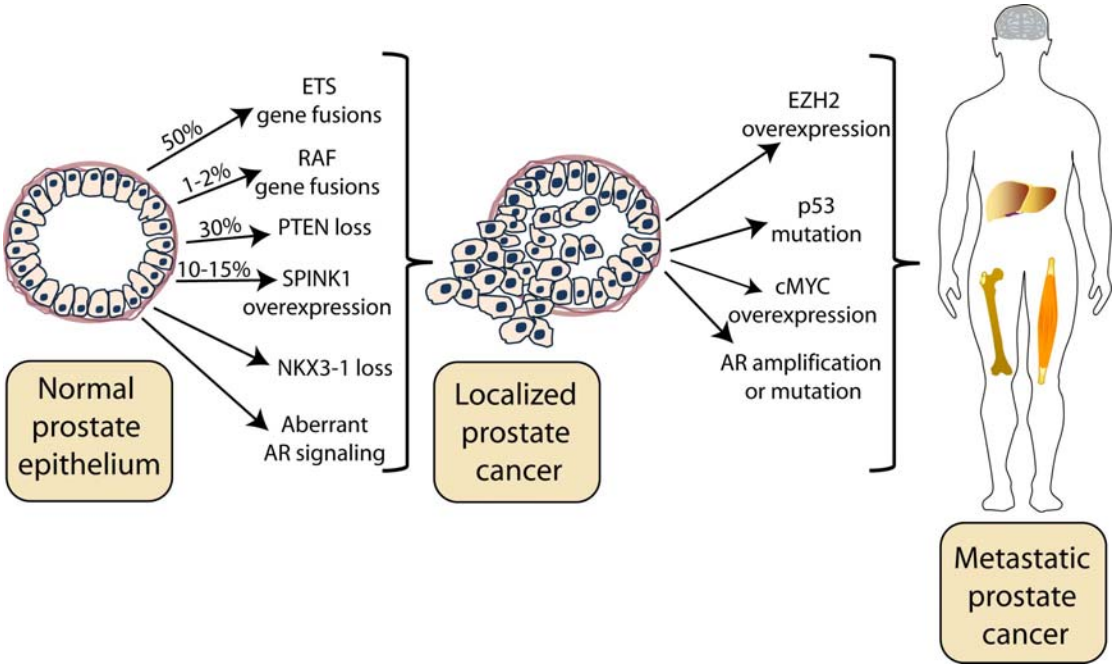


Figure 1.5: Major molecular events in prostate cancer progression.

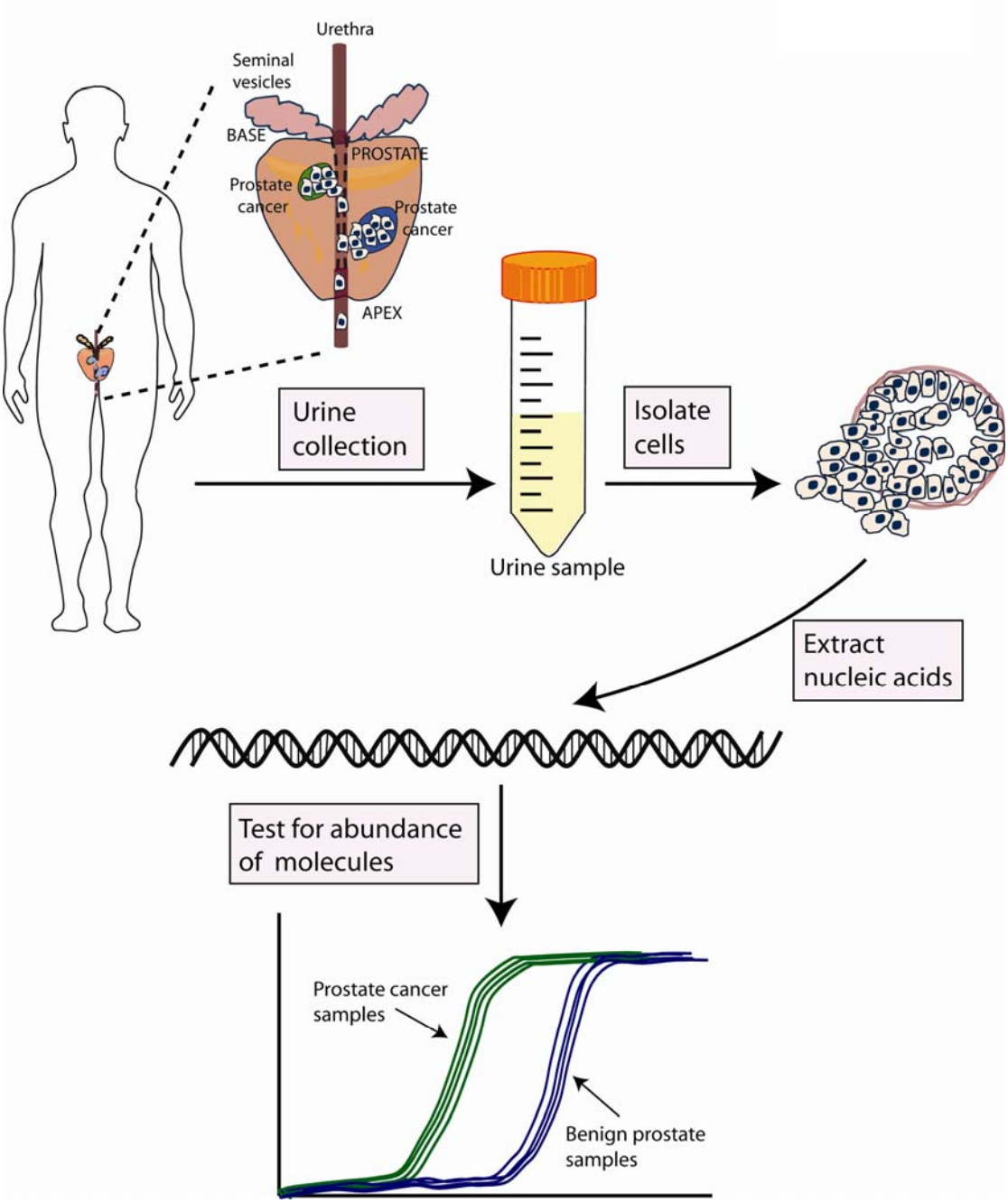


Figure 1.6: Schematic of urine sample analysis. Men selected for prostate cancer screening produce a urine sample for analysis. Urine samples are centrifuged to obtain cells shed into the urine, including cancer cells. Nucleic acids, such as RNA and DNA, are isolated from these cells, and molecular tests, such as PCA3 and TMPRSS2-ERG, are run on this genetic material to generate a molecular profile that distinguishes men with cancer from men without cancer.

TABLES

Table 1.1: Types of ncRNAs known in humans

Category	Name	Quality of supporting data	Specific role in carcinogenesis	Aberration in cancer
Housekeeping RNAs	Transfer RNAs	High	No	No
	Ribosomal RNAs	High	No	No
	Small nucleolar RNAs	High	No	No
	Small nuclear RNAs	High	No	No
Small ncRNAs (200 bp or less in size)	MicroRNAs	High	Yes	Amplification, deletion, methylation, gene expression
	Tiny transcription initiation RNAs	High	Not known	Not known
	Repeat associated small interfering RNAs	High	Not known	Not known
	Promoter-associated short RNAs	High	Not known	Not known
	Termini-associated short RNAs	High	Not known	Not known
	Antisense termini associated short RNAs	High	Not known	Not known
	Transcription start site antisense RNAs	Moderate	Not known	Not known
	Retrotransposon-derived RNAs	High	Not known	Not known
	3'UTR-derived RNAs	Moderate	Not known	Not known
	Splice-site RNAs	Poor	Not known	Not known
Long ncRNA (over 200 bp in size)	Long or large intergenic ncRNAs	High	Yes	Gene expression, translocation
	Transcribed ultraconserved regions	High	Yes	Gene expression
	Pseudogenes	High	Yes	Gene expression, deletion
	Enhancer RNAs	High	Yes	Not known
	Repeat-associated ncRNAs	High	Not known	Not known
	Long intronic ncRNAs	Moderate	Not known	Not known
	Antisense RNAs	High	Yes	Gene expression
	Promoter-associated long RNAs	Moderate	Not known	Not known
	Long stress-induced non-coding transcripts	Moderate	Yes	Gene expression

Table 1.2: Examples of lncRNAs in cancer

lncRNA	Function	Cancer Type	Cancer Phenotype	Molecular Interactors
HULC	Biomarker	Hepatocellular	Not known	Unknown
PCA3	Biomarker	Prostate	Not known	Unknown
ANRIL/p15AS	Oncogenic	Prostate, Leukemia	Suppression of senescence via INK4A	Binds PRC1 and PRC2
HOTAIR	Oncogenic	Breast, hepatocellular	Promotes metastasis	Binds PRC2 and LSD1
MALAT1/NEAT2	Oncogenic	Lung, prostate, breast, colon	Unclear	Contributory to nuclear paraspeckle function
PCAT-1	Oncogenic	Prostate	Promotes cell proliferation; inhibits BRCA2	Unknown
PCGEM1	Oncogenic	Prostate	Inhibits apoptosis; promotes cell proliferation	Unknown
TUC338	Oncogenic	Hepatocellular	Promotes cell proliferation and colony formation	Unknown
uc.73a	Oncogenic	Leukemia	Inhibits apoptosis; promotes cell proliferation	Unknown
H19	Oncogenic; Tumor suppressive	Breast, hepatocellular	Promotes cell growth and proliferation; activated by cMYC; downregulated by prolonged cell proliferation	Unknown
GAS5	Tumor suppressive	Breast	Induces apoptosis and growth arrest; Prevents GR-induced gene expression	Binds GR
linc-p21	Tumor suppressive	Mouse models of lung, sarcoma, lymphoma	Mediates p53 signaling; induces apoptosis	Binds hnRNP-k
MEG3	Tumor suppressive	Meningioma, hepatocellular, leukemia, pituitary tumors	Mediates p53 signaling; inhibits cell proliferation	Unknown
PTENP1	Tumor suppressive	Prostate, colon	Binds PTEN-suppressing miRNAs	Unknown

Abbreviations: Polycomb Repressive Complex 1 , PRC1; Polycomb Repressive Complex 2, PRC2; Glucocorticoid Receptor, GR

Table 1.3: Gene fusions of ETS transcription factors in prostate cancer.

5' Fusion Gene	3' Fusion Gene	Prevalence
TMPRSS2	ERG	~40-50%
NDRG1	ERG	<1%
TMPRSS2	ETV1	5-10%
HERV_K	ETV1	<1%
C15orf21	ETV1	<1%
HNRPA2B1	ETV1	<1%
SLC45A3	ETV1	<1%
ACSL3	ETV1	<1%
FLJ35294	ETV1	<1%
TMPRSS2	ETV4	1-2%
KLK2	ETV4	<1%
CANT1	ETV4	<1%
DDX5	ETV4	<1%
TMPRSS2	ETV5	1-2%
SLC45A3	ETV5	<1%

REFERENCES

1. F. H. Crick, *Symp Soc Exp Biol* **12**, 138 (1958).
2. F. H. Crick, L. Barnett, S. Brenner, R. J. Watts-Tobin, *Nature* **192**, 1227 (Dec 30, 1961).
3. E. Ochoa, *Nobel Lecture*, (1959).
4. D. Thieffry, S. Sarkar, *Trends Biochem Sci* **23**, 312 (Aug, 1998).
5. K. Matsubara, K. Okubo, *Curr Opin Biotechnol* **4**, 672 (Dec, 1993).
6. F. Liang *et al.*, *Nat Genet* **25**, 239 (Jun, 2000).
7. E. Birney *et al.*, *Nature* **447**, 799 (Jun 14, 2007).
8. E. L. van Dijk *et al.*, *Nature* **475**, 114 (Jul 7, 2011).
9. P. Kapranov *et al.*, *Science* **316**, 1484 (Jun 8, 2007).
10. J. Cheng *et al.*, *Science* **308**, 1149 (May 20, 2005).
11. P. Carninci *et al.*, *Science* **309**, 1559 (Sep 2, 2005).
12. M. Huarte, J. L. Rinn, *Hum Mol Genet* **19**, R152 (Oct 15, 2010).
13. A. Pauli, J. L. Rinn, A. F. Schier, *Nat Rev Genet* **12**, 136 (Feb, 2011).
14. J. R. Prensner *et al.*, *Nat Biotechnol*, (Jul 31, 2011).
15. E. A. Gibb, C. J. Brown, W. L. Lam, *Mol Cancer* **10**, 38 (2011).
16. R. J. Taft, K. C. Pang, T. R. Mercer, M. Dinger, J. S. Mattick, *J Pathol* **220**, 126 (Jan, 2010).
17. R. Garzon, G. A. Calin, C. M. Croce, *Annu Rev Med* **60**, 167 (2009).
18. Y. He, B. Vogelstein, V. E. Velculescu, N. Papadopoulos, K. W. Kinzler, *Science* **322**, 1855 (Dec 19, 2008).
19. G. J. Faulkner *et al.*, *Nat Genet* **41**, 563 (May, 2009).
20. M. Guttman *et al.*, *Nature* **458**, 223 (Mar 12, 2009).
21. T. K. Kim *et al.*, *Nature* **465**, 182 (May 13, 2010).
22. G. A. Calin *et al.*, *Cancer Cell* **12**, 215 (Sep, 2007).
23. C. Braconi *et al.*, *Proc Natl Acad Sci U S A* **108**, 786 (Jan 11, 2011).
24. C. J. Brown *et al.*, *Nature* **349**, 38 (Jan 3, 1991).
25. M. S. Bartolomei, S. Zemel, S. M. Tilghman, *Nature* **351**, 153 (May 9, 1991).
26. C. Trapnell *et al.*, *Nat Biotechnol* **28**, 511 (May, 2010).
27. M. Guttman *et al.*, *Nat Biotechnol* **28**, 503 (May, 2010).
28. M. G. Grabherr *et al.*, *Nat Biotechnol* **29**, 644 (2011).
29. R. A. Gupta *et al.*, *Nature* **464**, 1071 (Apr 15, 2010).
30. D. Wang *et al.*, *Nature* **474**, 390 (Jun 16, 2011).
31. I. J. Matouk *et al.*, *PLoS One* **2**, e845 (2007).
32. M. P. Lee *et al.*, *Proc Natl Acad Sci U S A* **96**, 5203 (Apr 27, 1999).
33. R. Weksberg *et al.*, *Hum Mol Genet* **10**, 2989 (Dec 15, 2001).
34. S. Lottin *et al.*, *Carcinogenesis* **23**, 1885 (Nov, 2002).
35. A. Gabory, H. Jammes, L. Dandolo, *Bioessays* **32**, 473 (Jun, 2010).
36. D. Barsyte-Lovejoy *et al.*, *Cancer Res* **66**, 5330 (May 15, 2006).
37. C. Pantoja, L. de Los Rios, A. Matheu, F. Antequera, M. Serrano, *Cancer Res* **65**, 26 (Jan 1, 2005).
38. J. Bliek *et al.*, *Hum Mol Genet* **10**, 467 (Mar 1, 2001).
39. J. Chow, E. Heard, *Curr Opin Cell Biol* **21**, 359 (Jun, 2009).
40. J. T. Lee, *Genes Dev* **23**, 1831 (Aug 15, 2009).

41. J. Zhao, B. K. Sun, J. A. Erwin, J. J. Song, J. T. Lee, *Science* **322**, 750 (Oct 31, 2008).
42. S. M. Sirchia *et al.*, *PLoS One* **4**, e5559 (2009).
43. A. L. Richardson *et al.*, *Cancer Cell* **9**, 121 (Feb, 2006).
44. S. Ganesan *et al.*, *Cell* **111**, 393 (Nov 1, 2002).
45. C. Xiao *et al.*, *Cell* **128**, 977 (Mar 9, 2007).
46. T. Kawakami, K. Okamoto, O. Ogawa, Y. Okada, *Lancet* **363**, 40 (Jan 3, 2004).
47. E. Pasmant *et al.*, *Cancer Res* **67**, 3963 (Apr 15, 2007).
48. K. L. Yap *et al.*, *Mol Cell* **38**, 662 (Jun 11, 2010).
49. Y. Kotake *et al.*, *Oncogene* **30**, 1956 (Apr 21, 2011).
50. C. E. Burd *et al.*, *PLoS Genet* **6**, e1001233 (2010).
51. E. Pasmant, A. Sabbagh, M. Vidaud, I. Bieche, *FASEB J* **25**, 444 (Feb, 2011).
52. X. Zhang *et al.*, *Blood* **113**, 2526 (Mar 12, 2009).
53. K. C. Wang *et al.*, *Nature* **472**, 120 (Apr 7, 2011).
54. D. Mancini-Dinardo, S. J. Steele, J. M. Levorse, R. S. Ingram, S. M. Tilghman, *Genes Dev* **20**, 1268 (May 15, 2006).
55. F. M. Pauler, M. V. Koerner, D. P. Barlow, *Trends Genet* **23**, 284 (Jun, 2007).
56. J. L. Rinn *et al.*, *Cell* **129**, 1311 (Jun 29, 2007).
57. M. C. Tsai *et al.*, *Science* **329**, 689 (Aug 6, 2010).
58. T. Kino, D. E. Hurt, T. Ichijo, N. Nader, G. P. Chrousos, *Sci Signal* **3**, ra8 (2010).
59. M. Mourtada-Maarabouni, M. R. Pickard, V. L. Hedge, F. Farzaneh, G. T. Williams, *Oncogene* **28**, 195 (Jan 15, 2009).
60. G. Bejerano *et al.*, *Science* **304**, 1321 (May 28, 2004).
61. P. Mestdagh *et al.*, *Oncogene* **29**, 3583 (Jun 17, 2010).
62. R. Margueron, D. Reinberg, *Nature* **469**, 343 (Jan 20, 2011).
63. T. Li *et al.*, *Mol Cell Biol* **28**, 6473 (Oct, 2008).
64. R. R. Pandey *et al.*, *Mol Cell* **32**, 232 (Oct 24, 2008).
65. A. M. Khalil *et al.*, *Proc Natl Acad Sci U S A* **106**, 11667 (Jul 14, 2009).
66. E. Bernstein *et al.*, *Mol Cell Biol* **26**, 2560 (Apr, 2006).
67. T. Nagano *et al.*, *Science* **322**, 1717 (Dec 12, 2008).
68. U. A. Orom *et al.*, *Cell* **143**, 46 (Oct 1, 2010).
69. W. Yu *et al.*, *Nature* **451**, 202 (Jan 10, 2008).
70. Y. Zhou *et al.*, *J Biol Chem* **282**, 24731 (Aug 24, 2007).
71. X. Zhang *et al.*, *Cancer Res* **70**, 2350 (Mar 15, 2010).
72. C. Braconi *et al.*, *Oncogene*, (May 30, 2011).
73. L. Benetatos *et al.*, *Leuk Res* **34**, 148 (Feb, 2010).
74. M. Huarte *et al.*, *Cell* **142**, 409 (Aug 6, 2010).
75. C. S. Bond, A. H. Fox, *J Cell Biol* **186**, 637 (Sep 7, 2009).
76. K. V. Prasanth *et al.*, *Cell* **123**, 249 (Oct 21, 2005).
77. D. Bernard *et al.*, *EMBO J* **29**, 3082 (Sep 15, 2010).
78. V. Tripathi *et al.*, *Mol Cell* **39**, 925 (Sep 24, 2010).
79. J. E. Wilusz, S. M. Freier, D. L. Spector, *Cell* **135**, 919 (Nov 28, 2008).
80. P. Ji *et al.*, *Oncogene* **22**, 8031 (Sep 11, 2003).
81. C. Gong, L. E. Maquat, *Nature* **470**, 284 (Feb 10, 2011).
82. L. Poliseno *et al.*, *Nature* **465**, 1033 (Jun 24, 2010).
83. M. N. Cabili *et al.*, *Genes Dev* **25**, 1915 (Sep 15, 2011).

84. J. R. Prensner *et al.*, *Nat Biotechnol* **29**, 742 (Aug, 2011).
85. E. S. Lander *et al.*, *Nature* **409**, 860 (Feb 15, 2001).
86. I. H. G. S. Consortium, *Nature* **431**, 931 (Oct 21, 2004).
87. M. D. Adams *et al.*, *Science* **252**, 1651 (Jun 21, 1991).
88. G. D. Schuler *et al.*, *Science* **274**, 540 (Oct 25, 1996).
89. F. Ozsolak, P. M. Milos, *Nat Rev Genet* **12**, 87 (Feb, 2011).
90. M. B. Clark *et al.*, *Genome Res*, (Mar 9, 2012).
91. L. J. Core, J. J. Waterfall, J. T. Lis, *Science* **322**, 1845 (Dec 19, 2008).
92. N. Hah *et al.*, *Cell* **145**, 622 (May 13, 2011).
93. J. Zhao *et al.*, *Mol Cell* **40**, 939 (Dec 22, 2010).
94. M. S. Steiner, *World J Urol* **21**, 15 (May, 2003).
95. M. D. Reis LAG, Krapcho M, Mariotto A, Miller BA, Feuer EJ, CleggL, Horner MJ, Howlader N, Eisner MP, Reichman M, Edwards BK, "SEER Cancer Statistics Review, 1975-2004" (National Cancer Institute, Bethesda, MD, 2007).
96. N. B. Delongchamps, A. Singh, G. P. Haas, *Cancer Control* **13**, 158 (Jul, 2006).
97. CDC. (Centers for Disease Control and Prevention, 2007), pp. http://www.cdc.gov/cancer/prostate/statistics/race.htm.
98. R. Etzioni, R. Cha, E. J. Feuer, O. Davidov, *Am J Epidemiol* **148**, 775 (Oct 15, 1998).
99. A. M. Wolf *et al.*, *CA Cancer J Clin* **60**, 70 (Mar-Apr, 2010).
100. T. Kahn *et al.*, *Radiology* **173**, 847 (Dec, 1989).
101. J. E. McNeal, E. A. Redwine, F. S. Freiha, T. A. Stamey, *Am J Surg Pathol* **12**, 897 (Dec, 1988).
102. D. Hanahan, R. A. Weinberg, *Cell* **100**, 57 (Jan 7, 2000).
103. Z. Kan *et al.*, *Nature* **466**, 869 (Aug 12, 2010).
104. C. S. Grasso *et al.*, *Nature*, (In press).
105. M. F. Berger *et al.*, *Nature* **470**, 214 (Feb 10, 2011).
106. J. D. Brooks, G. S. Bova, W. B. Isaacs, *Prostate* **26**, 35 (Jan, 1995).
107. M. M. Ittmann, R. Wieczorek, *Hum Pathol* **27**, 28 (Jan, 1996).
108. L. C. Trotman *et al.*, *PLoS Biol* **1**, E59 (Dec, 2003).
109. M. Yoshimoto *et al.*, *Br J Cancer* **97**, 678 (Sep 3, 2007).
110. B. Han *et al.*, *Mod Pathol* **22**, 1083 (Aug, 2009).
111. M. Yoshimoto *et al.*, *Mod Pathol* **21**, 1451 (Dec, 2008).
112. X. Wu, K. Senechal, M. S. Neshat, Y. E. Whang, C. L. Sawyers, *Proc Natl Acad Sci U S A* **95**, 15587 (Dec 22, 1998).
113. B. S. Carver *et al.*, *Cancer Cell* **19**, 575 (May 17, 2011).
114. W. W. He *et al.*, *Genomics* **43**, 69 (Jul 1, 1997).
115. C. Bowen *et al.*, *Cancer Res* **60**, 6111 (Nov 1, 2000).
116. C. J. Bieberich, K. Fujita, W. W. He, G. Jay, *J Biol Chem* **271**, 31779 (Dec 13, 1996).
117. C. R. Bethel *et al.*, *Cancer Res* **66**, 10683 (Nov 15, 2006).
118. C. D. Vocke *et al.*, *Cancer Res* **56**, 2411 (May 15, 1996).
119. M. R. Emmert-Buck *et al.*, *Cancer Res* **55**, 2959 (Jul 15, 1995).
120. W. A. Sakr *et al.*, *Cancer Res* **54**, 3273 (Jun 15, 1994).
121. Q. Lei *et al.*, *Cancer Cell* **9**, 367 (May, 2006).

122. X. Ouyang, T. L. DeWeese, W. G. Nelson, C. Abate-Shen, *Cancer Res* **65**, 6773 (Aug 1, 2005).
123. J. H. Kim *et al.*, *Cancer Res* **67**, 8229 (Sep 1, 2007).
124. K. Ellwood-Yen *et al.*, *Cancer Cell* **4**, 223 (Sep, 2003).
125. R. B. Jenkins, J. Qian, M. M. Lieber, D. G. Bostwick, *Cancer Res* **57**, 524 (Feb 1, 1997).
126. Huggins, *Cancer Res* **1**, (1941, 1941).
127. C. Huggins, *Cancer Res* **27**, 1925 (Nov, 1967).
128. P. Koivisto *et al.*, *Cancer Res* **57**, 314 (Jan 15, 1997).
129. T. Visakorpi *et al.*, *Nat Genet* **9**, 401 (Apr, 1995).
130. G. Buchanan, R. A. Irvine, G. A. Coetzee, W. D. Tilley, *Cancer Metastasis Rev* **20**, 207 (2001).
131. B. J. Feldman, D. Feldman, *Nat Rev Cancer* **1**, 34 (Oct, 2001).
132. M. Marcelli *et al.*, *Cancer Res* **60**, 944 (Feb 15, 2000).
133. M. E. Taplin *et al.*, *Cancer Res* **59**, 2511 (Jun 1, 1999).
134. M. E. Taplin *et al.*, *N Engl J Med* **332**, 1393 (May 25, 1995).
135. W. D. Tilley, G. Buchanan, T. E. Hickey, J. M. Bentel, *Clin Cancer Res* **2**, 277 (Feb, 1996).
136. R. Chmelar, G. Buchanan, E. F. Need, W. Tilley, N. M. Greenberg, *Int J Cancer* **120**, 719 (Feb 15, 2007).
137. G. N. Brooke, M. G. Parker, C. L. Bevan, *Oncogene*, (Nov 26, 2007).
138. C. Neslund-Dudas *et al.*, *Prostate* **67**, 1654 (Nov 1, 2007).
139. M. T. Salam, G. Ursin, E. C. Skinner, T. Dessissa, J. K. Reichardt, *Urol Oncol* **23**, 246 (Jul-Aug, 2005).
140. A. Shibata *et al.*, *Prostate* **52**, 269 (Sep 1, 2002).
141. F. Labrie *et al.*, *Endocr Rev* **7**, 67 (Feb, 1986).
142. J. P. Gaddipati *et al.*, *Cancer Res* **54**, 2861 (Jun 1, 1994).
143. Z. Culig, *Urology* **62**, 21 (Nov, 2003).
144. S. Godoy-Tundidor, A. Hobisch, K. Pfeil, G. Bartsch, Z. Culig, *Clin Cancer Res* **8**, 2356 (Jul, 2002).
145. S. A. Tomlins *et al.*, *Science* **310**, 644 (Oct 28, 2005).
146. S. A. Tomlins *et al.*, *Nature* **448**, 595 (Aug 2, 2007).
147. S. Perner *et al.*, *Cancer Res* **66**, 8337 (Sep 1, 2006).
148. M. J. Soller *et al.*, *Genes Chromosomes Cancer* **45**, 717 (Jul, 2006).
149. M. Yoshimoto *et al.*, *Neoplasia* **8**, 465 (Jun, 2006).
150. R. Mehra *et al.*, *Cancer Res* **68**, 3584 (May 15, 2008).
151. C. Lin *et al.*, *Cell* **139**, 1069 (Dec 11, 2009).
152. R. S. Mani *et al.*, *Science* **326**, 1230 (Nov 27, 2009).
153. S. A. Tomlins *et al.*, *Neoplasia* **10**, 177 (Feb, 2008).
154. S. Perner *et al.*, *Am J Surg Pathol* **31**, 882 (Jun, 2007).
155. N. Cerveira *et al.*, *Neoplasia* **8**, 826 (Oct, 2006).
156. J. M. Mosquera *et al.*, *Clin Cancer Res* **14**, 3380 (Jun 1, 2008).
157. S. A. Tomlins *et al.*, *Sci Transl Med* **3**, 94ra72 (Aug 3, 2011).
158. G. Attard *et al.*, *Oncogene* **27**, 253 (Jan 10, 2008).
159. F. Demichelis *et al.*, *Oncogene* **26**, 4596 (Jul 5, 2007).
160. S. W. Fine *et al.*, *Mod Pathol* **23**, 1325 (Oct, 2010).

161. A. Gopalan *et al.*, *Cancer Res* **69**, 1400 (Feb 15, 2009).
162. N. Palanisamy *et al.*, *Nat Med* **16**, 793 (Jul, 2010).
163. X. S. Wang *et al.*, *Cancer Discov* **1**, 35 (Jun 1, 2011).
164. M. J. Bussemakers *et al.*, *Cancer Res* **59**, 5975 (Dec 1, 1999).
165. J. B. de Kok *et al.*, *Cancer Res* **62**, 2695 (May 1, 2002).
166. D. Hessels, J. A. Schalken, *Nat Rev Urol* **6**, 255 (May, 2009).
167. M. J. Roobol *et al.*, *Eur Urol* **58**, 475 (Oct, 2010).
168. A. Haese *et al.*, *Eur Urol* **54**, 1081 (Nov, 2008).
169. R. Wang, A. M. Chinnaiyan, R. L. Dunn, K. J. Wojno, J. T. Wei, *Cancer* **115**, 3879 (Sep 1, 2009).
170. V. Srikantan *et al.*, *Proc Natl Acad Sci U S A* **97**, 12216 (Oct 24, 2000).
171. G. Petrovics *et al.*, *Oncogene* **23**, 605 (Jan 15, 2004).
172. X. Fu, L. Ravindranath, N. Tran, G. Petrovics, S. Srivastava, *DNA Cell Biol* **25**, 135 (Mar, 2006).

CHAPTER 2

Transcriptome sequencing across a cohort of prostate cancers identifies long non-coding RNAs associated with disease progression

SUMMARY

High-throughput sequencing of polyA+ RNA (RNA-Seq) in human cancer shows remarkable potential to identify both novel markers of disease and uncharacterized aspects of tumor biology, particularly non-coding RNA (ncRNA) species. We employed RNA-Seq on a cohort of 102 prostate tissues and cells lines and performed *ab initio* transcriptome assembly to discover unannotated ncRNAs. In total, we identified >6,000 unannotated transcripts, of which 1,857 were intergenic in nature. These transcripts were associated with histone marks supporting independent transcriptional start sites and active transcription, including H3K4me3, H3K36me3, and RNA Polymerase II. To identify novel transcripts that may be implicated in cancer biology, we nominated 121 such Prostate Cancer Associated Transcripts (PCATs) with cancer-specific expression patterns, several of which also scored as prominent outliers expressed in only a subset of samples. Strikingly, many PCATs exhibited prostate-specific expression, suggesting that lineage-restricted expression of lncRNAs may contribute to disease biology. These findings establish the utility of RNA-Seq to identify disease-associated ncRNAs that may improve the stratification of cancer subtypes.

INTRODUCTION

Recently, next generation transcriptome sequencing (RNA-Seq) has provided a method to delineate the entire set of transcriptional aberrations in a disease, including novel transcripts and non-coding RNAs (ncRNAs) not measured by conventional analyses (*1-5*). To facilitate interpretation of sequence read data, existing computational methods typically process individual samples using either short read gapped alignment followed by *ab initio* reconstruction (*2, 3*), or *de novo* assembly of read sequences followed by sequence alignment (*4, 5*). These methods provide a powerful framework to uncover uncharacterized RNA species, including antisense transcripts, short RNAs <250 bps, or long ncRNAs (lncRNAs) >250 bps.

While still largely unexplored, ncRNAs, particularly lncRNAs, have emerged as a new aspect of biology, with evidence suggesting that they are frequently cell-type specific, contribute important functions to numerous systems (*6, 7*), and may interact with known cancer genes such as *EZH2* (*8*). Indeed, several well-described examples, such as *HOTAIR* (*8, 9*) and *ANRIL* (*10, 11*), indicate that ncRNAs may be essential actors in cancer biology, typically facilitating epigenetic gene repression via chromatin modifying complexes (*12, 13*). Moreover, ncRNA expression may confer clinical information about patient outcomes and have utility as diagnostic tests (*9, 14*). The characterization of RNA species, their functions, and their clinical applicability is therefore a major area of biological and clinical importance.

Here, we describe a comprehensive analysis of lncRNAs in 102 prostate cancer tissue samples and cell lines by RNA-Seq. We employ *ab initio* computational approaches to delineate the annotated and unannotated transcripts in this disease, and we

find 121 ncRNAs, termed Prostate Cancer Associated Transcripts (PCATs), whose expression patterns distinguish benign, localized cancer, and metastatic cancer samples.

RESULTS

RNA-Seq analysis of the prostate cancer transcriptome

Over two decades of research has generated a genetic model of prostate cancer based on numerous neoplastic events, such as loss of the *PTEN* (*15*) tumor suppressor gene and gain of oncogenic ETS transcription factor gene fusions (*16-18*) in large subsets of prostate cancer patients. We hypothesized that prostate cancer similarly harbored disease-associated ncRNAs in molecular subtypes.

To pursue this hypothesis, we employed transcriptome sequencing on a cohort of 102 prostate tissues and cell lines (20 benign adjacent prostates (benign), 47 localized tumors (PCA), and 14 metastatic tumors (MET) and 21 prostate cell lines). From a total of 1.723 billion sequence fragments from 201 lanes of sequencing (108 paired-end, 93 single read on the Illumina Genome Analyzer and Genome Analyzer II), we performed short read gapped alignment (*19*) and recovered 1.41 billion mapped reads, with a median of 14.7 million mapped reads per sample. We used the Cufflinks *ab initio* assembly approach (*3*) to produce, for each sample, the most probable set of putative transcripts that served as the RNA templates for the sequence fragments in that sample (**Figures 2.1** and **2.2**).

As expected from a large tumor tissue cohort, individual transcript assemblies may exhibit sources of "noise", such as artifacts of the sequence alignment process, unspliced intronic pre-mRNA, and genomic DNA contamination. To exclude these from

our analyses, we trained a decision tree to classify transcripts as "expressed" versus "background" on the basis of transcript length, number of exons, recurrence in multiple samples, and other structural characteristics (**Figure 2.1b, left**). The classifier demonstrated a sensitivity of 70.8% and specificity of 88.3% when trained using transcripts that overlapped genes in the AceView database (*20*), including 11.7% of unannotated transcripts that were classified as "expressed" (**Figure 2.1b, right**). We then clustered the "expressed" transcripts into a consensus transcriptome and applied additional heuristic filters to further refine the assembly. The final *ab initio* transcriptome assembly yielded 35,415 distinct transcriptional loci (**Table 2.1**).

Discovery of prostate cancer non-coding RNAs

We compared the assembled prostate cancer transcriptome to the UCSC, Ensembl, Refseq, Vega, and ENCODE gene databases to identify and categorize transcripts (**Figure 2.1c**). While the majority of the transcripts (77.3%) corresponded to annotated protein coding genes (72.1%) and non-coding RNAs (5.2%), a significant percentage (19.8%) lacked any overlap and were designated "unannotated" (**Figure 2.3a**). These included partially intronic antisense (2.44%), totally intronic (12.1%), and intergenic transcripts (5.25%), consistent with previous reports of unannotated transcription (*21-23*). Due to the added complexity of characterizing antisense or partially intronic transcripts without strand-specific RNA-Seq libraries, we focused on totally intronic and intergenic transcripts.

Global characterization of novel intronic and intergenic transcripts demonstrated that they were more highly expressed (**Figure 2.3b**), had greater overlap with expressed

sequence tags (ESTs) (**Figure 2.4**), and displayed a clear but subtle increase in conservation over randomly permuted controls (novel intergenic transcripts $p = 2.7 \times 10^{-4} \pm 0.0002$ for $0.4 < \omega < 0.8$; novel intronic transcripts $p = 2.6 \times 10^{-5} \pm 0.0017$ for $0 < \omega < 0.4$, Fisher's exact test, **Figure 2.3c**). By contrast, unannotated transcripts scored lower than protein-coding genes for these metrics, which corroborates data in previous reports (*2, 24*). Interestingly, a small subset of novel intronic transcripts showed a profound degree of conservation (**Figure 2.3c**, insert). Finally, analysis of coding potential revealed that only 5 of 6,144 transcripts harbored a high quality open reading frame (ORF), indicating that the vast majority of these transcripts represent ncRNAs (**Figure 2.5**).

To determine whether our unannotated transcripts were supported by histone modifications defining active transcriptional units, we used published prostate cancer ChIP-Seq data for two prostate cell lines (*25*), VCaP and LNCaP. After filtering our dataset for transcribed repetitive elements known to display alternative patterns of histone modifications (*26*), we observed a strong enrichment for histone modifications characterizing transcriptional start sites (TSSs) and active transcription, including H3K4me2, H3K4me3, Acetyl-H3 and RNA polymerase II (**Figure 2.3d-g**) but not H3K4me1, which characterizes enhancer regions (*27*) (**Figures 2.6**). Interestingly, intergenic ncRNAs showed greater enrichment compared to intronic ncRNAs in these analyses (**Figure 2.3d-g**).

To elucidate global changes in transcript abundance in prostate cancer, we performed a differential expression analysis for all transcripts. We found 836 genes differentially-expressed between benign samples and localized tumors (FDR < 0.01),

with annotated protein-coding and ncRNA genes constituting 82.8% and 7.4% of differentially-expressed genes, respectively, including known prostate cancer biomarkers such *AMACR* (*28*), *HPN* (*29*), and *PCA3* (*14*) (**Figure 2.3h**). Finally, 9.8% of differentially-expressed genes corresponded to unannotated ncRNAs, including 3.2% within gene introns and 6.6% in intergenic regions.

Characterization of Prostate Cancer Associated Transcripts

As ncRNAs may contribute to human disease (*6-9*), we identified aberrantly expressed uncharacterized ncRNAs in prostate cancer. We found a total of 1,859 unannotated lncRNAs throughout the human genome. Overall, these intergenic RNAs resided approximately half-way between two protein coding genes (**Figures 2.7**), and over one-third (34.1%) were ≥10kb from the nearest protein-coding gene, which is consistent with previous reports (*30*) and supports the independence of intergenic ncRNAs genes.

A focused analysis of the 1,859 unannotated intergenic RNAs yielded 106 that were differentially expressed in localized tumors (FDR < 0.05, **Figure 2.8a**). A cancer outlier expression analysis similarly nominated numerous unannotated ncRNA outliers (**Figure 2.8b**) as well as known prostate cancer outliers, such as *ERG* (*18*), *ETV1 (17, 18)*, *SPINK1* (*31*) and *CRISP3* (*32*). Merging these results produced a set of 121 unannotated transcripts that accurately discriminated benign, localized tumor, and metastatic prostate samples by unsupervised clustering (**Figure 2.8a**). These 121 unannotated transcripts were ranked and named as Prostate Cancer Associated

Transcripts (PCATs) according to their fold change in localized tumor versus benign tissue (**Tables 2.2** and **2.3**).

Validation of novel ncRNAs

To gain confidence in our transcript nominations, we validated multiple unannotated transcripts *in vitro* by reverse transcription PCR (RT-PCR) and quantitative real-time PCR (qPCR) (**Figure 2.9**). qPCR for four transcripts (*PCAT-114*, *PCAT-14*, *PCAT-43*, *PCAT-1*) on an independent cohort of prostate tissues confirmed predicted cancer-specific expression patterns (**Figure 2.8c-f**). Interestingly, all four are prostate-specific, with minimal expression seen by qPCR in breast (n=14) or lung cancer (n=16) cell lines or in 19 normal tissue types. This is further supported by expression analysis of these transcripts in our RNA-Seq compendium of 13 tumor types, representing 325 samples (**Figure 2.10**). This tissue specificity was not necessarily due to regulation by androgen receptor signaling, as only *PCAT-14* expression was induced when androgen responsive VCaP and LNCaP cells were treated with the synthetic androgen R1881, consistent with previous data from this locus (*17*) (**Figure 2.11**). *PCAT-1* and *PCAT-14* also showed cancer-specific upregulation when tested on a panel of matched tumor-normal samples (**Figure 2.12**).

Of note, *PCAT-114*, which ranks as the #5 best outlier, just ahead of *ERG* (**Figure 2.8b**), appears as part of a large, >500 kb locus of expression in a gene desert in Chr2q31. We termed this region Second Chromosome Locus Associated with Prostate-1 (SChLAP1) (**Figure 2.13**). Careful analysis of the SChLAP1 locus revealed both

discrete transcripts and intronic transcription, highlighting this region as an intriguing aspect of the prostate cancer transcriptome.

DISCUSSION

This study represents the largest RNA-Seq analysis to date and the first to comprehensively analyze a common epithelial cancer from a large cohort of human tissue samples. As such, our study has adapted existing computational tools intended for small-scale use (*3*) and developed new methods in order to distill large numbers of transcriptome datasets into a single consensus transcriptome assembly that reflects a coherent biological picture.

Among the numerous uncharacterized ncRNA species detected by our study, we have focused on 121 prostate cancer-associated PCATs, which we believe represent a set of uncharacterized ncRNAs that may have important biological functions in this disease. In this regard, these data contribute to a growing body of literature supporting the importance of unannotated ncRNA species in cellular biology and oncogenesis (*6-12*), and broadly our study confirms the utility of RNA-Seq in defining functionally-important elements of the genome (*2-4*). These PCATs reflect a unique aspect of prostate cancer biology, as many of them exhibit prostate-specific expression (**Figure 2.10**), but this is typically not due to androgen receptor signaling. As such, PCATs and other similarly lineage-restricted lncRNAs in different cancers may represent rationale candidates for biomarker development, since they offer high specificity for distinguishing cancers. Furthermore, elucidation of the mechanisms responsible for this lineage-restricted expression will likely uncover novel aspects of cancer biology.

MATERIALS AND METHODS

RNA-Seq Library Preparation.

Next generation sequencing of RNA was performed on 21 prostate cell lines, 20 benign adjacent prostates, 47 localized tumors, and 14 metastatic tumors according to Illumina's protocol using 2ug of RNA. RNA integrity was measured using an Agilent 2100 Bioanalyzer, and only samples with a RIN score >7.0 were advanced for library generation. 2μg total RNA was selected for polyA+ RNA using Sera-Mag oligo(dT) beads (Thermo Scientific), fragmented with the Ambion Fragmentation Reagents kit (Ambion, Austin, TX). cDNA synthesis, end-repair, A-base addition, and ligation of the Illumina PCR adaptors (single read or paired-end where appropriate) were performed according to Illumina's protocol as previously described (*33-35*). Libraries were then size-selected for 250-300 bp cDNA fragments on a 3.5% agarose gel and PCR-amplified using Phusion DNA polymerase (Finnzymes) for 15 –18 PCR cycles. PCR products were then purified on a 2% agarose gel and gel-extracted. Library quality was credentialed by assaying each library on an Agilent 2100 Bioanalyzer of product size and concentration. Libraries were sequenced as 36-45mers on an Illumina Genome Analyzer I or Genome Analyzer II flowcell according to Illumina's protocol. All single read samples were sequenced on a Genome Analyzer I, and all paired-end samples were sequenced on a Genome Analyzer II.

Bioinformatic analyses

To achieve an *ab initio* prediction of the prostate cancer transcriptome we leveraged existing publicly tools for mapping, assembly, and quantification of transcripts,

and supplemented these tools with additional informatics filtering steps to enrich the results for the most robust transcript predictions (**Figure 2.1**). Transcripts were then identified and classified by comparing them against gene annotation databases (**Figure 2.1**). Details of the bioinformatics analyses are provided below.

Mapping reads with TopHat: Reads were aligned using TopHat v1.0.13 (Feb 5, 2010) (*19*), a gapped aligner capable of discovering splice junctions *ab initio*. Briefly, TopHat aligns reads to the human genome using Bowtie(*36*) to determine a set of "coverage islands" that may represent putative exons. TopHat uses these exons as well as the presence of GT-AG genomic splicing motifs to build a second set of reference sequences spanning exon-exon junctions. The unmapped reads from the initial genome alignment step are then remapped against this splice junction reference to discover all the junction-spanning reads in the sample. TopHat outputs the reads that successfully map to either the genome or the splice junction reference in SAM format for further analysis. For this study we used a maximum intron size of 500kb, corresponding to over 99.98% of RefSeq introns (*37*). This may limit sensitivity to detect read-through chimeras at the expense of increased specificity of splice junction calls. For sequencing libraries we determined the insert size using an Agilent 2100 Bioanalyzer prior to data analysis, and found that this insert size agreed closely with software predictions (data not shown). An insert size standard deviation of 20 bases was chosen in order to match the most common band size cut from gels during library preparation.

In total, 1.723 billion fragments were generated from 201 lanes of sequencing on the Illumina Genome Analyzer and Illumina Genome Analyzer II. Reads were mapped to the human genome (hg18) downloaded from the UCSC genome browser website (*38,*

39). We obtained 1.418 billion unique alignments including 114.4 million splice junctions for use in transcriptome assembly. Reads with multiple alignments with less than two mismatches were discarded.

Ab initio assembly and quantification with Cufflinks: Aligned reads from TopHat were assembled into sample-specific transcriptomes with Cufflinks version 0.8.2 (March 26, 2010) (*3*). Cufflinks assembles exonic and splice-junction reads into transcripts using their alignment coordinates. To limit false positive assemblies we used a maximum intronic length of 300kb, corresponding to the 99.93% percentile of known introns. After assembling transcripts, Cufflinks computes isoform-level abundances by finding a parsimonious allocation of reads to the transcripts within a locus. We filtered transcripts with abundance less than 15% of the major transcript in the locus, and minor isoforms with abundance less than 5% of the major isoform. Default settings were used for the remaining parameters.

The Cufflinks assembly stage yielded a set of transcript annotations for each of the sequenced libraries. We partitioned the transcripts by chromosome and used the Cuffcompare utility provided by Cufflinks to merge the transcripts into a combined set of annotations. The Cuffcompare program performs a union of all transcripts by merging transcripts that share all introns and exons. The 5' and 3' exons of transcripts were allowed to vary by up to 100nt during the comparison process.

Distinguishing transcripts from background signal: Cuffcompare reported a total of 8.25 million distinct transcripts. Manual inspection of these transcripts in known protein coding gene regions suggested that most of the transcripts were likely to be poor quality reconstructions of overlapping larger transcripts. Also, many of the transcripts

were unspliced and had a total length smaller than the size selected fragment length of approximately ~250nt. Furthermore, many of these transcripts were only present in a single sample. Therefore, we reasoned that the vast majority of predicted loci probably corresponded to stochastic transcriptional "noise", subtle amounts of genomic DNA present in the sample, or artifacts due to errors in read alignment.

Here, we designed a statistical classifier to predict transcripts over background signal with the goal of identifying highly recurrent transcripts that may be altered in prostate cancer. This approach rests on the premise that a manually curated gene database could represent a reliable set of "true" positives on which to train a statistical learning algorithm. We used the AceView annotations (*20*) which we believe have an adequate representation of low abundance ncRNA transcripts that may be cell-type specific in addition to protein coding gene isoforms. For each transcript predicted by Cufflinks we collected the following statistics: length (bp), number of exons, recurrence (number of samples in which the transcript was predicted), 95th percentile of abundance (measured in Fragments per Kilobase per Million reads (FPKM)) across all samples, and uniqueness of genomic DNA harboring the transcript transcript (measured using the Rosetta uniqueness track from UCSC (*40*)). Using this information, we used recursive partitioning and regression trees in R (package rpart) to predict, for each transcript, whether its expression patterns and structural properties resembled those of annotated genes. Classification was performed independently for each chromosome in order to incorporate the effect of gene density variability on expression thresholds. Transcripts that were not classified as annotated genes were discarded, and the remainder were subjected to additional analysis and filtering steps. By examining the decision tree results

we observed that the 95th percentile of expression across all samples as well as the recurrence of each transcript were most frequently the best predictors of expressed versus background transcripts (**Figure 2.1b,c**).

Refinement of transcript fragments: The statistical classifier predicted a total 2.88 million (34.9%) transcript fragments as "expressed" transcripts. We then developed a program to extend and merge intron-redundant transcripts to produce a minimum set of transcripts that could possibly explain the assemblies produced by Cufflinks. By merging transcripts in this manner we relinquished the ability to detect some types of alternative TSSs, but drastically reduced the total number of independent transcripts in our assembly. We believe merging all intron-redundant transcripts is suitable for qualitative detection of transcriptionally active regions, but more sophisticated methods would be necessary for the study of alternative splicing, alternative TSSs, and alternative poly-adenylation site usage within well characterized regions. The merging step produced a total of 123,554 independent transcripts. We then re-computed transcript abundance levels for these revised transcripts in Reads per Kilobase per Million (RPKM) units. These expression levels were used for the remainder of the study.

We applied several additional filtering steps to isolate the most robust transcripts. First, we discarded transcripts with a total length less than 200nt. Our size selection protocol isolates RNA molecules larger than this, and small RNA sequencing protocols would likely be needed to quantify smaller molecules with high confidence. Second, we discarded single exon transcripts with greater than 75% overlap to another longer transcript. We believe many of these are produced from unspliced pre-mRNA molecules and do not represent functional RNA products. Third, we removed transcripts that lacked

a completely unambiguous genomic DNA stretch of at least 40nt. We measured genomic uniqueness using the Rosetta uniqueness track downloaded from the UCSC genome browser website. We believe transcripts spanning poorly mappable regions are more likely to occur due to mapping artifacts and the availability of longer reads would alleviate the need for this filtering step. Finally, we retained transcripts that were not present in at least 5% of our cohort (>5 samples) at more than 5.0 RPKM. It is possible that certain subtypes of prostate cancer may express highly specific transcripts, and future studies to characterize these transcripts could provide additional insight into the biology of tumor subtypes.

In certain instances we observed transcripts that were interrupted by poorly mappable genomic regions. Additionally, for low abundance genes we observed fragmentation due to the lack of splice junction or paired-end read evidence needed to connect nearby fragments. We reasoned that expression profiles of these fragmented transcripts should be highly correlated. To demonstrate this, we measured the difference in the Pearson correlation between expression of randomly chosen exons on the same transcript versus expression of spatially proximal exons on different transcripts. We found that in our cohort, a Pearson correlation >0.8 had a positive predictive value (PPV) of >95% for distinct exons to be part of the same transcript. Using this criteria, we performed hierarchical agglomerative clustering to extend transcript fragments into larger transcriptional units. Pairs of transcripts further than 100kb apart, transcripts on opposite strands, and overlapping transcripts were not considered for clustering. Groups of correlated transcripts were merged, and introns <40nt in length were removed.

Comparison with gene annotation databases: The 44,534 transcripts produced by the bioinformatics pipeline were classified by comparison with a comprehensive list of "annotated" transcripts from UCSC, RefSeq, ENCODE, Vega, and Ensembl. First, transcripts corresponding to processed pseudogenes were separated. This was done to circumvent a known source of bias in the TopHat read aligner. TopHat maps reads to genomic DNA in its first step, predisposing exon-exon junction reads to align to their spliced retroposed pseudogene homologues. Future improvements to splice junction mapping algorithms could improve accuracy and eliminate this bias.

Next, transcripts with >1bp of overlap with at least one annotated gene on the correct strand were designated "annotated", and the remainder were deemed "unannotated". Transcripts with no overlap with protein coding genes were subdivided into intronic, intergenic, or partially intronic antisense categories based on their relative genomic locations.

Informatics filtering of unspliced pre-mRNA isoforms: We observed a significant increase in the percentage of intronic transcripts in our assembly relative to known intronic ncRNAs. This led us to observe that in many cases unspliced pre-mRNAs may appear at sufficient levels to escape the filtering steps employed by Cufflinks during the assembly stage. We then removed intronic and antisense transcripts that were correlated (Pearson correlation >0.5) to their overlapping protein coding genes in order to better approximate the true number of intronic or antisense transcripts in the transcriptome. This effectively removed transcripts within genes such as *PCA3* and *HPN* that were obvious pre-mRNA artifacts, while leaving truly novel intronic transcripts – such as those within *FBXL7* and *CDH13* – intact. While this informatics approach may conservatively

predict intronic and antisense transcription, it enriched our results for the most independent transcripts.

In effect, these steps produced a consensus set of 35,415 transcripts supporting long poly-adenylated RNA molecules in human prostate tissues and cell lines. Per chromosome transcript counts closely mirrored known transcript databases (**Table 2.2**), suggesting that the informatics procedures employed here compensate well for gene density variability across chromosomes. Overall we detected a similar number of transcripts as present in the either the RefSeq or UCSC databases (*37*).

Coding potential analysis: To analyze coding potential, we extracted the DNA sequences for each transcript and searched for open reading frames (ORFs) using the `txCdsPredict` program from the UCSC source tool set (*39*). This program produces a score corresponding to the protein coding capacity of a given sequence, and scores >800 are ~90% predictive of protein coding genes. We used this threshold to count transcripts with coding potential, and found only 5 of 6,641 unannotated genes with scores >800, compared with 1,669 of 25,414 protein coding transcripts. Additionally, we observed that protein coding genes possess consistently longer ORFs than either unannotated or annotated ncRNA transcripts, suggesting that these unannotated transcripts represent ncRNAs (**Figure 2.5**).

Separation of transcripts into repetitive and non-repetitive categories: To separate transcripts into "repeat" and "non-repeat" transcripts, we extracted the genomic DNA corresponding to the transcript exons and calculated the fraction of repeat-masked nucleotides in each sequence. For the designation of repeat classes, we used the RepMask 3.2.7 UCSC Genome Browser track (*39*). After examining various thresholds,

we chose to separate transcripts containing >25% repetitive DNA for the purposes of the ChIP-seq and conservation analyses discussed below.

Conservation Analysis: We used the SiPhy package (*41*) to estimate the locate rate of variation (ω) of all non-repetitive transcript exons across 29 placental mammals. The program was run as described on the SiPhy website: (http://www.broadinstitute.org/genome_bio/siphy/documentation.html).

ChIP-Seq datasets: We employed published ChIP-Seq datasets for H3K4me1, H3K4me2, H3K4me3, Acetylated H3, Pan-H3, and H3K36me3 generated in our laboratory (*25*). These data are publically available through the NCBI Geo Omnibus (GEO GSM353632). We analyzed the raw ChIP-Seq data using MACS (*42*) (H3K4me1, H3K4me2, H3K4me3, Acetylated H3, and Pan-H3) or SICER (*43*) (H3K36me3) peak finder programs using default settings. These peak finders were used based upon their preferential suitability to detect different types of histone modifications (*44*). The H3K4me3-H3K36me3 chromatin signature used to identify lncRNAs was determined from the peak coordinates by associating each H3K4me3 peak with the closest H3K36me3-enriched region up to a maximum of 10kb away. The enhancer signature (H3K4me1 but not H3K4me3) was determined by subtracting the set of overlapping H3K4me3 peaks from the entire set of H3K4me1 peaks. These analyses were performed with the bx-python libraries distributed as part of the Galaxy bioinformatics infrastructure (*45*).

Differential Expression Analysis: To predict differentially expressed transcripts we first prepared a matrix of log-transformed, normalized RPKM expression values. We used the base 2 logarithm after adding 0.1 to all RPKM values. The data were first

centered by subtracting the median expression of the benign samples for each transcript. We then used the Significance Analysis of Microarrays (SAM) method (*46*) with 250 permutations of the Tusher et al. S0 selection method to predict differentially expressed genes. We chose a delta value corresponding to the 90th percentile FDR desired for individual analyses. The MultiExperiment Viewer application (*47*) was used to run SAM and generate heatmaps. We confirmed that the results matched expected results through comparison with microarrays and known prostate cancer biomarkers (data not shown).

Outlier Analysis: A modified COPA analysis (*18, 48*) was performed on the 81 tissue samples in the cohort. RPKM expression values were used and shifted by 1.0 in order to avoid division by zero. The COPA analysis had the following steps: 1) gene expression values were median centered, using the median expression value for the gene across the all samples in the cohort. This sets the gene's median to zero. 2) The median absolute deviation (MAD) was calculated for each gene, and then each gene expression value was scaled by its MAD. 3) The 80, 85, 90, 98 percentiles of the transformed expression values were calculated for each gene and the average of those four values was taken. Then, genes were rank ordered according to this "average percentile", which generated a list of outliers genes arranged by importance. 4) Finally, genes showing an outlier profile in the benign samples were discarded. Six novel transcripts ranked as both outliers and differentially-expressed genes in our analyses. These six were manually classified either as differentially-expressed or outlier status based on what each individual's distribution across samples suggested.

FIGURES

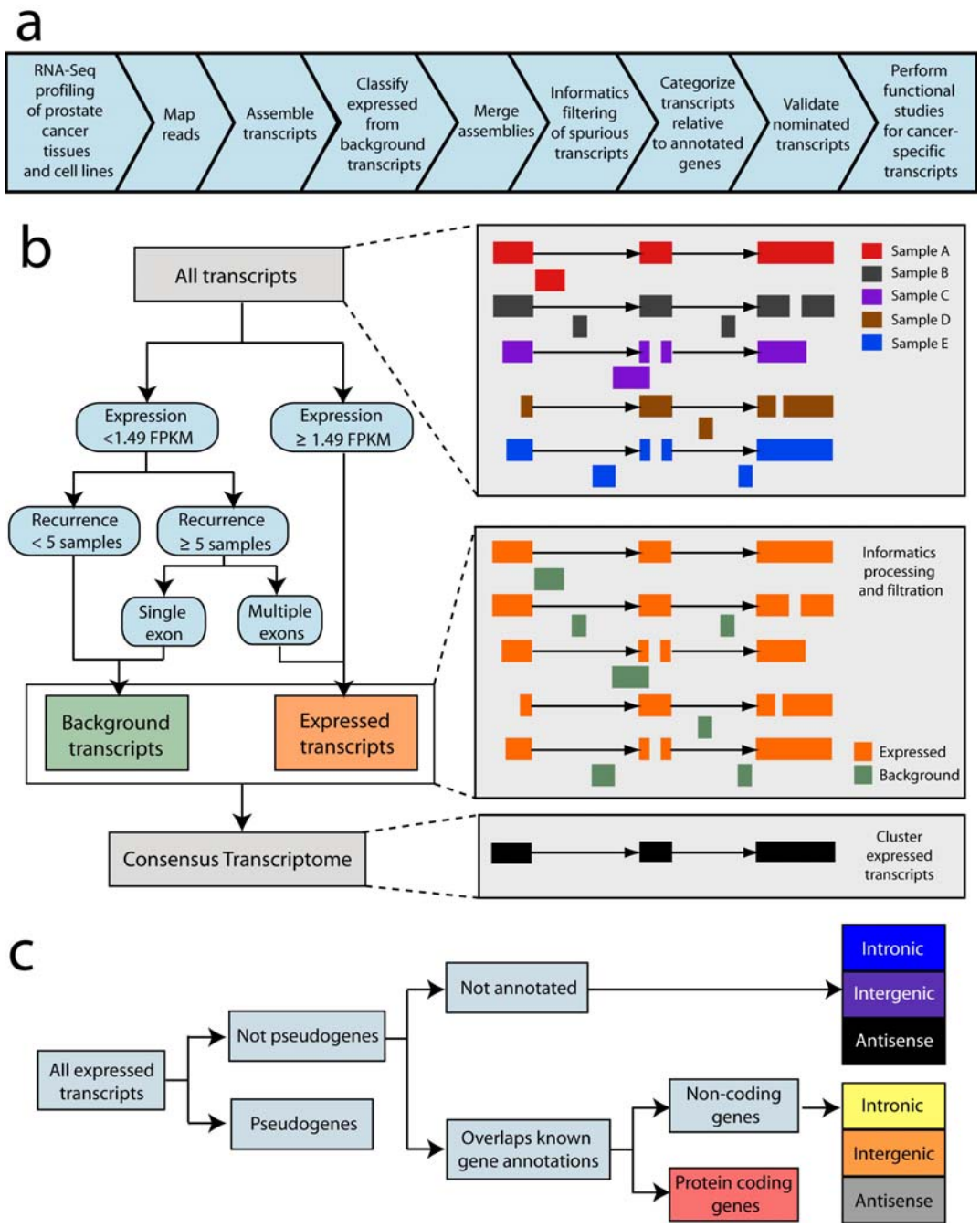


Figure 2.1: Analysis of transcriptome data for the detection of unannotated transcripts. (**a**) A schematic overview of the methodology employed in this study. (**b**) A graphical representation showing the bioinformatics filtration model used to merge individual transcriptome libraries into a single consensus transcriptome. (**c**) Following informatic processing and filtration of the sequencing data, transcripts were categorized in order to identify unannotated ncRNAs.

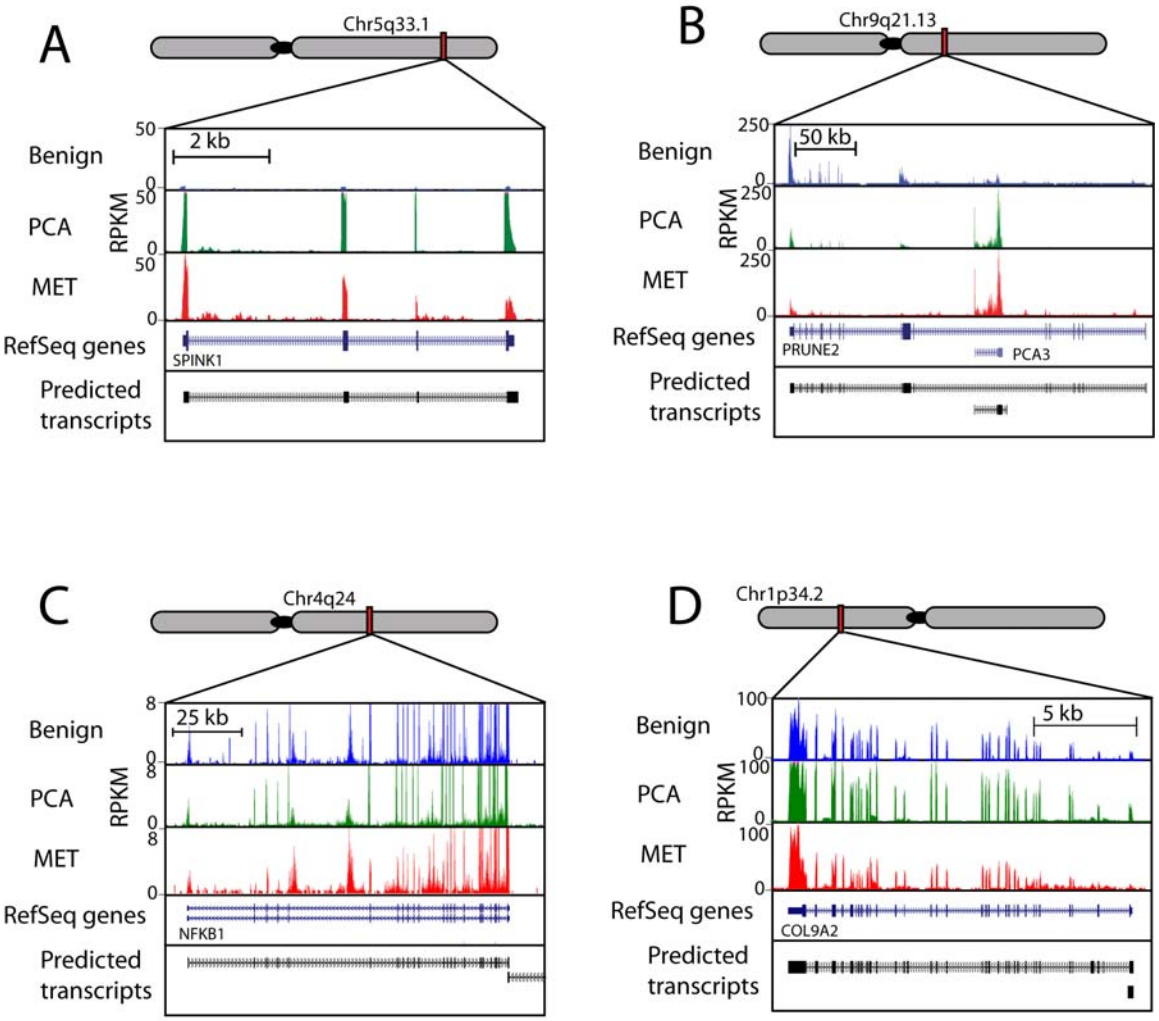


Figure 2.2: Transcript assembly of known genes. We employed *ab initio* transcript assembly on prostate transcriptome sequencing data to reconstruct the known prostate transcriptome. Four examples of transcriptome reassembly are displayed above. (**a**) *SPINK1*, a biomarker for prostate cancer. (**b**) *PRUNE2* with the *PCA3* non-coding RNA within its intronic regions. Note that *PCA3* is a prostate cancer biomarker while *PRUNE2* is not. The two transcripts remain independent. (**c**) *NFKB1*. (**d**) *COL9A2*.

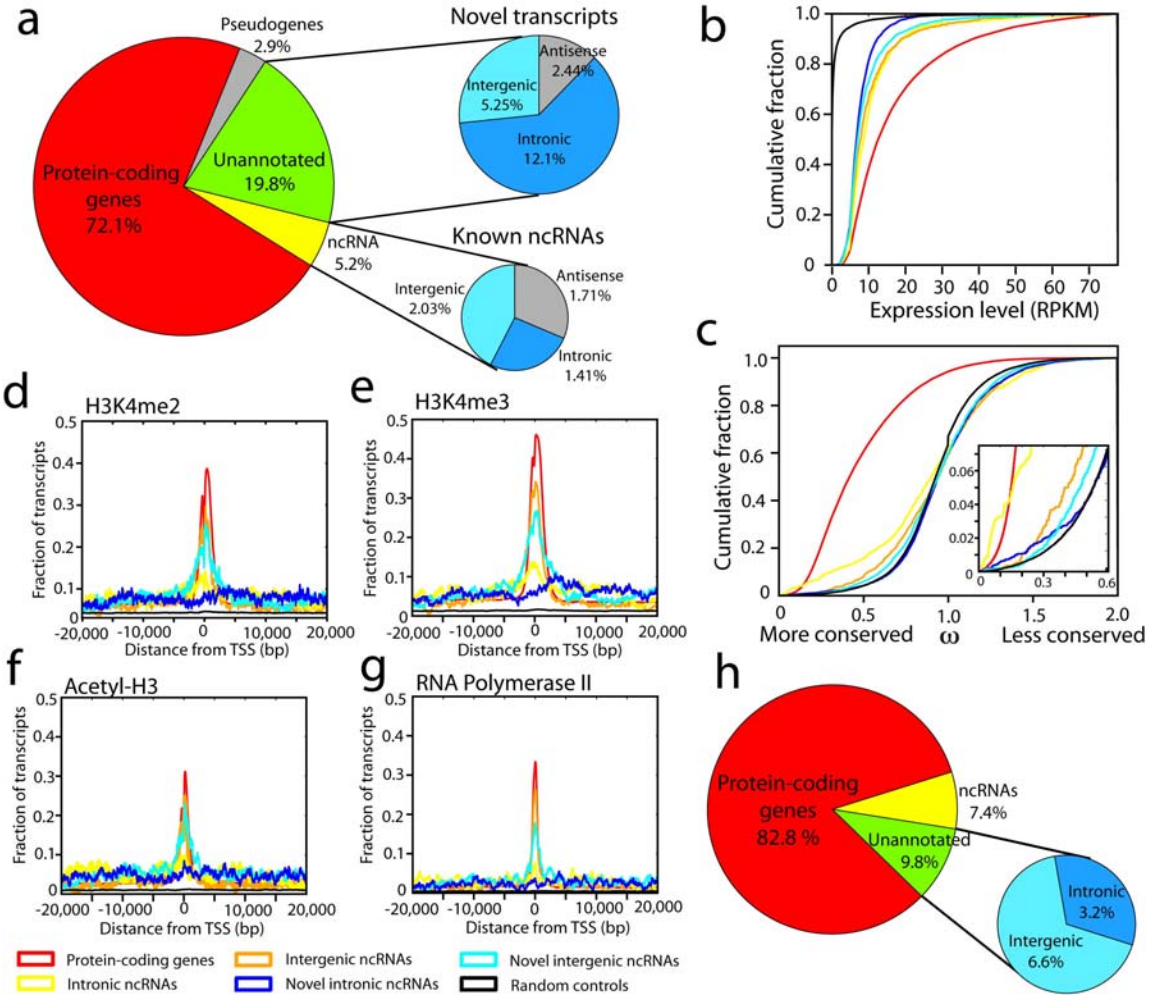
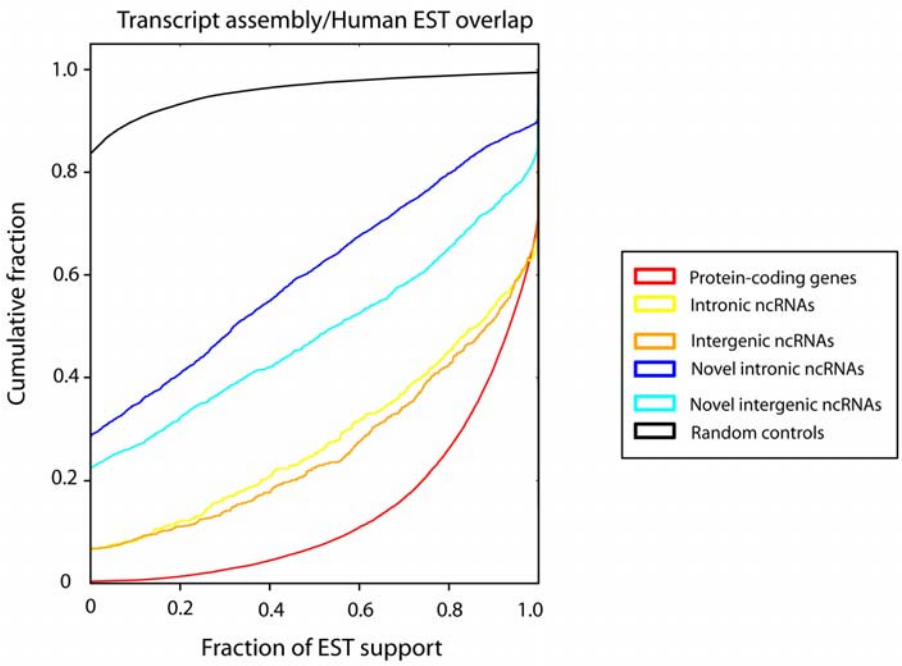


Figure 2.3: Prostate cancer transcriptome sequencing reveals dysregulation of novel transcripts. (**a**) A global overview of transcription in prostate cancer. The left pie chart displays transcript distribution in prostate cancer. The upper and lower right pie charts display unannotated or annotated ncRNAs, respectively categorized as sense transcripts (intergenic and intronic) and antisense transcripts. (**b**) A line graph showing that unannotated transcripts are more highly expressed (RPKM) than control regions. Negative control intervals were generated by randomly permuting the genomic positions of the transcripts. (**c**) Conservation analysis comparing unannotated transcripts to known genes and intronic controls shows a subtle degree of purifying selection among unannotated transcripts. The insert on the right shows an enlarged view. (**d-g**) Intersection plots displaying the fraction of unannotated transcripts enriched for H3K4me2 (**d**), H3K4me3 (**e**), Acetyl-H3 (**f**) or RNA polymerase II (**g**) at their transcriptional start site (TSS) using ChIP-Seq and RNA-Seq data for the VCaP prostate cancer cell line. The legend for these plots (**b-g**) is shared and located below (**f**) and (**g**). (**h**) A pie chart displaying the distribution of differentially expressed transcripts in prostate cancer (FDR < 0.01).

A



B

Category	Transcripts	EST hits	Percent of all ESTs	ESTs per Transcript
Annotated proteins	25550	4564852	56.43%	178.6634834
Intergenic ncRNA	720	32891	0.41%	45.68194444
Intronic ncRNA	500	57015	0.70%	114.03
Unannotated intergenic ncRNA	1859	17478	0.22%	9.40182894
Unannotated intronic ncRNA	4285	197142	2.44%	46.00746791
Total	**35415**	**4869378**		

Figure 2.4: Analysis of EST support for novel transcripts. ESTs from the UCSC database table "Human ESTs" were used to evaluate the amount of overlap between ESTs and novel transcripts. (**a**) A line graph showing the fraction of genes whose transcripts are supported by a particular fraction of ESTs. Over 20% of novel transcripts have no support by ESTs. (**b**) A table displaying the number of ESTs supporting each class of transcripts. Percent of all ESTs was calculated using the total number of annotated ESTs (8,089,356), not the total number of observed ESTs in the transcriptome data (4,869,378).

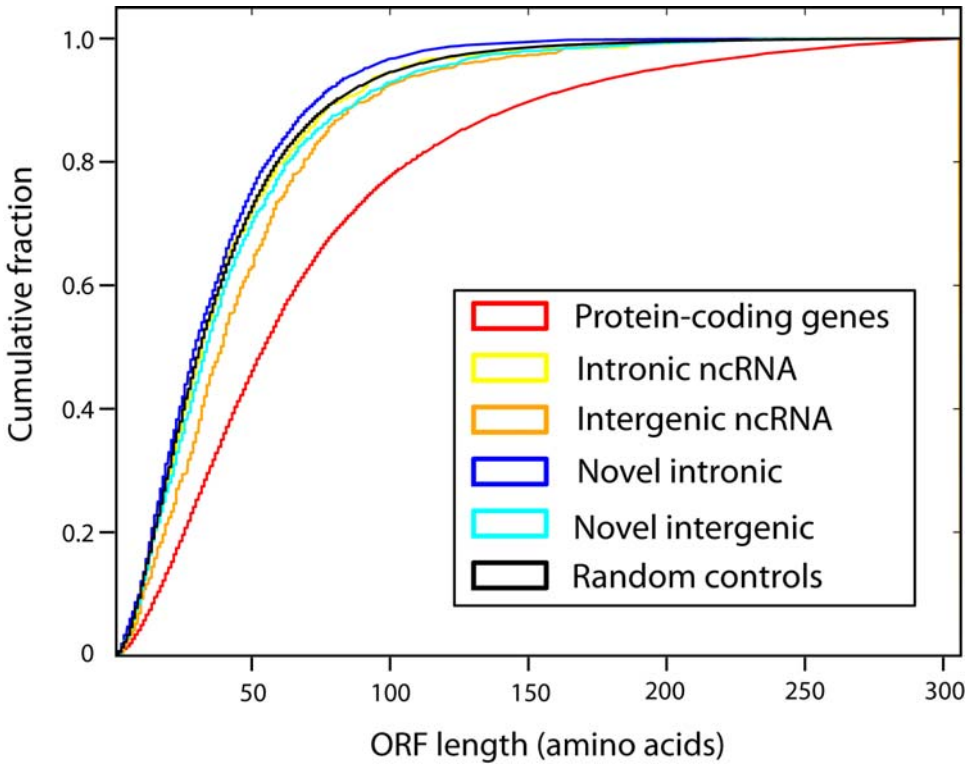


Figure 2.5: Analysis of coding potential of unannotated transcripts. DNA sequences for each transcript were extracted and searched for open reading frames (ORFs) using the txCdsPredict program from the UCSC source tool set. Using these methods, novel transcripts showed poor protein-coding capacity compared to protein-coding genes, and novel transcripts scored similarly to known ncRNAs, suggesting that the vast majority of unannotated transcripts in prostate cancer represent ncRNAs.

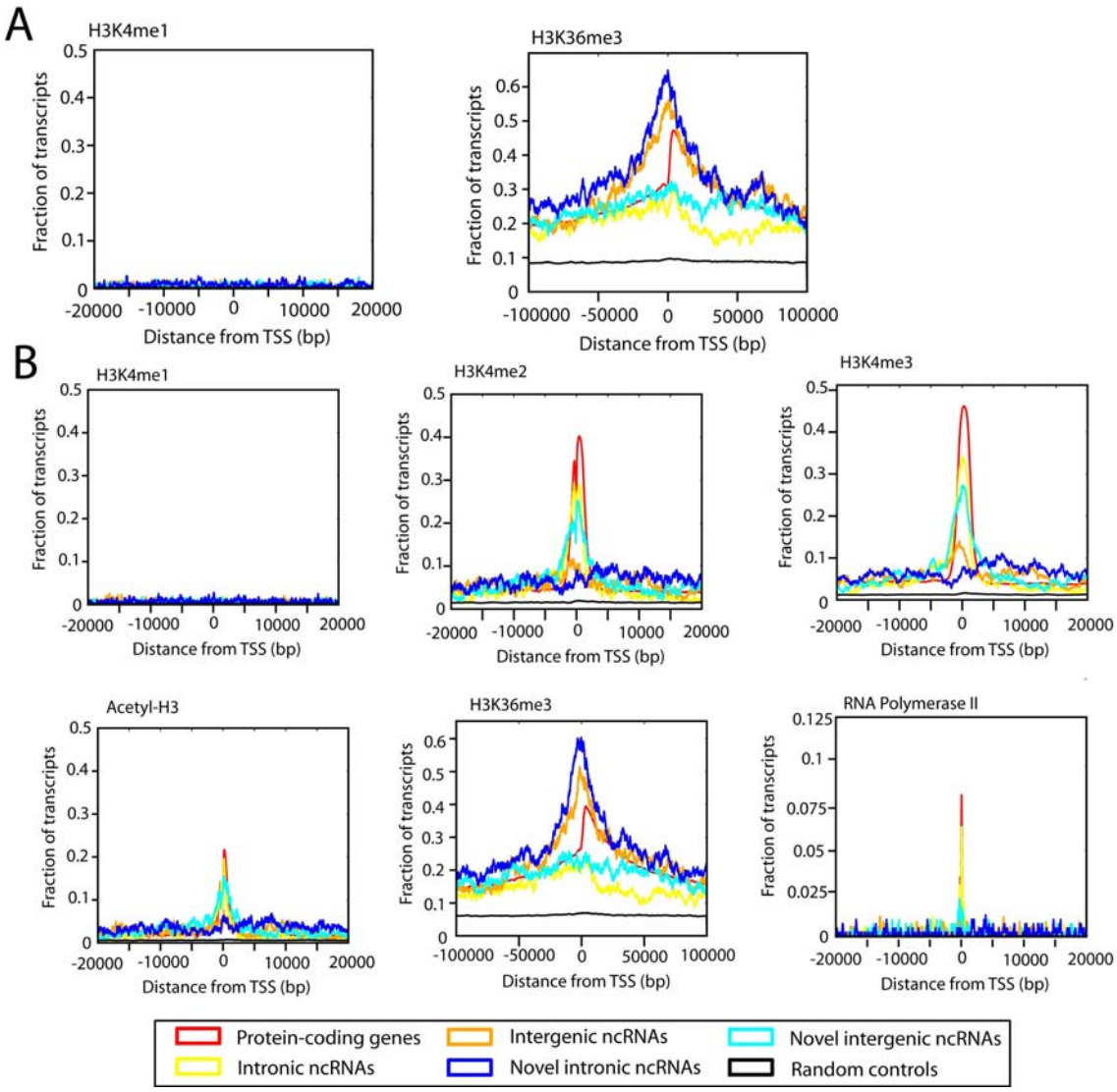


Figure 2.6: Overlap of unannotated transcripts with ChIP-Seq data. Perviously published ChIP-Seq data for VCaP and LNCaP prostate cancer cells were intersected with unannotated prostate cancer transcripts and annotated control genes. Non-coding RNAs were divided into intergenic and intronic. (**a**) H3K4me1 (left) and H3K36me3 (right) in VCaP. H3K4me1 was not enriched in any samples. (b) ChIP-seq data in LNCaP cells. Overlap between transcription start sites and H3K4me2, H3K4me3, Acetyl-H3, H3K36me3, and RNA polymerase II. Intergenic ncRNAs tend to show higher enrichment than intronic ncRNAs for both annotated and unannotated transcripts. H3K4me1 was not enriched in any samples, and H3K36me3 showed enrichment over a broader distance.

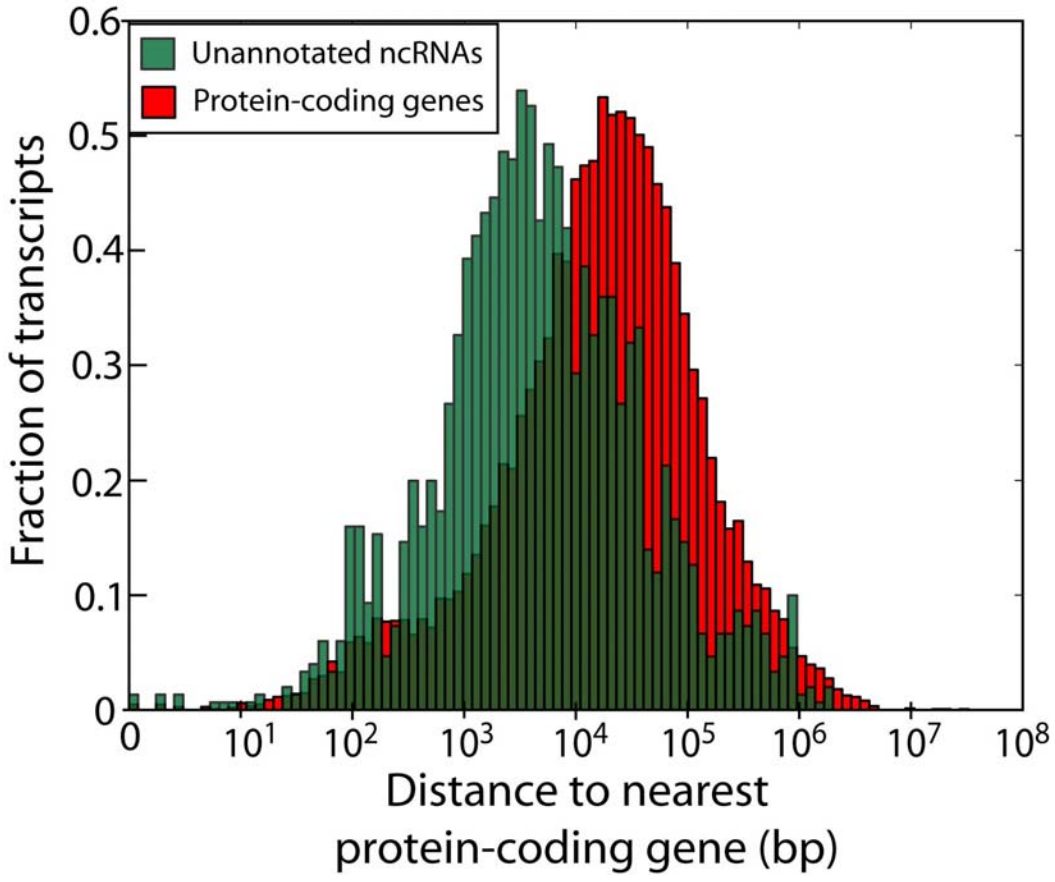


Figure 2.7: Intergenic unannotated genes are located halfway between annotated genes. The distance between intergenic unannotated ncRNAs to the closest protein-coding gene was calculated, forming a normal distribution around a mean of 4,292 kb. For comparison the distance between protein-coding genes was likewise calculated, forming a normal distribution around a mean of 8,559 kb. These data suggest that, on average, novel intergenic genes are located approximately halfway between protein-coding genes.

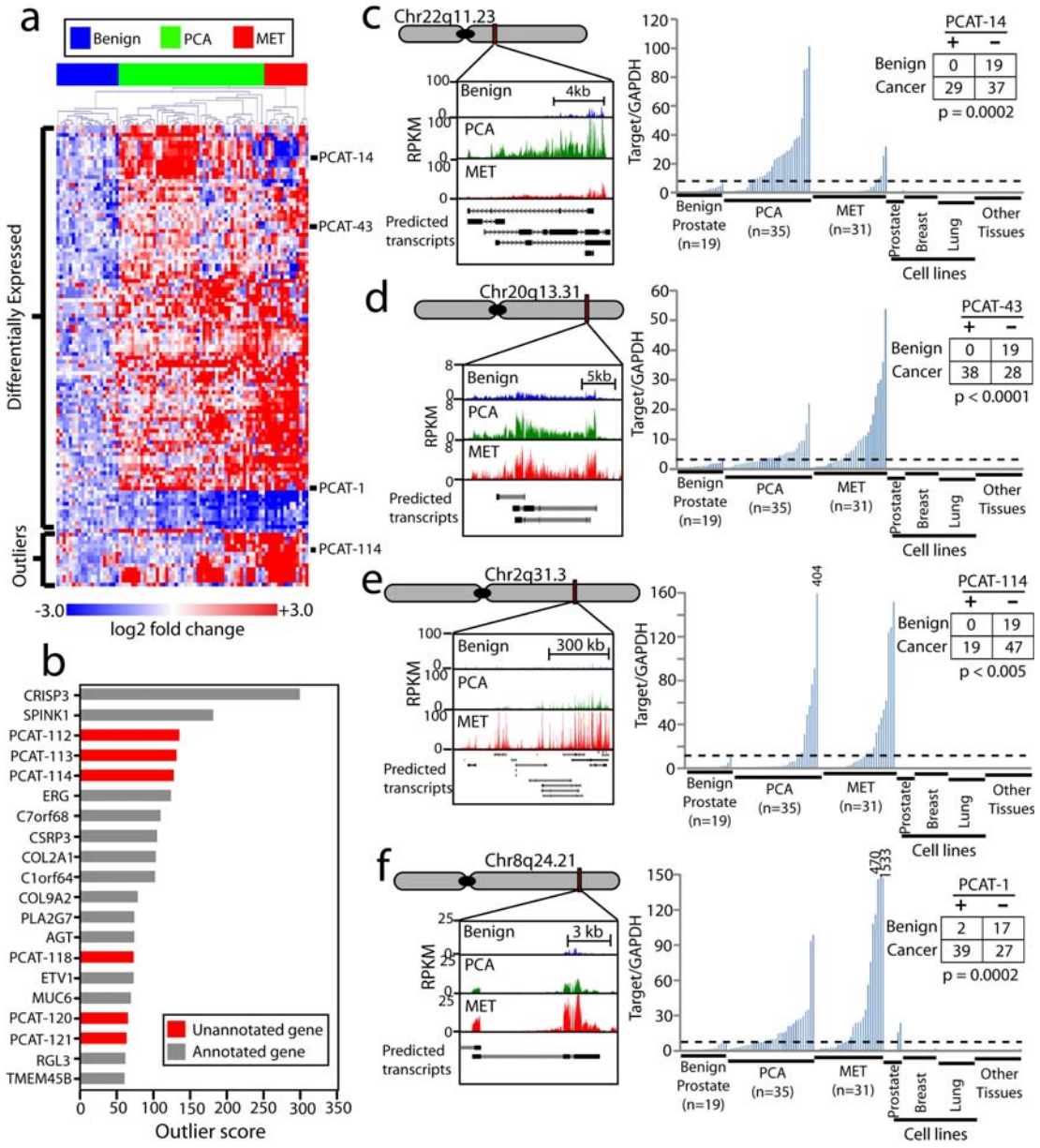


Figure 2.8: Unannotated intergenic transcripts differentiate prostate cancer and benign prostate samples. (**a**) Unsupervised clustering analyses of differentially-expressed or outlier unannotated intergenic transcripts clusters benign samples, localized tumors, and metastatic cancers. Expression is plotted as log2 fold change relative to the median of the benign samples. The four transcripts detailed in this study are indicated on the side. (**b**) Cancer outlier expression analysis for the prostate cancer transcriptome ranks unannotated transcripts prominently. (**c-f**) qPCR on an independent cohort of prostate and non-prostate samples (Benign (n=19), PCA (n=35), MET (n=31), prostate cell lines (n=7), breast cell lines (n=14), lung cell lines (n=16), other normal samples

(n=19)) measures expression levels of four nominated ncRNAs—*PCAT-1*, *PCAT-43*, *PCAT-114*, and *PCAT-14*—upregulated in prostate cancer. Inset tables on the right quantify "positive" and "negative" expressing samples using the cut-off value (shown as a black dotted line). Statistical significance was determined using a Fisher's exact test. (**c**) *PCAT-14*. (**d**) *PCAT-43*. (**e**) *PCAT-114* (SChLAP1). (**f**) *PCAT-1*. qPCR analysis was performed by normalizing to *GAPDH* and the median expression of the benign samples.

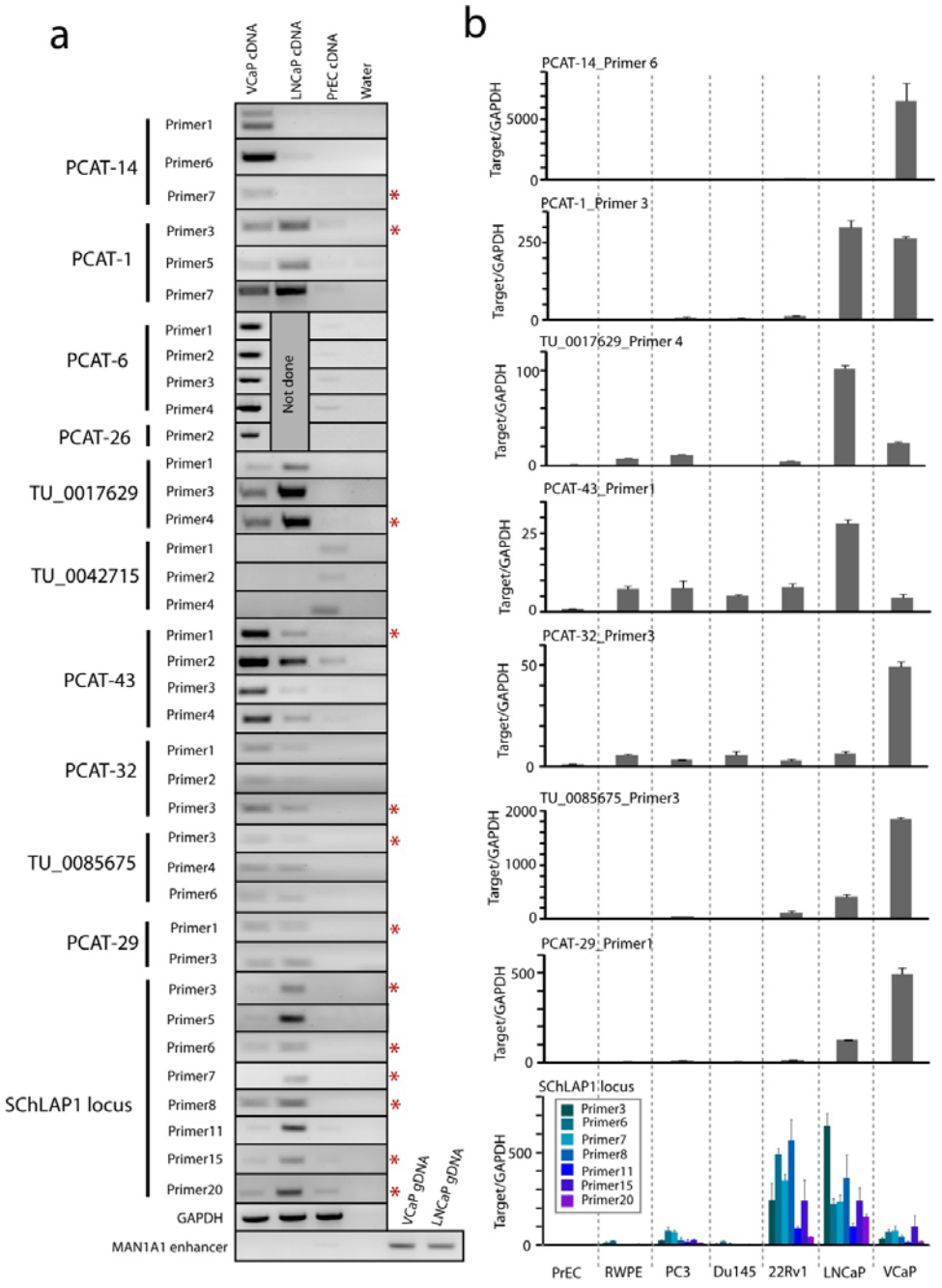


Figure 2.9: Validation of novel transcripts in prostate cell lines. 11/14 unannotated transcripts selected for validation by RT-PCR and qPCR were confirmed in cell line models. (**a**) RT-PCR gels showing expected bands for the 11 transcripts that validated. *GAPDH* and a genomic region upstream of *MAN1A1* were used as positive and negative control, respectively. Representative PCR bands from all validated transcripts were Sanger-sequenced. Transcripts were confirmed by multiple independent primer pairs (See table S10) (**b**) Representative qPCR results using primers selected from (**a**) The primers used in (**b**) are indicated by a red asterisk in (**a**). All error bars are mean ± S.E.M.

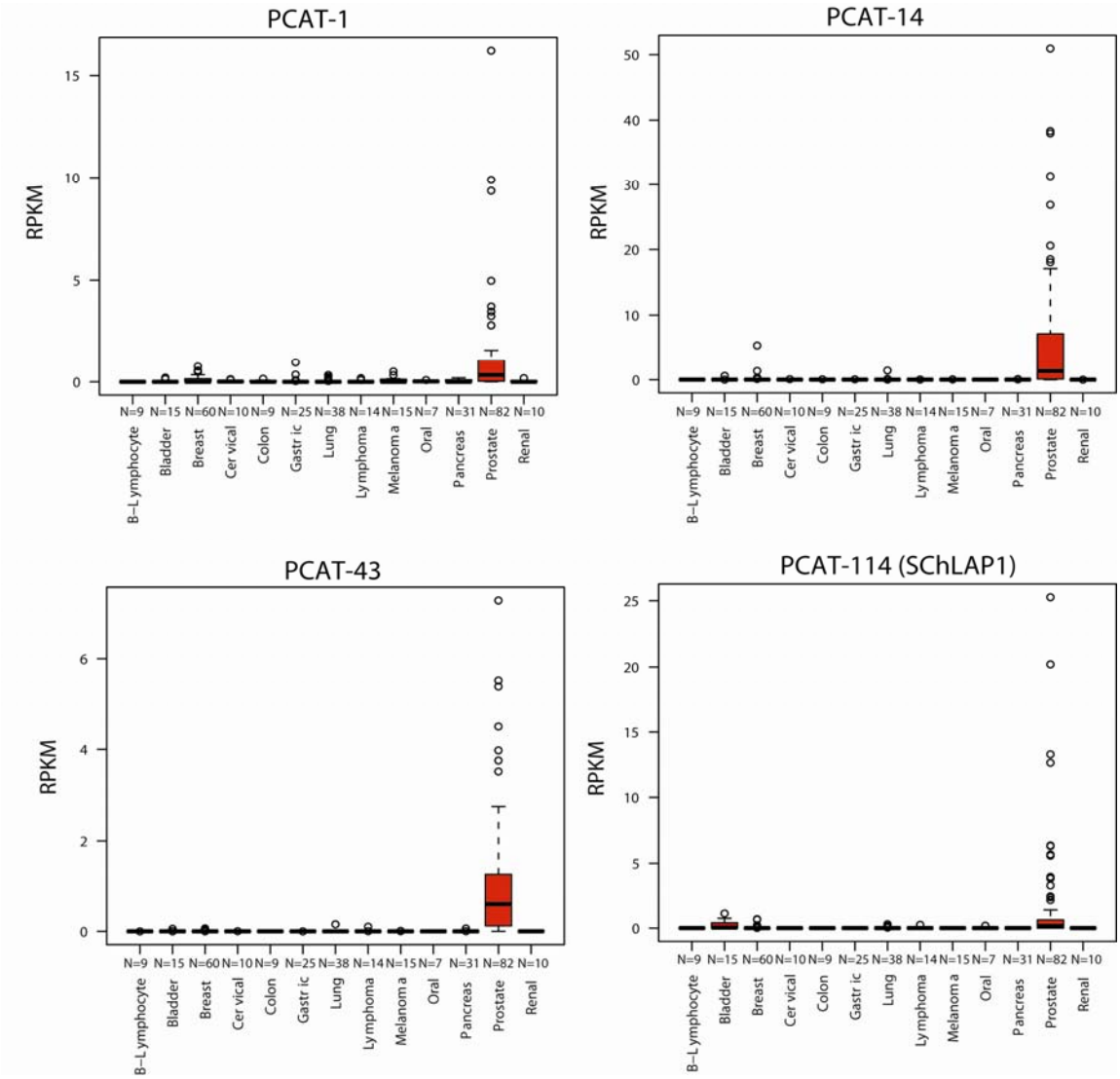


Figure 2.10: Prostate-specificity of PCATs in RNA-Seq compendium. RNA-Seq was performed on a compendium of 325 cell lines and tissues, and gene expression was quantified in RPKM. Evaluation of expression levels for *PCAT-1*, *PCAT-14*, *PCAT-114*, or *PCAT-43* indicates prostate-specific expression of these transcripts.

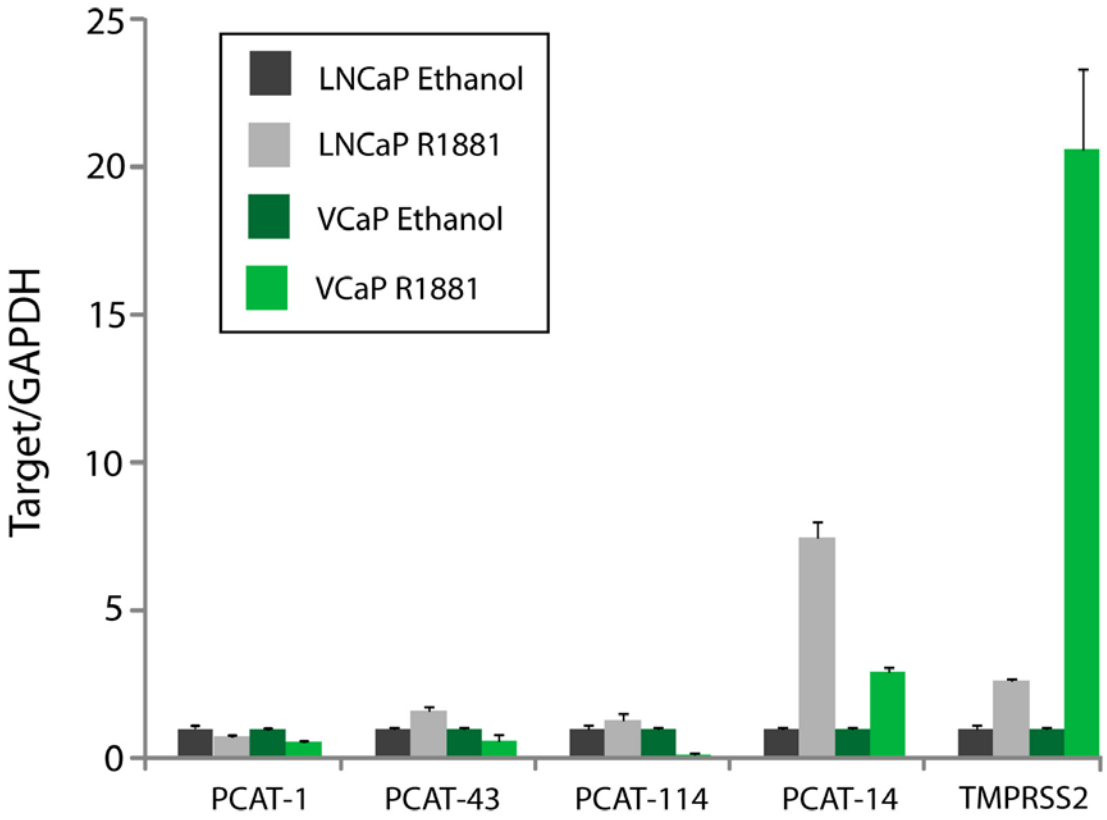


Figure 2.11: *PCAT-14* is upregulated by androgen signaling. VCaP and LNCaP cells were grown in charcoal-stripped serum for 48 hours prior to treatment with 5nM R1881 or vehicle (ethanol) control. 24 hours after treatment, cells were harvested, total RNA and cDNA were generated, and cells were assayed for expression levels of unannotated non-coding RNAs. No consistent change is seen in *PCAT-1*, *PCAT-43* and *PCAT-114* expression upon addition of R1881. However, *PCAT-14* shows consistent upregulation in both VCaP and LNCaP cells treated with R1881. *TMPRSS2* serves as a control androgen-regulated gene. All experiments are normalized to *GAPDH* and the relative expression of the corresponding ethanol-treated sample. All error bars are mean ± S.E.M.

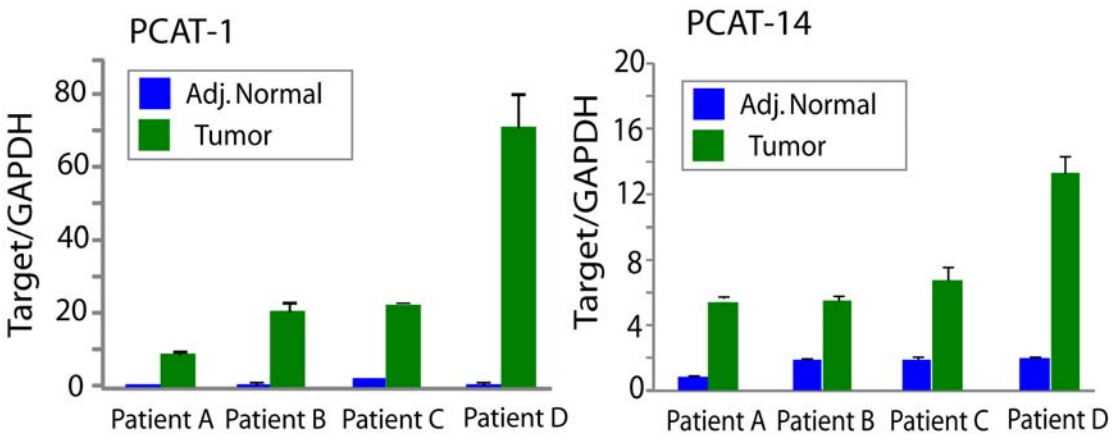


Figure 2.12: *PCAT-1* and *PCAT-14* are upregulated in matched tumor tissues. Four matched tumor-normal patient tissue samples were assayed for *PCAT-1* and *PCAT-14* expression by qPCR. Expression levels were normalized to *GAPDH* and to the median of the benign samples. All error bars are mean ± S.E.M.

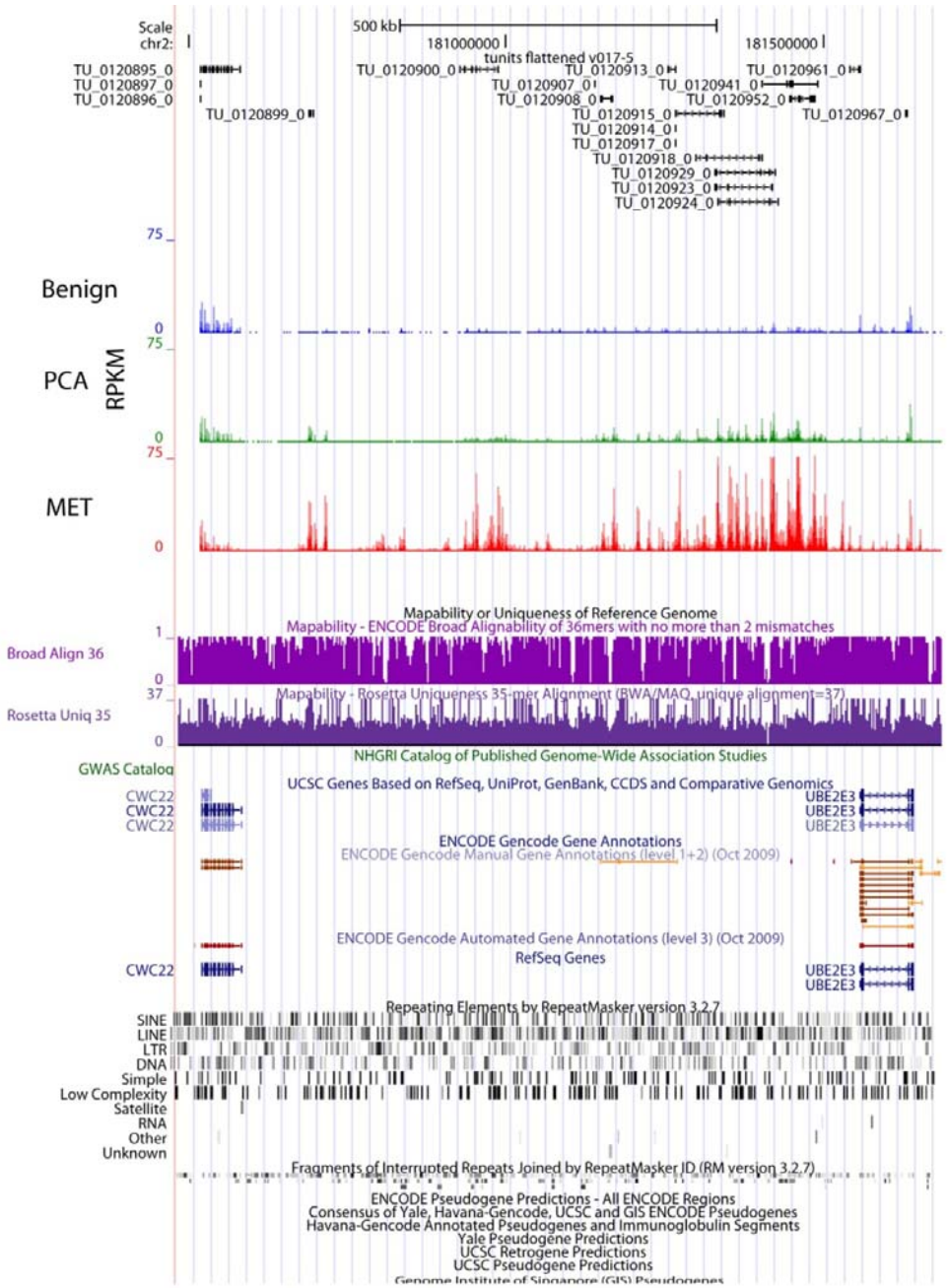


Figure 2.13: The SChLAP1 locus spans >500 kb. Visualization of transcriptome sequencing data in the UCSC genome browser indicates that a large, almost 1 Mb section of chromosome 2 is highly activated in cancer, contributing to many individual transcripts regulated in a coordinated fashion. For reference, the two flanking protein-coding genes, *UBE2E3* and *CWC22*, are shown. Neither *UBE2E3* nor *CWC22* appears differentially regulated in prostate cancer.

TABLES

Table 2.1: Total transcript counts during the processing pipeline.

Chromosome	Cuffcompare	Classification tree filter	Merge intron-redundant transcripts	Informatic filters	Join transcript fragments	**Filter intronic pre-mRNA**	UCSC	Refseq
chr1	759121	272072	12701	5030	4489	3652	2499	3334
chr2	581574	206281	9353	3224	2856	2361	1579	2023
chr3	518621	167071	5706	2917	2560	2053	1312	1816
chr4	329950	103113	5160	2019	1731	1444	977	1238
chr5	380613	126139	5833	2365	2067	1694	1104	1465
chr6	396848	145607	7580	2590	2309	1874	1370	1667
chr7	432152	134051	6432	2355	2132	1703	1326	1583
chr8	308935	97724	4226	1729	1529	1243	848	1210
chr9	359300	122626	4069	1937	1767	1402	1114	1272
chr10	354625	103512	3509	1672	1508	1226	998	1382
chr11	424606	165211	6909	2922	2640	2102	1566	2023
chr12	425280	138650	6872	2653	2373	1858	1233	1668
chr13	159649	68284	3616	1118	908	751	425	549
chr14	261497	123741	4842	1806	1619	1308	885	1102
chr15	291241	108058	5816	1884	1626	1321	1362	1127
chr16	364747	124182	3968	2002	1835	1386	1093	1311
chr17	473261	168469	5581	2780	2582	1950	1480	1907
chr18	144300	49112	2504	785	682	539	377	459
chr19	494738	189411	7209	3543	3239	2269	1668	2314
chr20	217223	70308	3059	1243	1158	907	659	926
chr21	113368	29728	939	495	436	354	306	427
chr22	223385	73509	2401	1156	1068	798	633	771
chrX	222743	94591	4997	1516	1349	1161	959	1841
chrY	15190	4039	272	81	71	59	148	254
Total	**8E+06**	**2885489**	**123554**	**49822**	**44534**	**35415**	**25921**	**33669**

Table 2.2: A list of Prostate Cancer Associated Transcripts (PCATs) differentially expressed in prostate cancer.

PCAT ID	Gene	Chromosomal location	Expected score (dExp)	Observed score(d)	Fold change (PCA vs Benign	q-value (%)
PCAT-1	TU_0099865_0	chr8:128087842-128095202	-2.265	5.444	6.907	0
PCAT-10	TU_0078296_0	chr12:32394534-32405549	-1.594	3.690	3.085	0
PCAT-100	TU_0049213_0	chr4:102461960-102476087	-1.273	2.367	1.888	0.71247
PCAT-101	TU_0093070_0	chr11:64945809-64961189	-1.295	2.365	1.913	0.71247
PCAT-102	TU_0051063_0	chr4:187244297-187244767	1.892	-2.849	0.510	0.73226
PCAT-103	TU_0098190_0	chr8:61704765-61708199	1.983	-2.861	0.403	0.73226
PCAT-104	TU_0038811_0	chr3:57890130-57890834	1.962	-2.884	0.444	0.73226
PCAT-105	TU_0020914_0	chr19:9718612-9721799	1.643	-2.924	0.506	0.73226
PCAT-106	TU_0112056_0	chr15:69658838-69678469	1.838	-3.036	0.462	0
PCAT-107	TU_0036396_0	chr14:104617328-104619095	1.850	-3.119	0.455	0
PCAT-108	TU_0095765_0	chr11:117640504-117642734	2.100	-3.263	0.382	0
PCAT-109	TU_0050224_0	chr4:147115887-147190781	2.198	-3.298	0.286	0
PCAT-11	TU_0078290_0	chr12:32394534-32410898	-1.534	3.675	3.157	0
PCAT-110	TU_0112059_0	chr15:69667695-69691724	1.815	-3.382	0.437	0
PCAT-111	TU_0098382_0	chr8:68494189-68495887	2.541	-4.059	0.308	0
PCAT-12	TU_0002597_0	chr6:34335202-34338521	-1.626	3.647	3.352	0
PCAT-13	TU_0049368_0	chr4:106772318-106772770	-1.689	3.608	2.830	0
PCAT-14	TU_0106548_0	chr22:22209111-22212055	-1.939	3.591	5.963	0
PCAT-15	TU_0078293_0	chr12:32396393-32414822	-1.521	3.571	2.922	0
PCAT-16	TU_0099884_0	chr8:128301493-128307576	-1.445	3.566	2.517	0
PCAT-17	TU_0112014_0	chr15:67722165-67739990	-1.633	3.562	3.659	0
PCAT-18	TU_0084306_0	chr5:15896315-15947088	-1.845	3.560	5.747	0
PCAT-19	TU_0114240_0	chr2:1534883-1538193	-1.697	3.523	4.340	0
PCAT-2	TU_0090142_0	chr11:4748677-4760303	-2.440	4.678	11.397	0
PCAT-20	TU_0008499_0	chr7:24236191-24236455	-1.830	3.507	6.682	0
PCAT-21	TU_0078299_0	chr12:32290896-32292169	-1.730	3.506	3.292	0
PCAT-22	TU_0000033_0	chr6:1619606-1668581	-1.768	3.494	2.247	0
PCAT-23	TU_0096472_0	chr11:133844590-133862924	-1.878	3.410	5.985	0

PCAT-24	TU_0114259_0	chr2:1606782-1607314	-1.666	3.392	5.061	0
PCAT-25	TU_0096473_0	chr11:133844590-133862995	-1.896	3.386	6.107	0
PCAT-26	TU_0100361_0	chr8:144914456-144930753	-1.652	3.381	3.842	0
PCAT-27	TU_0040394_0	chr3:133418632-133441282	-1.621	3.320	2.972	0
PCAT-28	TU_0043432_0	chr13:34032994-34050503	-1.674	3.204	3.209	0
PCAT-29	TU_0112020_0	chr15:67764259-67801825	-1.560	3.194	3.594	0
PCAT-3	TU_0054603_0	chr16:82380933-82394836	-2.179	4.461	5.892	0
PCAT-30	TU_0042717_0	chr13:23149908-23200198	-2.065	3.169	4.970	0
PCAT-31	TU_0078292_0	chr12:32290485-32406307	-1.450	3.151	2.891	0
PCAT-32	TU_0084146_0	chr5:14025126-14062770	-1.645	3.126	2.619	0
PCAT-33	TU_0056168_0	chr18:22477042-22477666	-1.538	3.056	3.195	0
PCAT-34	TU_0040383_0	chr3:133360541-133429262	-1.556	3.042	3.748	0
PCAT-35	TU_0112025_0	chr15:67780574-67782345	-1.682	3.041	3.433	0
PCAT-36	TU_0041688_0	chr3:186741299-186741933	-1.475	3.006	2.543	0
PCAT-37	TU_0103642_0	chr9:109187089-109187455	-1.739	2.999	6.612	0
PCAT-38	TU_0040375_0	chr3:133280694-133394609	-1.547	2.975	3.907	0
PCAT-39	TU_0047312_0	chr4:39217669-39222163	-1.639	2.912	3.612	0
PCAT-4	TU_0090140_0	chr11:4748163-4759145	-2.115	4.435	7.200	0
PCAT-40	TU_0106545_0	chr22:22218478-22219162	-1.759	2.890	3.736	0
PCAT-41	TU_0054541_0	chr16:79408800-79435066	-1.749	2.870	6.648	0
PCAT-42	TU_0060446_0	chr1:28438629-28450156	-1.488	2.857	1.982	0
PCAT-43	TU_0054537_0	chr16:79406933-79430041	-1.780	2.847	7.261	0
PCAT-44	TU_0072907_0	chr20:55759486-55771563	-1.525	2.797	2.812	0
PCAT-45	TU_0043403_0	chr13:33844637-33845921	-1.579	2.792	3.640	0
PCAT-46	TU_0038678_0	chr3:53515951-53517078	-1.705	2.786	3.691	0
PCAT-47	TU_0101706_0	chr9:3408690-3415374	-1.478	2.782	3.307	0
PCAT-48	TU_0074308_0	chr10:42652247-42653596	-1.993	2.773	8.825	0
PCAT-49	TU_0054538_0	chr16:79408946-79450819	-1.713	2.763	7.094	0
PCAT-5	TU_0078288_0	chr12:32393283-32405731	-1.916	4.313	3.566	0
PCAT-50	TU_0101709_0	chr9:3411967-3415374	-1.465	2.762	3.189	0
PCAT-51	TU_0106544_0	chr22:22210421-22220506	-1.615	2.758	3.742	0
PCAT-52	TU_0046121_0	chr4:766363-766599	-1.570	2.757	1.436	0
PCAT-53	TU_0106542_0	chr22:22211315-22220506	-1.610	2.756	3.378	0
PCAT-54	TU_0106541_0	chr22:22209111-22219162	-1.659	2.734	3.664	0
PCAT-55	TU_0044453_0	chr13:51505777-51524522	-1.342	2.732	2.537	0
PCAT-56	TU_0054540_0	chr16:79419351-79423673	-1.817	2.728	7.285	0
PCAT-57	TU_0104717_0	chr9:130697833-130698832	-1.294	2.722	2.334	0
PCAT-58	TU_0089014_0	chr5:176014905-176015351	-1.397	2.705	1.780	0
PCAT-59	TU_0108452_0	chr15:19344745-19362916	-1.584	2.676	1.848	0

PCAT-6	TU_0099864_0	chr8:128094589-128103681	-1.721	4.266	3.900	0
PCAT-60	TU_0112003_0	chr15:67645590-67775246	-1.439	2.668	3.045	0
PCAT-61	TU_0078286_0	chr12:32395588-32405731	-1.358	2.666	2.612	0
PCAT-62	TU_0078303_0	chr12:32274210-32274530	-1.502	2.659	3.331	0
PCAT-63	TU_0112004_0	chr15:67644390-67650387	-1.518	2.651	2.993	0
PCAT-64	TU_0071087_0	chr20:21428679-21429454	-1.492	2.649	4.648	0
PCAT-65	TU_0072906_0	chr20:55759768-55770657	-1.506	2.645	2.958	0
PCAT-66	TU_0054240_0	chr16:70155175-70173873	-1.472	2.644	3.531	0
PCAT-67	TU_0047330_0	chr4:39217641-39222163	-1.514	2.628	3.070	0
PCAT-68	TU_0055435_0	chr18:6718938-6719172	-1.605	2.617	2.922	0
PCAT-69	TU_0054542_0	chr16:79420131-79423590	-1.791	2.610	7.664	0
PCAT-7	TU_0084308_0	chr5:15938753-15949124	-1.964	4.124	4.748	0
PCAT-70	TU_0079791_0	chr12:54971063-54971481	-1.442	2.601	2.014	0
PCAT-71	TU_0043411_0	chr13:33918267-33926769	-1.495	2.599	3.386	0
PCAT-72	TU_0056121_0	chr18:20196762-20197522	-1.253	2.594	1.719	0
PCAT-73	TU_0043412_0	chr13:33918267-33935946	-1.589	2.590	4.280	0
PCAT-74	TU_0065837_0	chr1:149791252-149795934	-1.385	2.588	2.934	0
PCAT-75	TU_0043401_0	chr13:33825711-33845275	-1.599	2.585	4.346	0
PCAT-76	TU_0006463_0	chr6:144659819-144660143	-1.499	2.574	2.201	0
PCAT-77	TU_0048506_0	chr4:80329017-80348259	-1.574	2.569	2.802	0
PCAT-78	TU_0084140_0	chr5:14003669-14054874	-1.404	2.547	2.598	0
PCAT-79	TU_0082982_0	chr12:121776584-121777370	-1.529	2.546	2.620	0
PCAT-8	TU_0084303_0	chr5:15899476-15955226	-2.025	4.052	7.104	0
PCAT-80	TU_0013212_0	chr7:138990883-139001515	-1.230	2.544	1.688	0
PCAT-81	TU_0072912_0	chr20:55779532-55780817	-1.430	2.541	3.865	0
PCAT-82	TU_0112281_0	chr15:70586704-70590792	-1.459	2.538	2.429	0
PCAT-83	TU_0048767_0	chr4:88120066-88124880	-1.374	2.532	2.233	0
PCAT-84	TU_0108455_0	chr15:19358326-19365341	-1.565	2.526	1.946	0
PCAT-85	TU_0091997_0	chr11:58560356-58573012	-1.315	2.519	2.118	0
PCAT-86	TU_0121655_0	chr2:202985284-202998634	-1.401	2.476	2.219	0.85961
PCAT-87	TU_0071798_0	chr20:33775260-33778511	-1.336	2.465	1.657	0.85037
PCAT-88	TU_0049200_0	chr4:102469973-102476087	-1.322	2.457	1.946	0.84132
PCAT-89	TU_0121714_0	chr2:203295212-203314868	-1.346	2.450	1.762	0.83247
PCAT-9	TU_0082746_0	chr12:120197102-120197416	-1.861	3.755	5.143	0
PCAT-90	TU_0098937_0	chr8:95748751-95751321	-1.453	2.422	2.233	0.8238
PCAT-91	TU_0108453_0	chr15:19356996-19364013	-1.803	2.409	3.840	0.76781
PCAT-92	TU_0114170_0	chr15:99659312-99669199	-1.436	2.406	2.125	0.76781
PCAT-93	TU_0089906_0	chr11:1042845-1045705	-1.390	2.402	2.639	0.76781
PCAT-94	TU_0001559_0	chr6:30283700-30286011	-1.352	2.399	1.511	0.76781

PCAT-95	TU_0050557_0	chr4:159976338-160016453	-1.175	2.399	2.052	0.76781
PCAT-96	TU_0078294_0	chr12:32395632-32413064	-1.456	2.397	2.186	0.76781
PCAT-97	TU_0044933_0	chr13:94755992-94760688	-1.291	2.397	2.190	0.76781
PCAT-98	TU_0017730_0	chr17:52346638-52346880	-1.417	2.387	1.471	0.76043
PCAT-99	TU_0039020_0	chr3:66578329-66607777	-1.266	2.372	1.711	0.71247

Table 2.3: A list of Prostate Cancer Associated Transcripts (PCATs) nominated by outlier analysis of prostate cancer.

PCAT ID	Gene	Chromosomal location	Outlier Score	Median Expression (RPKM)	Maximum Expression (RPKM)
PCAT-112	TU_0029004_0	chrX:66691350-66692032	130.7349	1.000	90.921
PCAT-113	TU_0054542_0	chr16:79420131-79423590	127.0431	5.610	135.850
PCAT-114	TU_0120899_0	chr2:180689090-180696402	123.5416	1.053	94.693
PCAT-115	TU_0054540_0	chr16:79419351-79423673	119.0908	4.162	94.446
PCAT-116	TU_0120918_0	chr2:181297540-181400892	112.7101	1.453	92.180
PCAT-117	TU_0054538_0	chr16:79408946-79450819	98.0185	1.830	93.121
PCAT-118	TU_0059541_0	chr1:20685471-20686432	68.3573	1.783	1375.150
PCAT-119	TU_0120924_0	chr2:181331111-181427485	63.9546	1.389	365.202
PCAT-120	TU_0074308_0	chr10:42652247-42653596	60.9184	1.394	65.771
PCAT-121	TU_0049192_0	chr4:102257900-102306678	59.2500	1.385	69.242
PCAT-122	TU_0054537_0	chr16:79406933-79430041	58.0448	1.853	42.751
PCAT-123	TU_0120900_0	chr2:180926864-180985967	55.8439	1.000	67.658
PCAT-124	TU_0114527_0	chr2:10858318-10858530	54.7646	1.297	35.006
PCAT-125	TU_0112020_0	chr15:67764259-67801825	53.8828	2.028	88.990
PCAT-126	TU_0120923_0	chr2:181328093-181419226	52.9793	1.282	232.556
PCAT-127	TU_0049231_0	chr4:102257900-102259695	52.7700	1.340	67.628

REFERENCES

1. M. L. Metzker, *Nat Rev Genet* **11**, 31 (Jan, 2010).
2. M. Guttman *et al.*, *Nat Biotechnol* **28**, 503 (May, 2010).
3. C. Trapnell *et al.*, *Nat Biotechnol* **28**, 511 (May, 2010).
4. G. Robertson *et al.*, *Nat Methods* **7**, 909 (Nov, 2010).
5. D. R. Zerbino, E. Birney, *Genome Res* **18**, 821 (May, 2008).
6. M. Huarte *et al.*, *Cell* **142**, 409 (Aug 6, 2010).
7. U. A. Orom *et al.*, *Cell* **143**, 46 (Oct 1, 2010).
8. J. L. Rinn *et al.*, *Cell* **129**, 1311 (Jun 29, 2007).
9. R. A. Gupta *et al.*, *Nature* **464**, 1071 (Apr 15, 2010).
10. E. Pasmant *et al.*, *Cancer Res* **67**, 3963 (Apr 15, 2007).
11. K. L. Yap *et al.*, *Mol Cell* **38**, 662 (Jun 11, 2010).
12. M. C. Tsai *et al.*, *Science* **329**, 689 (Aug 6, 2010).
13. Y. Kotake *et al.*, *Oncogene*, (Dec 13, 2010).
14. J. B. de Kok *et al.*, *Cancer Res* **62**, 2695 (May 1, 2002).
15. J. Li *et al.*, *Science* **275**, 1943 (Mar 28, 1997).
16. J. R. Prensner, A. M. Chinnaiyan, *Curr Opin Genet Dev* **19**, 82 (Feb, 2009).
17. S. A. Tomlins *et al.*, *Nature* **448**, 595 (Aug 2, 2007).
18. S. A. Tomlins *et al.*, *Science* **310**, 644 (Oct 28, 2005).
19. C. Trapnell, L. Pachter, S. L. Salzberg, *Bioinformatics* **25**, 1105 (May 1, 2009).
20. D. Thierry-Mieg, J. Thierry-Mieg, *Genome Biol* **7 Suppl 1**, S12 1 (2006).
21. E. Birney *et al.*, *Nature* **447**, 799 (Jun 14, 2007).
22. P. Carninci *et al.*, *Science* **309**, 1559 (Sep 2, 2005).
23. Y. He, B. Vogelstein, V. E. Velculescu, N. Papadopoulos, K. W. Kinzler, *Science* **322**, 1855 (Dec 19, 2008).
24. M. Guttman *et al.*, *Nature* **458**, 223 (Mar 12, 2009).
25. J. Yu *et al.*, *Cancer Cell* **17**, 443 (May 18, 2010).
26. D. S. Day, L. J. Luquette, P. J. Park, P. V. Kharchenko, *Genome Biol* **11**, R69 (2010).
27. T. K. Kim *et al.*, *Nature* **465**, 182 (May 13, 2010).
28. M. A. Rubin *et al.*, *JAMA* **287**, 1662 (Apr 3, 2002).
29. S. M. Dhanasekaran *et al.*, *Nature* **412**, 822 (Aug 23, 2001).
30. H. van Bakel, C. Nislow, B. J. Blencowe, T. R. Hughes, *PLoS Biol* **8**, e1000371 (May, 2010).
31. S. A. Tomlins *et al.*, *Cancer Cell* **13**, 519 (Jun, 2008).
32. A. S. Bjartell *et al.*, *Clin Cancer Res* **13**, 4130 (Jul 15, 2007).
33. C. A. Maher *et al.*, *Nature* **458**, 97 (Mar 5, 2009).
34. C. A. Maher *et al.*, *Proc Natl Acad Sci U S A* **106**, 12353 (Jul 28, 2009).
35. N. Palanisamy *et al.*, *Nat Med* **16**, 793 (Jul, 2010).
36. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, *Genome Biol* **10**, R25 (2009).
37. D. L. Wheeler *et al.*, *Nucleic Acids Res* **28**, 10 (Jan 1, 2000).
38. D. Karolchik *et al.*, *Nucleic Acids Res* **31**, 51 (Jan 1, 2003).
39. W. J. Kent *et al.*, *Genome Res* **12**, 996 (Jun, 2002).

40. B. Rhead *et al.*, *Nucleic Acids Res* **38**, D613 (Jan, 2010).
41. M. Garber *et al.*, *Bioinformatics* **25**, i54 (Jun 15, 2009).
42. Y. Zhang *et al.*, *Genome Biol* **9**, R137 (2008).
43. C. Zang *et al.*, *Bioinformatics* **25**, 1952 (Aug 1, 2009).
44. S. Pepke, B. Wold, A. Mortazavi, *Nat Methods* **6**, S22 (Nov, 2009).
45. D. Blankenberg *et al.*, *Curr Protoc Mol Biol* **Chapter 19**, Unit 19 10 1 (Jan, 2010).
46. V. G. Tusher, R. Tibshirani, G. Chu, *Proc Natl Acad Sci U S A* **98**, 5116 (Apr 24, 2001).
47. V. T. Chu, R. Gottardo, A. E. Raftery, R. E. Bumgarner, K. Y. Yeung, *Genome Biol* **9**, R118 (2008).
48. J. W. MacDonald, D. Ghosh, *Bioinformatics* **22**, 2950 (Dec 1, 2006).

CHAPTER 3

Characterization of *PCAT-1* in prostate cancer

SUMMARY

High-throughput methodologies have enabled a detailed dissection of the cancer transcriptome, nominating thousands of long non-coding RNAs (lncRNAs) in disease. However, the functional role, if any, of these lncRNAs is largely unexplored. Here, we describe the functional characterization of *PCAT-1*, a prostate-specific lncRNA discovered by transcriptome sequencing of prostate cancers. We find that *PCAT-1* expression in prostate cancer promotes cell proliferation by coordinating a repressive signature of gene expression regulation. Interestingly, we identified *EZH2* as a negative regulator of *PCAT-1* and observed that *PCAT-1*-regulated genes were antagonized by *EZH2 in vitro* and this gene signature stratified human prostate cancer tissues into *PCAT-1*-like and *EZH2*-like biological subgroups. This gene signature was enriched for genes participating in DNA structural maintenance and DNA damage repair, suggesting a functional contribution of *PCAT-1* to these biological processes. Taken together, the findings presented herein identify *PCAT-1* as a novel transcriptional repressor implicated in subset of prostate cancer patients.

INTRODUCTION

High-throughput sequencing of RNA (RNA-seq) has facilitated rapid discovery of thousands of unannotated non-coding RNA transcripts. Yet, the biological function of many of these ncRNAs is unknown. One class of ncRNAs, long non-coding RNAs (lncRNAs) >200 bp in length, has been implicated in widespread biological roles, including cancer biology. While several types of ncRNAs appear to be derived from protein-coding genes, thus complicating the interpretation of their biological significance, lncRNAs are characterized by independent gene promoters and their gene structure mirrors that of protein-coding genes, commonly exhibiting splicing of multiple exons and a 3' polyadenylation signal.

We employed RNA-seq on >100 prostate cancer tissues and cell lines and identified >1,800 intergenic lncRNAs, of which 121 demonstrated dysregulated expression in prostate cancer and were termed PCATs. Here, we identify PCAT-1 as a novel prostate cancer gene that serves to promote cell proliferation by coordinating repression of specific target genes. Moreover, we describe an antagonist PCAT-1-EZH2 axis that competitively regulates gene expression signatures *in vitro* and in human cancer tissues.

RESULTS

PCAT-1, a novel prostate cancer lncRNA

We previously nominated and validated numerous novel lncRNAs found in prostate cancer tissues and cell lines. To explore several transcripts more closely, we performed 5' and 3' rapid amplification of cDNA ends (RACE) for *PCAT-1* and *PCAT-14*. Interestingly, the *PCAT-14* locus contained components of viral ORFs from the HERV-K endogenous retrovirus family, whereas *PCAT-1* incorporates portions of a *mariner* family transposase (*1, 2*), an Alu, and a viral long terminal repeat (LTR) promoter region (**Figure 3.1** and **3.2a**). While *PCAT-14* was upregulated in localized prostate cancer but largely absent in metastases (**Figure 2.8**), *PCAT-1* was strikingly upregulated in a subset of metastatic and high-grade localized (Gleason score ≥7) cancers (**Figure 2.8f**). Because of this notable profile, we hypothesized that *PCAT-1* may have coordinated expression with the oncoprotein *EZH2*, a core PRC2 protein that is upregulated in solid tumors and contributes to a metastatic phenotype (*3, 4*). Surprisingly, we found that *PCAT-1* and *EZH2* expression were nearly mutually exclusive (**Figure 3.2b**), with only one patient showing outlier expression of both. This suggests that outlier *PCAT-1* and *EZH2* expression may define two subsets of high-grade disease.

PCAT-1 is located in the chromosome 8q24 gene desert approximately 725 kb upstream of the *c-MYC* oncogene. To confirm that *PCAT-1* is a non-coding gene, we cloned the full-length *PCAT-1* transcript and performed *in vitro* translational assays, which were negative as expected (**Figure 3.3**). Next, since Chr8q24 is known to harbor prostate cancer-associated single nucleotide polymorphisms (SNPs) and to exhibit

frequent chromosomal amplification (*5-10*), we evaluated whether the relationship between *EZH2* and *PCAT-1* was specific or generalized. To address this, we measured expression levels of *c-MYC* and *NCOA2*, two proposed targets of Chr8q amplification (*7, 10*), by qPCR. Neither *c-MYC* nor *NCOA2* levels showed striking expression relationships to *PCAT-1*, *EZH2*, or each other (**Figure 3.4**). Likewise, *PCAT-1* outlier expression was not dependent on Chr8q24 amplification, as highly expressing localized tumors often did not have 8q24 amplification and high copy number gain of 8q24 was not sufficient to upregulate *PCAT-1* (**Figure 3.5**).

PCAT-1 Function and Regulation

Despite reports showing that upregulation of the ncRNA *HOTAIR* participates in PRC2 function in breast cancer(*11*), we do not observe strong expression of this ncRNA in prostate (**Figure 3.6**), suggesting that other ncRNAs may be important in this cancer. To determine the mechanism for the expression profiles of *PCAT-1* and *EZH2*, we inhibited *EZH2* activity in VCaP cells, which express low-to-moderate levels of *PCAT-1*. Knockdown of *EZH2* by shRNA or pharmacologic inhibition of *EZH2* with the inhibitor 3-deazaneplanocin A (DZNep) caused a dramatic upregulation in *PCAT-1* expression levels (**Figure 3.2c,d**), as did treatment of VCaP cells with the demethylating agent 5'deoxyazacytidine, the histone deacetylase inhibitor SAHA, or both (**Figure 3.2e**). Chromatin immunoprecipitation (ChIP) assays also demonstrated that SUZ12, a core PRC2 protein, directly binds the *PCAT-1* promoter approximately 1kb upstream of the TSS (**Figure 3.2f**). Interestingly, RNA immunoprecipitation (RIP) similarly showed binding of *PCAT-1* to SUZ12 protein in VCaP cells (**Figure 3.7a**). RIP assays followed

by RNase A, RNase H, or DNase I treatment either abolished, partially preserved, or totally preserved this interaction, respectively (**Figure 3.7b**). This suggests that *PCAT-1* exists primarily as a single-stranded RNA and secondarily as a RNA/DNA hybrid.

To explore the functional role of *PCAT-1* in prostate cancer, we stably overexpressed full length *PCAT-1* or controls in RWPE benign immortalized prostate cells. We observed a modest but consistent increase in cell proliferation when *PCAT-1* was overexpressed at physiological levels (**Figure 3.8**). Next, we designed siRNA oligos to *PCAT-1* and performed knockdown experiments in LNCaP cells, which lack PRC2-mediated regulation of *PCAT-1* (**Figure 3.9**). Supporting our overexpression data, knockdown of *PCAT-1* with three independent siRNA oligos resulted in a 25% - 50% decrease in cell proliferation in LNCaP cells (**Figure 3.8**), but not control DU145 cells lacking *PCAT-1* expression (**Figure 3.8**) or VCaP cells, in which *PCAT-1* is expressed but repressed by PRC2 (**Figure 3.8**).

Gene expression profiling of LNCaP knockdown samples on cDNA microarrays indicated that *PCAT-1* modulates the transcriptional regulation of 370 genes (255 upregulated, 115 downregulated; FDR ≤ 0.01) (**Appendix 2**). Gene ontology analysis of the upregulated genes showed preferential enrichment for cellular processes such as mitosis and cell cycle, whereas the downregulated genes had no concepts showing statistical significance (**Figure 3.8**and **Table 3.1**). These results suggest that *PCAT-1*'s function is predominantly repressive in nature, similar to other lncRNAs. We next validated expression changes in three key *PCAT-1* target genes (*BRCA2*, *CENPE* and *CENPF*) whose expression is upregulated upon *PCAT-1* knockdown (**Figure 3.10a**) in

LNCaP and VCaP cells, the latter of which appear less sensitive to *PCAT-1* knockdown likely due to lower overall expression levels of this transcript.

PCAT-1 signatures in prostate cancer

Because of the regulation of *PCAT-1* by PRC2 in VCaP cells, we hypothesized that knockdown of *EZH2* would also downregulate *PCAT-1* targets as a secondary phenomenon due to the subsequent upregulation of *PCAT-1*. Simultaneous knockdown of *PCAT-1* and *EZH2* would thus abrogate expression changes in *PCAT-1* target genes. Performing this experiment in VCaP cells demonstrated that *PCAT-1* target genes were indeed downregulated by *EZH2* knockdown, and that this change was either partially or completely reversed using siRNA oligos to *PCAT-1* (**Figure 3.10a**), lending support to the role of *PCAT-1* as a transcriptional repressor. Taken together, these results suggest that *PCAT-1* biology may exhibit two distinct modalities: one in which PRC2 represses *PCAT-1* and a second in which active *PCAT-1* promotes cell proliferation. *PCAT-1* and PRC2 may therefore characterize distinct subsets of prostate cancer.

To examine our clinical cohort, we used qPCR to measure expression of *BRCA2*, *CENPE*, and *CENPF* in our tissue samples. Consistent with our model, we found that *PCAT-1*-expressing samples tended to have low expression of *PCAT-1* target genes (**Figure 3.10b**). Moreover, comparing *EZH2*-outlier and *PCAT-1*-outlier patients (see **Figure 3.2b**), we found that two distinct patient phenotypes emerged: those with high *EZH2* tended to have high levels of *PCAT-1* target genes; and those with high *PCAT-1* expression displayed the opposite expression pattern (**Figure 3.10c**). Network analysis of the top 20 upregulated genes following *PCAT-1* knockdown with the HefaLMP tool

(*12*) further suggested that these genes form a coordinated network (**Figure 3.10d**), corroborating our previous observations. Taken together, these results provide initial data into the composition and function of the prostate cancer ncRNA transcriptome.

DISCUSSION

Our discovery of the prostate-specific ncRNA gene *PCAT-1*, which is markedly overexpressed in a subset of prostate cancers, particularly metastases, provides the first example of a novel lncRNA being discovered by large-scale RNA-seq analyses of human tissues. Our results suggest that *PCAT-1* may contribute to cell proliferation in a subset of prostate cancers. It is also notable that *PCAT-1* resides in the 8q24 "gene desert" locus, in the vicinity of well-studied prostate cancer risk SNPs and the *c-MYC* oncogene, suggesting that this locus—and its frequent amplification in cancer—may be linked to additional aspects of cancer biology. In addition, the interplay between PRC2 and *PCAT-1* further suggests that this ncRNA may have an important role in prostate cancer progression (**Figure 3.10e**). Other ncRNAs identified by this analysis may similarly contribute to prostate cancer as well. Furthermore, recent pre-clinical efforts to detect prostate cancer non-invasively through the collection of patient urine samples have shown promise for several urine-based prostate cancer biomarkers, including the ncRNA *PCA3* (*13, 14*). While additional studies are needed, our identification of ncRNA biomarkers for prostate cancer suggests that urine-based assays for these ncRNAs may also warrant investigation, particularly for those that may stratify patient molecular subtypes.

Taken together, our findings support an important role for tissue-specific ncRNAs in prostate cancer and suggest that cancer-specific functions of these ncRNAs may help to "drive" tumorigenesis. We further speculate that specific ncRNA signatures may occur universally in all disease states and applying these methodologies to other diseases may reveal key aspects of disease biology and clinically-important biomarkers.

MATERIALS AND METHODS

Experimental studies

Cell lines: The benign immortalized prostate cell line RWPE as well as PC3, Du145, LNCaP, VCaP, 22Rv1, CWR22, C4-2B, NCI-660, MDA PCa 2b, WPMY-1, and LAPC-4 prostate cell lines were obtained from the American Type Culture Collection (Manassas, VA). Benign non-immortalized prostate epithelial cells (PrEC) and prostate smooth muscle cells (PrSMC) were obtained from Lonza (Basel, Switzerland).

PC3, Du145, LNCaP, 22Rv1, and CRW22 cells were grown in RPMI 1640 (Invitrogen) and supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin. LNCaP CDS parent cells were grown in RPMI 1640 lacking phenol red (Invitrogen) supplemented with 10% charcoal-dextran stripped FBS (Invitrogen) and 1% penicillin-streptomycin. LNCaP CDS 1, 2, and 3 are androgen-independent subclones derived from extended cell culture in androgen-depleted media. VCaP and WPMY-1 cells were grown in DMEM (Invitrogen) and supplemented with 10% fetal bovine serum (FBS) with 1% penicillin-streptomycin. NCI-H660 cells were grown in RPMI 1640 supplemented with 0.005 mg/ml insulin, 0.01 mg/ml transferrin, 30 nM sodium selenite, 10 nM hydrocortisone, 10 nM beta-estradiol, 5% FBS and an extra 2 mM of L-glutamine (for a final concentration of 4 mM). MDA PCa 2b cells were grown in F-12K medium (Invitrogen) supplemented with 20% FBS, 25 ng/ml cholera toxin, 10ng/ml EGF, 0.005 mM phosphoethanolamine, 100 pg/ml hydrocortisone, 45 nM selenious acid, and 0.005 mg/ml insulin. LAPC-4 cells were grown in Iscove's media (Invitrogen) supplemented with 10% FBS and 1 nM R1881. C4-2B cells were grown in 80% DMEM supplemented with 20% F12, 5% FBS, 3 g/L NaCo3, 5ug/ml insulin, 13.6 pg/ml triiodothyonine,

5ug/ml transferrin, 0.25ug/ml biotin, and 25 ug/ml adenine. PrEC cells were grown in PrEGM supplemented with 2ml BPE, 0.5 ml hydrocortisone, 0.5ml EGF, 0.5 ml epinephrine, 0.5 ml transferrin, 0.5 ml insulin, 0.5 ml retinoic acid, and 0.5 ml triiodothyronine, as part of the PrEGM BulletKit (Lonza). PrSMC cells were grown in SmGM-2 media supplemented with 2ml BPE, 0.5 ml hydrocortisone, 0.5ml EGF, 0.5 ml epinephrine, 0.5 ml transferrin, 0.5 ml insulin, 0.5 ml retinoic acid, and 0.5 ml triiodothyronine, as part of the SmGM-2 BulletKit (Lonza).

For androgen treatment experiments, LNCaP and VCaP cells were grown in androgen-depleted media lacking phenol red and supplemented with 10% charcoal-stripped serum and 1% penicillin-streptomycin. After 48 hours, cells were treated with 5nM methyltrienolone (R1881, NEN Life Science Products) or an equivalent volume of ethanol. Cells were harvested for RNA at 6, 24, and 48 hours post-treatment.

Prostate tissues were obtained from the radical prostatectomy series and Rapid Autopsy Program (*15*) at the University of Michigan tissue core. These programs are part of the University of Michigan Prostate Cancer Specialized Program Of Research Excellence (S.P.O.R.E.). All tissue samples were collected with informed consent under an Institutional Review Board (IRB) approved protocol at the University of Michigan. (SPORE in Prostate Cancer (Tissue/Serum/Urine) Bank Institutional Review Board # 1994-0481).

RNA isolation and cDNA synthesis: Total RNA was isolated using Trizol and an RNeasy Kit (Invitrogen) with DNase I digestion according to the manufacturer's instructions. RNA integrity was verified on an Agilent Bioanalyzer 2100 (Agilent Technologies, Palo

Alto, CA). cDNA was synthesized from total RNA using Superscript III (Invitrogen) and random primers (Invitrogen).

Quantitative Real-time PCR: Quantitative Real-time PCR (qPCR) was performed using Power SYBR Green Mastermix (Applied Biosystems, Foster City, CA) on an Applied Biosystems 7900HT Real-Time PCR System. All oligonucleotide primers were obtained from Integrated DNA Technologies (Coralville, IA) and are listed in **Appendix 1.** The housekeeping gene, *GAPDH*, was used as a loading control. Fold changes were calculated relative to *GAPDH* and normalized to the median value of the benign samples.

Reverse-transcription PCR: Reverse-transcription PCR (RT-PCR) was performed for primer pairs using Platinum Taq High Fidelity polymerase (Invitrogen). PCR products were resolved on a 1.5% agarose gel. PCR products were either sequenced directly (if only a single product was observed) or appropriate gel products were extracted using a Gel Extraction kit (Qiagen) and cloned into pcr4-TOPO vectors (Invitrogen). PCR products were bidirectionally sequenced at the University of Michigan Sequencing Core using either gene-specific primers or M13 forward and reverse primers for cloned PCR products. All oligonucleotide primers were obtained from Integrated DNA Technologies (Coralville, IA) and are listed in **Appendix 2.**

RNA-ligase-mediated rapid amplification of cDNA ends (RACE): 5' and 3' RACE was performed using the GeneRacer RLM-RACE kit (Invitrogen) according to the manufacturer's instructions. RACE PCR products were obtained using Platinum Taq

High Fidelity polymerase (Invitrogen), the supplied GeneRacer primers, and appropriate gene-specific primers indicated in **Appendix 2**. RACE-PCR products were separated on a 1.5% agarose gels. Gel products were extracted with a Gel Extraction kit (Qiagen), cloned into pcr4-TOPO vectors (Invitrogen), and sequenced bidirectionally using M13 forward and reverse primers at the University of Michigan Sequencing Core. At least three colonies were sequenced for every gel product that was purified.

siRNA knockdown studies: Cells were plated in 100mM plates at a desired concentration and transfected with 20uM experimental siRNA oligos or non-targeting controls twice, at 12 hours and 36 hours post-plating. Knockdowns were performed with Oligofectamine in OptiMEM media. Knockdown efficiency was determined by qPCR. siRNA sequences (in sense format) for *PCAT-1* knockdown were as follows: siRNA 1 UUAAAGAGAUCCACAGUUAUU; siRNA 2 GCAGAAACACCAAUGGAUAUU; siRNA 3 AUACAUAAGACCAUGGAAAU; siRNA 4 GAACCUAACUGG-ACUUUAAUU. For *EZH2* siRNA, the following sequence was used: GAGGUUCAGACGAGCUGAUUU. 72 hours post-transfection, cells were trypsinized, counted with a Coulter counter, and diluted to 1 million cells/mL.

For proliferation assays, 10,000 cells were plated in 24-well plates and grown in regular media. 48 and 96 hours post-plating, cells were harvested and counted using a Coulter counter.

shRNA knockdown: The prostate cancer cell line VCaP was seeded at 50-60% confluency and allowed to attach over night. Cells were transfected with *EZH2* or non-targeting

shRNA lentiviral constructs as described previously for 48 hours. GFP+ cells were drug-selected using 1 ug/mL puromycin for 72 hours. 48 hours post-selection cells were harvested for protein and RNA using RIPA buffer or trizol, respectively. RNA was processed as described above. Protein lysates were boiled in sample buffer, and 10 ug protein was loaded onto a SDS-PAGE gel and run for separation of proteins. Proteins were transferred onto Polyvinylidene Difluoride membrane (GE Healthcare) and blocked for 90 minutes in blocking buffer (5% milk, 0.1% Tween, Tri-buffered saline (TBS-T)). Membranes were incubated overnight at 4C with either *EZH2* mouse monoclonal (1:1000, in dilution buffer, BD Biosciences, no. 612666), or *B-Actin* (Abcam, ab8226) for equal loading. Following 3 washes with TBS-T, and one wash with TBS, the blot was incubated with horseradish peroxidase-conjugated secondary antibody and the signals visualized by enhanced chemiluminescence system as described by the manufacturer (GE Healthcare).

Gene expression profiling: Expression profiling was performed using the Agilent Whole Human Genome Oligo Microarray (Santa Clara, CA), according to previously published protocols (*16*). All samples were run in technical triplicates comparing knockdown samples treated with PCAT1 siRNA compared to treatments with non-targeting control siRNA. Expression array data were processed using the SAM method (*17*) with an FDR ≤ 0.01. For gene ontology analysis, up- and down-regulated probes were separated and probe lists were analyzed using the DAVID bioinformatics platform (*18*) with the functional annotation clustering tool.

Drug treatments: VCaP cells were plated at equal cell densities in 100mm2 plates and allowed to attach for 48 hours. Cells were then treated with DMSO (vehicle control), 20uM 5'deoxyazacytidine (Sigma), 500nM HDAC inhibitor suberoylanilide hydroxamic acid (SAHA) (Biovision Inc.), or both 5'deoxyazacytidine and SAHA. 5'deoxyazacytidine treatments were performed for 6 days with media and drug re-applied every 48 hours. SAHA treatments were performed for 48 hours. DMSO treatments were performed for 6 days. Total RNA was collected using Trizol and extracted as described above.

For pharmacological inhibition of EZH2, VCaP cells were plated and grown overnight in 6 well plates. Following plating and attachment, cells were treated with either 0.1uM 3-deazaneplanocin A (DZNep) or DMSO vehicle control. Cell media and drug treatments were refreshed every 48 hours. RNA from cells was harvested at 72 hours and 144 hours post-treatment.

*Chromatin immunoprecipitation :*ChIP assays were performed as described previously (*19*). Antibodies used for ChIP were rabbit polyclonal IgG (Millipore, PP64), SUZ12 (Cell Signaling, 3737), and SUZ12 (Abcam, ab12073), using 4 - 7 ug of antibody per ChIP reaction. ChIP-PCR reactions were performed using SybrGreen as described above using 1:150th of the ChIP product per reaction. All ChIP-PCR reactions were performed in triplicate. Primer oligo sequences are found in **Appendix 2**.

RNA immunoprecipitation: RIP assays were performed using a Millipore EZ-Magna RIP RNA-Binding Protein Immunoprecipitation kit (Millipore, #17-701) according to the manufacturer's instructions. Input lysates were fractionated into cytoplasmic and nuclear fractions using a NE-PER Nuclear and Cytoplasmic Extraction kit (Thermo-Scientific/Pierce, #78835), and nuclear fractions were used for RIP assays. RIP-PCR was performed as RT-PCR, as described above, using total RNA as input controls. 1:150th of RIP RNA product was used per PCR reaction. Antibodies used for RIP were Rabbit polyclonal IgG (Millipore, PP64), SNRNP70 (Millipore, CS203216), and SUZ12 (Abcam, ab12073), using 4 – 7 ug of antibody per RIP reaction. Treatment of RIP RNA product was performed using 5 ug of RNase A (Invitrogen), 5 ug of RNase H (Invitrogen), or 5 ug of DNase I (Qiagen) and incubating at 37C for 15-20 minutes. All RIP assays were performed in biological duplicate.

Overexpression studies: PCAT-1 full length transcript was amplified from LNCaP cells and cloned into the pLenti6 vector (Invitrogen) along with RFP and LacZ controls. Insert sequences were confirmed by Sanger sequencing at the University of Michigan Sequencing Core. Lentiviruses were generated at the University of Michigan Vector Core. The benign immortalized prostate cell line RWPE was infected with lentiviruses expressing *PCAT-1*, RFP, or LacZ and stable pools and clones were generated by selection with blasticidin (Invitrogen). For cell proliferation assays, 10,000 cells were plated in 12-well plates and allowed to attach overnight. Cells were harvested and counted at day 2, day 4, and day 6 post-plating with a Coulter counter.

In vitro translation: Full length *PCAT-1,* Halo-tagged *ERG* or *GUS* positive control were cloned into the PCR2.1 entry vector (Invitrogen). Insert sequences were confirmed by Sanger sequencing at the University of Michigan Sequencing Core. *In vitro* translation assays were performed with the TnT Quick Coupled Transcription/Translation System (Promega) with 1mM methionine and Transcend Biotin-Lysyl-tRNA (Promega) according to the manufacturer's instructions.

Statistical analyses for experimental studies: All data are presented as means ± S.E.M. All experimental assays were performed in duplicate or triplicate. Statistical analyses shown in figures represent Fisher's exact tests or two-tailed t-tests, as indicated.

FIGURES

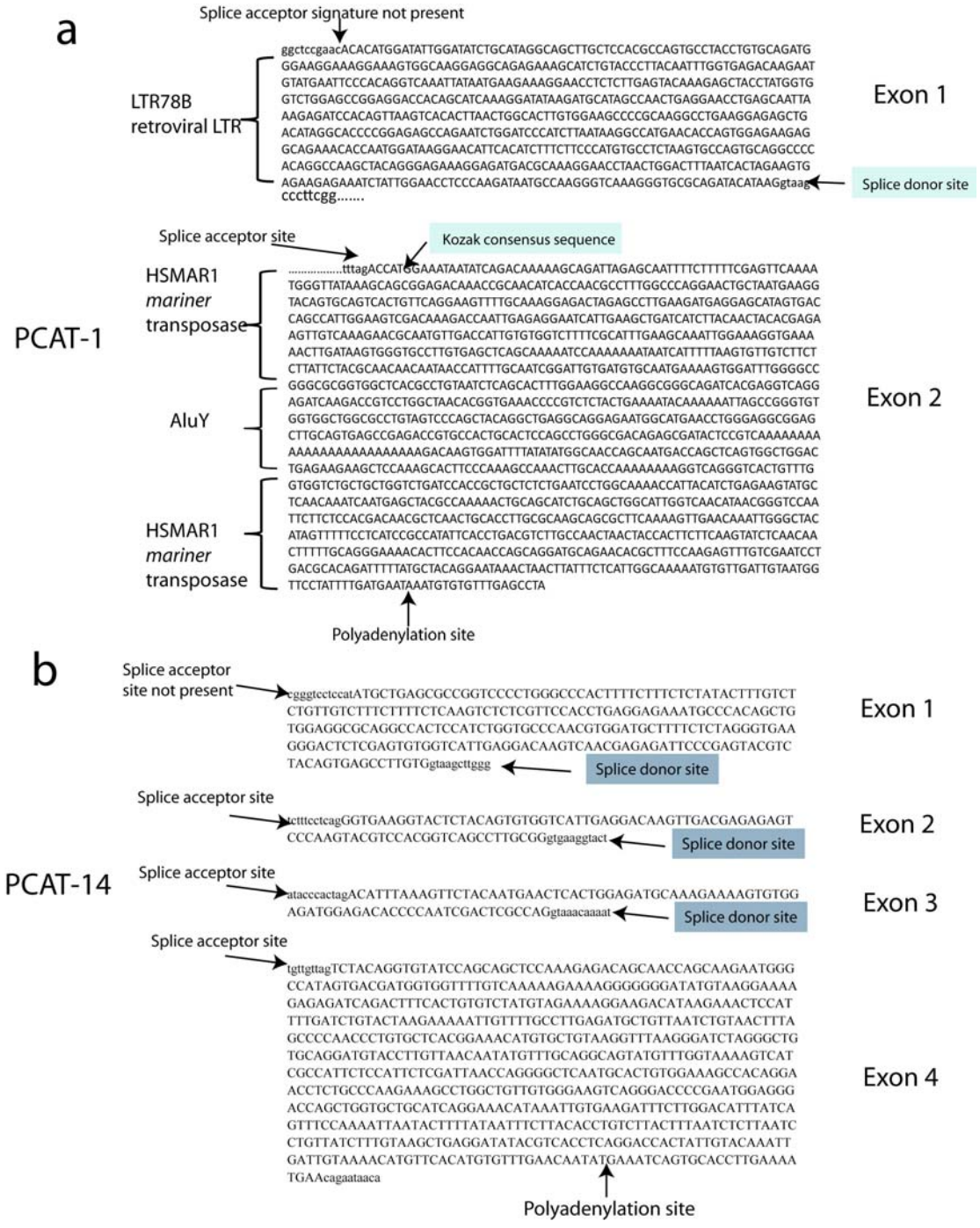


Figure 3.1: Analysis of *PCAT-1* and *PCAT-14* transcript structure. (**a**) 5' and 3' RACE experiments for *PCAT-1* showed a ncRNA transcript containing two exons, each of which harbored portions of ancestral repetitive elements. (**b**) 5' and 3' RACE analysis of *PCAT-14.* Expressed nucleotides are indicated in uppercase; genomic nucleotides are indicated in lowercase.

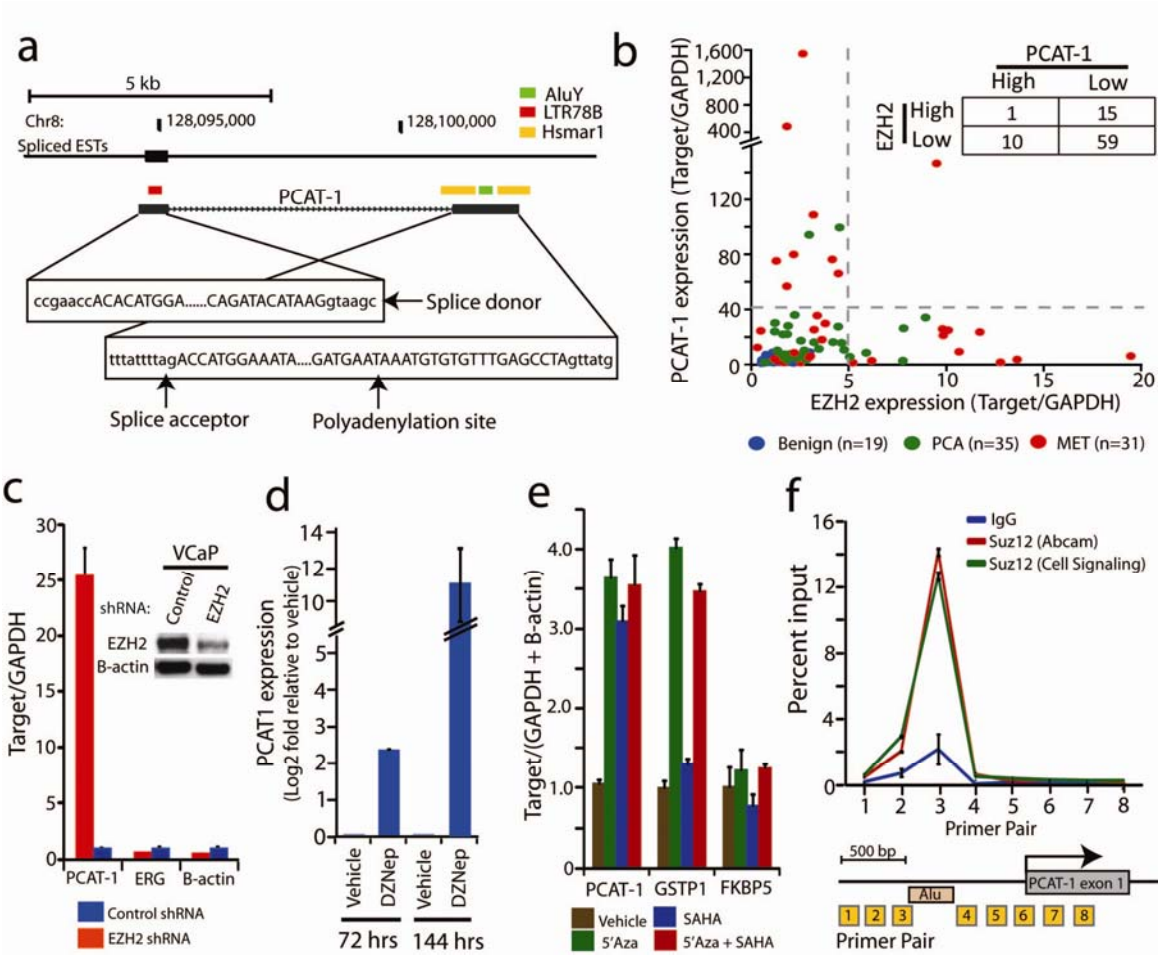


Figure 3.2: *PCAT-1* is a marker of aggressive cancer and a PRC2-repressed ncRNA. (**a**) The genomic location of *PCAT-1* determined by 5' and 3' RACE, with DNA sequence features indicated by the colored boxes (**b**) qPCR for *PCAT-1* (Y-axis) and *EZH2* (X-axis) on a cohort of benign (n=19), localized tumor (n=35) and metastatic cancer (n=31) samples. The inset table quantifies patient subsets demarcated by the gray dotted lines. (**c**) Knockdown of *EZH2* in VCaP resulted in upregulation of *PCAT-1*. Data were normalized to *GAPDH* and represented as fold change. *ERG* and *B-Actin* serve as negative controls. The inset Western blot indicates *EZH2* knockdown. (**d**) Treatment of VCaP cells with 0.1 μM of the *EZH2* inhibitor DZNep or vehicle control (DMSO) shows increased expression of *PCAT-1* transcript following *EZH2* inhibition. (**e**) *PCAT-1* expression is increased upon treatment of VCaP cells with the demethylating agent 5'Azacytidine, the histone deacetylase inhibitor SAHA, or a combination of both. qPCR data were normalized to the average of (*GAPDH+B-Actin)* and represented as fold change. *GSTP1* and *FKBP5* are positive and negative controls, respectively. (**f**) ChIP assays for *SUZ12* demonstrated direct binding of *SUZ12* to the *PCAT-1* promoter. Primer locations are indicated (boxed numbers) in the *PCAT-1* schematic.

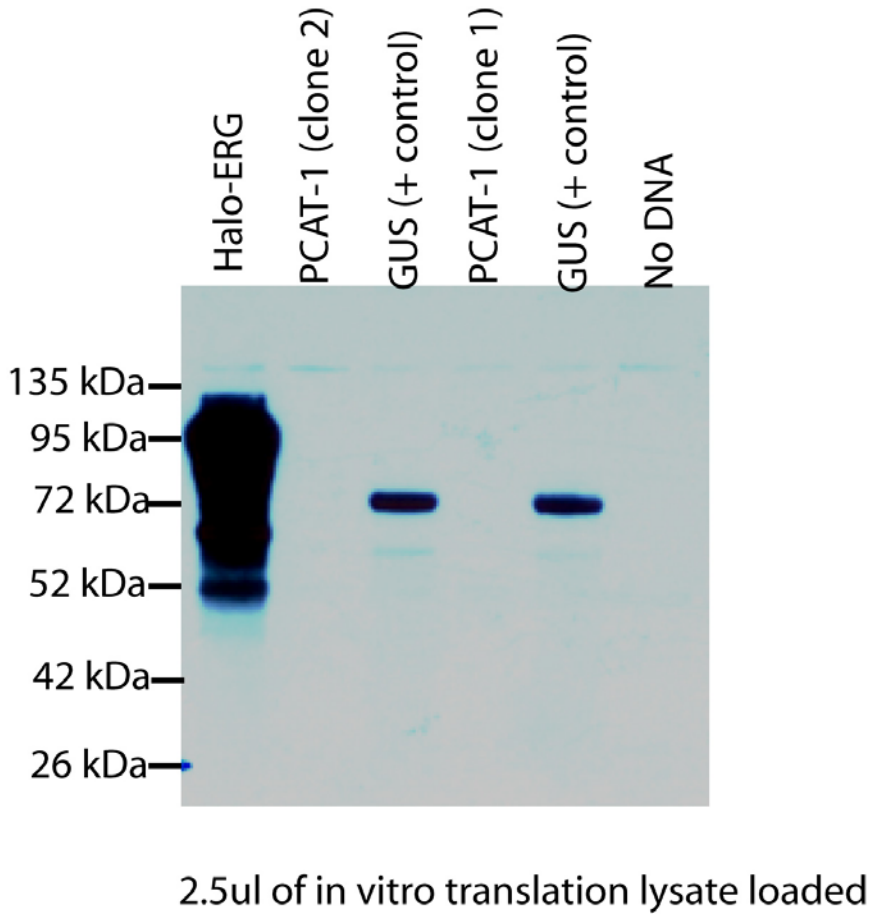


Figure 3.3: *In vitro* translation of *PCAT-1* confirms ncRNA status. Full length *PCAT-1* transcript was cloned into the PCR2.1 vector (Invitrogen) and expressed using the TnT Quick Coupled Transcription/Translation System (Promega). Western blot analysis resolving the proteins by SDS-PAGE indicated that *PCAT-1* is a non-coding RNA with no protein-coding capacity. GUS and ERG protein *in vitro* translation served as positive controls.

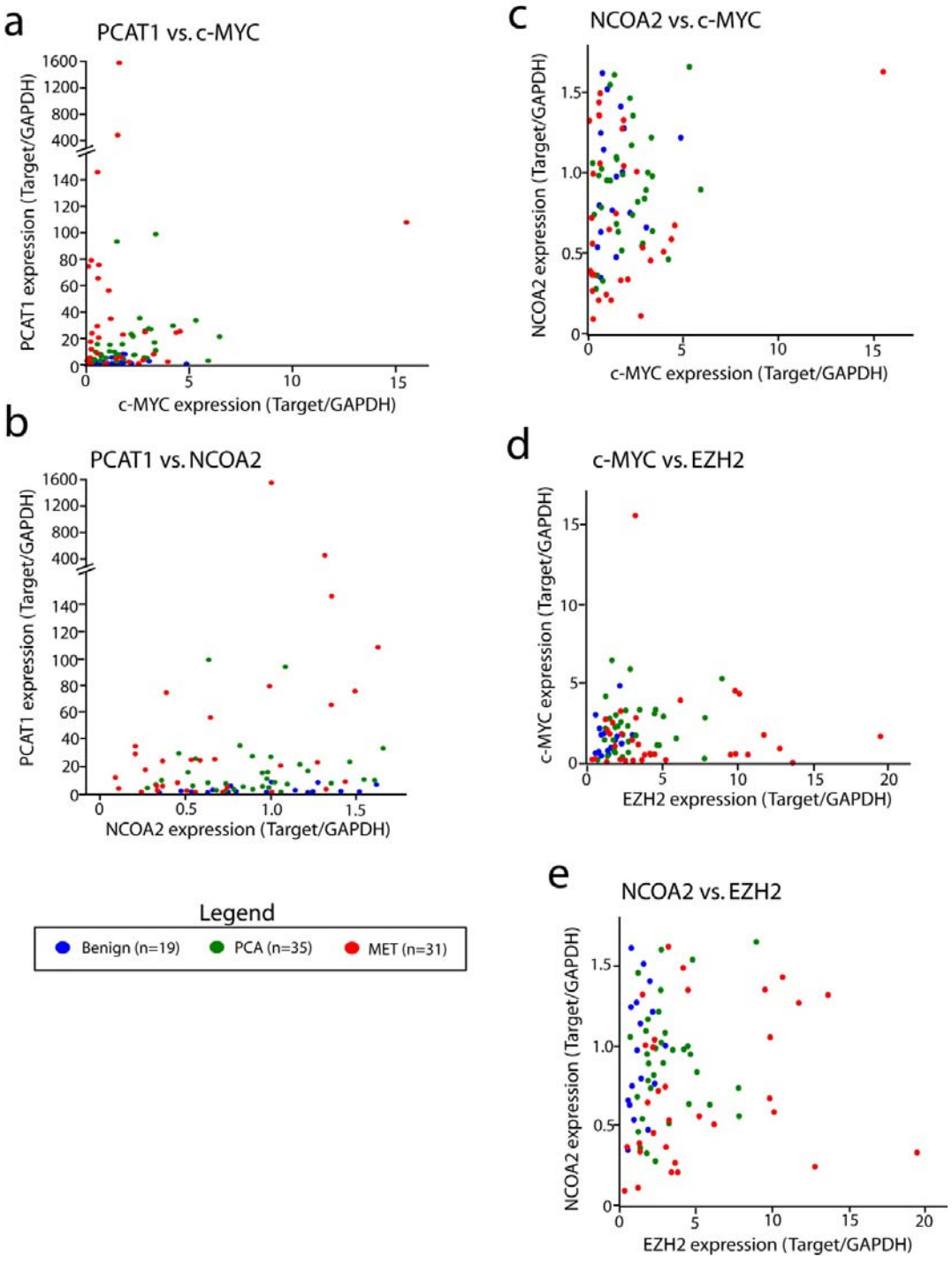


Figure 3.4: *PCAT-1* and *EZH2* expression is not exclusive with other Chr8q genes. qPCR was performed for *c-MYC*, *NCOA2*, *PCAT-1*, and *EZH2*. Patterns of expression were compared as above: (**a**) *PCAT-1* vs, *c-MYC*; (**b**) *PCAT-1* vs *NCOA2*; (**c**) *NCOA2* vs. *c-MYC*; (**d**) *c-MYC* vs. *EZH2*; (**e**) *NCOA2* vs. *EZH2*.

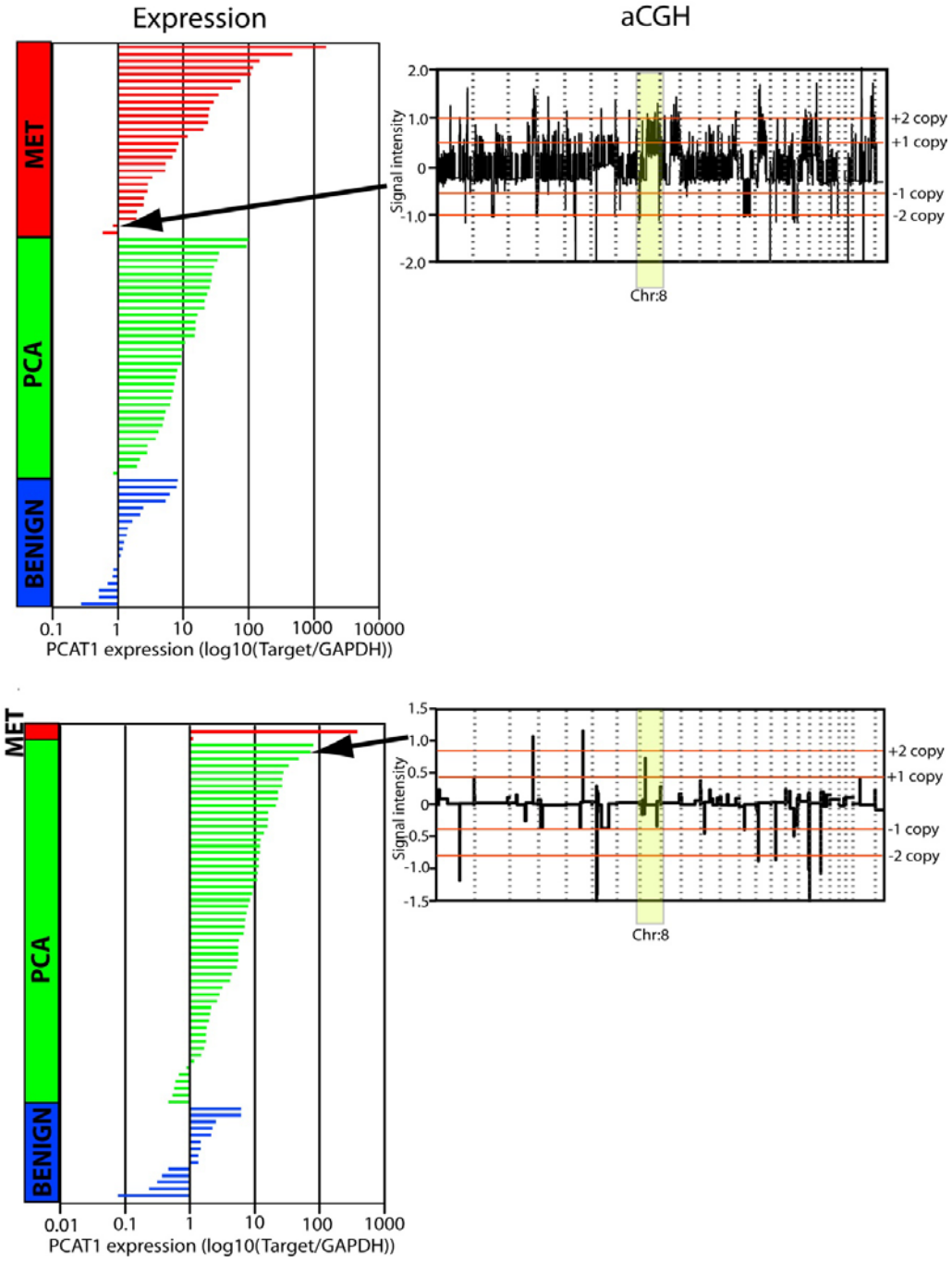


Figure 3.5: Genomic amplification of Chr8q is not required for *PCAT-1* upregulation. Left: qPCR for *PCAT-1* in the two prostate tissue cohorts shown in log10 scale. qPCR data is normalized to *GAPDH* and the median of the benign samples. Right: Matched copy-number data (array CGH) for two patients. The top panel indicates a patient with high Chr8q copy number gain and low expression of *PCAT-1*. The lower panel indicates a patient with high expression of *PCAT-1* but no copy number gain of

Chr8q. Samples were run against genomic DNA from a normal human reference pool on Agilent Human Genome CGH 105K oligo arrays (Agilent Technologies, Santa Clara, CA, USA).

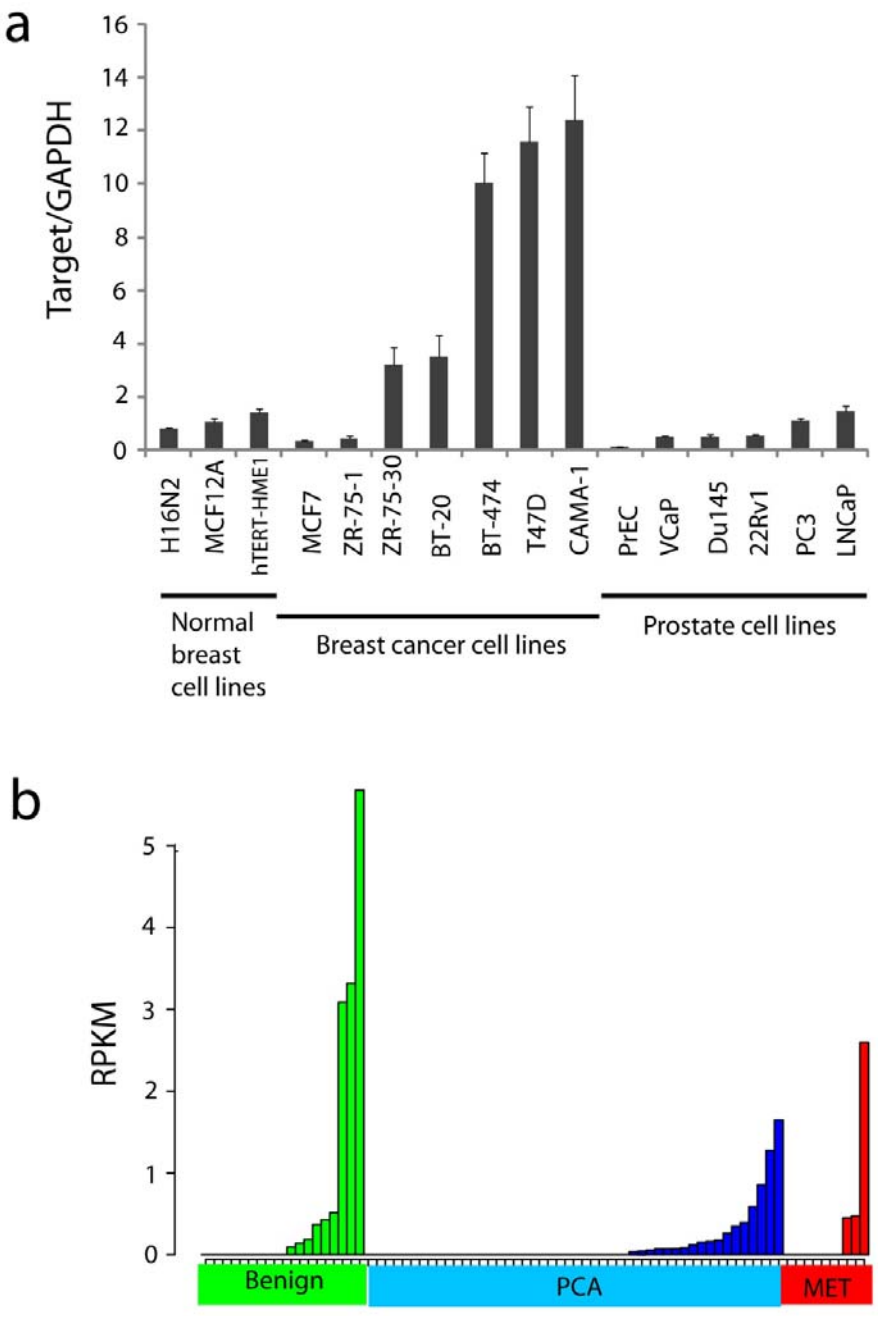


Figure 3.6: *HOTAIR* is not upregulated in prostate cancer. (**a**) qPCR analysis of a panel of breast cell lines and prostate cell lines for *HOTAIR* expression shows upregulation of *HOTAIR* in numerous breast cancer cell lines but not prostate cancer cell lines. Expression levels are normalized to *GAPDH* and the median expression of benign breast cell lines. (**b**) RPKM expression levels of *HOTAIR* in the prostate RNA-Seq cohort. Preferential upregulation of *HOTAIR* is not observed in prostate cancer and metastases samples.

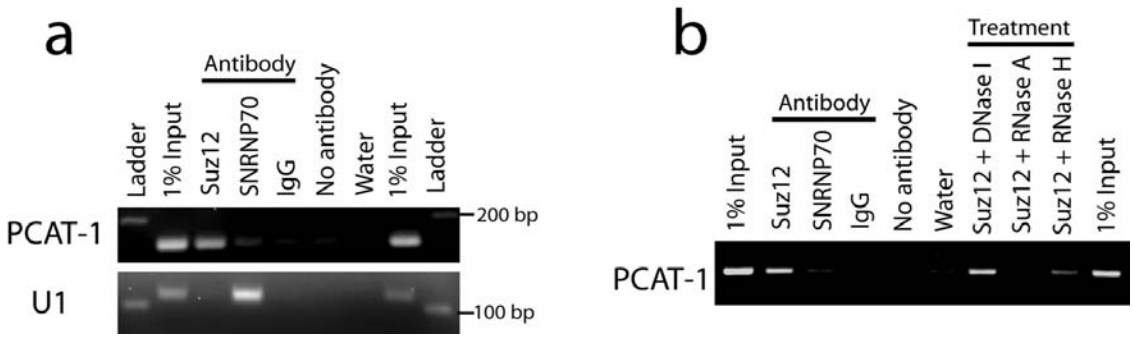


Figure 3.7: *PCAT-1* transcript associates with the PRC2 complex in VCaP cells. (**a**) RIP assays for *SUZ12* indicate binding of *PCAT-1* transcript to SUZ12 protein. *SNRNP70* serves as a control specifically binding to the *U1* RNA. Treating cells with IgG or no antibody serve as the negative controls. (**b**) The PRC2/*PCAT-1* interaction is stable to DNase I treatment but is abolished by RNase A treatment and partially degraded by RNase H treatment.

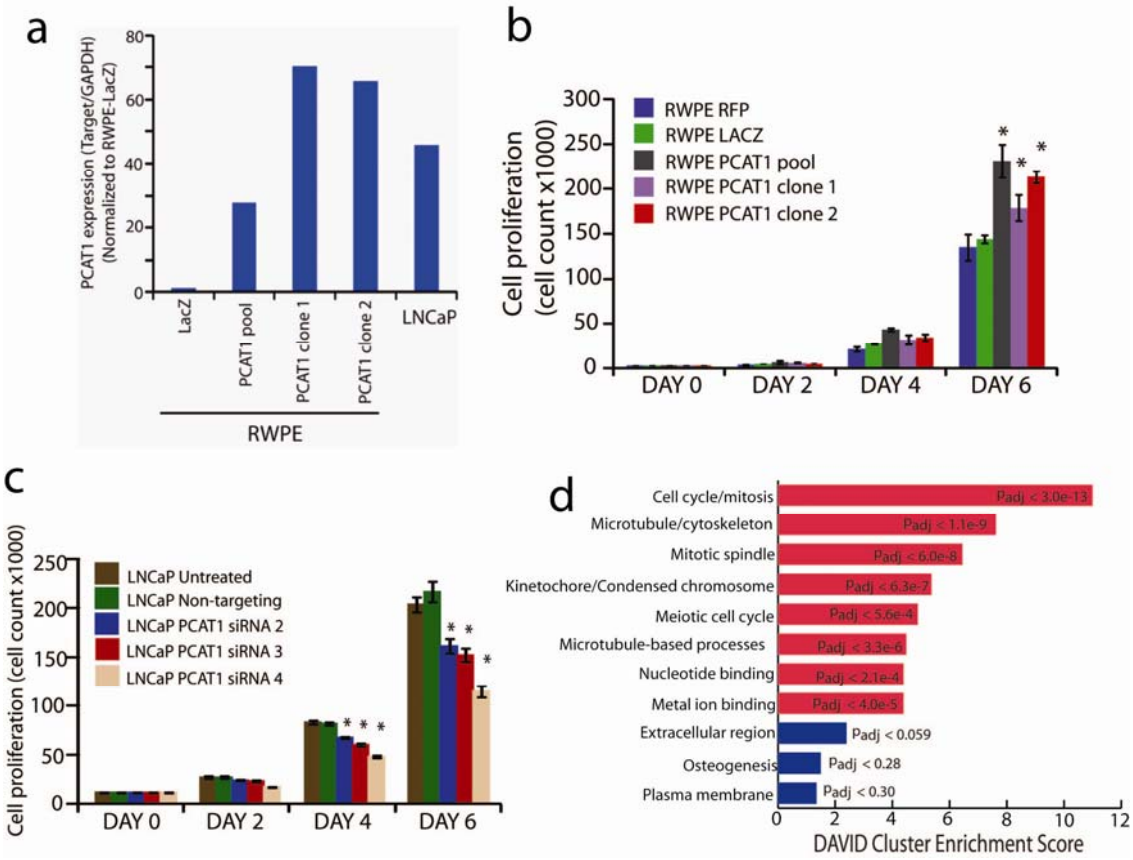


Figure 3.8: *PCAT-1* promotes cell proliferation. (**a**) Full length *PCAT-1* transcript was cloned into a lentiviral vector, and, following lentivirus production, RWPE benign immortalized prostate cells stably overexpressing *PCAT-1* were generated by selection with blasticidin. *PCAT-1* overexpression compared to the LacZ control cells was confirmed by qPCR. LNCaP serves as a positive control. (**b**) Cell proliferation assays for RWPE benign immortalized prostate cells stably infected with *PCAT-1* lentivirus or RFP and LacZ control lentiviruses. An asterisk (*) indicates $p \leq 0.02$ by a two-tailed Students t-test. (**c**) Cell proliferation assays in LNCaP using *PCAT-1* siRNAs. An asterisk (*) indicates $p \leq 0.005$ by a two-tailed Students t-test. (**d**) Gene ontology analysis of *PCAT-1* knockdown microarray data using the DAVID program. Blue bars represent the top hits for upregulated genes. Red bars represent the top hits for downregulated genes. All error bars in this figure are mean ± S.E.M.

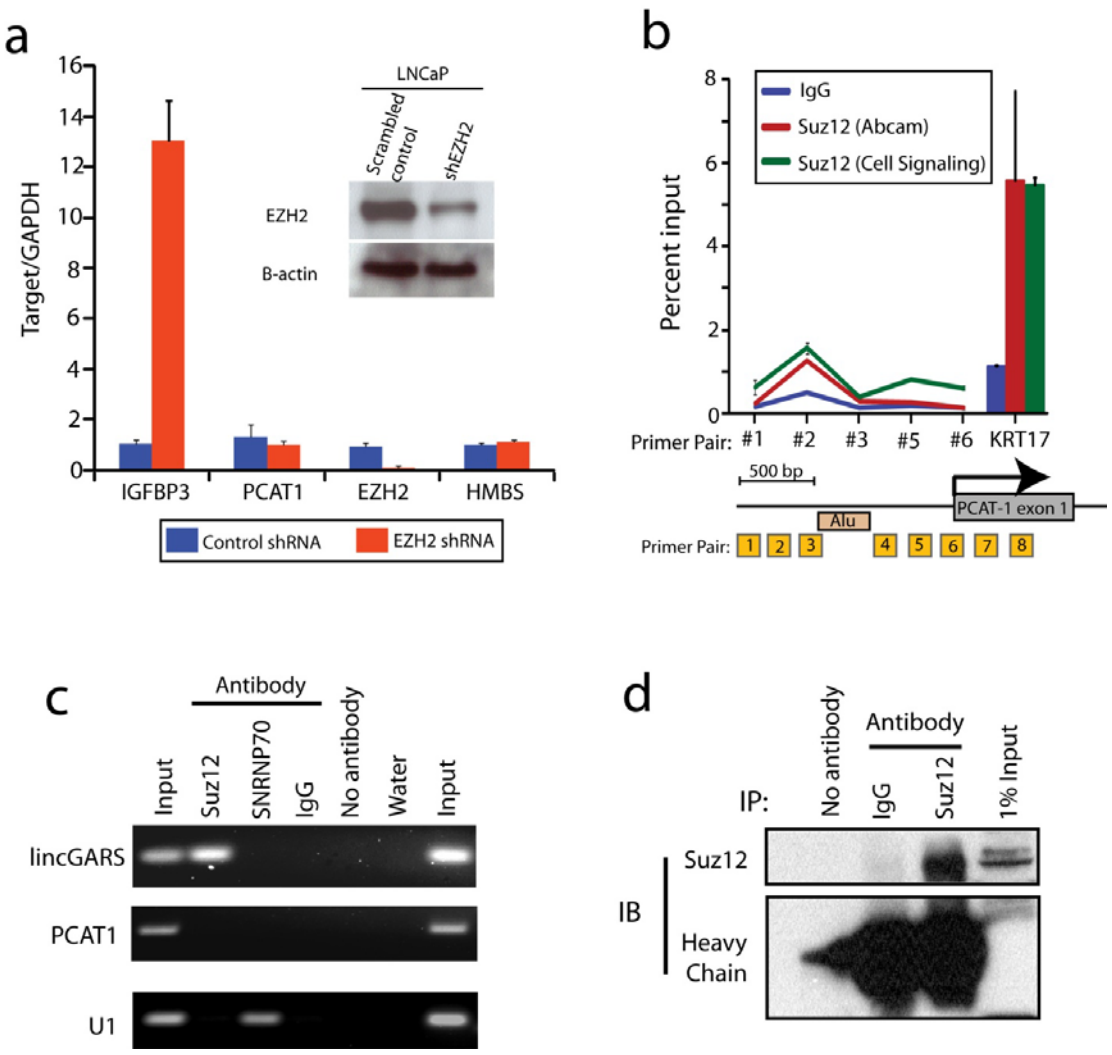


Figure 3.9: *PCAT-1* is not a PRC2 target in LNCaP cells. (**a**) LNCaP cells were infected with lentivirus for *EZH2* or scrambled control. qPCR showed no change in *PCAT-1* expression. *IGFBP3* and *HMBS* serve as positive and negative controls, respectively. (**b**) ChIP analysis of *SUZ12* in LNCaP cells does not show direct binding of PRC2 proteins upstream of the *PCAT-1* transcriptional start site (refer to fig. 3E for comparison). *KRT17* serves as a positive control. (**c**) RNA-IP analysis of in LNCaP cells does not indicate binding of PCAT-1 transcript to PRC2. lincGARS serves as a positive control. (**d**) A representative image of *SUZ12* RNA-IP pulldown efficiency. Equal fractions of LNCaP nuclear lysate were treated with either IgG or *SUZ12* antibodies and, following washing, probed for SUZ12 protein. Treating nuclear lysates with no antibody serves as a negative control.

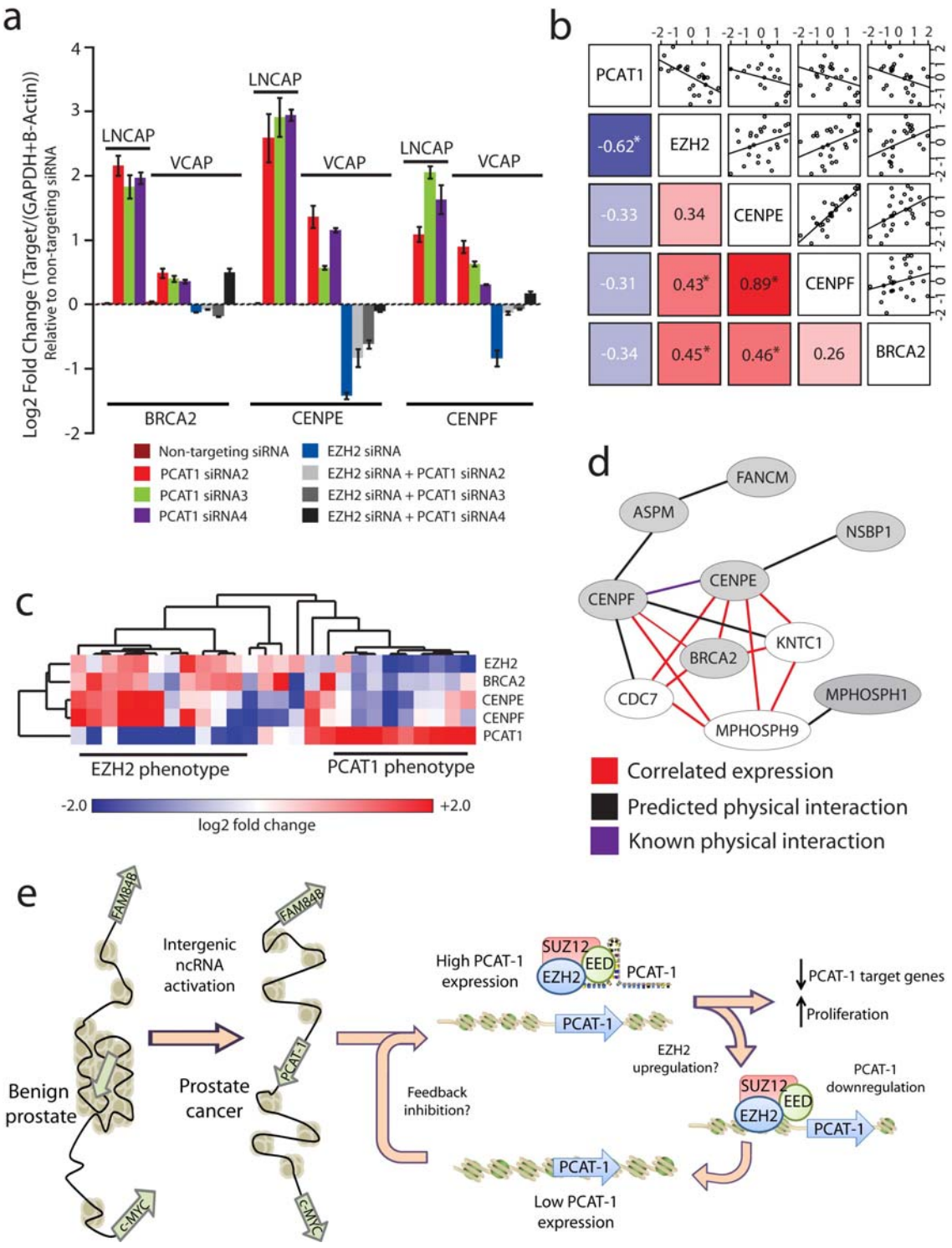


Figure 3.10: Prostate cancer tissues recapitulate *PCAT-1* signaling. (**a**) qPCR expression of three *PCAT-1* target genes after *PCAT-1* knockdown in VCaP and LNCaP cells, as well as following *EZH2* knockdown or dual *EZH2* and *PCAT-1* knockdown in VCaP cells. qPCR data were normalized to the average of (*GAPDH+B-Actin)* and

represented as fold change. Error bars represent mean ± S.E.M. (**b**) Standardized log2-transformed qPCR expression of a set of tumors and metastases with outlier expression of either *PCAT-1* or *EZH2*. The shaded squares in the lower left show Spearman correlation values between the indicated genes (* indicates $p < 0.05$). Blue and red indicate negative or positive correlation, respectively. The upper squares show the scatter plot matrix and fitted trendlines for the same comparisons. (**c**) A heatmap of *PCAT-1* target genes (*BRCA2*, *CENPF*, *CENPE*) in *EZH2*-outlier and *PCAT-1*-outlier patient samples (see **Fig. 4b**). Expression was determined by qPCR and normalized as in (**b**). (**d**) A predicted network generated by the HefaLMP program for 7 of 20 top upregulated genes following *PCAT-1* knockdown in LNCaP cells. Gray nodes are genes found following *PCAT-1* knockdown. Red edges indicate co-expressed genes; black edges indicate predicted protein-protein interactions; and purple edges indicate verified protein-protein interactions. **(e)** A proposed schematic representing *PCAT-1* upregulation, function, and relationship to PRC2.

TABLES

Table 3.1: Gene ontology concepts from PCAT-1 knockdown samples

Annotation Cluster 1	**Enrichment Score: 10.997759098136163**					
Category	Term	PValue	Fold Enrichment	Bonferroni	Benjamini	FDR
GOTERM_BP_FAT	GO:0022403~cell cycle phase	2.71E-16	6.02522351	3.03E-13	3.03E-13	3.66E-13
GOTERM_BP_FAT	GO:0000279~M phase	2.25E-15	6.67546375	3.03E-12	1.51E-12	3.64E-12
GOTERM_BP_FAT	GO:0022402~cell cycle process	1.01E-14	4.80692371	1.38E-11	4.59E-12	1.66E-11
SP_PIR_KEYWORDS	mitosis	6.82E-13	9.12491342	1.66E-10	8.29E-11	8.82E-10
GOTERM_BP_FAT	GO:0000278~mitotic cell cycle	2.57E-12	5.49083562	3.51E-09	8.77E-10	4.23E-09
GOTERM_BP_FAT	GO:0007049~cell cycle	3.80E-12	3.72268142	5.18E-09	1.04E-09	6.25E-09
GOTERM_BP_FAT	GO:0051301~cell division	6.75E-12	6.10506469	9.21E-09	1.53E-09	1.11E-08
SP_PIR_KEYWORDS	cell cycle	7.18E-12	5.11554227	1.75E-09	5.82E-10	9.29E-09
SP_PIR_KEYWORDS	cell division	7.26E-12	6.95774648	1.76E-09	4.41E-10	9.39E-09
GOTERM_CC_FAT	GO:0015630~microtubule cytoskeleton	1.01E-11	4.68954996	2.22E-09	1.11E-09	1.29E-08
GOTERM_BP_FAT	GO:0000280~nuclear division	8.33E-11	6.85309773	1.13E-07	1.62E-08	1.37E-07
GOTERM_BP_FAT	GO:0007067~mitosis	8.33E-11	6.85309773	1.13E-07	1.62E-08	1.37E-07
GOTERM_BP_FAT	GO:0000087~M phase of mitotic cell cycle	1.14E-10	6.73017221	1.55E-07	1.94E-08	1.87E-07
GOTERM_BP_FAT	GO:0048285~organelle fission	1.67E-10	6.58258071	2.28E-07	2.53E-08	2.74E-07
GOTERM_CC_FAT	GO:0005819~spindle	1.37E-09	8.77846791	3.00E-07	6.01E-08	1.74E-06
GOTERM_CC_FAT	GO:0044430~cytoskeletal part	3.53E-09	3.23343567	7.74E-07	9.67E-08	4.50E-06
GOTERM_BP_FAT	GO:0007059~chromosome segregation	1.17E-06	9.26437285	0.00159144	1.45E-04	0.00191904

Annotation Cluster 2	**Enrichment Score: 7.629201951158724**					
Category	Term	PValue	Fold Enrichment	Bonferroni	Benjamini	FDR
GOTERM_CC_FAT	GO:0043232~intracellular non-membrane-bounded organelle	2.96E-12	2.28855826	6.49E-10	6.49E-10	3.77E-09
GOTERM_CC_FAT	GO:0043228~non-membrane-bounded organelle	2.96E-12	2.28855826	6.49E-10	6.49E-10	3.77E-09
GOTERM_CC_FAT	GO:0015630~microtubule cytoskeleton	1.01E-11	4.68954996	2.22E-09	1.11E-09	1.29E-08
GOTERM_CC_FAT	GO:0005856~cytoskeleton	3.14E-09	2.71949761	6.89E-07	9.84E-08	4.00E-06
GOTERM_CC_FAT	GO:0044430~cytoskeletal part	3.53E-09	3.23343567	7.74E-07	9.67E-08	4.50E-06
SP_PIR_KEYWORDS	cytoskeleton	1.95E-06	3.18068411	4.75E-04	5.27E-05	0.00252634
GOTERM_CC_FAT	GO:0044450~microtubule organizing center part	3.38E-06	12.2898551	7.41E-04	6.17E-05	0.00430253
GOTERM_CC_FAT	GO:0005813~centrosome	1.01E-05	5.01514265	0.00221202	1.70E-04	0.01286134
GOTERM_CC_FAT	GO:0005814~centriole	2.54E-05	16.6507714	0.00554459	3.71E-04	0.03228875

GOTERM_CC_FAT	GO:0005815~microtubule organizing center	3.04E-05	4.49147314	0.00663476	3.92E-04	0.03865727

Annotation Cluster 3	**Enrichment Score: 6.459571965787644**					
Category	Term	PValue	Fold Enrichment	Bonferroni	Benjamini	FDR
GOTERM_CC_FAT	GO:0005819~spindle	1.37E-09	8.77846791	3.00E-07	6.01E-08	1.74E-06
GOTERM_CC_FAT	GO:0005874~microtubule	1.69E-06	4.98237368	3.69E-04	3.36E-05	0.00214593
GOTERM_CC_FAT	GO:0005876~spindle microtubule	1.81E-05	17.7991005	0.00395089	2.83E-04	0.02299052

Annotation Cluster 4	**Enrichment Score: 5.3750790495781615**					
Category	Term	PValue	Fold Enrichment	Bonferroni	Benjamini	FDR
GOTERM_CC_FAT	GO:0044427~chromosomal part	4.96E-11	5.47664629	1.09E-08	3.62E-09	6.31E-08
GOTERM_CC_FAT	GO:0005694~chromosome	3.02E-10	4.76879077	6.62E-08	1.66E-08	3.85E-07
GOTERM_CC_FAT	GO:0000775~chromosome, centromeric region	1.67E-09	9.71294998	3.66E-07	6.09E-08	2.12E-06
GOTERM_CC_FAT	GO:0000793~condensed chromosome	2.60E-08	8.73731884	5.70E-06	6.34E-07	3.31E-05
SP_PIR_KEYWORDS	kinetochore	7.88E-07	11.9275654	1.92E-04	2.74E-05	0.00101965
GOTERM_CC_FAT	GO:0000779~condensed chromosome, centromeric region	8.38E-07	11.7312253	1.83E-04	1.83E-05	0.00106551
GOTERM_BP_FAT	GO:0007059~chromosome segregation	1.17E-06	9.26437285	0.00159144	1.45E-04	0.00191904
GOTERM_CC_FAT	GO:0000776~kinetochore	2.87E-05	8.93807642	0.00626152	3.92E-04	0.03647611
GOTERM_CC_FAT	GO:0000777~condensed chromosome kinetochore	5.23E-05	10.3828086	0.0113946	6.03E-04	0.06654048
GOTERM_CC_FAT	GO:0000940~outer kinetochore of condensed chromosome	1.20E-04	38.2351047	0.02594508	0.00131352	0.15256829
GOTERM_BP_FAT	GO:0051493~regulation of cytoskeleton organization	0.100330664	2.82110602	1	0.87987522	82.3838413
GOTERM_BP_FAT	GO:0070507~regulation of microtubule cytoskeleton organization	0.106560286	5.36010144	1	0.87800896	84.2836176
GOTERM_BP_FAT	GO:0032886~regulation of microtubule-based process	0.137277936	4.59437266	1	0.91151043	91.1525468

Annotation Cluster 5	**Enrichment Score: 4.893802251111501**					
Category	Term	PValue	Fold Enrichment	Bonferroni	Benjamini	FDR
GOTERM_BP_FAT	GO:0051327~M phase of meiotic cell cycle	4.90E-06	7.8168146	0.0066582	5.57E-04	0.00804899
GOTERM_BP_FAT	GO:0007126~meiosis	4.90E-06	7.8168146	0.0066582	5.57E-04	0.00804899
GOTERM_BP_FAT	GO:0051321~meiotic cell cycle	5.82E-06	7.65728777	0.00789578	6.10E-04	0.00955096
SP_PIR_KEYWORDS	meiosis	1.90E-04	11.1323944	0.04520584	0.00288704	0.24589866

Annotation Cluster 6	**Enrichment Score: 4.490795812791411**					
Category	Term	PValue	Fold Enrichment	Bonferroni	Benjamini	FDR
GOTERM_BP_FAT	GO:0007017~microtubule-based process	2.41E-08	5.53584247	3.29E-05	3.29E-06	3.96E-05
GOTERM_BP_FAT	GO:0000226~microtubule cytoskeleton organization	1.19E-04	5.24765176	0.14921239	0.0089372	0.1945143
GOTERM_BP_FAT	GO:0007010~cytoskeleton organization	0.011785455	2.27398243	0.9999999	0.36164495	17.6918611

Annotation Cluster 7	**Enrichment Score: 4.398895114360735**					
Category	Term	PValue	Fold Enrichment	Bonferroni	Benjamini	FDR
SP_PIR_KEYWORDS	atp-binding	5.12E-06	2.29972952	0.00124353	1.24E-04	0.00662221

SP_PIR_KEYWORDS	nucleotide-binding	1.04E-05	2.0784525	0.00252414	2.11E-04	0.01344999
UP_SEQ_FEATURE	nucleotide phosphate-binding region:ATP	2.40E-05	2.46666468	0.0183188	0.00614394	0.03666554
GOTERM_MF_FAT	GO:0005524~ATP binding	3.49E-05	2.02959734	0.01136867	0.01136867	0.04720456
GOTERM_MF_FAT	GO:0032559~adenyl ribonucleotide binding	4.61E-05	2.00323894	0.01499727	0.00752696	0.06238072
GOTERM_MF_FAT	GO:0032553~ribonucleotide binding	5.56E-05	1.86793055	0.01806136	0.00454625	0.0752377
GOTERM_MF_FAT	GO:0032555~purine ribonucleotide binding	5.56E-05	1.86793055	0.01806136	0.00454625	0.0752377
GOTERM_MF_FAT	GO:0030554~adenyl nucleotide binding	6.12E-05	1.95071141	0.01987313	0.0040066	0.08285798
GOTERM_MF_FAT	GO:0000166~nucleotide binding	7.07E-05	1.7424561	0.02291439	0.00385605	0.09567995
GOTERM_MF_FAT	GO:0017076~purine nucleotide binding	7.12E-05	1.82823052	0.02309376	0.00333224	0.0964374
GOTERM_MF_FAT	GO:0001883~purine nucleoside binding	8.50E-05	1.92083453	0.02748603	0.00347779	0.11502619
GOTERM_MF_FAT	GO:0001882~nucleoside binding	9.85E-05	1.90744469	0.03177824	0.00358181	0.13326945

Annotation Cluster 8	**Enrichment Score: 4.398595309269413**					
Category	Term	PValue	Fold Enrichment	Bonferroni	Benjamini	FDR
SP_PIR_KEYWORDS	chelation	1.32E-06	52.1830986	3.22E-04	4.02E-05	0.0017119
PIR_SUPERFAMILY	PIRSF002564:metallothionein	3.84E-06	41.6969697	3.42E-04	3.42E-04	0.00417325
UP_SEQ_FEATURE	region of interest:Alpha	3.93E-06	41.6690141	0.00301834	0.00301834	0.00599575
UP_SEQ_FEATURE	region of interest:Beta	3.93E-06	41.6690141	0.00301834	0.00301834	0.00599575
INTERPRO	IPR003019:Metallothionein superfamily, eukaryotic	5.75E-06	38.3942688	0.00227046	0.00227046	0.00801144
INTERPRO	IPR000006:Metallothionein, vertebrate	5.75E-06	38.3942688	0.00227046	0.00227046	0.00801144
INTERPRO	IPR018064:Metallothionein, vertebrate, metal binding site	5.75E-06	38.3942688	0.00227046	0.00227046	0.00801144
SP_PIR_KEYWORDS	metal-thiolate cluster	6.07E-06	37.9513444	0.00147319	1.34E-04	0.00784606
UP_SEQ_FEATURE	metal ion-binding site:Divalent metal cation; cluster B	6.11E-06	37.8809219	0.00469478	0.00235015	0.00933359
UP_SEQ_FEATURE	metal ion-binding site:Divalent metal cation; cluster A	6.11E-06	37.8809219	0.00469478	0.00235015	0.00933359
SP_PIR_KEYWORDS	metal binding	6.61E-05	21.971831	0.01593785	0.00114693	0.08547079
SP_PIR_KEYWORDS	cadmium	8.95E-05	41.7464789	0.02151373	0.00144885	0.11568263
GOTERM_MF_FAT	GO:0046870~cadmium ion binding	2.18E-04	31.6836601	0.06889543	0.00711295	0.29434429
SP_PIR_KEYWORDS	acetylated amino end	0.006365991	5.11181374	0.78815082	0.06261547	7.92750279
SP_PIR_KEYWORDS	copper	0.029163321	5.9637827	0.99924748	0.20128604	31.8033936
GOTERM_MF_FAT	GO:0005507~copper ion binding	0.051786926	4.7289045	0.99999997	0.5474271	51.3371962

Annotation Cluster 9	**Enrichment Score: 3.6988709642933344**					
Category	Term	PValue	Fold Enrichment	Bonferroni	Benjamini	FDR
GOTERM_BP_FAT	GO:0007017~microtubule-based process	2.41E-08	5.53584247	3.29E-05	3.29E-06	3.96E-05
GOTERM_CC_FAT	GO:0005874~microtubule	1.69E-06	4.98237368	3.69E-04	3.36E-05	0.00214593
GOTERM_BP_FAT	GO:0007018~microtubule-based movement	1.29E-05	6.94827964	0.01744449	0.00125624	0.02120218
SP_PIR_KEYWORDS	microtubule	1.56E-05	4.86730247	0.00377475	2.91E-04	0.0201259
GOTERM_MF_FAT	GO:0003777~microtubule motor activity	4.90E-05	8.22952211	0.01595905	0.00534824	0.06641223

UP_SEQ_FEATURE	domain:Kinesin-motor	0.001694818	9.69046839	0.72913019	0.22989114	2.55741208
GOTERM_MF_FAT	GO:0003774~motor activity	0.001980066	4.49413619	0.47800924	0.05273411	2.6488895
SP_PIR_KEYWORDS	motor protein	0.004866667	4.46145576	0.6944027	0.05245967	6.11445689
INTERPRO	IPR019821:Kinesin, motor region, conserved site	0.013166214	8.04451346	0.99467421	0.52663331	16.8500322
INTERPRO	IPR001752:Kinesin, motor region	0.013166214	8.04451346	0.99467421	0.52663331	16.8500322
SMART	SM00129:KISc	0.018083379	7.0952381	0.8606657	0.48157491	18.5745625

Annotation Cluster 10	**Enrichment Score: 3.0093099494098143**					
Category	Term	PValue	Fold Enrichment	Bonferroni	Benjamini	FDR
SP_PIR_KEYWORDS	meiosis	1.90E-04	11.1323944	0.04520584	0.00288704	0.24589866
GOTERM_CC_FAT	GO:0044454~nuclear chromosome part	4.95E-04	5.68786681	0.10269743	0.00491344	0.62741764
GOTERM_CC_FAT	GO:0000228~nuclear chromosome	5.47E-04	4.80907372	0.1129476	0.00519737	0.69370818
GOTERM_CC_FAT	GO:0000794~condensed nuclear chromosome	0.017815631	7.16908213	0.98048938	0.13567574	20.4400109

Annotation Cluster 11	**Enrichment Score: 2.875446628462769**					
Category	Term	PValue	Fold Enrichment	Bonferroni	Benjamini	FDR
GOTERM_BP_FAT	GO:0000226~microtubule cytoskeleton organization	1.19E-04	5.24765176	0.14921239	0.0089372	0.1945143
GOTERM_BP_FAT	GO:0007052~mitotic spindle organization	9.24E-04	20.0110454	0.7163991	0.05567155	1.50693476
GOTERM_BP_FAT	GO:0007051~spindle organization	0.021578503	6.67034846	1	0.49916371	30.1108428

REFERENCES

1. T. Oosumi, W. R. Belknap, B. Garlick, *Nature* **378**, 672 (Dec 14, 1995).
2. H. M. Robertson, K. L. Zumpano, A. R. Lohe, D. L. Hartl, *Nat Genet* **12**, 360 (Apr, 1996).
3. C. G. Kleer *et al.*, *Proc Natl Acad Sci U S A* **100**, 11606 (Sep 30, 2003).
4. S. Varambally *et al.*, *Nature* **419**, 624 (Oct 10, 2002).
5. N. Ahmadiyeh *et al.*, *Proc Natl Acad Sci U S A* **107**, 9742 (May 25, 2010).
6. A. A. Al Olama *et al.*, *Nat Genet* **41**, 1058 (Oct, 2009).
7. R. Beroukhim *et al.*, *Nature* **463**, 899 (Feb 18, 2010).
8. J. Gudmundsson *et al.*, *Nat Genet* **39**, 631 (May, 2007).
9. J. Sotelo *et al.*, *Proc Natl Acad Sci U S A* **107**, 3001 (Feb 16, 2010).
10. B. S. Taylor *et al.*, *Cancer Cell* **18**, 11 (Jul 13, 2010).
11. R. A. Gupta *et al.*, *Nature* **464**, 1071 (Apr 15, 2010).
12. C. Huttenhower *et al.*, *Genome Res* **19**, 1093 (Jun, 2009).
13. B. Laxman *et al.*, *Cancer Res* **68**, 645 (Feb 1, 2008).
14. D. Hessels *et al.*, *Eur Urol* **44**, 8 (Jul, 2003).
15. M. A. Rubin *et al.*, *Clin Cancer Res* **6**, 1038 (Mar, 2000).
16. S. A. Tomlins *et al.*, *Neoplasia* **10**, 177 (Feb, 2008).
17. V. G. Tusher, R. Tibshirani, G. Chu, *Proc Natl Acad Sci U S A* **98**, 5116 (Apr 24, 2001).
18. G. Dennis, Jr. *et al.*, *Genome Biol* **4**, P3 (2003).
19. J. Yu *et al.*, *Cancer Res* **67**, 10657 (Nov 15, 2007).

CHAPTER 4

A lineage-specific long non-coding RNA controls homologous recombination in cancer

SUMMARY

The dysregulation of repair pathways that monitor and process double-stranded DNA breaks (DSBs) is essential to tumorigenesis in many cancers. However, while direct mutation of components of the DSB repair machinery is common in hereditary cancer, mechanisms of impaired DSB repair in sporadic cancers remains incompletely understood (*1-3*). Here, we describe the first role for a lineage-specific long noncoding RNA (lncRNA) in DSB repair. We identify *PCAT-1*, a prostate cancer-specific lncRNA, which regulates prostate cell response to genotoxic stress. *PCAT-1* expression results in a functional deficiency in homologous recombination (HR) through its repression of the *BRCA2* tumor suppressor, which, in turn, imparts a high sensitivity to small molecule inhibitors of *PARP1*, both *in vitro* and *in vivo*. These effects were mediated through repression of the *BRCA2* 3'UTR by *PCAT-1* in a microRNA-like manner. Importantly, these findings represent lineage-specific functions of *PCAT-1* that were not observed in breast cancer cells. Our observations thus identify *PCAT-1* as the first example of a lineage-specific lncRNA contributing to regulation of the DSB repair machinery in cancer.

INTRODUCTION

The uncontrolled accumulation of DSBs represents a putative Achilles heel for cancer cells, since these lesions are toxic and their repair requires re-ligation of completely detached genetic material (*4*). Several mechanisms, such as non-homologous end joining (NHEJ), microhomology-mediated end joining (MMEJ) and homologous recombination (HR), contribute to DSB repair and are employed variously during the cell cycle, depending on whether a specific DSB harbors either large, small, or no stretches (NHEJ, MMEJ, and HR, respectively) of complementary DNA sequences on the two fragments of broken DNA (*4, 5*). In particular, the lethality of excess DSBs has been exploited for the therapeutic treatment of breast and ovarian cancers harboring *BRCA1/2* mutations, which leads to defective HR and increased DSBs (*6*). These cancers exhibit synthetic lethality when treated with small molecule inhibitors of the *PARP1* DNA repair enzyme, whose inhibition prevents a second method of DNA repair and leads to gross collapse of cellular DNA maintenance (*7-9*).

Recently, long noncoding RNAs (lncRNAs) have emerged as new layer of cell biology, contributing to diverse biological processes such as pluripotency maintenance in stem cells (*10*), X chromosome inactivation in women (*11*), and neurodegenerative disease (*12*). In cancer, aberrant expression of lncRNAs is associated with cancer progression (*13*), and overexpression of oncogenic lncRNAs can promote tumor cell proliferation and metastasis through transcriptional regulation of target genes (*14-16*). Yet, while previous studies in cancer have identified lncRNAs induced by genotoxic stress (*17, 18*), no study to date has described a role for lncRNAs in the regulation of DSB repair. We have previously reported the systematic discovery of novel lncRNAs in

prostate cancer (*13*). Of note, we characterized *PCAT-1* as a prostate-specific lncRNA upregulated in subsets of prostate cancer patients and implicated in tumor growth (*13*). *PCAT-1* was found to repress *in vitro* genes associated with DNA maintenance, DNA repair, and mitotic spindle assembly, and negative correlations between *PCAT-1* and its DNA repair gene targets were similarly observed in human prostate cancer tissues. Here, we describe a direct link between *PCAT-1* and the capacity for prostate cells to engage in DSB repair.

RESULTS

We noted that *PCAT-1* was a repressor of *BRCA2*, a tumor suppressor gene integral to the proper induction of the HR pathway (*6*). We hypothesized that *PCAT-1*-mediated repression of *BRCA2* would lead to impaired HR upon genotoxic stress and increased cell susceptibility to PARP-1 inhibitors (**Figure 4.1a**). To pursue this hypothesis, we generated a panel of four *in vitro* cell culture model systems: *PCAT-1* overexpression in Du145 prostate cancer cells (which lack endogenous expression of this lncRNA), *PCAT-1* overexpression in RWPE benign prostate cells (which lack endogenous expression of this lncRNA), stable knockdown of *PCAT-1* in LNCaP prostate cancer cells (which harbor high endogenous levels of *PCAT-1* expression), and *PCAT-1* overexpression in MCF7 breast cancer cells (which lack endogenous expression of *PCAT-1*) (**Figure 4.1b**).

Western blot analysis of these four isogenic models uniformly revealed strong downregulation of BRCA2 protein levels in RWPE and Du145 prostate cells and upregulation of BRCA2 in LNCaP sh-*PCAT-1* cells (**Figure 4.1c**). Consistent with the

lineage-specific expression of *PCAT-1* exclusive in prostate(*13*), MCF7-*PCAT-1* cells showed no change in BRCA2 levels, suggesting that *PCAT-1* also functions in a lineage-specific manner (**Figure 4.1c**).

Importantly, *BRCA2* inactivation impairs HR of DSBs and serves as a predictive biomarker for response to treatment with inhibitors of the *PARP1* DNA repair enzyme through synthetic lethality that results from joint inactivation of two DNA repair pathways (HR via *BRCA2* inactivation, and base excision repair via *PARP1* inhibition) (*19*). Accordingly, treatment of our four isogenic cell lines with either a PARP1 inhibitor, Olaparib (*9*) or ABT-888 (*20*), or radiation, resulted in modulation of RAD51 foci formation, which is a component of the HR pathway and a marker for engagement of the HR machinery (*21*). Specifically, *PCAT-1* overexpression decreased RAD51 foci formation post-therapy and *PCAT-1* knockdown increased RAD51 foci formation post-therapy in prostate cells (**Figure 4.1d** and **Figure 4.2**).

Because *PCAT-1* impairs HR, genotoxic stress of *PCAT-1*-expressing cells should lead to an accumulation of DSBs, which can be visualized using phospho-gammaH2AX (p-gH2AX) foci, a marker of double-stranded DNA breaks that have not been repaired (*4*). To test this, we treated our isogenic Du145 and LNCaP cell line models with Olaparib, ABT-888, or radiation. As predicted, *PCAT-1* overexpression in Du145 led to an increase in p-gH2AX foci under stress conditions (**Figure 4.3**), indicating that *PCAT-1* impairs DSB repair in these cells. Similarly, LNCaP cells with *PCAT-1* knockdown displayed decreased levels of p-gH2AX foci (**Figure 4.3**). Finally, we also evaluated the ability of our isogenic cell lines to sustain growth in clonogenic survival assays, a gold-standard assay for cell viability following genotoxic stress, after treatment of cells with

PARP1 inhibition or radiation. We found that *PCAT-1* expression led to decreased cell survival in Du145 and RWPE cells, whereas *PCAT-1* knockdown increased LNCaP cell survival, in these assays (**Figure 4.4** and **Figure 4.5**). MCF7 cells overexpressing *PCAT-1* did not show a change in clonogenic survival in these conditions. Together, these data indicate that *PCAT-1* expression may impart cell sensitivity to genotoxic stress by decreasing the HR response through downregulation of BRCA2.

Because *PCAT-1* expressing cells exhibit reduced HR efficiency when challenged, we investigated whether PARP1 inhibition selectively killed *PCAT-1*-expressing cells. Following treatment with two PARP1 inhibitors (Olaparib or ABT-888), we observed that knockdown of *PCAT-1* in LNCaP cells prevented cell death, whereas overexpression of *PCAT-1* in Du145 and RWPE prostate cells, but not MCF7 breast cells, increased cell death in response to PARP inhibition (**Figure 4.6a-f**). This change in cell sensitivity to PARP1 inhibitors was quite profound, as the IC50 for LNCaP and Du145 cells changed by at least five-fold when *PCAT-1* expression was either exogenously introduced or depleted by knockdown (**Figure 4.6a-d**). Similar results were observed in RWPE cells overexpressing *PCAT-1* (**Figure 4.7**), and treatment with the mTOR inhibitor rapamycin did not produce this effect (**Figure 4.8**), supporting specificity of *PCAT-1* in inducing sensitivity to PARP1 inhibitors. Prostate cell sensitivity to PARP1 inhibitors was mediated through BRCA2, as knockdown of *BRCA2* in LNCaP sh*PCAT-1* cells (which have increased levels of BRCA2) rescued the sensitivity of these cells to PARP1 inhibition in a dose-dependent manner according to the efficiency of the knockdown (**Figure 4.6g**) and reduced RAD51 foci post-treatment (**Figure 4.6g**). To rule out the possibility that altered cell cycle distributions contribute to

our observed phenotypes, we performed flow cytometry on our isogenic RWPE, LNCaP, and Du145 cell models, which demonstrated no change in cell cycle in our model systems (**Figure 4.9**).

To evaluate the contribution of *PCAT-1* to PARP inhibitor response *in vivo*, we generated xenografts of Du145 cells expressing either empty vector control or *PCAT-1*. We observed that Du145-*PCAT-1* cells grew significantly more rapidly in severe combined immunodeficient (SCID) mice, consistent with our previous findings that *PCAT-1* accelerates prostate cell proliferation *in vitro* (**Figure 4.10a**) (*13*). Moreover, Du145-*PCAT-1* xenografts showed marked susceptibility and tumor regression following intra-peritoneal (IP) administration of Olaparib, whereas Du145-control cells showed only a subtle change in growth while the drug was administered, indicating that the background effect of Olaparib therapy—possibly due to its effects on other members of the PARP family (*22*)—is small (**Figure 4.10a**). Mice in all groups of treatment maintained their body weights and showed no evidence of weight loss (**Figure 4.10b**).

Importantly, Du145 xenografts retained both *PCAT-1* expression and BRCA2 repression (**Figure 4.10c**). To investigate *PCAT-1* signaling under control-treated (DMSO) and Olaparib-treated conditions, we also observed *in vivo* upregulation of *PCAT-1*-induced target genes (**Figure 4.10c**) also observed in RWPE-*PCAT-1* overexpressing cells (**Figure 4.11**) (*13*). These data suggest that *PCAT-1* is mechanistically linked to increased prostate cell sensitivity to PARP1 inhibitors via its repression of BRCA2 both *in vitro* and *in vivo*.

While many lncRNAs are noted to regulate gene transcription through epigenetic mechanisms (*14-16, 23, 24*), we did not observe evidence for this possibility with *PCAT-*

1. Although *PCAT-1* regulated *BRCA2* transcript levels *in vitro* (**Figure 4.12a**), treatment of RWPE-*LacZ* and RWPE-*PCAT-1* cells with the DNA methylation inhibitor 5-azacytidine (5-aza), the histone deacetylase inhibitor TSA, or both did not reveal enhanced epigenetic regulation of *BRCA2* mRNA in *PCAT-1*-expressing cells (**Figure 4.12b**), although there was a baseline regulation of *BRCA2* in both cell lines when 5-aza and TSA were combined. Together, these results suggest that epigenetic repression of *BRCA2* is not the primary mechanism of *PCAT-1*. Similarly, a recent report suggests that lncRNAs containing Alu elements in their transcript sequence may utilize these repetitive sequences to regulate target gene mRNAs via STAU1-dependent degradation (*25*). Although *PCAT-1* harbors an Alu element from bps 1103 – 1402, knockdown of STAU1 in LNCaP or VCaP cells, which endogenously harbor *PCAT-1*, did not alter *BRCA2* levels (**Figure 4.12c**).

To determine whether *PCAT-1* may function in a manner more analogous to microRNAs, which regulate mRNA levels post-transcriptionally (*26*), we generated a luciferase construct of the *BRCA2* 3'UTR (**Figure 4.13a**). Surprisingly, we found that RWPE-*PCAT-1* cells, but not control RWPE-*LacZ* cells, were able to directly repress the activity of the wild-type *BRCA2* 3'UTR construct (**Figure 4.13a**). Further, to map a region of *PCAT-1* required for repression of the *BRCA2* 3'UTR, we additionally generated a series of *PCAT-1* deletion constructs and overexpressed these in RWPE cells (**Figure 4.13b**). We generated these constructs to establish whether the 3' end of *PCAT-1*, which contains portions of ancestral transposase and Alu repeat elements (**Figure 4.13b**) (*13*), or the 5' end of *PCAT-1*, which consists of non-repetitive DNA sequences, was required for *BRCA2* repression. We observed that the 5' end of *PCAT-1* was

sufficient to downregulate the *BRCA2* 3'UTR luciferase signal as well as endogenous *BRCA2* transcript levels (**Figure 4.13c,d**), and for this regulation, the first 250 bps of the *PCAT-1* gene were required. By contrast, the 3' end of *PCAT-1* was expendable for this regulation. Importantly, the 5' end of *PCAT-1* was similarly sufficient to sensitize RWPE cells to Olaparib treatment *in vitro*, resulting in decreased cell survival in cells expressing the 5' end of *PCAT-1* (**Figure 4.13e**). Together, these results indicate that *PCAT-1* overexpression is able to directly repress the activity of the *BRCA2* 3'UTR, and that this repression required the 5' end of *PCAT-1*.

Of note, we previously reported that the *TMPRSS2-ERG* gene fusion in prostate cancer may also serve as a biomarker for PARP1 inhibitor therapy, due to physical interaction between the ERG protein and PARP-1 (*27*). Interestingly, we observed that prostate cancer tissue samples with the highest levels of *PCAT-1* tended to be ERG-negative, although this analysis did not reach statistical significance (**Figure 4.14**). Thus, *PCAT-1* may identify a unique set of patients for whom PARP1 inhibitor therapy may have a mechanistic basis.

DISCUSSION

To our knowledge, this is the first report of a lineage-specific lncRNA being involved in the DSB repair process. While several lncRNAs have been reported to be transcriptionally induced by genotoxic stress (*17, 18*), *PCAT-1* is the first example of a lncRNA implicated in the functional regulation of the DNA repair machinery (**Figure 4.15)**. These results expand the potential roles for lncRNAs in cancer biology and contrast strikingly with previous reports that lncRNAs operate epigenetically through

chromatin modifying complexes (*28*). Indeed, epigenetic regulation likely represents only a one of numerous mechanisms for lncRNA function (*15, 25, 29-31*). Supporting this notion, we do not observe compelling evidence that *PCAT-1* functions in an epigenetic manner, but rather it may exhibit microRNA-like features. Our characterization of a *PCAT-1*-BRCA2 axis therefore provides the first report of novel mechanism of HR regulation in human cancer.

These data suggest that lncRNAs may have a more widespread role in genome maintenance and DNA repair than previously appreciated. A report in fission yeast found that locations of meiotic recombination are enriched for the transcription of local ncRNAs (*32*). Moreover, lncRNAs such as *Xist* and *HOTTIP* have been shown to contribute to large-scale genome architecture (*23, 33*). A detailed analysis of the *CDKN1A* promoter uncovered a p53-induced lncRNA that contributed to transcriptional regulation of pro-apoptotic genes (*34*), but this gene was not shown to be involved in DSB repair. Of note, however, Adamson et al. nominated the RNA-binding protein *RBMX* as a novel component of the homologous recombination pathway (*35*), suggesting that RNA-protein interactions may be integral to this process.

This work sheds insight onto potential mechanisms of impaired DSB repair in cancers lacking an inactivating mutation in canonical DSB repair proteins. Thus, our studies have uncovered a novel mechanism of "BRCA-ness"—the clinical observation that many cancers lacking BRCA1/BRCA2 mutations exhibit the clinical features of impaired DSB repair (*3, 36*). While *PCAT-1* expression and function appear to be restricted to prostate cancer cells, we hypothesize that other cancers with a BRCA-like phenotype may harbor lineage-specific lncRNAs involved in the regulation and execution

of proper HR and other forms of DSB repair. Finally, future clinical trials examining the efficacy of PARP1 inhibitors in prostate cancer will provide critical information as to whether *PCAT-1* may serve as a predictive biomarker for patient response to PARP1 inhibitor therapy.

MATERIALS AND METHODS

Cell lines

All cell lines were obtained from the American Type Culture Collection (Manassas, VA). Cell lines were maintained using standard media and conditions. Du145-derived cell lines were maintained in DMEM supplemented with 10% FBS (Invitrogen) and 1% penicillin-streptomycin (Invitrogen) in a 5% CO_2 cell culture incubator. RWPE-derived cell lines were maintained in KSF (Invitrogen) supplemented with Bovine Pituitary Extract, Epidermal Growth Factor and 1% penicillin-streptomycin in a 5% CO_2 cell culture incubator. MCF7-derived cell lines were maintained in DMEM supplemented with 10% FBS, 10mg/mL insulin, and 1% penicillin-streptomycin in a 5% CO_2 cell culture incubator. LNCAP-derived cell lines were maintained in RPMI 1640 (Invitrogen) supplemented with 10% FBS and 1% penicillin-streptomycin in a 5% CO_2 cell culture incubator.

PCAT-1 or control-expressing cell lines were generated by cloning PCAT-1 or control LacZ into the pLenti6 vector (Invitrogen). After confirmation of the insert sequence, lentiviruses were generated at the University of Michigan Vector Core and transfected intoMCF7, RWPE, or Du145 cells. Stably-transfected cells were selected using blasticidin (Invitrogen).

For LNCAP cells with stable knockdown of PCAT-1, cells were seeded at 50-60% confluency, incubated overnight, and transfected with PCAT1 or non-targeting shRNA lentiviral constructs for 48 hours. GFP+ cells were drug-selected using 1 ug/mL puromycin. PCAT1 shRNAs were customed generated by Systems Biosciences using the

following sequences: shRNA 1 GCAGAAACACCAAUGGAUAUU; shRNA 2 AUACAUAAGACCAUGGAAAU.

Cell line treatments

Cells were plated in 6-well plates and allowed to grow for 24 hours. For radiation experiments, after 24 hours, cells were subjected to 2 Gray (Gy), 4 Gy, or 6 Gy of radiation (Philips RT250, Kimtron Medical). RAD51 and gamma-H2AX staining was performed 6 hours or 24 hours post-treatment, and RNA was collected 24 hours post-treatment.

For Olaparib or ABT-888 treatments, cells were treated with the indicated doses of drug or DMSO control. RAD51 and gamma-H2AX staining was performed 6 hours post-treatment, and RNA was collected 24 hours post-treatment.

For epigenetic drug treatments, RWPE-LacZ and RWPE-PCAT1 cells were treated with 20uM 5'deoxyazacytidine (Sigma), 500 nM HDAC inhibitor Trichostatin A (TSA) (Sigma), or both 5'deoxyazacytidine and TSA. 5'deoxyazacytidine treatments were performed for 6 days with media and drug re-applied every 48 hours. TSA treatments were performed for 48 hours. DMSO treatments were performed for 6 days.

RNA isolation; cDNA synthesis; and PCR experiments

Total RNA was isolated using Trizol and an RNeasy Kit (Invitrogen) with DNase I digestion according to the manufacturer's instructions. RNA integrity was verified on an Agilent Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA). cDNA was synthesized from total RNA using Superscript III (Invitrogen) and random primers

(Invitrogen). Quantitative Real-time PCR (qPCR) was performed using Power SYBR Green Mastermix (Applied Biosystems, Foster City, CA) on an Applied Biosystems 7900HT Real-Time PCR System.

The relative quantity of the target gene was completed for each sample using the ΔΔCt method by the comparing mean Ct of the gene to the average Ct of the geometric mean of two housekeeping genes, GAPDH and HMBS. All oligonucleotide primers were synthesized by Integrated DNA Technologies (Coralville, IA). The primer sequences for the transcript analyzed are provided in **Appendix 2**.

Xenografts

Five week-old male SCID mice (CB.17. SCID), were purchased from Charles River, Inc. (Charles River Laboratory, Wilmington, MA). 1 x 10^6 Du145-control or Du145-PCAT1 stable cells were resuspended in 100µl of saline with 50% Matrigel (BD Biosciences, Becton Drive, NJ) and were implanted subcutaneously into the left and right flank regions of the mice. Mice were anesthetized using a cocktail of Xylazine (80-120 mg/kg, IP) and Ketamine (10mg/kg, IP) for chemical restraint before tumor implantation. All tumors were staged for two weeks before starting the drug treatment. At the beginning of the third week, mice with tumors (10 tumors per treatment group, average size 150-200 mm3) were treated with Olaparib (100mg/kg, IP twice daily five times per week) or an equal volume of DMSO control. Olaparib was obtained from Axon Medchem (Groningen, The Netherlands). Growth in tumor volume was recorded weekly by using digital calipers and tumor volumes were calculated using the formula (π/6) (L × W2), where L = length of tumor and W = width. Loss of body weight during the course

of the study was also monitored weekly; all changes in mouse weights occurred within tolerable limits as set by the University Committee on Use and Care of Animals (UCUCA) at the University of Michigan. At the end of the xenograft studies, mice were sacrificed, and mouse xenograft tumors were subsequently harvested, maintained in formalin, and subsequently embedded in paraffin. All procedures involving mice were approved by the UCUCA at the University of Michigan and conform to their relevant regulatory standards.

Immunofluorescence

1 x 10^5cells were plated on sterile coverslips in 12-well plates and cultured for 24 hours. After treated for 6 hour or 24 hour, cells were fixed with 4% paraformaldehyde for 15min at room temperature, and then permeabilized with 0.5% Triton X-100 for 20min. Cells were blocked for 1 hour in blocking buffer (0.5% BSA, 2.5% goat serum and 0.05% Triton X-100 in PBS), and then incubated with primary antibody against Rad51(Santa Cruz sc-8349) at 1:400 or gamma-H2AX (Millipore JBW301) at 1:1,000, followed by secondary antibody Alexa Fluor 488 goat anti-rabbit IgG(Invitrogen A11008) or Alexa Fluor 594 goat anti-mouse IgG(Invitrogen A11005) at 1:2,000. Nuclei were stained with DAPI(4',6-diamidino-2-phenylindole)(Invitrogen P36931) and coverslips were kept in dark overnight. The cells were observed and images were recorded by fluorescent microscope (Olympus 1 X 71), with the software Olympus DP Controller 2002 (Olympus Optical Co. Ltd). A minimum of 100 cells were analyzed on each coverslip and the number of foci on digital images was determined using the image analysis software package Image J(version 1.42, available from Research Services

Branch of NIH). The quantitative results were shown as the average cell fraction with more than 5 or 10 foci of at least three independent experiments.

Luciferase Assays

RWPE-LacZ, RWPE-PCAT1 pool, and RWPE-PCAT1 clone 2 cells were transfected with full length BRCA2 3'UTR or BRCA2 3'UTR deletion luciferase constructs as well as pRL-TK vector as internal control for luciferase activity. Following 2 days of incubation, the cells were lysed and luciferase assays conducted using the dual luciferase assay system (Promega, Madison, WI, USA). Each experiment was performed in quadruplicate.

Immunoblot Analysis

Cells were lysed in RIPA lysis buffer (Sigma, St. Louis, MO) and briefly sonicated for homogenization. Aliquots of each protein extract were boiled in sample buffer, size fractionated by SDS-PAGE at 4C, and transferred onto Polyvinylidene Difluoride membrane (GE Healthcare, Piscataway, NJ). The membrane was then incubated at room temperature for 1-2 hours in blocking buffer [Tris-buffered saline, 0.1% Tween (TBS-T), 5% nonfat dry milk] and incubated at 4C with the appropriate antibody. Following incubation, the blot was washed 4 times with TBS-T and incubated with horseradish peroxidase-conjugated secondary antibody. The blot was then washed 4 times with TBS-T and twice with TBS, and the signals visualized by enhanced chemiluminescence system as described by the manufacturer (GE Healthcare).

The following antibodies were used for immunoblot analysis: BRCA2 (EMD OP95) and B-actin (Sigma A5441).

Chemosensitivity Assays

Five thousand cells were plated in each well of a 96-well plate in sets of ten. Cells were then treated with a single dose of Olaparib or ABT-888 as indicated for 72 hours. WST assays (Roche) were performed according to company protocol. Briefly, 10% WST was added to each well, the plates were incubated at 37C with 5% CO2 for 4 hours and the OD at 450nM was measured.

Cell cycle analysis

1 x 10^6 cells were plated on 12-well plates and cultured for 24 hours. Cells were collected and fixed with ice-cold 70% ethanol at -20C overnight. For analyzing cell cycle, fixed cells were centrifuged at 600g for 6 min and washed with ice-cold PBS, resuspended in PBS containing 50ug/ml propidium iodide and 100ug/ml RNAse A (Invitrogen), then incubated at room temperature for at least 20min in dark. Samples were analyzed at the University of Michigan Flow Cytometry Core. Minimum of 10,000 cells were counted for each analysis.

Clonogenic survival assay

1 x 10^5 or 1 x 10^6 cells were plated on 6-well plates or 100mm dish and cultured for 24h. After treated with Radiation for 24 hours or PARP inhibitor (Olaparib or ABT-888) for 72 hours at indicated dose, cells were trypsinized and re-plated at different

concentrations on 60mm dishes or Poly-D-Lysine (Sigma P6407) treated 6-well plates in fresh media. After cultured for 7-14 days, cell colonies were fixed (Vacetic acid:Vmethonal=1:3) and stained with crystal violet. Colony number was counted for each condition in at least 3 independent experiments.

siRNA knockdown analysis

1 x 10^5 cells were plated on 6-well plates and allowed to attach overnight in serum-free media. Cells were transfected with siRNAs mixed with Optimem (Invitrogen) and Oligofectamine (Invitrogen) according to standard protocols for a final concentration of siRNA of 10nM. 24 hours post-transfection, media containing 10% FBS was added. RNA was harvested 48 hours later. siRNAs used were: STAU1-5: GCAGGGAGUUUGUGAUGCA; STAU1-7: CGAGUAAAGCCUAGAAUCA; STAU1-8: CGGAUGCAGUCCACCUAUA

Statistical analyses for experimental studies

All data are presented as means ± standard deviation or S.E.M, as indicated. All experimental assays were performed in duplicate or triplicate. Statistical analyses shown in figures represent Fisher's exact tests or Student t-tests, as indicated.

FIGURES

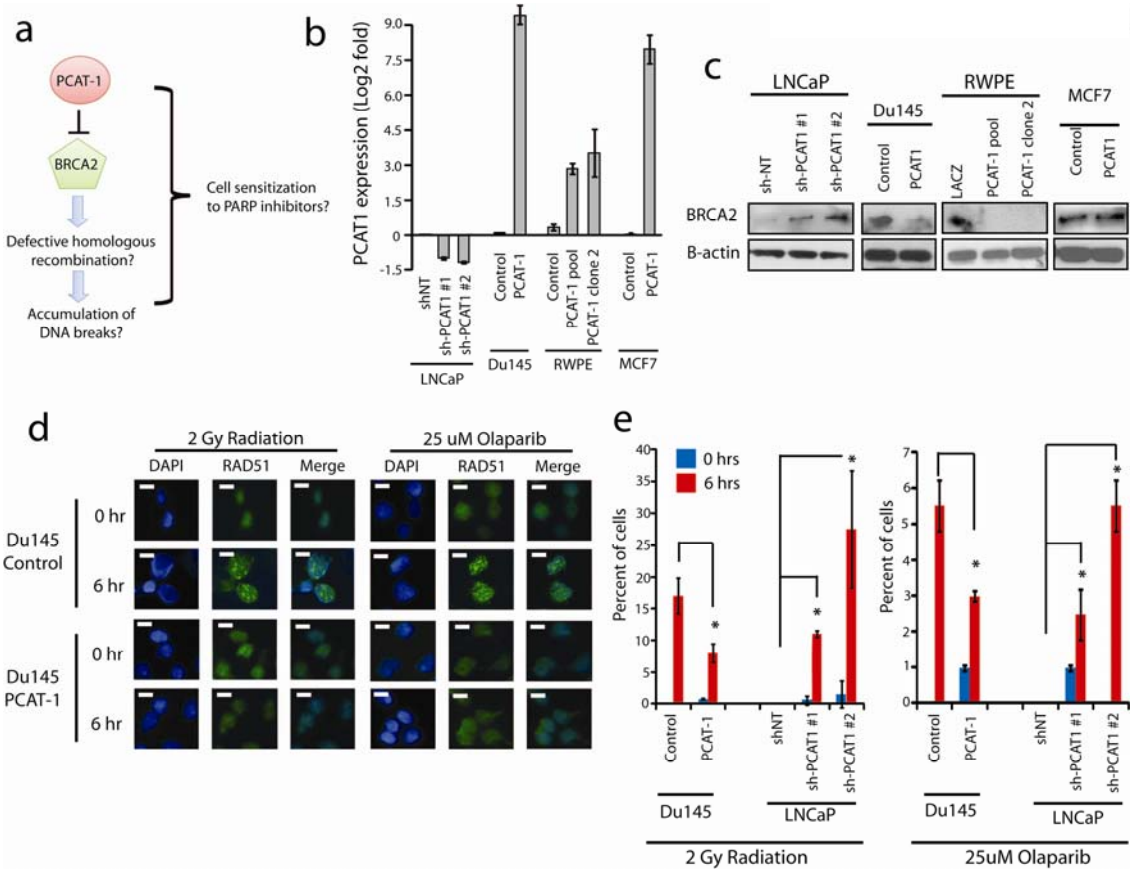


Figure 4.1: *PCAT-1* expression leads to defective homologous recombination in prostate cells. (**a**) A schematic illustrating *PCAT*-1 repression of BRCA2 leading to impaired homologous recombination (HR) and PARP inhibitor sensitivity. (**b**) Expression level of *PCAT-1* by qPCR in four isogenic cell lines with overexpression (Du145, RWPE, MCF7) or knockdown (LNCaP) of *PCAT-1*. Error bars indicate S.E.M. (**c**) Western blot analysis of BRCA2 in four isogenic cell lines with overexpression (Du145, RWPE, MCF7) or knockdown (LNCaP) of *PCAT-1*. (**d**) Induction of RAD51 foci in Du145-*PCAT-1* cells following 2 Gy of ionizing radiation or treatment with 25uM Olaparib. (**e**) Quantification of RAD51 foci in isogenic Du145 and LNCaP cell lines following 2 Gy of radiation or treatment with 25uM Olaparib. For LNCaP cell line models, cells with > 5 foci per cell were quantified. For Du145 cell line models, cells with > 10 foci per cell were quantified. Error bars represent standard deviation. An asterisk (*) indicates $p < 0.05$ by Student's t-test.

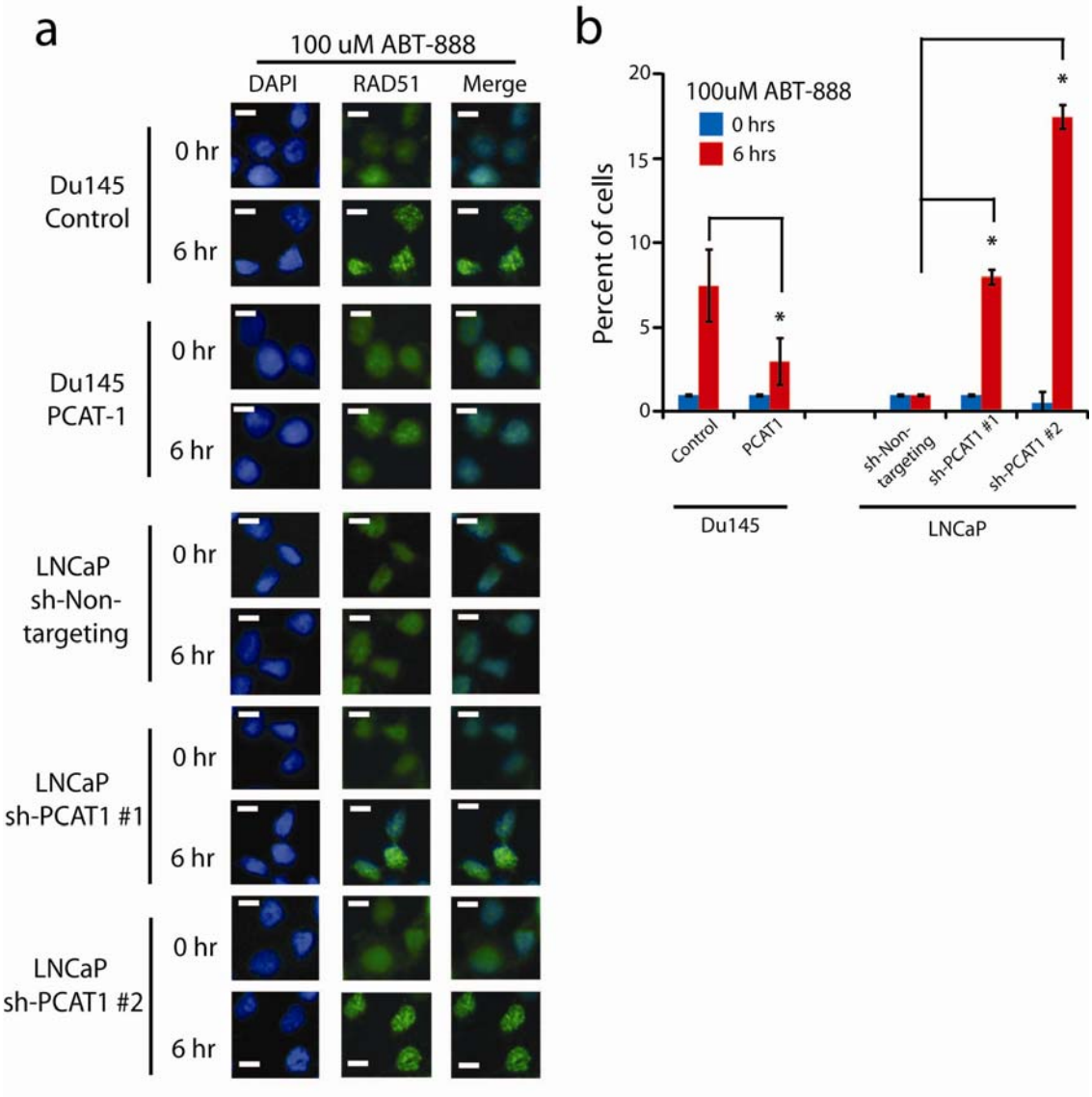


Figure 4.2: Effect of PCAT-1 on RAD51 foci formation post-ABT-888 treatment. Isogenic Du145-PCAT1 cells or LNCAP shPCAT1 knockdown cells were treated with DMSO control or 100uM ABT-888. After 6 hours, cells were fixed and stained for RAD51 and counterstained with DAPI. The number of RAD51 foci were counted. (**a**) Immunofluorescence images for RAD51 foci post-ABT-888 treatment. (**b**) Quantification of RAD51 foci post-treatment in the respective cell populations. Error bars indicate standard deviation.

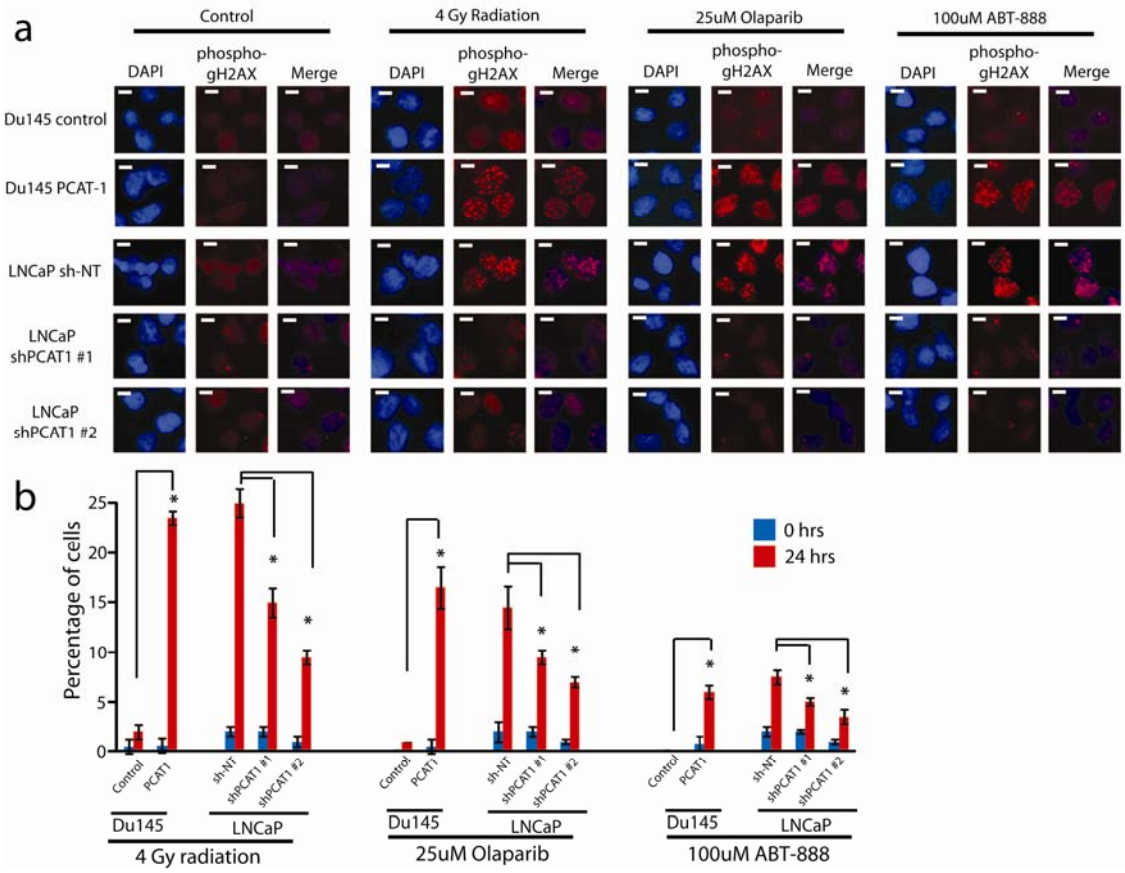


Figure 4.3: *PCAT-1* modulates g-H2AX foci formation following genotoxic stress. (**a**) LNCaP or Du145 cells with knockdown or overexpression for *PCAT-1* were subjected to 4 Gy of ionizing radiation, 25uM Olaparib, 100uM of ABT-888, or control DMSO. 24 hours post-treatment, cells were fixed and stained for phospho-gamma-H2AX and counterstained for DAPI. (**b**) Quantification of phospho-gamma-H2AX foci in LNCaP and Du145 isogenic *PCAT-1* cells treated with radiation or PARP inhibitors. For LNCaP cell line models, cells with > 5 foci per cell were quantified. For Du145 cell line models, cells with > 10 foci per cell were quantified. Error bars represent the standard deviation. An asterisk (*) indicates $p < 0.05$ by Student's t-test.

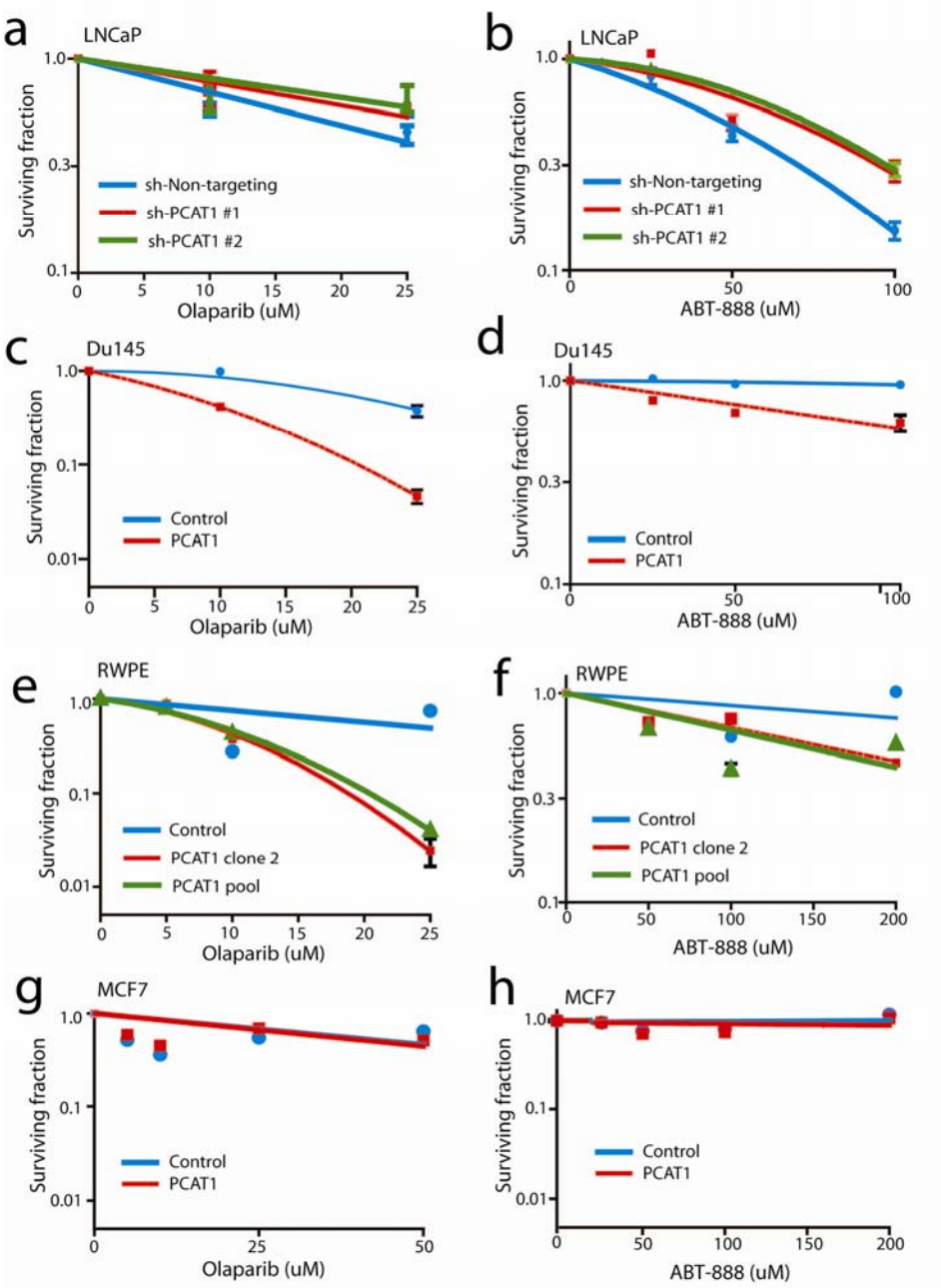


Figure 4.4: PCAT-1 expression alters clonogenic survival following PARP inhibition. Isogenic cell line models were treated with increasing doses of Olaparib or ABT-888 and assayed by clonogenic colony formation. (**a,b**) LNCaP PCAT1 knockdown cells show increased resistance to PARP inhibitors in the clonogenic assay. (**c,d**) Du145 PCAT1 overexpressing cells show increased susceptibility to PARP inhibitors in the clonogenic assay. (**e,f**) RWPE cells overexpressing PCAT1 show increased susceptibility to PARP inhibitors in the clonogenic assay. (**g,h**) MCF7 cells overexpressing PCAT1 show no change in the susceptibility to PARP inhibitors in the clonogenic assay.

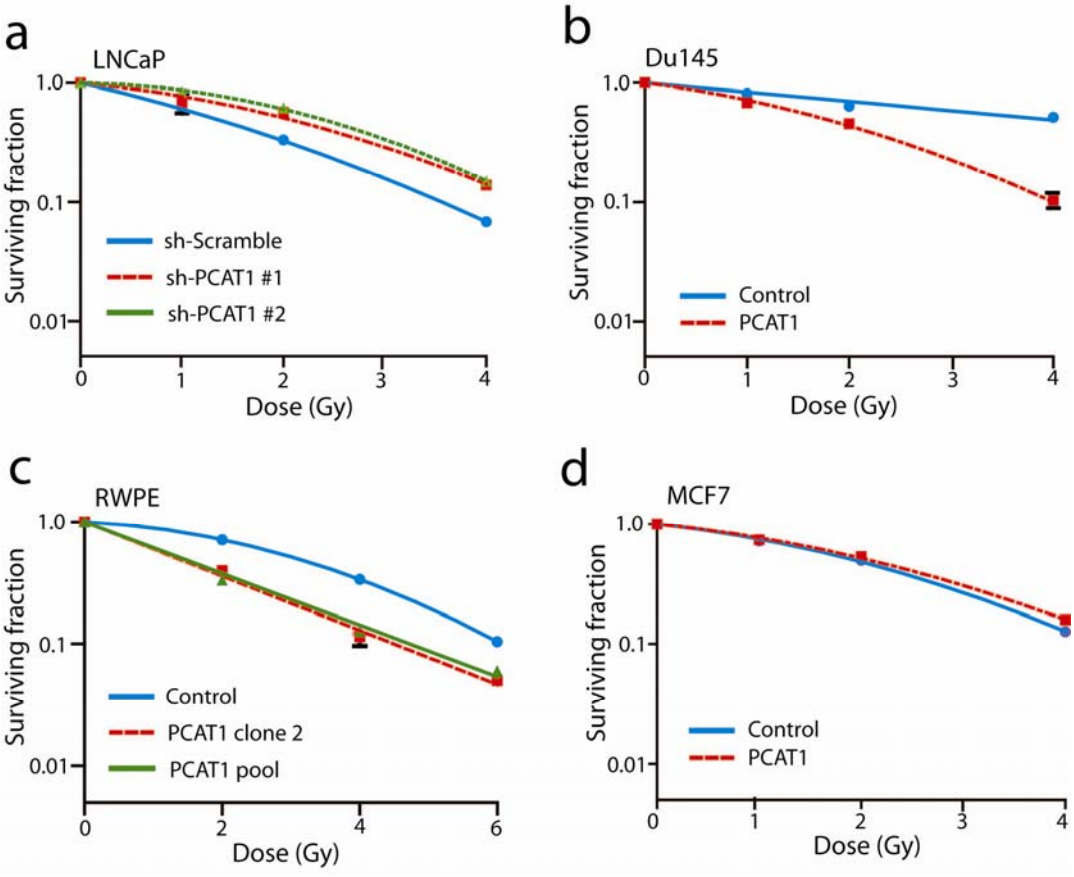


Figure 4.5: PCAT-1 expression alters clonogenic survival following radiation. Isogenic cell line models were treated with increasing doses of ionizing radiation and assayed by clonogenic colony formation. (**a**) LNCaP PCAT1 knockdown cells show increased resistance to radiation in the clonogenic assay. (**b**) Du145 PCAT1 overexpressing cells show increased susceptibility to radiation in the clonogenic assay. (**c**) RWPE cells overexpressing PCAT1 show increased susceptibility to radiation in the clonogenic assay. (**d**) MCF7 cells overexpressing PCAT1 show no change in the susceptibility to radiation in the clonogenic assay.

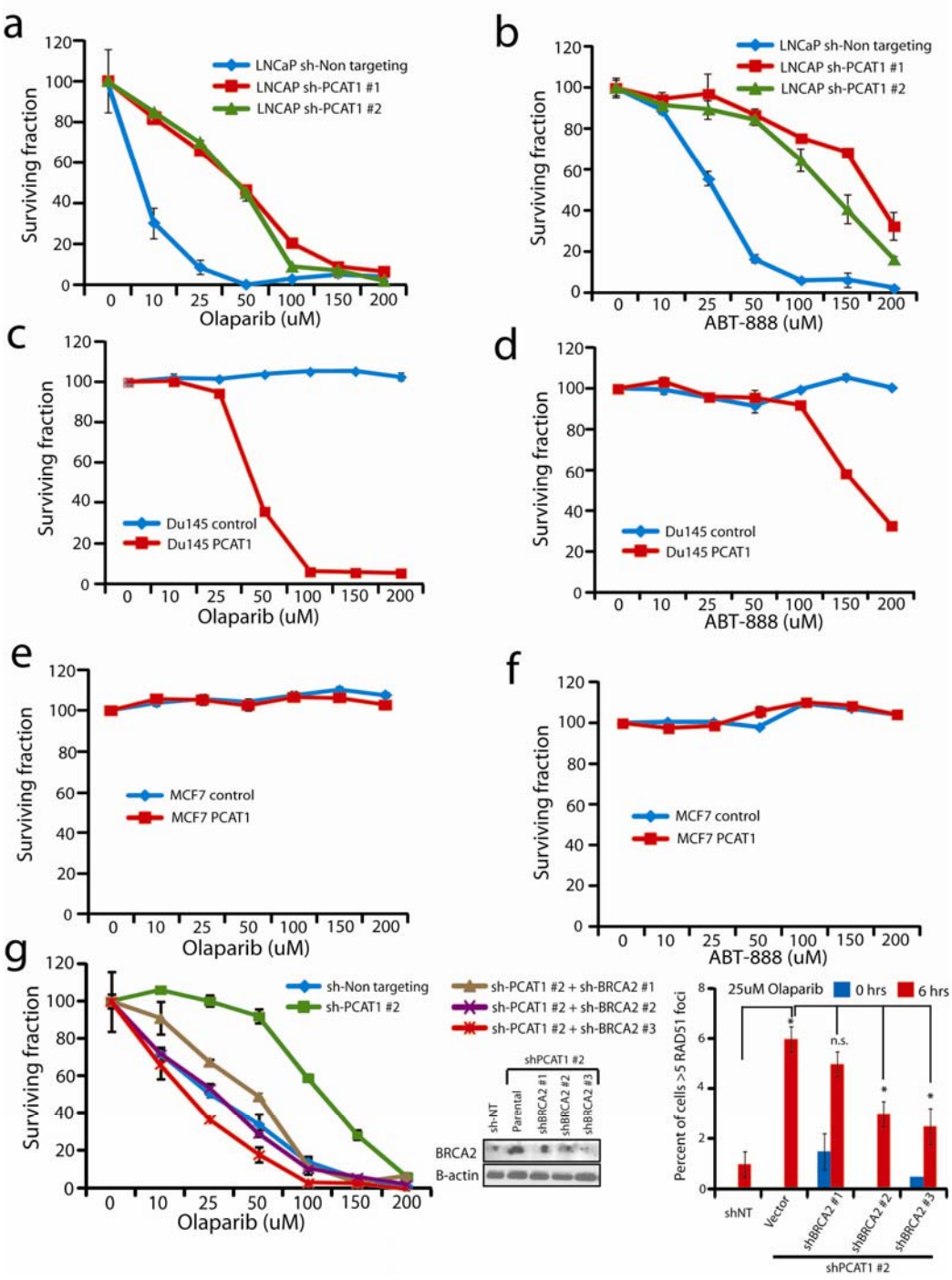


Figure 4.6: *PCAT-1* expression results in prostate cell sensitivity to PARP inhibition. (**a,b**) LNCaP cells with *PCAT-1* knockdown exhibit enhanced cell survival 72 hrs post-treatment with Olaparib (**a**) or ABT-888 (**b**). Cell survival is determined via WST assays. (**c,d**) Du145 cells with *PCAT-1* overexpression exhibit reduced cell survival 72 hrs post-treatment with Olaparib (**c**) or ABT-888 (**d**). MCF7 cells with *PCAT-1* overexpression exhibit no change in cell survival 72 hrs post-treatment with Olaparib (**e**)

or ABT-888 (**f**). (**g**) BRCA2 knockdown in LNCaP sh*PCAT-1* cells rescues cell sensitivity to Olaparib. A Western blot showing efficiency of BRCA2 knockdown is included. 72 hours post-treatment with 25uM Olaparib, cells were fixed and stained for RAD51 foci. Quantification of RAD51 foci shows a dose-dependent decrease in RAD51 foci following BRCA2 knockdown. Error bars represent standard deviation. An asterisk (*) indicates $p < 0.05$ by Student's t-test.

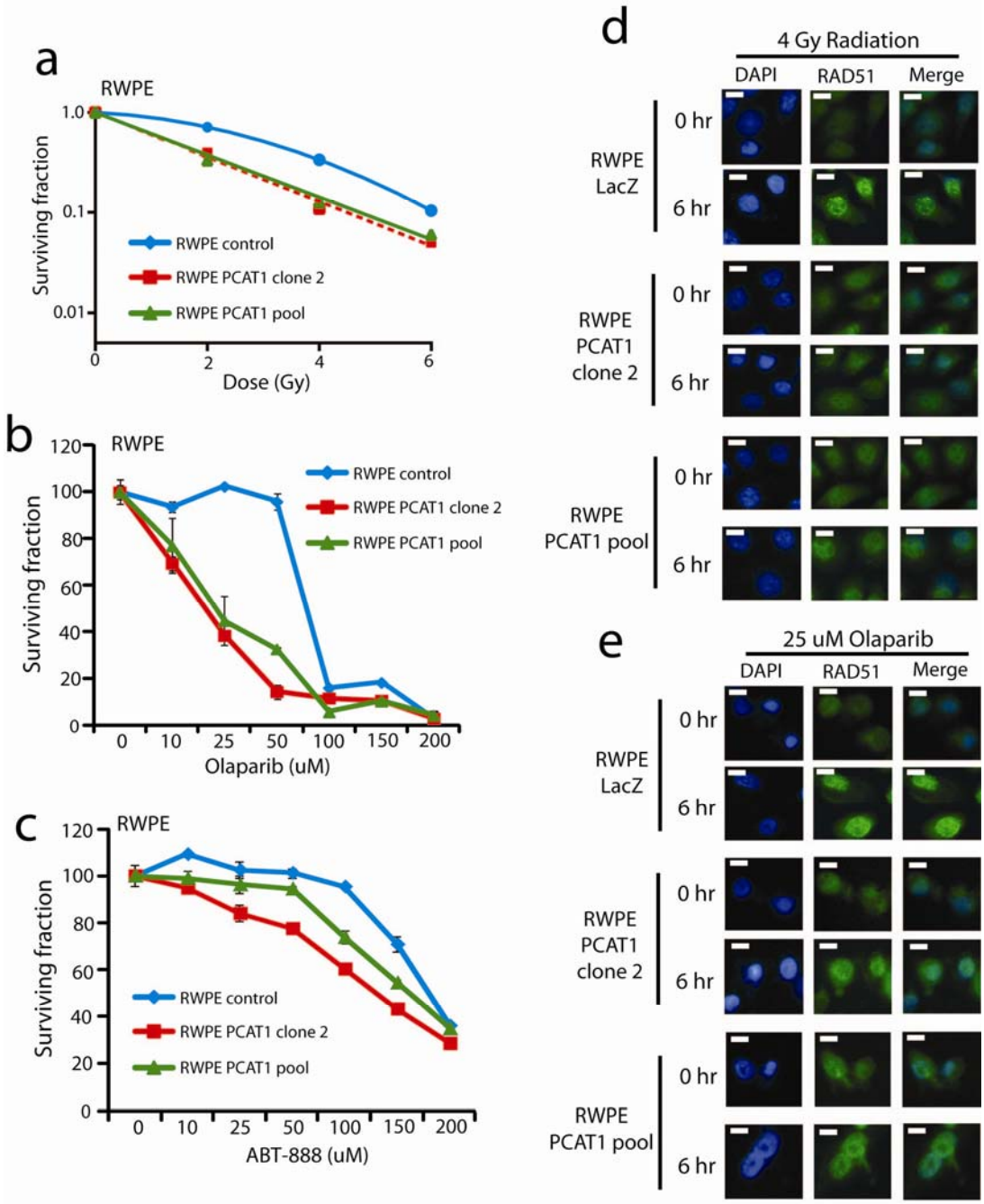


Figure 4.7: PCAT-1 overexpression in RWPE engenders sensitivity to genotoxic stress. (**a**) RWPE cells overexpressing PCAT1 were treated with increasing doses of ionizing radiation and subjected to a clonogenic assay. RWPE-PCAT1 cells show increased cell death following radiation compared to RWPE-LacZ controls. (**b,c**) RWPE-LacZ and RWPE-PCAT1 cells were treated with increasing doses of Olaparib (**b**) or ABT-888 (**c**). 72 hours later, cell survival was determined using WST. (**d,e**) Isogenic RWPE-PCAT-1 cells were treated with 4 Gy of ionizing radiation (**d**) or 25uM of Olaparib (**e**). After 6 hours, cells were fixed and stained for RAD51 and counterstained with DAPI. Error bars indicate standard deviation.

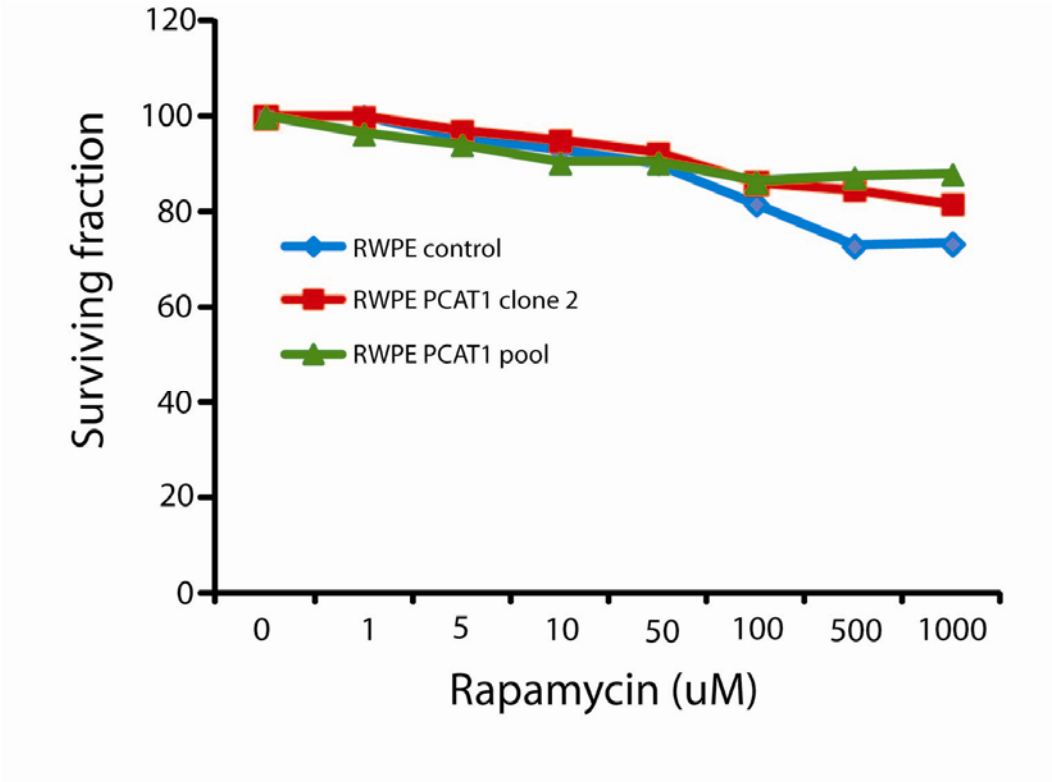


Figure 4.8: RWPE-PCAT1 cells do not show altered sensitivity to rapamycin. RWPE-LacZ, RWPE-PCAT1 pool, and RWPE-PCAT1 clone 2 cells were treated to increasing doses of rapamycin. 72 hours post-treatment, cell survival was measured using WST.

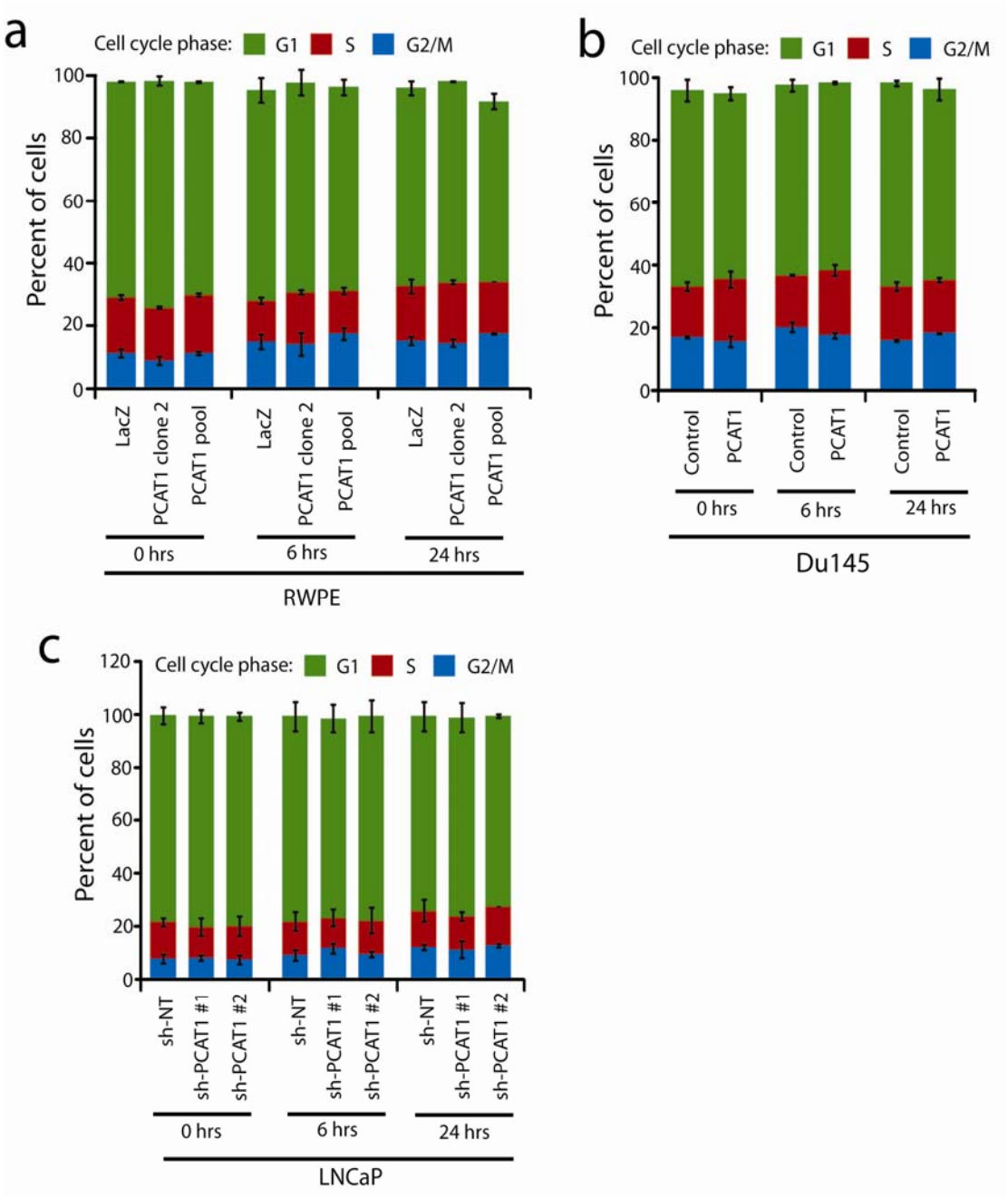


Figure 4.9: Modulation of PCAT1 expression does not impact cell cycle distribution. (**a**) RWPE isogenic cell lines were treated with DMSO and cell cycle was monitored over 24 hrs. (**b**) Du145 isogenic cell lines were treated with DMSO and cell cycle was monitored over 24 hrs. (**c**) LNCaP isogenic cell lines were treated with DMSO and cell cycle was monitored over 24 hrs. Error bars represent S.E.M.

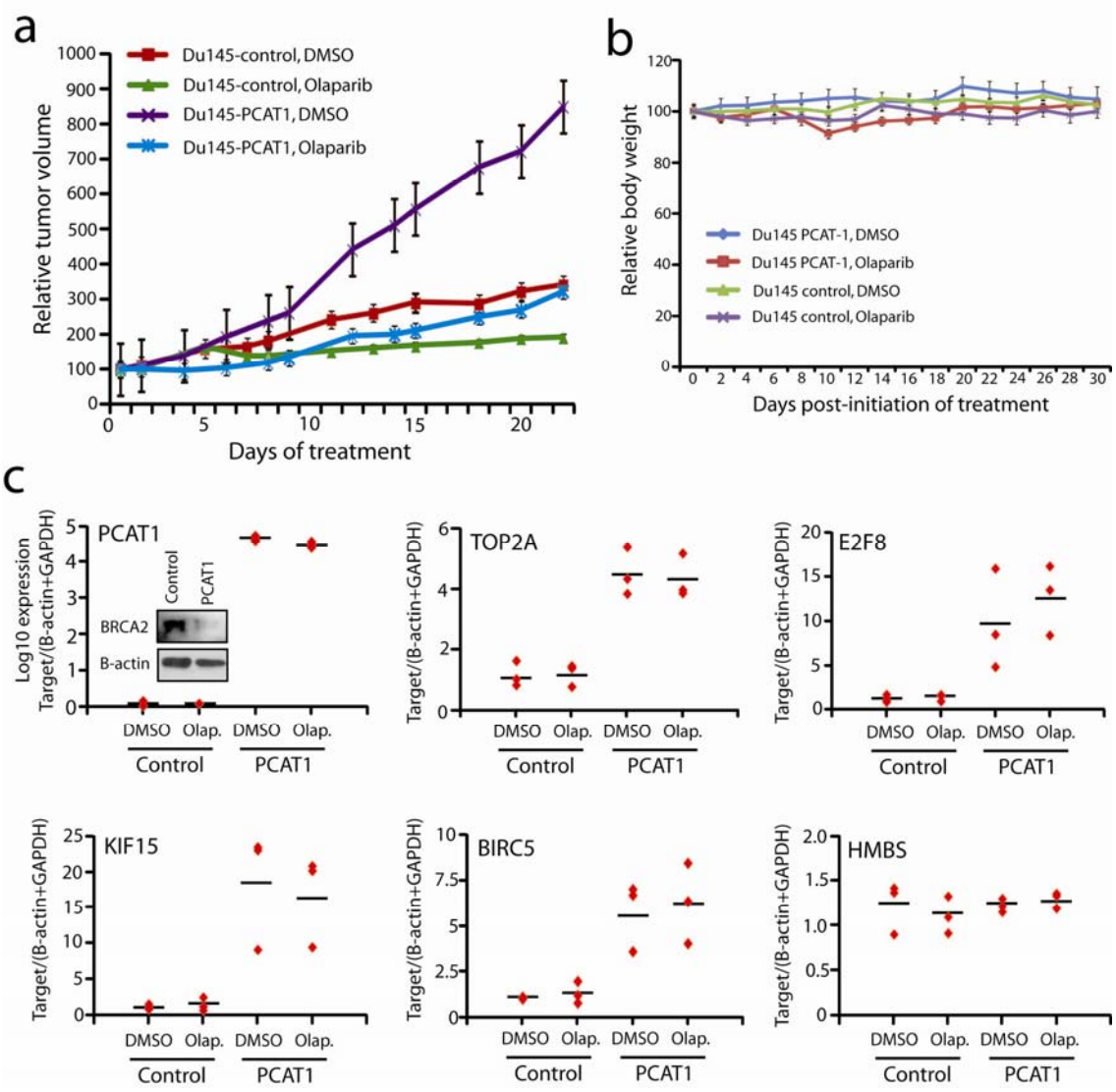


Figure 4.10: Treatment of *PCAT-1*-expressing tumors with PARP inhibitors leads to *in vivo* tumor regression. (**a**) Tumor growth curves for Du145-control and Du145-*PCAT-1* xenografts following initiation of treatment with DMSO control or Olaparib. Tumor volumes are normalized to 100, and time = 0 represents the start of treatment administration. Treatment was initiated three weeks after xenograft engraftment. Error bars represent S.E.M. (**b**) Treatment of Du145-control or Du145-*PCAT-1* mice had no effect on mouse body weight over time. (**c**) Expression levels of *PCAT-1* and *PCAT-1* target genes (*KIF15*, *TOP2A*, *E2F8*, and *BIRC5*) are elevated in Du145-*PCAT-1* xenografts compared to Du145-control xenografts in both DMSO-treated and Olaparib-treated mice. A Western blot illustrating BRCA2 levels in Du145-control and Du145-*PCAT-1* xenografts is included.

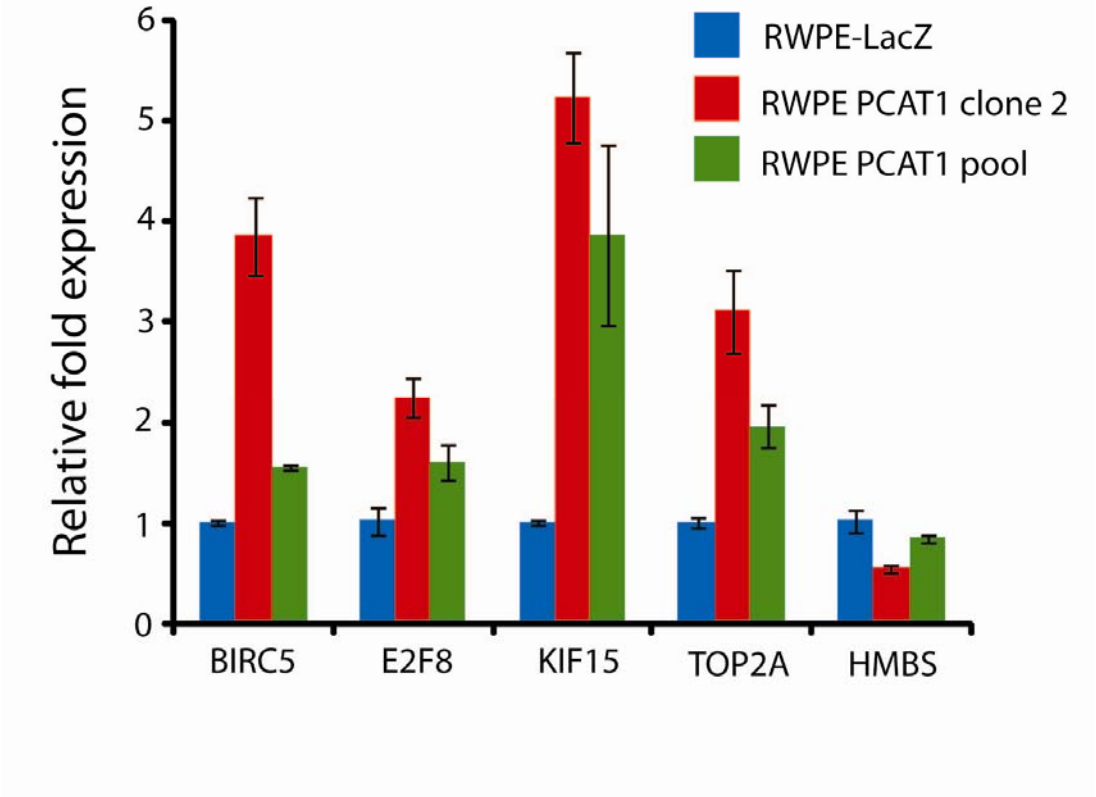


Figure 4.11: Expression of PCAT-1 regulated genes in RWPE isogenic cell lines. Four PCAT-1-induced genes (BIRC5, E2F8, KIF15, TOP2A) were measured by qPCR as well as the HMBS control housekeeping gene in the RWPE-LacZ, RWPE-PCAT1 pool, and RWPE-PCAT1 clone 2 isogenic model system. Error bars represent S.E.M.

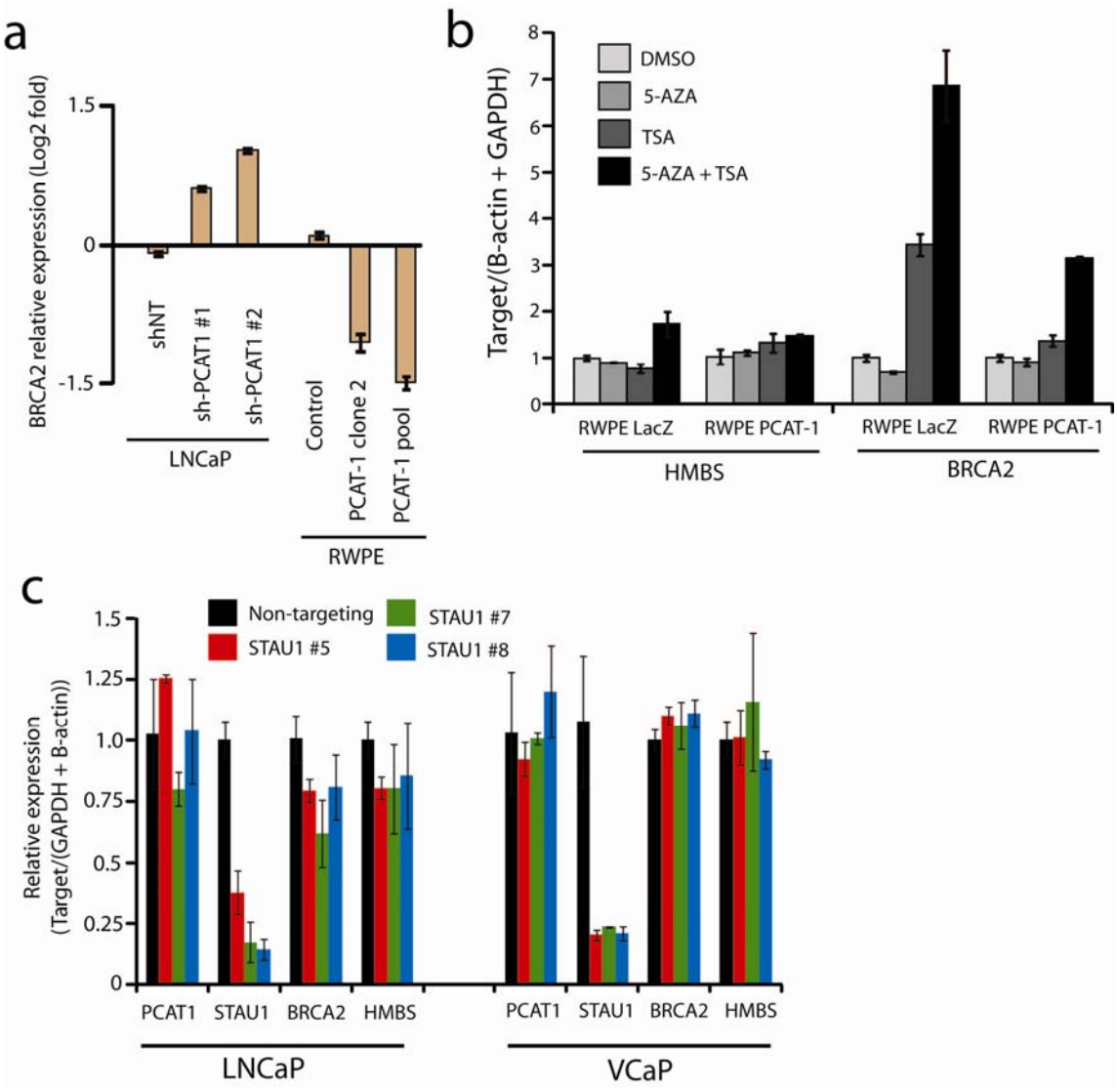


Figure 4.12: BRCA2 mRNA is repressed by PCAT-1 but not by an epigenetic or STAU-1 mediated mechanism. (**a**) Expression of BRCA2 mRNA in LNCaP shPCAT1 cells and RWPE PCAT-1 cells. BRCA2 is upregulated upon PCAT-1 knockdown in LNCaP and downregulated upon PCAT-1 overexpression. (**b**) RWPE-LacZ and RWPE PCAT-1 pool cells were treated with the demethylating agent, 5-azacytidine, the HDAC inhibitor, TSA, both or control DMSO. RNA was isolated and expression levels of BRCA2 were measured. BRCA2 mRNA levels were unchanged by 5-aza treatment, modestly upregulated by TSA in RWPE-LacZ cells, and modestly upregulated by TSA + 5-aza in both cell lines. BRCA2 regulation by PCAT-1 does not yield sensitivity to epigenetic inhibition pharmacologically. HMBS serves as a control gene unchanged by the treatments. Error bars represent S.E.M. (**c**) LNCaP and VCaP cells, which

endogenously expression PCAT-1, were transfected with independent siRNAs for STAU1 or non-targeting control. RNA was harvested and qPCR was used to analyze gene expression of STAU1, BRCA2, PCAT-1, and control HMBS. qPCR data were normalized to the average of B-actin + GAPDH. Error bars represent S.E.M.

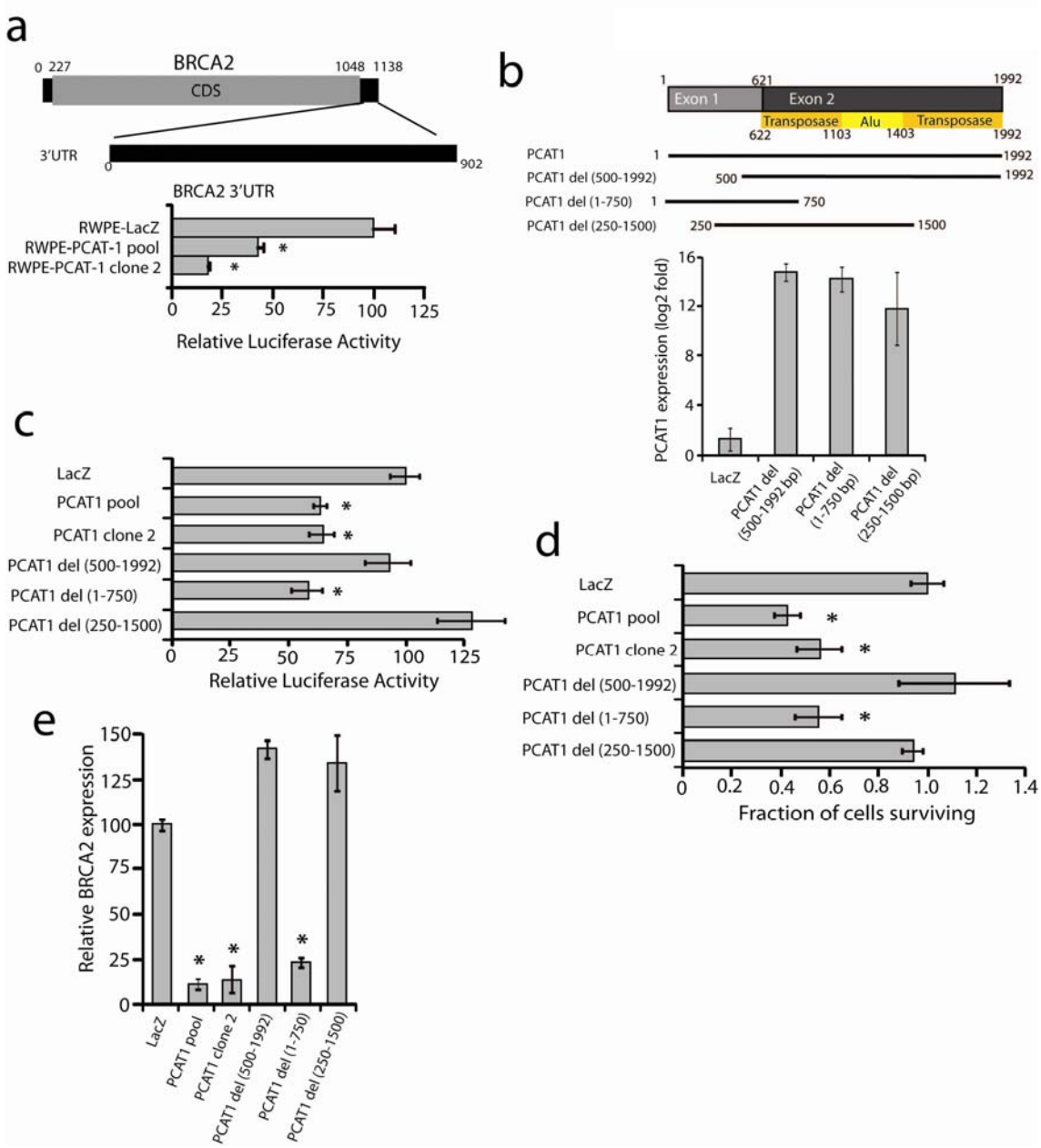
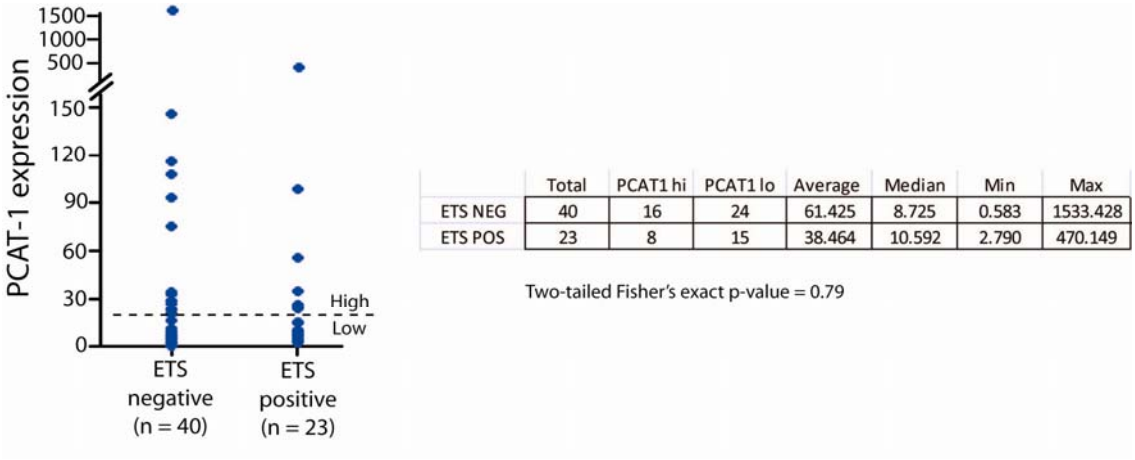


Figure 4.13: The 5' terminus of *PCAT-1* represses BRCA2 via the 3'UTR of BRCA2. (**a**) Transfection of a *BRCA2* full length 3'UTR luciferase construct in RWPE-*PCAT-1* cells results in repression of luciferase activity. A schematic of the *BRCA2* 3'UTR luciferase construct is included. (**b**) A schematic of *PCAT-1* deletion constructs overexpressed in RWPE cells and validation of overexpression by qPCR. (**c**) *PCAT-1* del (1-750 bps) was able to recapitulate repression of the *BRCA2* 3'UTR luciferase construct. (**d**) Endogenous *BRCA2* transcript levels in RWPE cells overexpressing *PCAT-1* deletion constructs. (**e**) Treatment of RWPE cells overexpressing *PCAT-1* deletion constructs with 25uM Olaparib. Cell survival was measured 72 hrs post-treatment with WST. Error bars represent standard error of the mean (S.E.M.). An asterisk (*) indicates $p < 0.05$ by Student's t-test.



	Total	PCAT1 hi	PCAT1 lo	Average	Median	Min	Max
ETS NEG	40	16	24	61.425	8.725	0.583	1533.428
ETS POS	23	8	15	38.464	10.592	2.790	470.149

Two-tailed Fisher's exact p-value = 0.79

Figure 4.14: PCAT-1 expression does not correlate with ETS status. 63 prostate cancer samples were assayed for PCAT-1 expression and determined for ETS fusion status (ERG and ETV1) by qPCR. Stratification of samples by ETS status reveals no significant association between ETS expression and PCAT1 expression, although the highest PCAT-1 sample was in an ETS-negative prostate cancer.

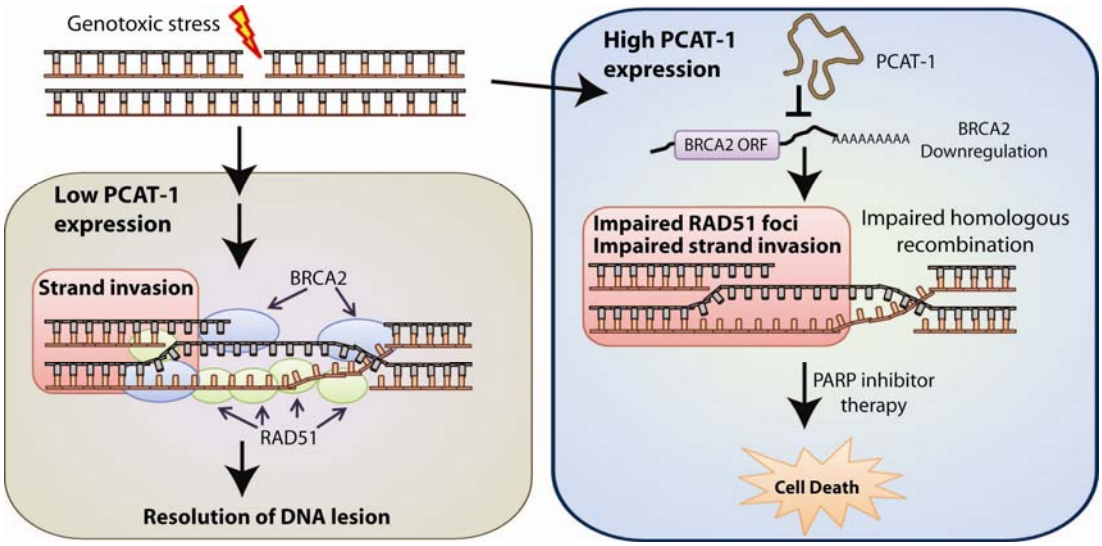


Figure 4.15: A model of *PCAT-1* function in homologous recombination.

REFERENCES

1. S. Negrini, V. G. Gorgoulis, T. D. Halazonetis, *Nat Rev Mol Cell Biol* **11**, 220 (Mar, 2010).
2. N. C. Turner *et al.*, *Oncogene* **26**, 2126 (Mar 29, 2007).
3. N. Turner, A. Tutt, A. Ashworth, *Nat Rev Cancer* **4**, 814 (Oct, 2004).
4. S. P. Jackson, J. Bartek, *Nature* **461**, 1071 (Oct 22, 2009).
5. J. San Filippo, P. Sung, H. Klein, *Annu Rev Biochem* **77**, 229 (2008).
6. R. Roy, J. Chun, S. N. Powell, *Nat Rev Cancer* **12**, 68 (Jan, 2012).
7. H. Farmer *et al.*, *Nature* **434**, 917 (Apr 14, 2005).
8. H. E. Bryant *et al.*, *Nature* **434**, 913 (Apr 14, 2005).
9. P. C. Fong *et al.*, *N Engl J Med* **361**, 123 (Jul 9, 2009).
10. M. Guttman *et al.*, *Nature*, (Aug 28, 2011).
11. J. T. Lee, *Genes Dev* **23**, 1831 (Aug 15, 2009).
12. M. A. Faghihi *et al.*, *Nat Med* **14**, 723 (Jul, 2008).
13. J. R. Prensner *et al.*, *Nat Biotechnol* **29**, 742 (Aug, 2011).
14. R. A. Gupta *et al.*, *Nature* **464**, 1071 (Apr 15, 2010).
15. J. R. Prensner, A. M. Chinnaiyan, *Cancer Discov* **1**, 391 (Oct, 2011).
16. J. L. Rinn *et al.*, *Cell* **129**, 1311 (Jun 29, 2007).
17. M. Huarte *et al.*, *Cell* **142**, 409 (Aug 6, 2010).
18. A. M. Khalil *et al.*, *Proc Natl Acad Sci U S A* **106**, 11667 (Jul 14, 2009).
19. S. A. Martin, C. J. Lord, A. Ashworth, *Curr Opin Genet Dev* **18**, 80 (Feb, 2008).
20. C. K. Donawho *et al.*, *Clin Cancer Res* **13**, 2728 (May 1, 2007).
21. P. Baumann, F. E. Benson, S. C. West, *Cell* **87**, 757 (Nov 15, 1996).
22. E. Wahlberg *et al.*, *Nat Biotechnol* **30**, 283 (2012).
23. K. C. Wang *et al.*, *Nature* **472**, 120 (Apr 7, 2011).
24. K. L. Yap *et al.*, *Mol Cell* **38**, 662 (Jun 11, 2010).
25. C. Gong, L. E. Maquat, *Nature* **470**, 284 (Feb 10, 2011).
26. D. P. Bartel, *Cell* **136**, 215 (Jan 23, 2009).
27. J. C. Brenner *et al.*, *Cancer Cell* **19**, 664 (May 17, 2011).
28. M. C. Tsai *et al.*, *Science* **329**, 689 (Aug 6, 2010).
29. T. K. Kim *et al.*, *Nature* **465**, 182 (May 13, 2010).
30. D. Wang *et al.*, *Nature* **474**, 390 (Jun 16, 2011).
31. M. Cesana *et al.*, *Cell* **147**, 358 (Oct 14, 2011).
32. W. P. Wahls, E. R. Siegel, M. K. Davidson, *PLoS One* **3**, e2887 (2008).
33. E. Splinter *et al.*, *Genes Dev* **25**, 1371 (Jul 1, 2011).
34. T. Hung *et al.*, *Nat Genet* **43**, 621 (2011).
35. B. Adamson, A. Smogorzewska, F. D. Sigoillot, R. W. King, S. J. Elledge, *Nat Cell Biol* **14**, 318 (2012).
36. A. M. Oonk *et al.*, *Ann Oncol*, (Feb 21, 2012).

CHAPTER 5

A novel, lineage-specific long non-coding RNA coordinates prostate cancer aggressiveness

SUMMARY

Prostate cancer is a clinically heterogenous disease, in which only a minority of patients have aggressive cancer. However, the molecular basis for this dichotomy remains incompletely understood. Here, we identified a novel polyadenylated, seven-exon, nuclear long noncoding RNA (lncRNA), termed *SChLAP-1 (S*econd *Ch*romosome *L*ocus *A*ssociated with *P*rostate-1) that is overexpressed in a subset of localized and metastatic cancers. *In vitro* and *in vivo* gain-of-function and loss-of-function experiments demonstrated that *SChLAP-1* is involved in prostate cancer cell invasiveness, a hallmark for aggressive cancers. Mechanistically, we found that *SChLAP-1* directly binds and antagonizes the SWI/SNF chromatin modifying complex, which functions as a tumor suppressor in many cancer types. Finally, we determined that outlier expression of SChLAP-1 in prostate cancer was associated with clinically significant and metastatic prostate cancer gene signatures. These results indicate that a *SChLAP-1*-SWI/SNF interaction coordinates cell invasiveness through competitive gene expression control. Our findings establish *SChLAP-1* as a novel lncRNA implicated in the development of aggressive prostate cancer.

INTRODUCTION

Long non-coding RNAs (lncRNAs) are polyadenylated RNA species >200bp in length commonly characterized by splicing of multiple exons, H3K4me3 promoter methylation, and transcription by RNA polymerase II (*1*). lncRNA-mediated biology has been implicated in a wide variety of cellular processes, including pluripotency in stem cells (*2*) and X chromosome inactivation in women (*3*). In cancer, lncRNAs are emerging as a prominent layer of previously unknown transcriptional regulation, contributing to tumor biology by modulating gene expression of key target genes, often by collaborating with key epigenetic complexes such as Polycomb Repressive Complex I (*4, 5*) (PRC1) and Polycomb Repressive Complex 2 (PRC2) (*4, 6-9*), among others. Clinically, overexpression of the *HOTAIR* lncRNA has been shown to correlate with aggressive breast (*6*), colon (*10*), hepatocellular (*11*), and gastrointestinal stromal tumors (*12*), suggesting that lncRNA-mediated biology plays a prominent role in cancer progression.

However, lncRNAs frequently display lineage-specific expression patterns (*13, 14*), making generalizations about their functional contributions to a variety of cancers difficult. For example, expression of *HOTAIR*, while an important component of breast cancer (*6*), is virtually absent in prostate cancer (*13*). Similarly, we have previously described *PCAT-1* as a prostate-cancer-specific regulator of tumor proliferations (*13*). We hypothesized that prostate-specific lncRNAs mediated prostate cancer aggressiveness. To explore this, we sought to discover novel prostate cancer outlier genes whose expression may impart aggressive features to a subset of cancers.

RESULTS

We have previously used RNA-seq to profile lncRNA expression in >100 prostate cell lines and tissues, including both localized and metastatic prostate cancers (*13*). In this analysis, we defined a set of unannotated lncRNAs, including *PCAT-1*, whose aberrant expression associated with prostate cancer. To uncover novel lncRNAs implicated in disease biology, we performed a modified cancer outlier profile analysis (*15*) (COPA) to nominate intergenic lncRNAs selectively upregulated in a subset of cancers. Of >1800 integenic lncRNAs, we observed that only two, *PCAT-109* and *PCAT-114*, showed striking outlier profiles and ranked among the top five outliers in prostate cancer when compared to protein-coding genes (**Figure 5.1A** and ref. 2). Notably, both are located in a "gene desert" on Chromosome 2q31.3, a region not previously examined in prostate cancer (**Figure 5.1B** and **Figure 5.2**).

Efforts to validate *PCAT-109* by PCR and rapid amplification of cDNA ends (RACE) failed, partly due to the fact that this gene is not robustly expressed in any prostate cell lines (**Figure 5.1B** and data not shown). By contrast, in the *PCAT-114* region, PCR experiments and 5' and 3' RACE defined a 1.4 kb, polyadenylated lncRNA encoded by up to seven exons spanning nearly 200kb on Ch2q31.3 (**Figure 5.1C** and **Figure 5.3**). We named this gene *S*econd *Ch*romosome *L*ocus *A*ssociated with *P*rostate-1 (*SChLAP-1*) after its genomic location. To further characterize this gene, we employed a published ChIP-Seq dataset of prostate cancer (*16*) and found that the transcriptional start site (TSS) of *SChLAP-1* was marked by trimethylation of H3K4 (H3K4me3) and its gene body harbored trimethylation of H3K36 (H3K36me3) (**Figure 5.1C**), an epigenetic signature consistent with canonical protein-coding genes and lncRNAs (*17*).

Interestingly, the *SChLAP-1* promoter does not contain a CpG island but rather remnants of an ancestral retroviral long terminal repeat (**Figure 5.2**). PCR assays defined numerous splicing isoforms of this gene, of which two constituted the vast majority (>90%) of *SChLAP-1* expression and were termed isoform #1 and #2 (**Figure 5.1D**). Other isoforms were minority species, the most common of which was termed isoform #3 (**Figure 5.1D**).

Using quantitative PCR (qPCR), we validated that *SChLAP-1* was upregulated in ~20% of prostate cancers (**Figure 5.1E**), including metastatic prostate cancer. Moreover, examination of *SChLAP-1* expression in our RNA-seq compendium of >500 samples, representing >13 cancer types, demonstrated that expression of *SChLAP-1* was absent from other tumor types (**Figure 5.4**), suggesting that expression of this gene is specific to prostate cancer biology. To establish *SChLAP-1* as a non-coding gene, we cloned three isoforms (isoforms 1, 2 and 3) and performed *in vitro* translation assays, which were negative (**Figure 5.5**). Consistent with this, we found that *SChLAP-1* RNA in prostate cell lines was predominantly located in the nucleus (**Figure 5.1F**), while most protein-coding mRNAs are located in the cytoplasm, where they engage the ribosomal translation machinery.

To elucidate the functional role for *SChLAP-1* in prostate cancer, we performed siRNA knockdowns of this gene using two independent siRNAs as well as siRNA to *EZH2*, which is essential for cancer cell invasion (*18, 19*), as a positive control. Remarkably, in three prostate cancer cell lines but not breast and lung cancer cells lacking *SChLAP-1* expression, knockdown of *SChLAP-1* dramatically impaired cell

invasion *in vitro* at a level comparable to *EZH2* (**Figure 5.6A**). *SChLAP-1* knockdown also impaired cell proliferation in prostate cells but not non-prostate cells (**Figure 5.7**).

To confirm this phenotype, we overexpressed the three most abundant *SChLAP-1* isoforms in RWPE benign immortalized prostate cells at physiologic levels similar to the LNCaP cell line (**Figure 5.8**). While *SChLAP-1* overexpression did not impact cell proliferation (**Figure 5.8**), we found that RWPE cells expressing all three *SChLAP-1* isoforms, but not control cells, exhibited a dramatic ability to invade through Matrigel model basement membrane matrix *in vitro* (**Figure 5.6B**). Although RWPE cells expressing *SChLAP-1* isoform 2 had a trend towards modestly increased invasiveness compared to cells expressing *SChLAP-1* isoforms 1 and 3, this difference was not statistically significant ($p > 0.05$). Overexpression of *SChLAP-1* in HME benign breast or MCF7 breast cancer cells did not induce cell invasion (**Figure 5.9**), suggesting that *SChLAP-1* functions in a lineage-specific manner. These data support a lineage-specific role for *SChLAP-1*, which is consistent with its tissue-specific expression pattern in prostate cancer (**Figure 5.4**).

To characterize specific regions of *SChLAP-1* essential for its function, we generated deletion constructs tiling every 250bp and overexpressed these in RWPE cells. This analysis identified a single 250bp region (bp 1001 – 1250 for *SChLAP-1* isoform #1) necessary for *SChLAP-1* mediated invasion in RWPE (**Figure 5.6C**). This region is shared by all three major isoforms, and *in silico* modeling of the *SChLAP-1* RNA structure using RNAfold suggested the presence of a RNA hairpin in this region that is lost specifically in deletion construct #5 (**Figure 5.6D** and **Figure 5.10**), potentially implicating this structure in the function of the molecule.

Xenograft analysis of 22Rv1 cells stably knocking down *SChLAP-1* further confirmed that this gene is necessary for appropriate cancer cell metastatic seeding *in vivo*. To test this, we performed intracardiac injection of tumor cells and monitored luciferase signal from mouse lungs and distant metastases. These experiments showed that 22Rv1 sh*SChLAP-1* cells displayed impaired metastatic seeding at both proximal (lungs) and distal sites (**Figure 5.6E** and **Figure 5.11**). Indeed, 22Rv1 shSChLAP-1 cells displayed both fewer gross metastatic sites overall (an average 3.66 metastatic sites in sh*NT* mice vs. 2.07 metastatic sites in sh*SChLAP-1* #1 and 1.07 sites in sh*SChLAP-1* #2 mice, $p < 0.05$, Student's t-test) as well as smaller metastatic tumors when they did form (**Figure 5.6F** and **Figure 5.11**). Interestingly, sh*SChLAP-1* subcutaneous xenografts displayed slower tumor progression due to delayed tumor engraftment *in vivo*, even though the tumor growth kinetics once tumor engraftment occurred were not affected (**Figure 5.12**). Together, these *in vitro* and *in vivo* data support a prostate-specific role for *SChLAP-1* in cancer cell invasion, metastasis, and aggressiveness.

To interrogate potential mechanisms of *SChLAP-1* function, we performed microarray profiling of 22Rv1 and LNCaP prostate cancer cells treated with *SChLAP-1* or control siRNAs. To facilitate interpretation of these findings, we employed GSEA (*20*), which assesses the statistical overlap between a given microarray dataset and those from other publicly-available studies, to analyze genes with statistically significance changes in gene expression following *SChLAP-1* knockdown. Among the best-scoring concepts, we noticed prostate-cancer-specific gene signatures for genes positively or negatively correlated with BRM (*21*), an enzymatic subunit of the SWI/SNF chromatin modifying complex (*22*) (Shen_BRM) whose expression decreases in prostate cancer

(**Figure 5.13A, left**). This signature was generated through an analysis of the SWI/SNF complex in human prostate cancer samples (*21*), making it an attractive biological insight due to the tissue-specific functions of this complex. We further confirmed this observation in our RNA-seq dataset by generating gene signatures positive and negatively correlated to BRM **(Figure 5.13A, right)**, which similarly demonstrated that *SChLAP-1*-regulated genes were also highly enriched with this set. Importantly, *SChLAP-1*-regulated genes were inversely correlated with both BRM datasets (**Figure 5.13A**), where BRM-correlated genes (which decrease in prostate cancer as BRM decreases) were upregulated when we decreased *SChLAP-1* by siRNA. These results suggested that *SChLAP-1* and SWI/SNF regulate gene transcription in opposing manners, leading to an antagonism of SWI/SNF activity by *SChLAP-1*.

The SWI/SNF complex operates as a large, multi-protein complex that utilizes ATPase enzymatic activity to physically move nucleosomes and, in doing so, to regulate gene transcription (*22*). Several SWI/SNF complex members are the target of recurrent, inactivating mutations in cancer, including *ARID1A* (*23, 24*), *PBRM1* (*25*), and *SNF5* (*26*), and numerous studies suggest that loss of the SWI/SNF complex promotes cancer progression (*22, 27*). While SWI/SNF mutations are not commonly observed in prostate cancer, several reports suggest that downregulation of SWI/SNF complex members, particularly *BRM*, characterize prostate cancer (*21, 28*), and mice with a prostate-specific *BRM* deletion exhibit prostatic hyperplasia and castration-resistant cellular proliferation (*21*). Thus, antagonism of SWI/SNF activity by *SChLAP-1* would be consistent with the oncogenic behavior of *SChLAP-1* and the tumor suppressive behavior of the SWI-SNF complex.

To test whether *SChLAP-1* antagonizes SWI/SNF-mediated gene expression regulation, we performed siRNA knockdown of three key components of the SWI-SNF complex: *BRM* (also known as *SMARCA2*), *BRG1* (also known as *SMARCA4*), and *SNF5* (also known as *SMARCB1*). Like BRM, BRG1 serves as an enzymatic subunit of SWI/SNF complex activity, and SNF5 is an essential subunit thought to bind histone proteins (*29*). Knockdown of *BRM*, *BRG1*, and *SNF5* in two prostate cell lines, 22Rv1 and LNCaP, followed by expression microarray profiling generated highly overlapping sets of up- and down-regulated genes (**Figure 5.13B** and **Figure 5.14**), demonstrating that these factors have broad commonalities in their function in prostate cells. Knockdown of *BRM*, *BRG1* and *SNF5* also increased the invasiveness and proliferation rate of 22Rv1 cells, consistent with the role of SWI/SNF in tumor suppression (**Figure 5.15**).

A heatmap analysis of genes commonly regulated by knockdown of all SWI/SNF proteins (*BRM*, *BRG1*, and *SNF5*) demonstrated a clear antagonism between SWI/SNF regulation of gene expression and *SChLAP-1*'s effects, whose knockdown regulated the same genes of SWI/SNF in the opposing direction (**Figure 5.13B**). To quantify this globally, we used these microarray data to generate custom GSEA signatures for genes up- and down-regulated upon knockdown of each SWI/SNF protein (*BRM*, *BRG1*, and *SNF5*) and *SChLAP-1*, and we queried the overlap of these GSEA signatures against each other. Performing this analysis across two cell lines (22Rv1 and LNCaP) showed that SWI/SNF and *SChLAP-1* antagonize each other's gene expression program in a highly significant manner in 23 of 24 total GSEA comparisons (FDR < 0.05) (**Figure 5.13C** and **Figure 5.16**). Here, a positive GSEA enrichment score (ES) indicates genes upregulated

upon *SChLAP-1* knockown, and a negative GSEA NES indicates genes downregulated upon *SChLAP-1* knockdown. Together, these data strongly argue that *SChLAP-1* functions to suppress SWI/SNF complex activity in prostate cancer.

To examine the mechanism of *SChLAP-1* regulation of the SWI/SNF complex, we examined whether *SChLAP-1* regulated SWI/SNF complex genes themselves. Using Western blots, we failed to detect any change in BRM, BRG1, or SNF5 protein abundance following *SChLAP-1* knockdown or overexpression (**Figure 5.17**), suggesting that *SChLAP-1* exerts its function post-translationally. Motivated by reports of lncRNAs coordinating the function of epigenetic complexes through direct RNA-protein binding, we performed RNA immunoprecipitation assays (RIP) for SNF5, a core subunit essential for both BRG1 and BRM function, in 22Rv1 and LNCaP cells. We found that endogenous *SChLAP-1*, but not other prostate-specific lncRNAs such as *PCA3* and *PCAT-1*, bound robustly SNF5 protein (**Figure 5.13D**). RIP for androgen receptor (AR) and SNRNP70, which specifically binds to the *U1* snRNP, served as additional negative controls for these experiments (**Figure 5.18**).

We further hypothesized that *SChLAP-1*-SWI/SNF interactions may underlay the functional role of *SChLAP-1* in inducing cell invasion. To test this, we evaluated *SChLAP-1*-SNF5 binding in our RWPE-*SChLAP-1* overexpression model, including overexpression of *SChLAP-1* deletion construct #5, which fails to induce cell invasion (**Figure 5.6D)**. Strikingly, we observed that overexpression of both *SChLAP-1* isoform #1 and isoform #2 robustly bound to SNF5, whereas the deletion construct 5 (which lacks bps 1001-1250 in SChLAP-1 isoform #1) failed to bind SNF5 (**Figure 5.13E**). As controls, we also measured *AK093002* and *LOC145837*, two lncRNAs upregulated in

prostate cancer that are endogenously expressed in RWPE (data not shown). Control RNA-IP experiments for SNRNP70 demonstrated uniformly strong binding of this protein to *U1* in all RWPE cell lines evaluated (**Figure 5.18**). Thus, *SChLAP-1* regulates SWI/SNF complex activity through directly binding to SWI/SNF proteins through an interaction via bps 1001-1250.

To establish a link between *SChLAP-1* and aggressive prostate cancer, we performed bioinformatics analysis of gene expression signatures in this disease (**Figure 5.19A**). We developed a bioinformatics method to assess whether an individual gene associates with the compendium of clinical phenotypes defined by expression profiling. To do this for *SChLAP-1*, we used *SChLAP-1* expression to classify our RNA-seq data for localized prostate cancers into *SChLAP-1*-negative and *SChLAP-1*-positive samples. This analysis generated a set of genes positive and negatively correlated with *SChLAP-1* expression only in localized tumors, to avoid bias when comparing this signature with the metastatic gene signatures. Random permutations of the data enabled us to generate p-values, FDR values, and Pearson correlations for all genes in the signature, and for our gene signature we used associated genes with a FDR < 0.05. Network analysis of our *SChLAP-1* signature using datasets from Oncomine (*30*) reveals a striking degree of overlap with concepts related to prostate cancer progression (**Figure 5.19A**). Here, genes lowly expressed in *SChLAP-1*-positive tumors were similarly lowly expressed in metastatic and high-grade localized tumors. Conversely, genes expressed highly in *SChLAP-1*-positive tumors were also high in metastatic cancer.

To provide context to these findings, we expanded our analysis to include four additional cancer genes: *EZH2*, a known aggressive cancer gene (*18, 19*), *PCA3*, a

lncRNA biomarker upregulated in prostate cancer and used in prostate cancer diagnosis (*31, 32*), *AMACR*, a known prostate cancer outlier (*33*), and BRM, a SWI/SNF enzymatic subunit downregulated in prostate cancer (*21*). We also analyzed nucleoporin genes, *NUP133* and *NUP155,* as well as *B-actin* (*ACTB*) in this way as controls. To facilitate interpretation of these data, we established a compendium of prostate cancer gene signatures for different stages in prostate cancer progression (cancer vs. normal, advanced vs. indolent cancer, metastatic vs. localized cancer) using the Oncomine database to extract data from published microarray studies. We also used our RNA-seq data to generate gene signatures for high grade localized prostate cancer (Gleason ≥8 vs. Gleason 6) and metastatic vs. localized prostate cancer in our individual dataset.

We next evaluated the statistical overlap between each dataset in our clinical phenotype compendium with the gene signature derived from *SChLAP-1* or other genes of interest (**Figure 5.19B**). We computed p-values, FDR values, and odds ratios for every analysis between two gene sets. Using this database of associations, we generated a heat-map to visualize the relationship between our genes of interest and prostate cancer phenotypes. This analysis confirmed a strong association of *SChLAP-1*-associated genes and high-grade and metastatic cancers as well as poor clinical outcomes. In this respect, *SChLAP-1* was highly similar to *EZH2*, our positive control, which is widely associated with aggressive, lethal prostate cancer. By contrast, *PCA3* and *AMACR*, two biomarkers not associated with disease progression, strongly associate with Cancer vs. Normal concepts but not concepts for aggressive disease. *B-actin*, *NUP133*, and *NUP155* show no systematic association with prostate cancer signatures (**Figure 5.19B)**.

To establish the predictive value of our *SChLAP-1* gene signature, we investigated whether this signature was able to stratify patient outcomes over time. We used two publicly-available datasets to test whether our *SChLAP-1* signature associated with biochemical recurrence (*34*) and overall survival (*35*). Here, we observed a statistically-significant ($p<0.01$) association between the SChLAP-1 signature and more rapid disease recurrence (**Figure 5.19C**) and lower overall survival (**Figure 5.19D**). Interestingly, we noted that theses analyses also included several patients that received a low-grade Gleason score (Gleason=6) but displayed worse outcomes, consistent with the clinical observation that 10 – 20% of men with low-grade cancers experience disease recurrence following radical prostatectomy (*36*) and 6% of men with low-grade cancers experience prostate-cancer-specific mortality from metastatic disease within 15 years (*37*). Finally, we assessed *SChLAP-1* expression in urine sediment RNA from 111 biopsy-confirmed prostate cancer patients with Gleason score. We found that *SChLAP-1* was significantly higher in Gleason 7 patients vs. Gleason 6 patients ($p = 0.01$) (**Figure 5.20**), although we were unable to evaluate Gleason 8-10 due to low numbers of patients presenting with high Gleason scores at the time of PSA screening.

DISCUSSION

Here, we have discovered a novel, oncogenic lncRNA involved in prostate cancer cell aggressiveness and the suppression of the SWI/SNF complex. We find that *SChLAP-1* directly binds proteins in the SWI/SNF complex through a hairpin-like structure, leading to reversion of SWI/SNF-mediated gene expression (**Figure 5.21**). We further find that *SChLAP-1* coordinates prostate cancer cell invasion *in vitro* and metastatic

spread *in vivo*, and *SChLAP-1* expression characterizes a metastatic-like gene expression profile associated with high-grade localized prostate cancers and poor clinical outcomes. *SChLAP-1* may therefore represent a promising prognostic biomarker for localized prostate cancers. Taken together, our discovery of *SChLAP-1* has broad implications for prostate cancer biology and provides supporting evidence for the role of lncRNAs in the progression of aggressive cancers.

In several respects, *SChLAP-1* resembles the lncRNA *HOTAIR*, which exerts oncogenic functions through the coordination of several epigenetic complexes such as PRC2 and the co-repressor NuRD (*6, 7, 9*). Like *HOTAIR*, which is not expressed in prostate cancer (*13*), *SChLAP-1* coordinates regulation of gene expression and characterizes aggressive disease. However, unlike *HOTAIR*, which assists PRC2 function, *SChLAP-1* impairs SWI/SNF function. It is intriguing to note that SWI/SNF and PRC2 have been show to play opposing roles in regulating gene expression (*38, 39*), suggesting that upregulation of *SChLAP-1* may phenocopy PRC2 behavior in the suppression of SWI/SNF-mediated gene regulation.

Moreover, *SChLAP-1* appears to operate in a tissue-specific manner, showing outlier expression exclusively in prostate cancer. This is in contrast to *HOTAIR*, which has been investigated in multiple cancer types (*6, 10, 11*), as well as benign cell types such as fibroblasts (*7, 9*), which show high levels of endogenous *HOTAIR* expression. In this respect, *SChLAP-1* provides a compelling argument for the importance of tissue-specific lncRNAs and suggests that other malignancies likewise employ tissue-specific RNA species to coordinate cancer biology.

MATERIALS AND METHODS

Cell lines

All cell lines were obtained from the American Type Culture Collection (Manassas, VA). Cell lines were maintained using standard media and conditions. SChLAP-1 or control-expressing cell lines were generated by cloning SChLAP-1 or control into the pLenti6 vector (Invitrogen). Stably-transfected RWPE, HME and MCF7 cells were selected using blasticidin (Invitrogen). For LNCAP and 22Rv1 cells with stable knockdown of SChLAP-1, cells were transfected with SChLAP-1 or non-targeting shRNA lentiviral constructs for 48 hours. GFP+ cells were drug-selected using puromycin.

RNA isolation; cDNA synthesis; and PCR experiments

RNA isolation and cDNA synthesis was performed according standard protocols. Quantitative PCR was performed using Power SYBR Green Mastermix (Applied Biosystems, Foster City, CA), using GAPDH and HMBS as housekeeping control genes. The relative quantity of the target gene was completed for each sample using the ΔΔCt method. The primer sequences for the transcript analyzed are provided in **Appendix 2**.

Murine intracardiac and subcutaneous *in vivo* models

All experimental procedures were approved by the University of Michigan Committee for the Use and Care of Animals (UCUCA). *Intracardiac injection model*: 5 x 10^5 cells were introduced to CB-17 severe combine immunodefiecient mice (CB-17 SCID) at 6 weeks of age. Beginning one week post injection, bioluminescent imaging of mice was performed weekly using a CCD IVIS system with a 50-mm lens (Xenogen Corp.) and the

results were analyzed using LivingImage software (Xenogen). See SOM for additional details. *Subcutaneous injection model*: 1 x 10^6 cells were introduced to mice (CB-17 SCID), ages 5-7 weeks, with a matrigel scaffold (BD Matrigel Matrix, BD Biosciences) in the posterior dorsal flank region (n = 10 per cell line). Tumors were measured weekly using a digital caliper, and endpoint was determined as a tumor volume of 1000 mm^3.

Immunoblot Analysis

Cells were lysed in RIPA lysis buffer (Sigma, St. Louis, MO) supplemented with HALT protease inhibitor (Fisher). Western blotting analysis was performed with standard protocols using Polyvinylidene Difluoride membrane (GE Healthcare, Piscataway, NJ) and the signals visualized by enhanced chemiluminescence system as described by the manufacturer (GE Healthcare).

siRNA knockdown, proliferation and invasion studies

Cells were plated in 100mM plates at a desired concentration and transfected with 20uM experimental siRNA oligos or non-targeting controls according to standard protocols. 72 hours post-transfection with siRNA, cells were trypsinized, counted with a Coulter counter, and diluted to 1 million cells/mL. Proliferation assays were performed with a Coulter counter, and invasion of cells through matrigel (BD Biosciences) was performed according to standard protocols.

Chromatin immunoprecipitation

ChIP assays were performed as described previously.

RNA immunoprecipitation

RIP assays were performed using a Millipore EZ-Magna RIP RNA-Binding Protein Immunoprecipitation kit (Millipore, #17-701) according to the manufacturer's instructions.

Statistical analyses for experimental studies

All data are presented as means ± S.E.M. All experimental assays were performed in duplicate or triplicate. Statistical analyses shown in figures represent Fisher's exact tests or two-tailed Student t-tests, as indicated.

FIGURES

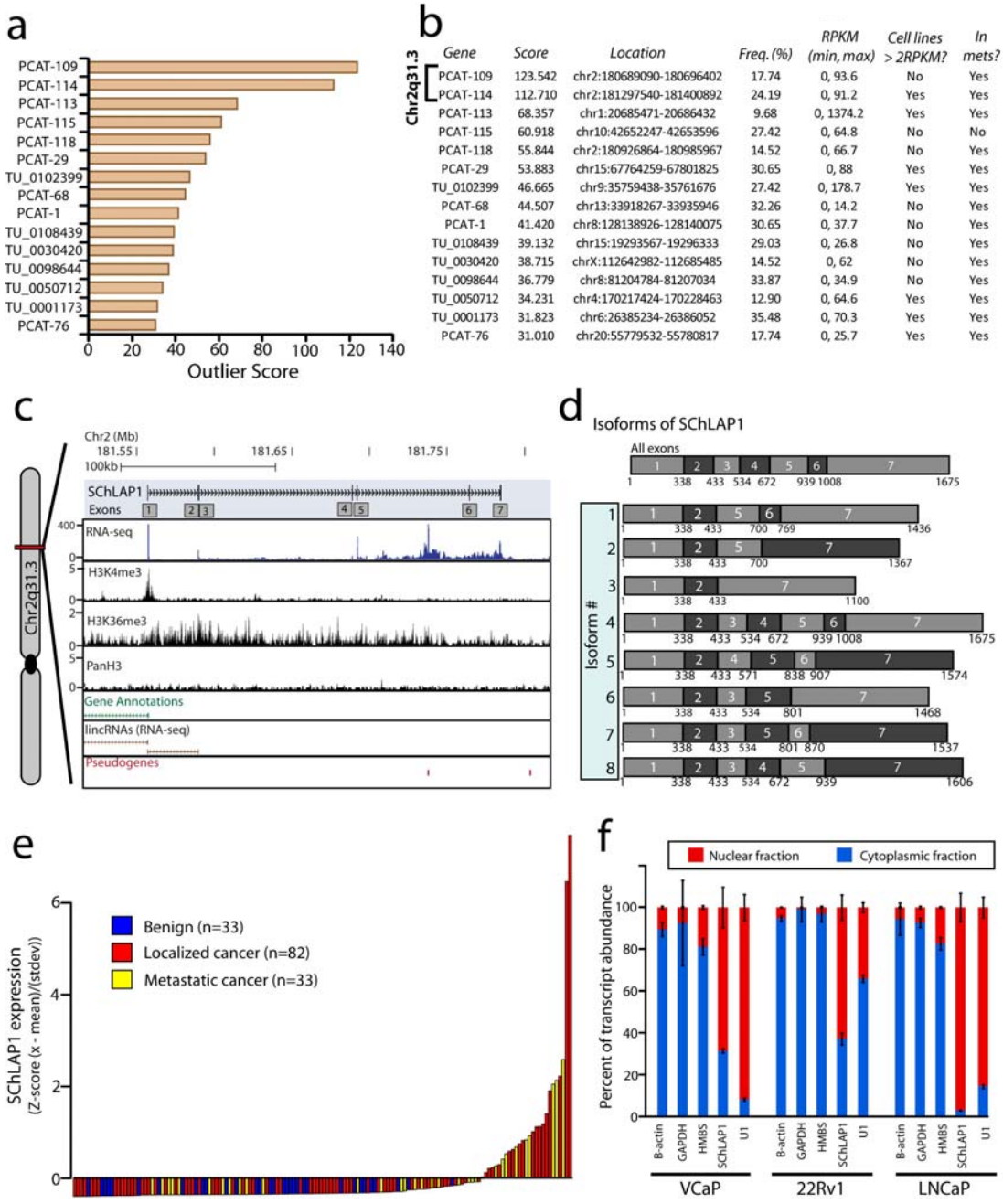


Figure 5. 1: Discovery of *SChLAP-1* as a prostate cancer lncRNA. (**a**) Cancer outlier profile analysis (COPA) for intergenic lncRNAs in prostate cancer nominates two transcripts, *PCAT-109* and *PCAT-114*, as prominent outliers. (**b**) A comparison of lncRNA outliers nominated by COPA, including their location, frequency in clinical samples, their expression in tissues and cell lines, and whether they occur in metastatic

prostate samples. (**c**) A representation of the *SChLAP-1* gene and its annotations in current databases. *SChLAP-1* may consist of up to seven exons on Chr2q31.3. An aggregated representation of current gene annotations for Ensembl, ENCODE, UCSC, Ref-Seq, and Vega shows no annotation for *SChLAP-1*. No spliced ESTs represent *SChLAP-1*. ChIP-Seq data for H3K4me3, RNA-Pol II, and H3K36me3 show enrichment at the *SChLAP-1* gene. Also, RNA-Seq data showing an outlier sample for *SChLAP-1* illustrates its expression. (**d**) A schematic summarizing the observed *SChLAP-1* isoforms. A total of 8 isoforms were observed, with isoform #1 and isoform #2 accounting for >90% of transcripts. (**e**) qPCR for *SChLAP-1* on a panel of benign prostate (n=33), localized prostate cancer (n=82), and metastatic prostate cancer (n=33) samples. qPCR data is normalized to the average of (*GAPDH* + *HMBS*) and represented as standardized expression values. (**f**) *SChLAP-1* expression is predominantly nuclear. Prostate cell lysates were fractionated and nuclear and cytoplasmic fractions were tested for RNA expression. *U1* is a positive control for nuclear gene expression.

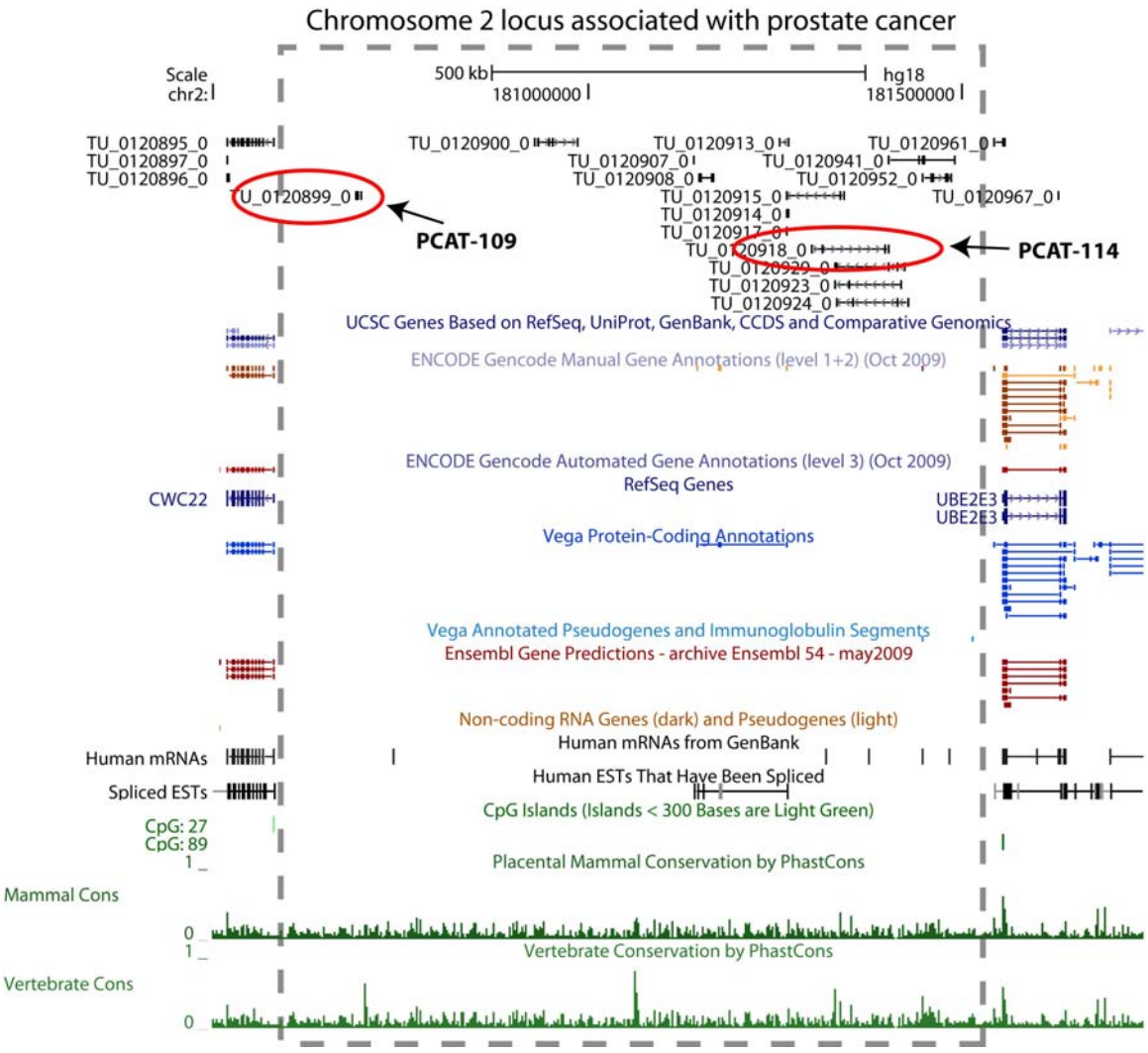


Figure 5.2: A chromosome 2 region contains prostate cancer-associated transcripts. A gene desert on chr2q31 between *CWC22* and *UBE2E3* contains multiple transcripts that are upregulated in prostate cancer, including the predicted outliers *PCAT-109* and *PCAT-114*. These transcripts are unannotated in major gene annotation databases.

Figure 5.3: The structure and sequence of *SChLAP1*. The sequences of the seven exons found in the SChLAP1 gene are detailed here, indicating the presence of splice donor and splice acceptor sites. Nucleotides comprising the SChLAP1 gene are capitalized, whereas genomic non-SChLAP1 basepairs are in lower case font.

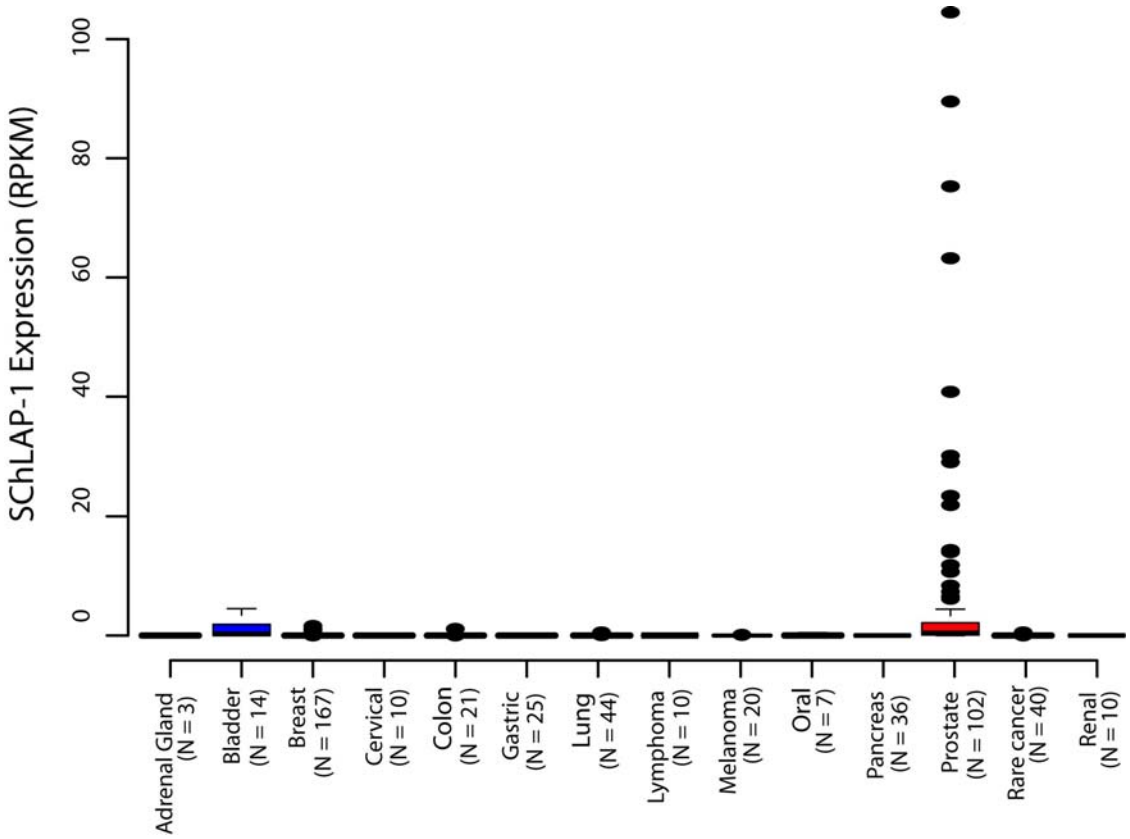


Figure 5.4: Expression of *SChLAP-1* across cancers. Over 500 cancer samples analyzed by RNA-seq were interrogated for *SChLAP-1* expression. High expression was observed in a subset of prostate cancer samples, whereas all other cancer types displayed minimal expression.

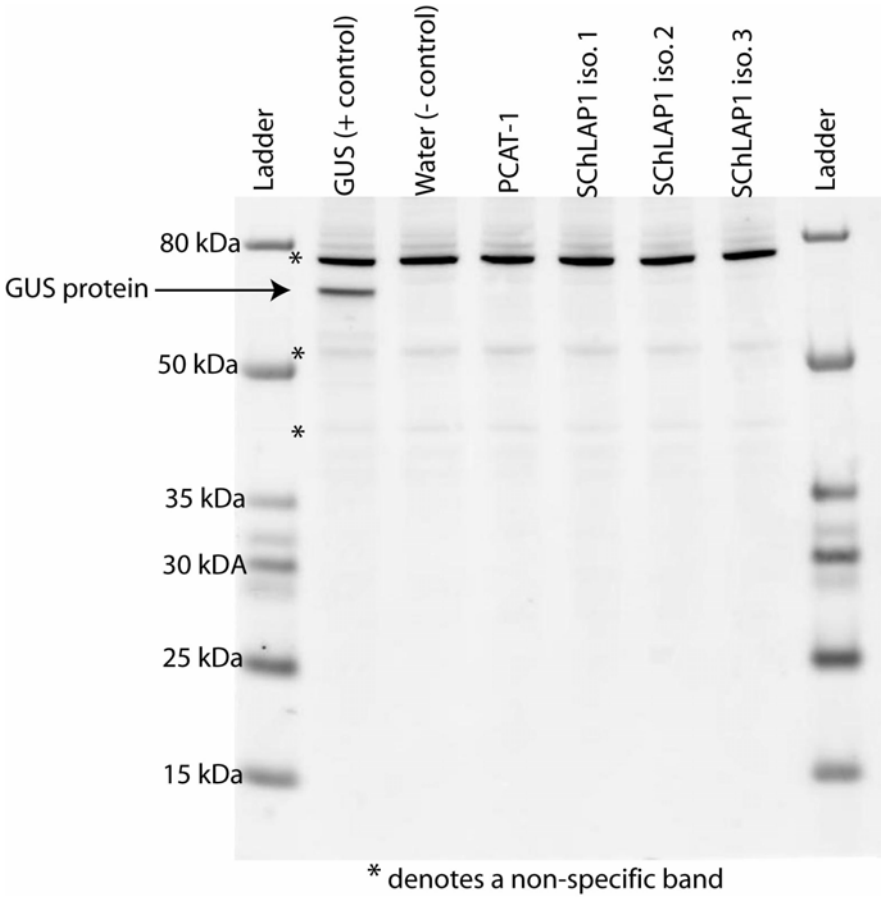


Figure 5.5: *In vitro* translation assays for *SChLAP-1*. Three isoforms of *SChLAP-1* were cloned and tested for protein-coding capacity using an *in vitro* translation assay (see SOM methods). GUS is used as a positive control. *PCAT-1* and water serve as negative controls. Non-specific bands are indicated with an asterisk. *SChLAP-1* isoforms do not generate a protein in this assay.

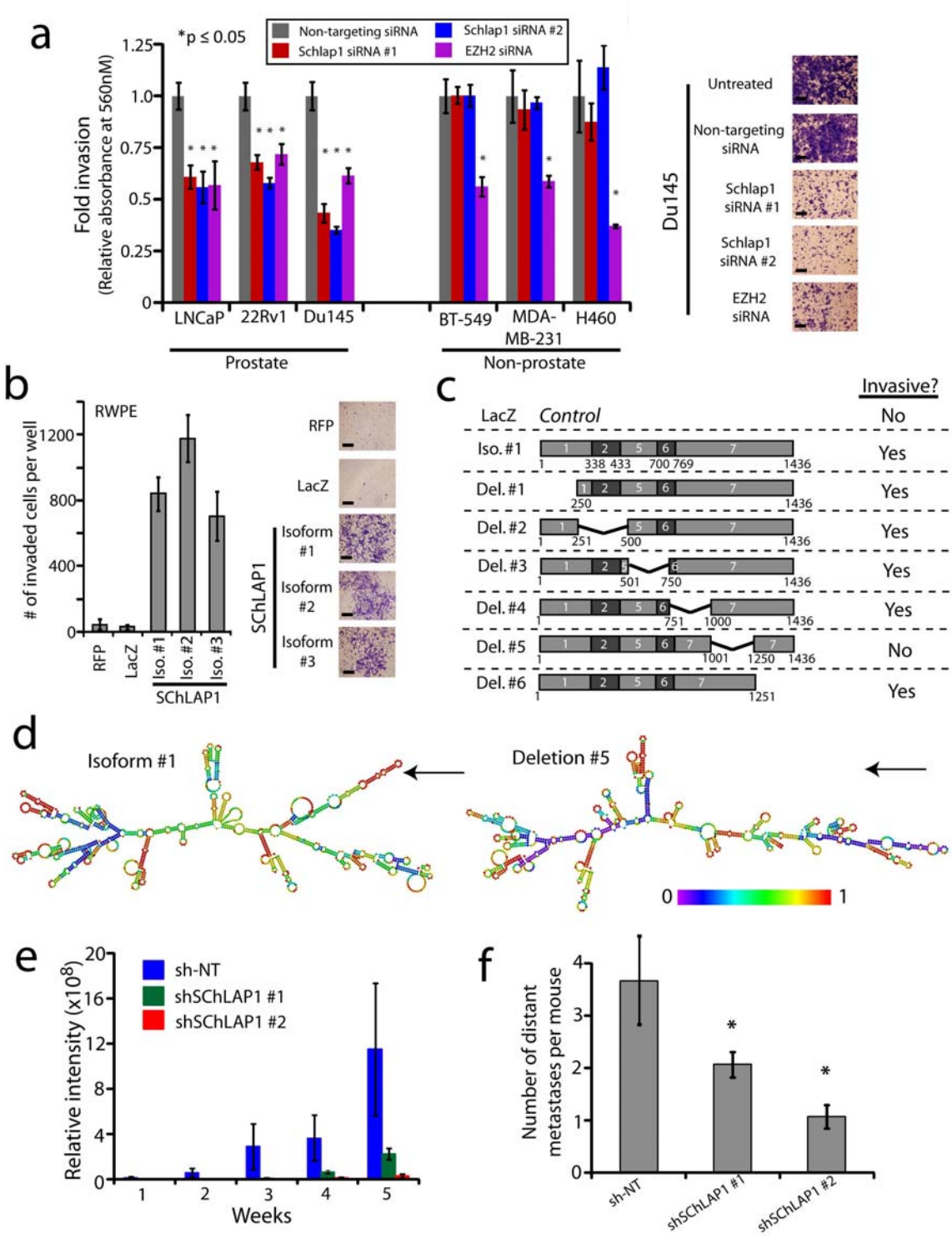


Figure 5.6: *SChLAP-1* coordinates prostate cancer cell invasion. (**a**) siRNA knockdown of *SChLAP-1 in vitro*. Three prostate cell lines (LNCaP, 22Rv1, Du145) and three non-prostate cell lines (BT-549, MDA-MB-231, H460) were treated with two

independent siRNAs for *SChLAP-1* and invasion through Matrigel in a Boyden chamber assay was monitored. *EZH2* siRNA serves as a positive control. (**b**) Overexpression of *SChLAP-1* in RWPE cells. Benign RWPE prostate cells overexpressing three isoforms of *SChLAP-1*, but not controls, demonstrate increased cellular invasion. (**c**) Deletion analysis of *SChLAP-1*. Deletion constructs of *SChLAP-1* were overexpressed in RWPE cells and the resulting cells were assayed for invasion in a Boyden chamber assay. Deletion of bps 1001-1250 impairs *SChLAP-1*-induced invasion. (**d**) RNA structural analysis of *SChLAP-1*. The RNA structure of wild-type *SChLAP-1* isoform #1 and deletion construct #5 were analyzed by RNAfold. (**e**) Tumor seeding with *SChLAP-1* knockdown *in vivo*. Intracardiac injection of 22Rv1 cells with stable *SChLAP-1* knockdown was performed in severe combined immunodeficient (SCID) mice, and metastatic seeding and growth of tumor cells was monitored. The relative intensity of whole-mouse luciferase signal is plotted. (**f**) The number of gross metastatic sites observed by luciferase signal in 22Rv1 sh*SChLAP-1* cells or shNT controls. Independent foci of luciferase signal were averaged for shNT (n=9), sh*SChLAP-1* #1 (n=14) and sh*SChLAP-1* #2 (n=14) mice. An asterisk (*) indicates $p < 0.05$ by Student's t-test.

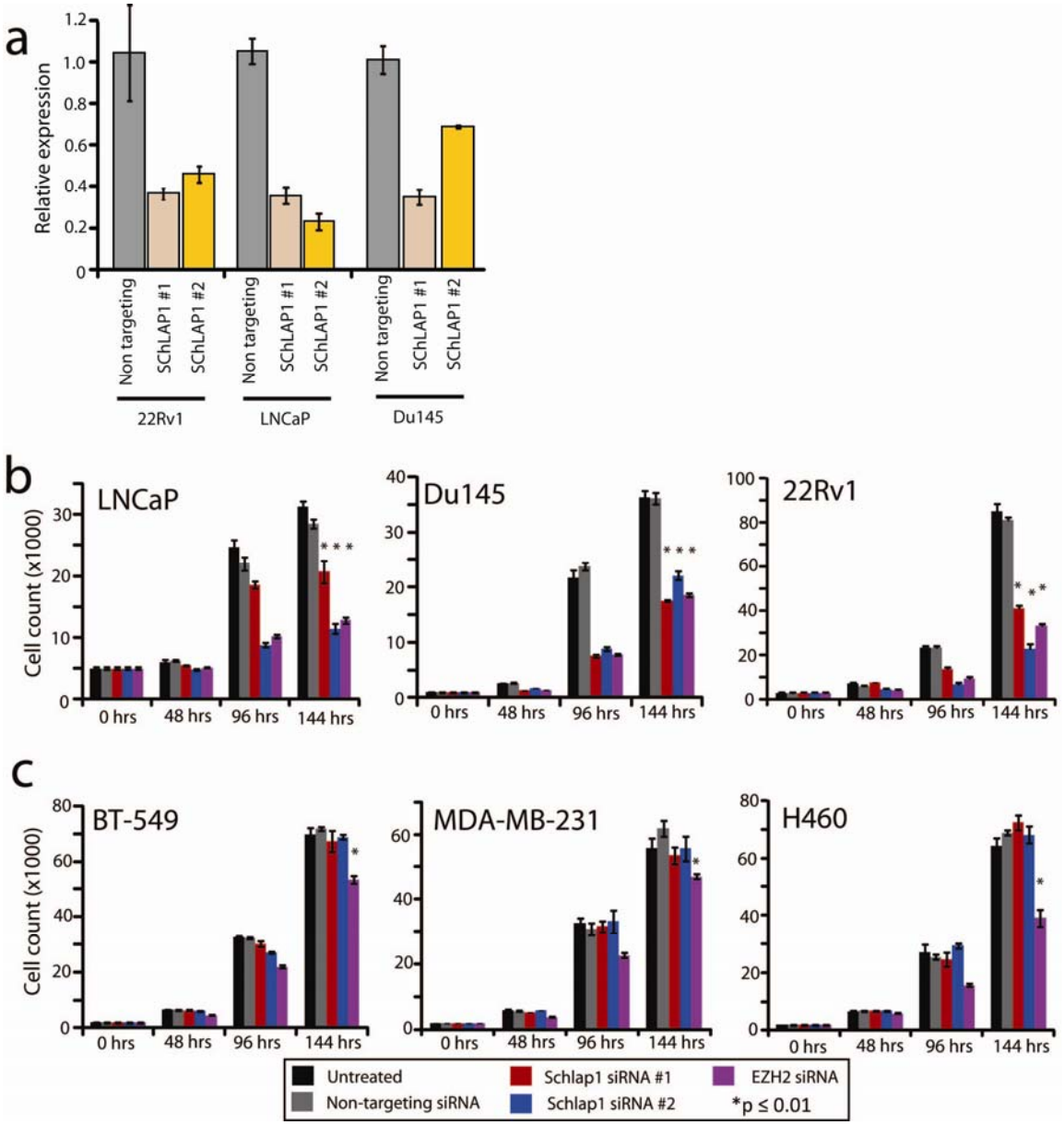


Figure 5.7: Knockdown of *SChLAP-1* impairs prostate cancer cell growth. (**a**) 22Rv1, LNCaP, and Du145 cells were treated with siRNAs against *SChLAP-1*. qPCR indicates relative knockdown efficiency in these cell lines. (**b**) Cell proliferation assays for prostate cancer cell lines (LNCaP, 22Rv1, Du145) treated with *SChLAP-1* siRNAs or non-targeting negative controls. *EZH2* siRNA serves as a positive control. (**c**) Cell proliferation assays for non-prostate cancer cell lines (BT-549, MDA-MB-231, H460) treated with *SChLAP-1* siRNAs or non-targeting negative controls. *EZH2* siRNA serves as a positive control. Error bars indicate S.E.M. An asterisk (*) indicates $p < 0.05$ by Student's t-test.

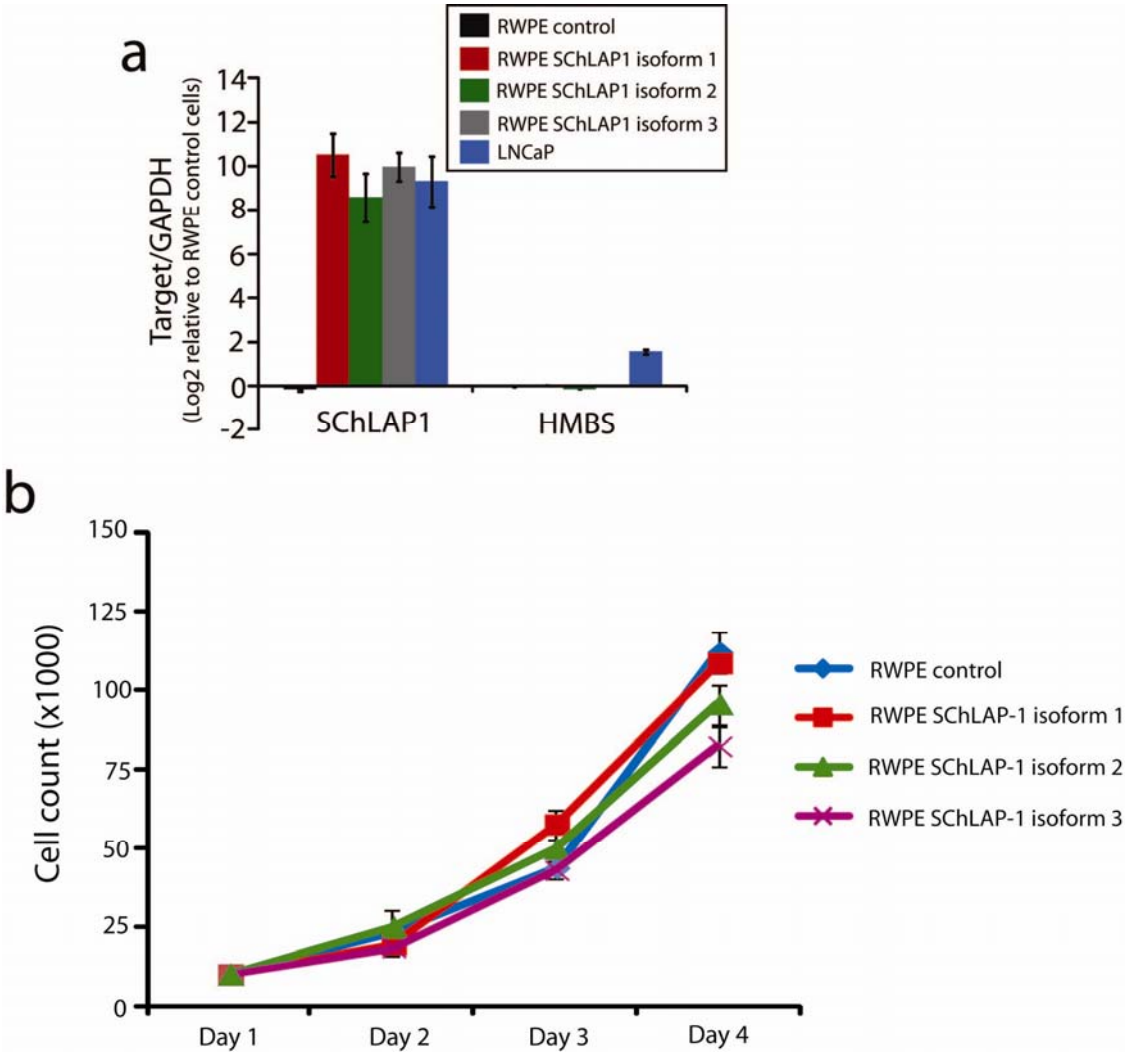


Figure 5.8: Overexpression of *SChLAP-1* does not increase cell proliferation. (**a**) Overexpression of *SChLAP-1* isoforms 1-3 in RWPE cells was confirmed using qPCR, which demonstrated that the overexpression resulted in comparable levels of *SChLAP-1* transcript to LNCaP cells that express this gene endogenously. HMBS serves as a negative control. (**b**) Cell proliferation assays for RWPE cells overexpressing *SChLAP-1* isoforms. No significant change in cell proliferation is observed. Error bars represent S.E.M.

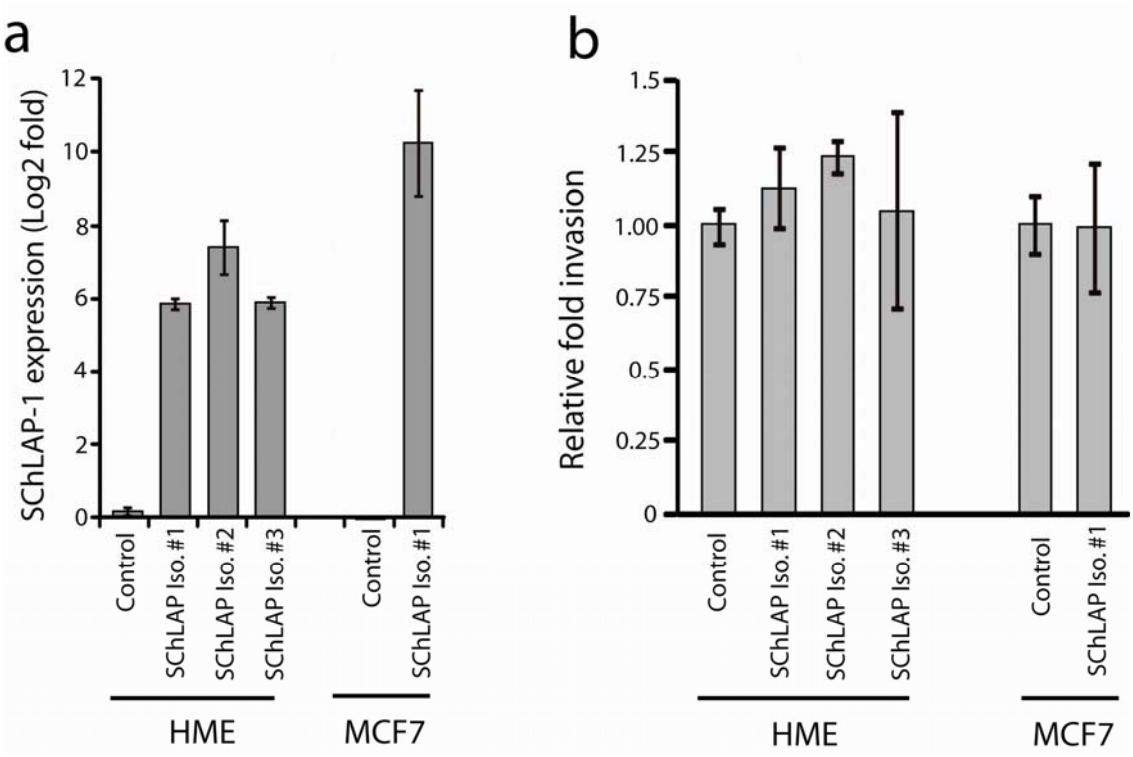


Figure 5.9: *SChLAP-1* overexpression in breast cells does not increase invasion. (**a**) HME immortalized benign breast cells or MCF7 breast cancer cells were infected with *SChLAP-1* lentivirus and stably-overexpressing cells were generated. *SChLAP-1* expression was measured by qPCR. (**b**) Invasion assays were performed for HME and MCF7 cells overexpressing *SChLAP-1*. Invaded cells were stained with crystal violet and the absorbance at 560nM was measured. Overexpression of *SChLAP-1* did not result in a significant increase in cell invasion ($p > 0.05$, Student's t-test). Error bars represent S.E.M.

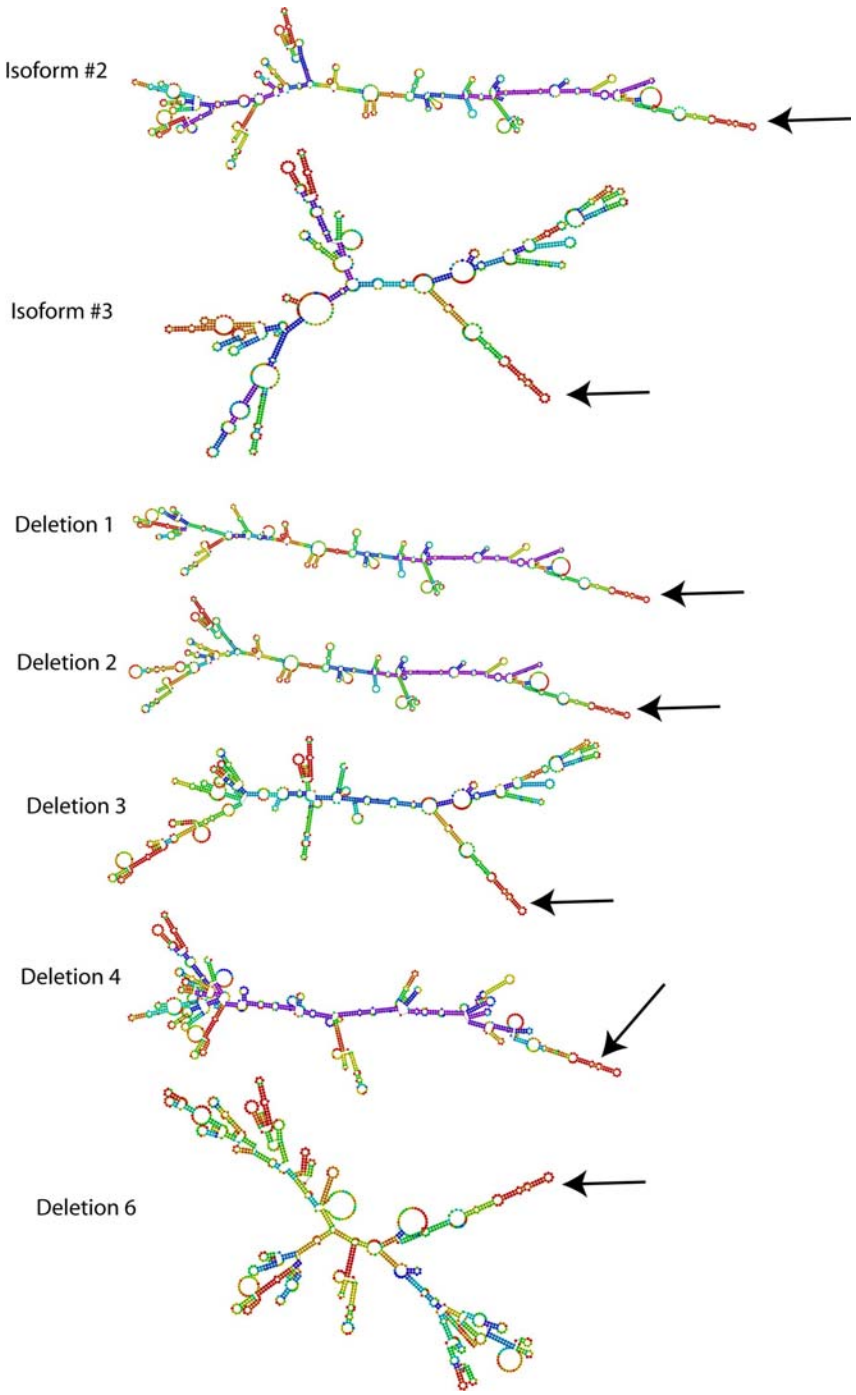


Figure 5.10: Structural predictions of *SChLAP-1* isoforms and deletion constructs. The sequences for *SChLAP-1* isoforms #2 and #3 and deletion constructs 1-4 and 6 were analyzed by RNAfold to predict secondary structures. Arrows identify the structural hairpin lost in *SChLAP-1* deletion construct #5.

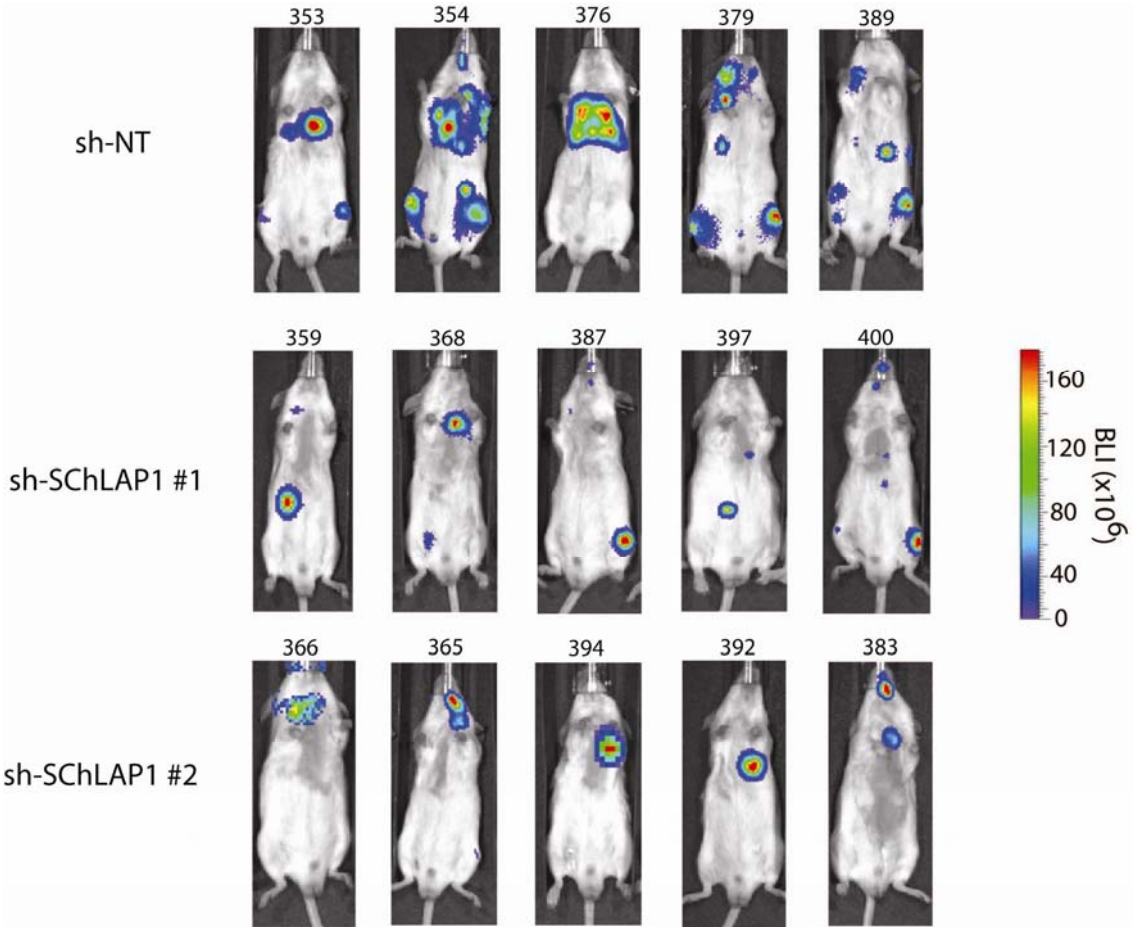


Figure 5.11: Intracardiac injection of *SChLAP-1* knockdown cells impairs cancer cell metastasis. 22Rv1-shNT, sh-*SChLAP-1* #1, and sh-*SChLAP-1* #2, cells were introduced into mice via intracardiac injection and luciferase signal was monitored for distant metastases. Whole-body images of mice (n=5 for each treatment in this figure) indicate that *SChLAP-1* knockdown impairs metastatic seeding. Above each image, the mouse ID is given.

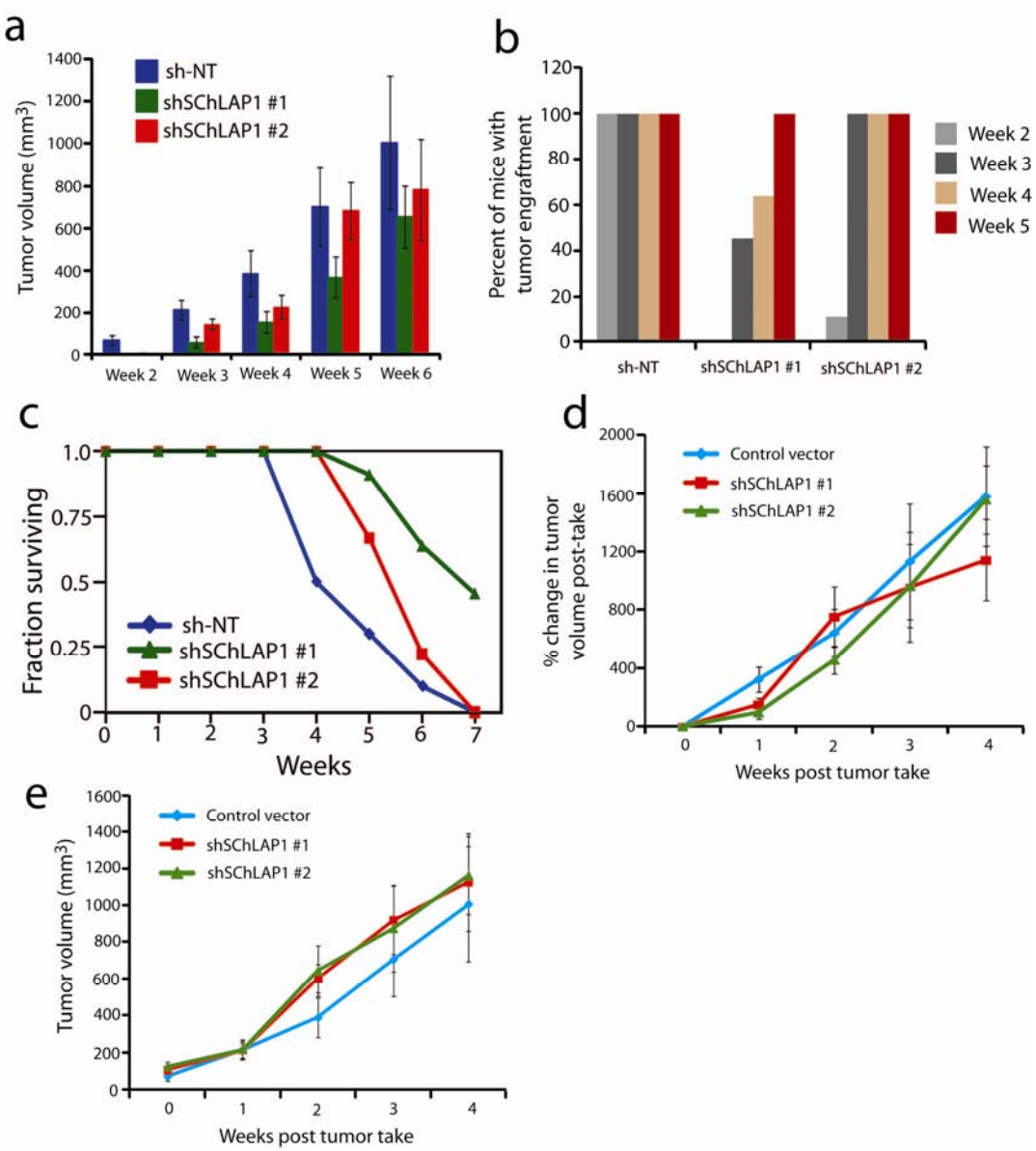


Figure 5.12: Knockdown of *SChLAP-1* delays tumor take but not tumor growth kinetics. 22Rv1 cells infected with lentivirus for shNT, sh-*SChLAP1* #1, and sh-*SChLAP1* #2 were injected subcutaneously in mouse flanks and tumor growth was monitored by caliper measurements. N = 10 mice for shNT cells, n = 12 mice for sh-*SChLAP1* #1 cells, n = 9 mice for sh-*SChLAP1* #2 cells. (**a**) Absolute tumor volume for 22Rv1 shNT, sh-SChLAP1 #1 and sh-SChLAP1 #2 cells. (**b**) Percent of mice with tumor engraftment over time. Knockdown of SChLAP-1 delays the onset of tumor engraftment. (**c**) The fraction of mice surviving following subcutaneous injection of the 22Rv1 cell lines. This plot represents tumor-specific death of mice sacrificed when the tumor volume reached the maximum allowable volume. (**d**) The percent change in tumor volume per cell line normalized to the time of tumor engraftment. (**e**) Tumor volume normalized to the time of tumor engraftment. Error bars indicate S.E.M.

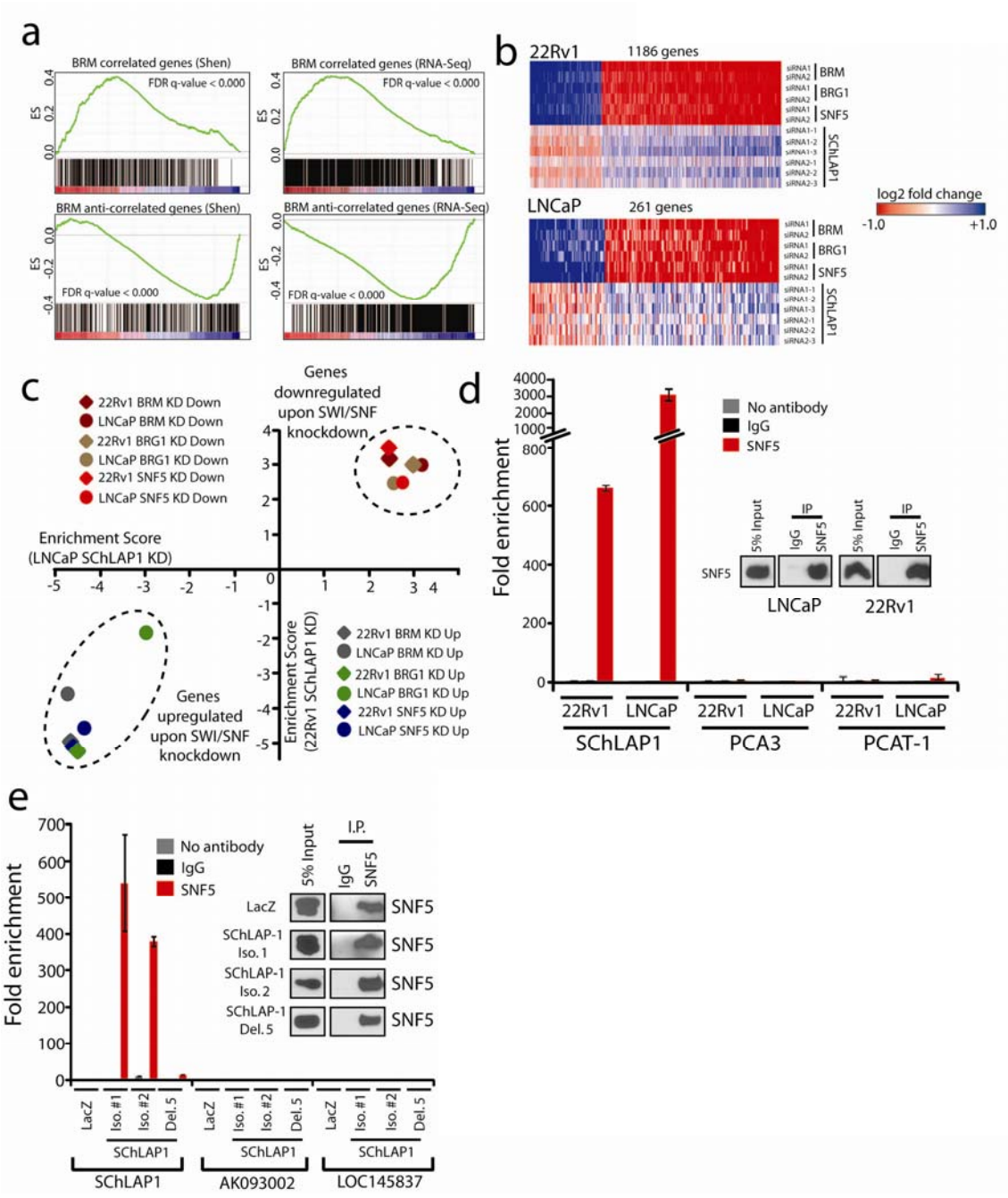


Figure 5.13: *SChLAP-1* antagonizes SWI/SNF complex function. (**a**) Gene set enrichment analysis (GSEA) of LNCaP and 22Rv1 cells treated with *SChLAP-1* siRNAs. GSEA results indicate that *SChLAP-1* knockdown results are inversely correlated with *BRM*-associated genes. Left, GSEA associations using BRM signatures from Shen *et al.* Right, GSEA associations using RNA-seq data. (**b**) Heatmap results for *SChLAP-1* or SWI/SNF knockdown in LNCaP and 22Rv1 cells. *SChLAP-1* and SWI/SNF proteins (BRM, BRG1, SNF5) regulate expression of the same genes in opposing manners. (**c**)

GSEA analysis of *SChLAP-1* and SWI/SNF knockdowns. Samples from (**b**) were analyzed by GSEA for overlap in gene expression regulation. Across two cell lines (LNCaP and 22Rv1), *SChLAP-1* knockdown had the opposite effect on gene expression as knockdown of *BRM*, *BRG1* or *SNF5*. (**d**) RNA immunoprecipitation (RIP) of SNF5 demonstrates *SChLAP-1* binding to SNF5 in 22Rv1 and LNCaP cells. *PCA3* and *PCAT-1* serve as negative controls. (**e**) RIP analysis of SNF5 in RWPE cells overexpressing *LacZ*, *SChLAP-1* isoform #1, *SChLAP-1* isoform #2, or *SChLAP-1* deletion construct 5. *AK093002* and *LOC145837* serve as negative controls.

a

UP						
	22Rv1-BRM	22Rv1-BRG1	22Rv1-SNF5	LNCaP-BRM	LNCaP-BRG1	LNCaP-SNF5
22Rv1-BRM	1276	723	630	270	203	601
22Rv1-BRG1	723	1055	583	245	173	544
22Rv1-SNF5	630	582	922	204	155	511
LNCaP-BRM	270	245	204	559	157	271
LNCaP-BRG1	203	173	154	157	785	198
LNCaP-SNF5	601	544	511	271	198	1010

DOWN						
	22Rv1-BRM	22Rv1-BRG1	22Rv1-SNF5	LNCaP-BRM	LNCaP-BRG1	LNCaP-SNF5
22Rv1-BRM	481	265	230	23	30	122
22Rv1-BRG1	265	442	231	23	28	114
22Rv1-SNF5	230	231	396	27	27	120
LNCaP-BRM	23	23	27	133	30	57
LNCaP-BRG1	30	28	26	30	110	44
LNCaP-SNF5	122	114	120	58	44	413

b

Two-tailed Fisher's exact test p-value		
	UP	DOWN
22Rv1 BRM vs 22Rv1 BRG1	<0.0001	<0.0001
22Rv1 BRM vs 22Rv1 SNF5	<0.0001	<0.0001
22Rv1 BRM vs LNCaP BRM	<0.0001	<0.0001
22Rv1 BRM vs LNCaP BRG1	<0.0001	<0.0001
22Rv1 BRM vs LNCaP SNF5	<0.0001	<0.0001
22Rv1 BRG1 vs 22Rv1 SNF5	<0.0001	<0.0001
22Rv1 BRG1 vs LNCaP BRM	<0.0001	<0.0001
22Rv1 BRG1 vs LNCaP BRG1	<0.0001	<0.0001
22Rv1 BRG1 vs LNCaP SNF5	<0.0001	<0.0001
22Rv1 SNF5 vs LNCaP BRM	<0.0001	<0.0001
22Rv1 SNF5 vs LNCaP BRG1	<0.0001	<0.0001
22Rv1 SNF5 vs LNCaP SNF5	<0.0001	<0.0001
LNCaP BRM vs LNCaP BRG1	<0.0001	<0.0001
LNCaP BRM vs LNCaP SNF5	<0.0001	<0.0001
LNCaP BRG1 vs LNCaP SNF5	<0.0001	<0.0001

Figure 5.14: Knockdown of SWI/SNF components produces overlapping sets of regulated genes. (**a**) A table tabulating the number of genes upregulated or downregulated by knockdown of SWI/SNF components (>2-fold change), and the overlap between gene sets is indicated. (**b**) Statistical significance of the overlapping gene sets for SWI/SNF knockdowns. Significance is determined by a two-tailed Fisher's exact test.

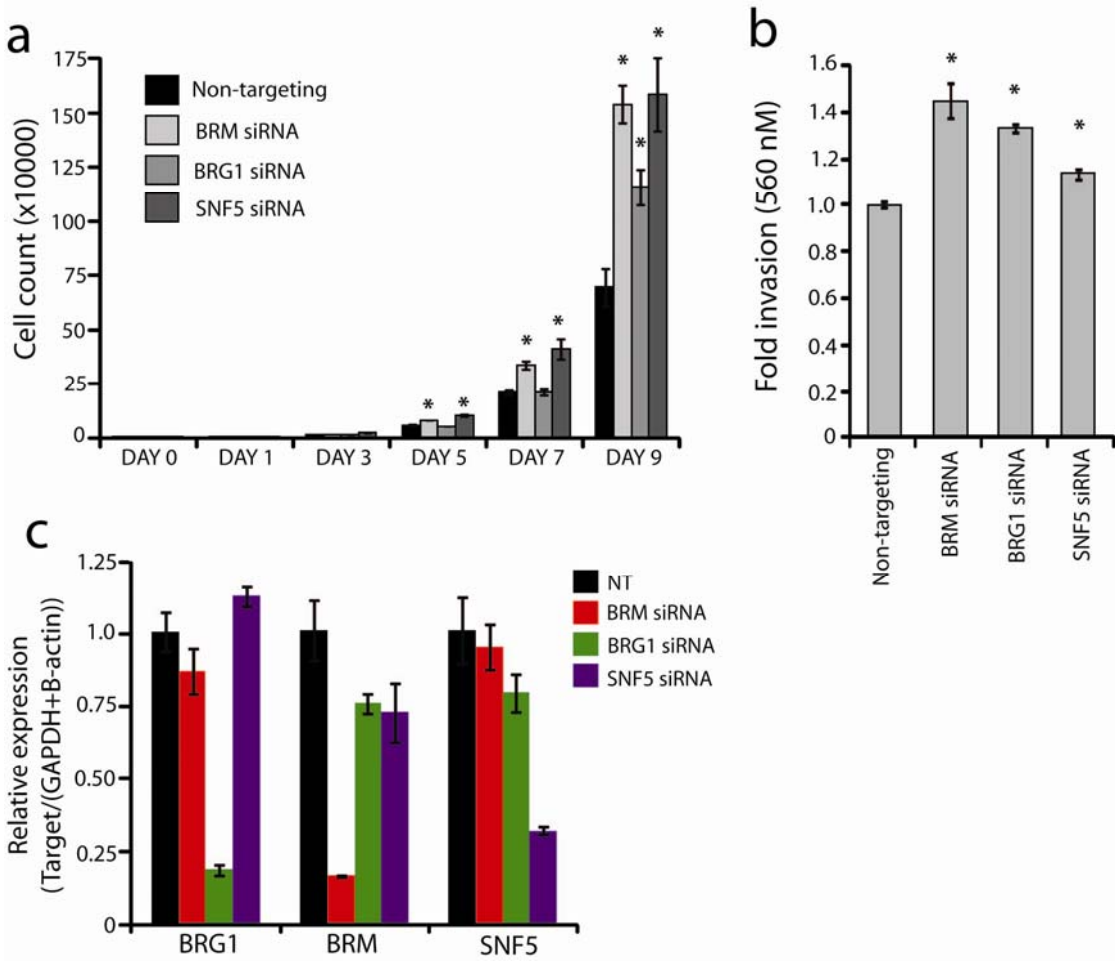


Figure 5.15: Knockdown of SWI/SNF components increases cell proliferation and invasion in 22Rv1 cells. (**a**) Cell proliferation assays for 22Rv1 cells treated with *BRG1*, *BRM*, *SNF5* or control non-targeting siRNA were examined. Knockdown of SWI/SNF members increases cell proliferation. (**b**) Boyden chamber cell invasion assays for 22Rv1 cells treated with *BRG1*, *BRM*, *SNF5* or control non-targeting siRNA were examined for invasiveness through matrigel. Invaded cells were stained with crystal violet and absorbance at 560nM was measured. Knockdown of SWI/SNF components modestly increases cell invasiveness. (**c**) Relative expression level of *SNF5*, *BRG1*, and *BRM* in 22Rv1 cells treated with siRNAs targeting those genes. Error bars represent S.E.M. An asterisk (*) indicates $p < 0.05$ by Student's t-test.

			LNCAP SChLAP-1 siRNA			
	Cell line	Concept	ES	NES	NOM p-val	FDR q-val
SWI/SNF siRNA	22Rv1	BRM_KD_DOWN_2-FOLD	0.326	2.434	0	0.0084
	22Rv1	BRM_KD_UP_2-FOLD	-0.368	-4.685	0	0
	22Rv1	BRG1_KD_DOWN_2-FOLD	0.396	2.965	0	0
	22Rv1	BRG1_KD_UP_2-FOLD	-0.399	-4.600	0	0
	22Rv1	SNF5_KD_DOWN_2-FOLD	0.326	2.403	0	0.0115
	22Rv1	SNF5_KD_UP_2-FOLD	-0.394	-4.610	0	0
	LNCaP	BRM_KD_DOWN_2-Fold	0.502	3.136	0	0
	LNCaP	BRM_KD_UP_2-FOLD	-0.441	-4.749	0	0
	LNCaP	BRG1_KD_DOWN_2-FOLD	0.441	2.493	0	0.0077
	LNCaP	BRG1_KD_UP_2-FOLD	-0.280	-3.014	0	0.0002
	LNCaP	SNF5_KD_DOWN_2-FOLD	0.356	2.779	0	0
	LNCaP	SNF5_KD_UP_2-FOLD	-0.381	-4.376	0	0

			22Rv1 SChLAP-1 siRNA			
	Cell line	Concept	ES	NES	NOM p-val	FDR q-val
SWI/SNF siRNA	22Rv1	BRM_KD_DOWN_2-FOLD	0.358	3.202	0	0
	22Rv1	BRM_KD_UP_2-FOLD	-0.395	-4.982	0	0
	22Rv1	BRG1_KD_DOWN_2-FOLD	0.347	3.019	0	0
	22Rv1	BRG1_KD_UP_2-FOLD	-0.418	-5.130	0	0
	22Rv1	SNF5_KD_DOWN_2-FOLD	0.402	3.461	0	0
	22Rv1	SNF5_KD_UP_2-FOLD	-0.426	-5.161	0	0
	LNCaP	BRM_KD_DOWN_2-Fold	0.486	3.061	0	0
	LNCaP	BRM_KD_UP_2-FOLD	-0.322	-3.532	0	0
	LNCaP	BRG1_KD_DOWN_2-FOLD	0.401	2.481	0	0.0041
	LNCaP	BRG1_KD_UP_2-FOLD	-0.157	-1.831	0	0.1711
	LNCaP	SNF5_KD_DOWN_2-FOLD	0.291	2.506	0	0.0026
	LNCaP	SNF5_KD_UP_2-FOLD	-0.372	-4.571	0	0

Figure 5.16: Summarized GSEA results for *SChLAP-1* and SWI/SNF knockdowns. *SChLAP-1*, *BRM*, *BRG1*, and *SNF5* were knocked-down using siRNAs in 22Rv1 LNCaP cells. Gene expression changes were compared using GSEA to expression changes. GSEA enrichment scores, as well as p-values and FDR q-values, are reported for all pair-wise comparisons performed. 23 of 24 comparisons show a significant GSEA gene expression overlap at FDR < 0.05.

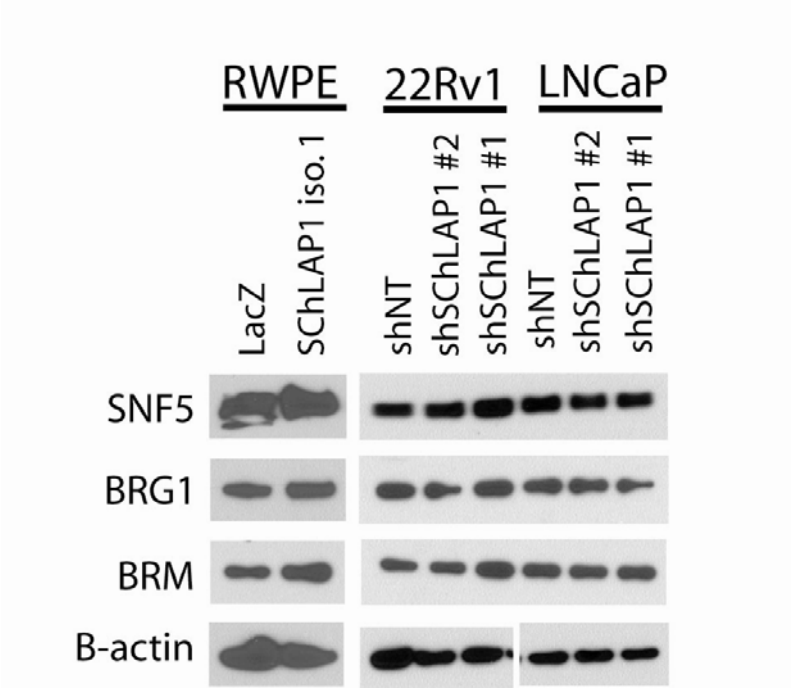


Figure 5.17: *SChLAP-1* does not impact SWI/SNF protein expression. Western blot analysis of 22Rv1 and LNCaP cells with stable knockdown of *SChLAP-1* or non-targeting control demonstrates no change in the protein levels of SNF5, BRG1, or BRM upon *SChLAP-1* knockdown. Similarly, overexpression of SChLAP1 in RWPE cells does not alter SWI/SNF protein expression.

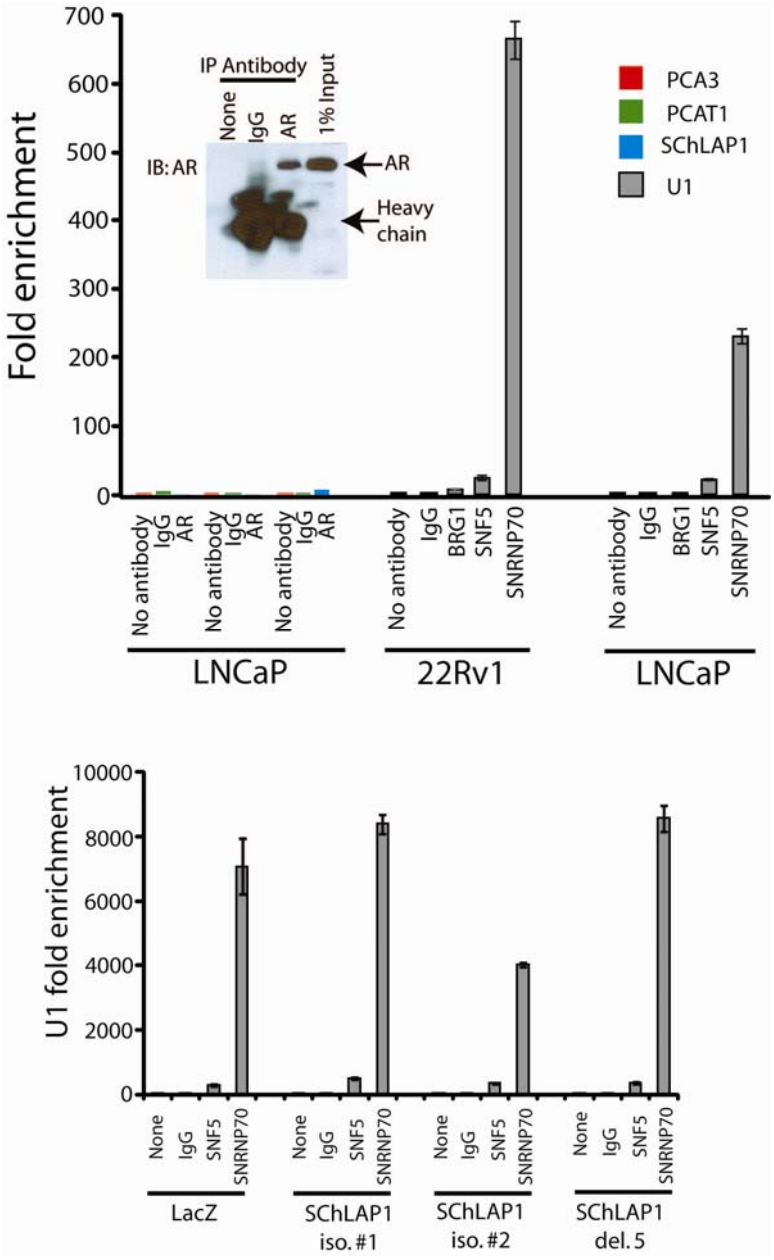


Figure 5.18: Immunoprecipitation of AR and SNRNP70 controls for RNA-IP experiments. *Left*, Androgen receptor (AR) was immunoprecipitated from LNCaP cells, and presence of SChLAP-1, PCAT1, or PCA3 RNA transcripts was measured by qPCR. No significant enrichment was detected. A Western blot for AR confirms efficient IP of AR protein in this experiment. *Right*, RNA-IP experiments for SWI/SNF complex members as well as SNRNP70 shows specific binding of SNRNP70 to the U1 ncRNA, indicating specificity of the RNA-IP experiments. Error bars indicate S.E.M.

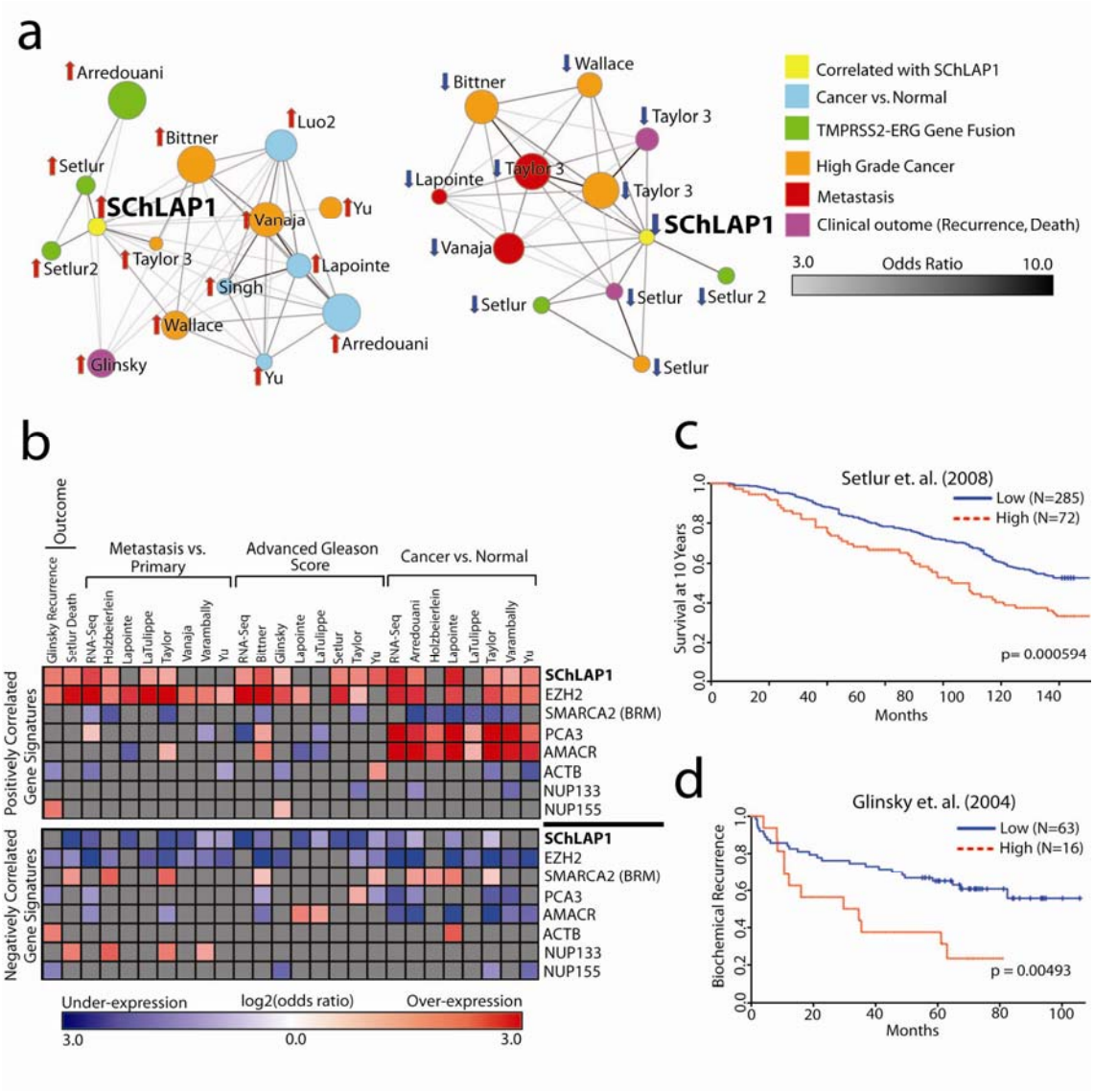
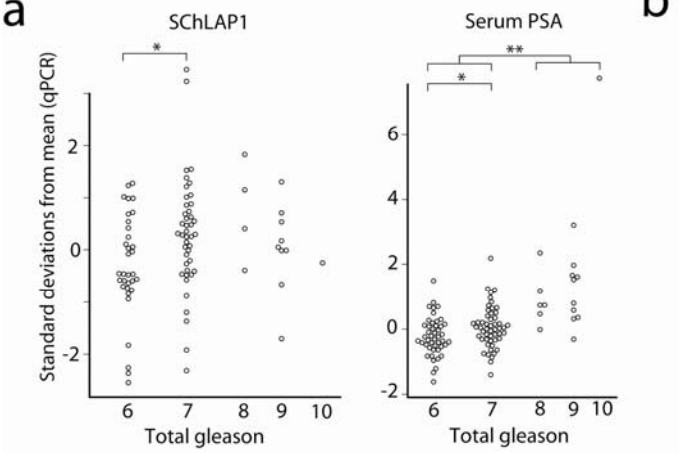


Figure 5.19: *SChLAP-1* expression characterizes aggressive prostate cancers. (a) Network analysis of a *SChLAP-1* gene signature using publicly-available data from Oncomine. The *SChLAP-1* gene signature was generated using genes positively or negatively correlated with *SChLAP-1* in our RNA-seq dataset (FDR<0.05). **Left**, a network analysis of genes positively correlated with *SChLAP-1*. **Right**, a network analysis of genes negatively correlated with *SChLAP-1*. (**b**) Meta-analysis of associated gene signatures for other genes using Oncomine gene datasets. (**c**) Stratification of overall survival in prostate cancer using the SChLAP-1 gene signature. (**d**) Stratification of biochemical recurrence using the SChLAP-1 gene signature.

a



b

	Biopsy Positive Patients
Number of Patients	114
Age (yr ± St. dev.)*	64.55 ± 8.44
PSA^**	
Mean (ng/ml ± St. dev.)	8.7 ± 11.11
0 - 2.5 ng/ml (%)	6 (5.7%)
2.6 - 4 ng/ml (%)	19 (17.9%)
4 - 10 ng/ml (%)	56 (52.8%)
>10 ng/ml (%)	25 (23.6)
Gleason grade***	
6	45 (40.6%)
7	48 (43.2%)
8	5 (4.5%)
9	12 (10.8%)
10	1 (0.9%)
TMPRSS2-ERG Status****	
Positive	45 (39.5%)
Negative	69 (60.5%)
Ethnicity******	
White (%)	100 (87.7%)
Asian or Pacific Islander (%)	1 (0.9%)
Black (%)	9 (7.9%)
Unknown/Other (%)	4 (3.5%)

* Data were available for 221 patients
** Data were available for 222 patients
*** Data were available for 111 patients
**** Data were available for 231 patients

^One outlier patient with biopsy-positive prostate cancer was excluded from the analysis due to a PSA level

Figure 5.20: *SChLAP-1* expression urine prostate cancer urine sediments. (a) *SChLAP-1* expression stratified by Gleason score in a cohort of urine sediments from prostate cancer patients. Serum PSA is shown for comparison. An asterisk (*) indicates $p < 0.05$. Two asterisks (**) indicates $p < 0.01$. (**b**) A table detailing the clinical characteristics of the urine cohort patients.

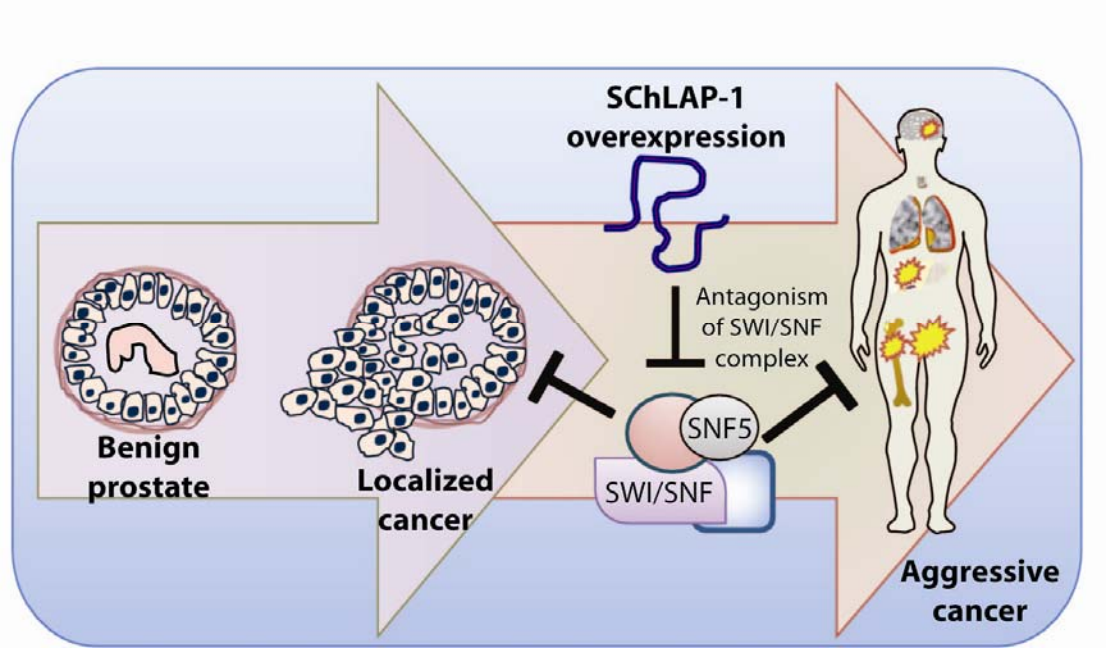


Figure 5.21: A model of *SChLAP-1* function in prostate cancer.

REFERENCES

1. J. R. Prensner, A. M. Chinnaiyan, *Cancer Discov* **1**, 391 (Oct, 2011).
2. M. Guttman *et al.*, *Nature*, (Aug 28, 2011).
3. J. T. Lee, *Genes Dev* **23**, 1831 (Aug 15, 2009).
4. Y. Kotake *et al.*, *Oncogene* **30**, 1956 (Apr 21, 2011).
5. K. L. Yap *et al.*, *Mol Cell* **38**, 662 (Jun 11, 2010).
6. R. A. Gupta *et al.*, *Nature* **464**, 1071 (Apr 15, 2010).
7. M. C. Tsai *et al.*, *Science* **329**, 689 (Aug 6, 2010).
8. J. Zhao, B. K. Sun, J. A. Erwin, J. J. Song, J. T. Lee, *Science* **322**, 750 (Oct 31, 2008).
9. J. L. Rinn *et al.*, *Cell* **129**, 1311 (Jun 29, 2007).
10. R. Kogo *et al.*, *Cancer Res* **71**, 6320 (Oct 15, 2011).
11. Z. Yang *et al.*, *Ann Surg Oncol* **18**, 1243 (May, 2011).
12. T. Niinuma *et al.*, *Cancer Res* **72**, 1126 (Mar 1, 2012).
13. J. R. Prensner *et al.*, *Nat Biotechnol* **29**, 742 (Aug, 2011).
14. M. N. Cabili *et al.*, *Genes Dev* **25**, 1915 (Sep 15, 2011).
15. S. A. Tomlins *et al.*, *Science* **310**, 644 (Oct 28, 2005).
16. J. Yu *et al.*, *Cancer Cell* **17**, 443 (May 18, 2010).
17. M. Guttman *et al.*, *Nature* **458**, 223 (Mar 12, 2009).
18. C. G. Kleer *et al.*, *Proc Natl Acad Sci U S A* **100**, 11606 (Sep 30, 2003).
19. S. Varambally *et al.*, *Nature* **419**, 624 (Oct 10, 2002).
20. A. Subramanian *et al.*, *Proc Natl Acad Sci U S A* **102**, 15545 (Oct 25, 2005).
21. H. Shen *et al.*, *Cancer Res* **68**, 10154 (Dec 15, 2008).
22. C. W. Roberts, S. H. Orkin, *Nat Rev Cancer* **4**, 133 (Feb, 2004).
23. K. C. Wiegand *et al.*, *N Engl J Med* **363**, 1532 (Oct 14, 2010).
24. S. Jones *et al.*, *Science* **330**, 228 (Oct 8, 2010).
25. I. Varela *et al.*, *Nature* **469**, 539 (Jan 27, 2011).
26. I. Versteege *et al.*, *Nature* **394**, 203 (Jul 9, 1998).
27. D. Reisman, S. Glaros, E. A. Thompson, *Oncogene* **28**, 1653 (Apr 9, 2009).
28. A. Sun *et al.*, *Prostate* **67**, 203 (Feb 1, 2007).
29. M. L. Dechassa *et al.*, *Mol Cell Biol* **28**, 6010 (Oct, 2008).
30. D. R. Rhodes *et al.*, *Neoplasia* **9**, 166 (Feb, 2007).
31. J. B. de Kok *et al.*, *Cancer Res* **62**, 2695 (May 1, 2002).
32. S. A. Tomlins *et al.*, *Sci Transl Med* **3**, 94ra72 (Aug 3, 2011).
33. M. A. Rubin *et al.*, *JAMA* **287**, 1662 (Apr 3, 2002).
34. G. V. Glinsky, A. B. Glinskii, A. J. Stephenson, R. M. Hoffman, W. L. Gerald, *J Clin Invest* **113**, 913 (Mar, 2004).
35. S. R. Setlur *et al.*, *J Natl Cancer Inst* **100**, 815 (Jun 4, 2008).
36. M. Han *et al.*, *J Urol* **169**, 517 (Feb, 2003).
37. A. J. Stephenson *et al.*, *J Clin Oncol* **27**, 4300 (Sep 10, 2009).
38. B. G. Wilson *et al.*, *Cancer Cell* **18**, 316 (Oct 19, 2010).
39. A. H. Shain *et al.*, *Proc Natl Acad Sci U S A* **109**, E252 (Jan 31, 2012).

CHAPTER 6

Conclusion: Future directions for long non-coding RNA research in cancer

SUMMARY OF THESIS WORK

This thesis sought to understand the role of lncRNAs in cancer, using prostate cancer as a primary model disease. While the findings presented herein reflect aspects of prostate cancer biology, I speculate that these principles may be generalized to other disease types, including non-cancerous disease states. In this regard, this work reflects merely the potential of lncRNA research, as the discovery and characterization of lncRNAs in other biological systems remains an active area of research. Indeed, this thesis work underscores two fundamental dictums: 1) the elucidation of all lncRNA species in all normal and diseased tissue/cell types is of critical importance; and 2) the scope of lncRNA biology in other tissues and diseases, including cancers, will likely uncover mechanisms of tissue-specific processes.

For this work, we have focused on prostate cancer as a primary model system. We began with the supposition that the annotated protein-coding genome represents only a partial list of the total number of transcribed regions, and we were motivated by reports of functionally-relevant lncRNAs (*1*-*3*), which resemble mRNAs structurally but do not

encode for a protein (*4*). As such, we hypothesized that lncRNAs were an uncharacterized (and largely undiscovered) set of genes. To pursue this hypothesis in the context of cancer, we employed transcriptome sequencing of prostate cancer tissues. We used cell lines as supportive research tools, recognizing that cell lines represent imperfect model systems that may differ substantially from human cancer (*5*).

We used *ab initio* transcriptome assembly as a computational method to predict novel transcripts bioinformatically. We found >6,000 total unannotated transcripts expressed in prostate cancer, of which approximately 1,800 were intergenic. These transcripts resembled protein-coding mRNAs structurally by exhibiting polyadenylation and promoter enrichment for H3K4me3 and RNA polymerase II; however, they lacked robust open reading frames in >98% of cases defining a large set on novel lncRNAs in prostate cancer. Among these, we refined a list of 121 lncRNAs that exhibited dysregulated expression levels in cancer, often in a tissue-specific manner exclusively expressed in prostate cancer. These 121 genes were termed PCATs.

PCAT-1 and SChLAP-1 as novel prostate cancer genes

From this list of candidate PCATs, we have validated many novel prostate-cancer associated genes and investigated two in depth: *PCAT-1* and *SChLAP-1*, which both exhibit high upregulation in subsets of prostate cancer (**Figure 6.1**). We defined *PCAT-1* as a two-exon gene, and *in vitro* and *in vivo* experiments demonstrated that this lncRNA promotes cell proliferation. *PCAT-1* regulates expression of a core set of ~250 genes that associate with DNA repair, DNA maintenance, and mitotic spindle formation. Of note, *PCAT-1* represses BRCA2, a tumor suppressor gene essential for homologous

recombination (HR) (*6*). We demonstrated *in vitro* and *in vivo* that *PCAT-1* expression leads to impaired repair of DNA double strand breaks by repressing HR following genotoxic stress. Importantly, these studies revealed that *PCAT-1* expression sensitizes cells to treatment with small molecule inhibitors of the PARP-1 DNA repair enzyme, whose inhibition in cells with BRCA1 or BRCA2 inactivation leads to gross collapse of genome integrity (*7-10*). This repression of BRCA2 was due to microRNA-like effects of *PCAT-1* in a post-transcriptional manner and required bps 1 – 750 of the *PCAT-1* gene.

By contrast, *SChLAP-1* exhibits multiple isoforms containing up to seven exons in total. *SChLAP-1* is expressed in ~20% of prostate cancers as a nuclear lncRNA and coordinates prostate cancer cell invasion and metastasis via a 250 bp fragment between bp 1001-1250. We found that *SChLAP-1* associates with an aggressive cancer phenotype and coordinates a gene expression program associated with metastatic disease. This gene expression program was antagonistic to the regulatory functions of the SWI/SNF nucleosome remodeling complex. SWI/SNF serves as a tumor suppressor complex in multiple cancer types (*11-17*), and we observe that *SChLAP-1* directly binds to SWI/SNF proteins but does not directly regulate their expression nor the expression of other major epigenetic complexes, suggesting that *SChLAP-1* modulates the activity of epigenetic complexes rather than controlling their expression.

EMERGING DIRECTIONS IN LONG NON-CODING RNA RESEARCH

This thesis work has defined novel aspects of prostate cancer biology and discovered genes not previously known. As such, the implications of this work are broad and suggest that numerous aspects of cancer biology remain unrevealed. While it is now

well-accepted that lncRNAs characterize all major cell lineages (*18*) and many disease states (*1, 19, 20*), the wide-ranging reports of lncRNAs in multiple cell types raise new questions and challenges for the field of lncRNA research and cancer biology. Just as the cancer research community has spent decades understanding the functional importance of a variety of proteins in cancer biology, now this same community must undertake a comprehensive investigation of lncRNAs. Here, I will discuss major areas of research that will be necessary to understand the role of lncRNAs in cancer.

Defining the lncRNA component of the human genome

Going forward, it is clear that the systematic identification and annotation of lncRNAs, and their expression patterns in human tissues and disease, is important to clarifying the molecular biology underlying cancer. These efforts will be facilitated by large-scale RNA-Seq studies followed by *ab initio* or *de novo* sequence data assembly to discover lncRNAs in an unbiased manner (*21, 22*).

However, it is increasingly appreciated that a number of annotated but uncharacterized transcripts are important lncRNAs—*HOTTIP* is one such example (*23*). Similarly, the STAU1-interacting lncRNAs described by Gong and colleagues were also found by screening for annotated transcripts that contained prominent Alu repeats (*24*). While these examples were annotated as non-coding genes, it is also possible that other annotated genes, enumerated in early studies as protein-coding but not studied experimentally, are mislabeled ncRNA genes. These may include the generic "open-reading frame" genes (such as LOCxxx or CxxORFxx genes) that have not received detailed study.

Supporting this, Dinger et al. recently argued that bioinformatically distinguishing between protein-coding and non-coding genes can be difficult and that traditional computational methods for doing this may have been inadequate in many cases (*25*). For example, *XIST* was initially identified as a protein-coding gene because it has a potential, unused open reading frame (ORF) of nearly 300 amino acids (*26*). Additional complications further include an increasing appreciation of mRNA transcripts that function both by encoding a protein and at the RNA level, which would support miRNA sequestration hypotheses posited by Pandolfi and colleagues (*27*).

Elucidating the role of lncRNA sequence conservation

In general, most protein-coding exons are highly conserved and most lncRNAs are poorly conserved (*28-30*). This is not always true, as T-UCRs are prime examples of conserved ncRNAs (*31-34*). However, the large majority of lncRNAs exhibit substantial sequence divergence among species, and lncRNAs that do show strong conservation frequently only exhibit this conservation in a limited region of the transcript, and not the remainder of the gene.

This conundrum has sparked many hypotheses, many of which have merit. Small regions of conservation could indicate function domains of a given ncRNA, such as a binding site for proteins, microRNAs, mRNAs, or genomic DNA. Development of abundant ncRNA species could also suggest evolutionary advancement as species develop. In support of this latter proposition, many have commented that complex mammalian genomes (such as the human genome) have a vastly increased non-coding DNA component of their genome compared to single-celled organisms and nematodes,

whereas the complement of protein-coding genes varies less throughout evolutionary time (*35*).

For lncRNAs, the issue of sequence conservation is paramount. However, it is now well established that poorly-conserved lncRNAs, including *PCAT-1* and *SChLAP1*, can be biologically important, but it is unclear whether these represent species-specific evolutionary traits or whether functional homologs have simply not been found. For example, *AIR* was initially described in mice in the 1980s, but a human homolog was not identified until 2008 (*36*).

Moreover, even lncRNAs with relatively high conservation, such as *HOTAIR*, may have species-specific function. Indeed, a study of murine *HOTAIR* (*mHOTAIR*) showed that *mHOTAIR* did not regulate the HoxD locus and did not recapitulate the functions observed in human cells (*37*). Other ncRNAs observed in mice, such as *linc-p21*, also show only limited sequence homology to their human forms and may have divergent functions as well (*2*). This may support hypotheses of rapid evolution of lncRNAs during the course of mammalian development. Moreover, this may suggest either that lncRNAs may have functions independent of conserved protein complexes (which have comparatively static functions throughout evolution) or that lncRNAs may adapt to cooperate with different protein complexes in different species.

Determining somatic alterations of lncRNAs in cancer

To date, somatic mutation of lncRNAs in cancer is not well explored. While numerous lncRNAs display altered expression levels in cancer, it is unclear to what

extent cancers specifically target lncRNAs for genomic amplification/deletion, somatic point mutations, or other targeted aberrations.

In several examples, data suggest that lncRNAs may be a target for somatic aberrations in cancer. For example, approximately half of prostate cancers harbor gene fusions of the ETS family transcription factors (*ERG*, *ETV1*, *ETV4*, *ETV5*), which generally result in the translocation of an androgen-regulated promoter to drive upregulation of the ETS gene (*38*). One patient was initially found to have an *ETV1* translocation to an intergenic androgen-regulated region (*39*) which was subsequently found to encode a prostate-specific lncRNA (*PCAT-14*) (*22*), thereby creating a gene fusion between the lncRNA and *ETV1*. Similarly, a *GAS5-BCL6* gene fusion, resulting from a chromosomal translocation and retaining the full coding sequence of *BCL6*, has been reported in a patient with B-cell lymphoma (*40*). Finally, Poliseno and colleagues demonstrated that the *PTEN* pseudogene, *PTENP1*, is genomically deleted in prostate and colon cancers, leading to aberrant expression levels of these genes (*27*).

These initial data suggest that somatic aberrations of lncRNAs do contribute to their dysregulated function in cancer, although most studies to date identify gene expression changes as the primary alteration in lncRNA function. Yet, the study of mutated lncRNAs in cancer will be an area of high importance in future investigations, as several prominent oncogenes, such as *KRAS*, show no substantial change in protein expression level in mutated compared to non-mutated cases. It is likely that some lncRNAs exhibit altered function in cancer due to structural aberrations (mutation, small insertions/deletions) without exhibiting overt changes in expression level.

Characterizing RNA structural motifs

Just as protein-coding genes harbor specific domains of amino acids that mediate distinct functions (e.g. a kinase domain), RNA molecules also have intricate and specific structures. Among the most well-known RNA structures is the stem-loop-stem design of a hairpin, which is integral for miRNA generation (*41*). RNA structures are also known to be essential for binding to proteins, particularly PRC2 proteins (*42*). However, global profiles of lncRNA structures are poorly understood. While it is clear that lncRNA structure is important to lncRNA function, few RNA domains are well-characterized. Moreover, it is likely that RNA domains occur at the level of secondary structure, as lncRNA sequences are highly diverse yet may form similar secondary structures following RNA folding (*43*).

To this end, both computational and experimental advancements are beginning to address these topics. While numerous computational algorithms have been proposed to predict RNA structures (*43*), perhaps the most dramatic advance in this area has been the development of RNA-Seq methods to interrogate aspects of RNA structure globally. Recently, Frag-Seq and PARS-Seq have demonstrated the unbiased evaluation of RNA structures by treating RNA samples with specific RNAses that cleave RNA at highly selective structural positions (*44, 45*). These RNA fragments are then processed and sequenced to determine the nucleotide sites where RNA transcripts were cleaved, indirectly implying a secondary structure. This area of research promises to yield tremendous insight into the overall mechanics of lncRNA function.

IMPLICATIONS OF LONG NON-CODING RNAs IN CANCER MANAGEMENT

lncRNA diagnostic biomarkers

For clinical medicine, lncRNAs offer several possible benefits. lncRNAs, such as *PCAT-1* and *SChLAP1*, commonly demonstrate restricted tissue-specific and cancer-specific expression patterns (*22*). This tissue-specific expression distinguishes lncRNAs from miRNAs and protein-coding mRNAs, which are frequently expressed from multiple tissue types. While the underlying mechanism for this is unclear, recent studies of chromatin confirmation show tissue-specific patterns, which may impact ncRNA transcription (*23, 46*). Given this specificity, ncRNAs may be superior biomarkers than many current protein-coding biomarkers, both for tissue-of-origin tests as well as cancer diagnostics.

A prominent example is *PCA3*, a lncRNA that is a prostate-specific gene and markedly overexpressed in prostate cancer. Although the biological function of *PCA3* is unclear, its utility as a biomarker has led to the development of a clinical *PCA3* diagnostic assay for prostate cancer, and this test is already being employed for clinical uses (*47, 48*). In this test, *PCA3* transcript is detected in prostate cancer patient urine samples, which contain prostate cancer cells shed into the urethra. Thus, monitoring *PCA3* does not require invasive procedures (**Figure 6.2A**) (*47*). The *PCA3* test represents the most effective clinical translational of a cancer-associated ncRNA gene to date, and the rapid timeline these developments—only 10 years from between its initial description and a clinical test—suggests that the use of ncRNAs in clinical medicine is only beginning. Non-invasive detection of other aberrantly expressed lncRNAs, such as upregulation of *HULC*, which occurs in hepatocellular carcinomas, has also been

observed in patient blood sera (*49*); however other lncRNA-based diagnostics have not been developed for widespread use.

lncRNA-based therapies

The transition from ncRNA-based diagnostics to ncRNA-based therapies is also showing initial signs of development. Although the implementation of therapies targeting ncRNAs is still remote for clinical oncology, experimental therapeutics employing RNA interference (RNAi) to target mRNAs have been tested in mice, cynomolgus monkeys, and humans (*50*), as part of a phase I clinical trial for patients with advanced cancer (**Figure 6.2B**). Davis and colleagues found that systemic administration of RNAi-based therapy was able to effectively localize to human tumors and reduce expression of its target gene mRNA and protein (*50*). Currently, ongoing clinical trials are further evaluating the safety and efficacy of RNAi-based therapeutics in patients with a variety of diseases, including cancer (*51*), and these approaches could be adapted to target lncRNA transcripts.

Other studies investigate an intriguing approach that employs modular assembly of small molecules to adapt to aberrant RNA secondary structure motifs in disease (*52*). This approach could potentially target aberrant ncRNAs, mutant mRNAs, as well as nucleotide triplet-repeat expansions seen in several neurological diseases (such as Huntington's disease). However, most RNA-based research remains in the early stages of development, and the potential for RNAi therapies targeting lncRNAs in cancer is still far from use in oncology clinics.

lncRNAs in genomic epidemiology

In the past decade, genome-wide association studies (GWAS) have become a mainstream way to identify germline SNPs that may predispose to myriad human diseases. In prostate cancer, over 20 GWAS have reported 31 SNPs with reproducible allele-frequency changes in prostate cancer patients compared to men without prostate cancer (*53*), and these 31 SNPs cluster into 14 genomic loci (*53*). In principle, profiling of these SNPs could represent an epidemiological tool to assess patient populations with a high risk of prostate cancer.

Of the 14 genomic loci, the most prominent by far is the "gene desert" region upstream of the *cMYC* oncogene on chromosome 8q24, which harbors 10 of the 31 reproducible SNPs associated with prostate cancer (**Figure 6.2C**). Several SNPs in the 8q24 region have been studied for their effect on enhancers (*54*), particularly for enhancers of *cMYC* (*55*), and chromosome looping studies have shown that many regions within 8q24 may physically interact with the genomic position of the *cMYC* gene (*56*).

Recently, our identification of *PCAT-1* as a novel chr8q24 gene implicated in prostate cancer pathogenesis further highlights the importance and complexity of this region (**Figure 6.2C**) (*22*). Although the relationship between *PCAT-1* and the 8q24 SNPs is not clear at this time, this discovery suggests that previously-termed "gene deserts" may, in fact, harbor critical lncRNA genes, and that SNPs found in these regions may impact uncovered aspects of biology. Relatedly, GWAS analyses of atherosclerosis, coronary artery disease, and type 2 diabetes have all highlighted *ANRIL* on Chr9p21 as a ncRNA gene harboring of disease-associated SNPs (*57*).

Clinically, the use of GWAS data may identify patient populations at risk for cancer and may stratify patient disease phenotypes, such as aggressive versus indolent cancer, and patient outcomes (*58*). SNP profiles may also be use to predict a patient's response to a given therapy (*59*). As such, the clinical translation of GWAS data remains an area of interest for cancer epidemiology.

ARE LNCRNAs REALLY NON-CODING?

A major assumption of this work is that computational biology can predict both the sequence identity of all transcripts in a cell and their protein coding potential. While number studies have now addressed the first question (*29, 30, 60, 61*), few studies have rigorous tested the second: are predicted lncRNAs really non-coding transcripts? On an individual level, single genes can be tested by *in vitro* translation assays, and these assays indicate that *PCAT-1* and *SChLAP-1*, for example, are non-coding. However, these assays are not suitable for large-scale analysis of thousands of genes, and thus the evidence that all predicted lncRNAs are truly non-coding remains anecdotal.

Recently, a novel method has been developed which may help to address this problem. Ingolia et al. precipitated ribosomal fractions of cell lysates and, following partial digestion with RNAse, isolated RNA protected by the bulky ribosome machinery, which prevents access of the RNAse to the RNA engaged in the ribosomes (*62*). Using a modified form of RNA-seq, the authors then comprehensively determined which RNA species were directly engaged in the translational machinery, presumably to make proteins. While this method assumes that all RNAs bound within the ribosome are

generating proteins, it provides the first global manner to identify putatively protein-coding RNAs.

This method has been applied to multiple model systems, including yeast (*62*), *E. coli* (*63*) and mouse cells (*64*), and the findings from these studies are striking. These studies suggest that the proteome is vastly more complex than previously appreciated. For example, numerous protein-coding genes are observed to give rise to many alternative proteins in addition to known isoforms, including novel peptides arising from the canonical 5' and 3' untranslated regions (*64*). While it is unclear whether this new abundance of peptides represents functionally-relevant proteins or the imprecision of an imperfect translational machinery, their presence is intriguing and suggests that this layer of biology deserves exploration.

With respect to lncRNAs, Ingolia et al. found that nearly half of reported lncRNAs in mouse embryonic stem cells demonstrated translational efficiency consistent with protein-coding genes (*4, 64*), suggesting that these genes may not be non-coding at all. Although these lncRNA-derived proteins have not been functionally characterized, it is likely that some of the presumed lncRNAs discovered as part of this thesis work are also translated and impart biological functions as a protein.

Why, then, did we not predict these proteins previously? The answer is that these novel proteins are almost uniformly <100 amino acids in length, which is considered as a "noise" range below the level of accurate prediction computationally (*65-68*). The DNA sequence of nearly any lncRNA can produce a predicted ORF of 10-100 amino acids, and conventionally these predicted ORFs have been ignored, since an analysis of random non-transcribed regions of DNA of similar lengths will also yield such predicted ORFs

(*67, 68*). However, the field of small proteins has recently been explored in model systems such as *Drosophila* and bacteria (*H. pylori*) (*69, 70*); these studies show functional roles for small peptides produced by presumed ncRNAs and suggest that numerous cell processes, such as cell differentiation (*71*) may be regulated by this novel mechanism. Thus, future efforts are needed to determine whether the newly-described lncRNA species in human biology and cancer also encode functional small proteins.

CONCLUSION

In the past decade, the rapid discovery of ncRNA species by high-throughput technologies has accelerated current conceptions of transcriptome complexity. While a biological understanding of these ncRNAs has proceeded more slowly, increasing recognition of lncRNAs has defined them as critical actors of numerous cellular processes. In cancer, dysregulated lncRNA expression characterizes the entire spectrum of disease and aberrant lncRNA function drives cancer through disruption of normal cell processes, typically by facilitating epigenetic repression of downstream target genes. lncRNAs thus represent a novel, poorly-characterized layer of cancer biology. This thesis work, in particular, has examined the presence and role of lncRNAs in prostate cancer, defining 121 prostate-cancer-associated lncRNAs (PCATs) by global transcriptome profiling, and characterizing two novel lncRNAs, *PCAT-1* and *SChLAP-1*, with oncogenic properties. In the near term, clinical translation of lncRNAs such as *PCAT-1* and *SChLAP-1* may assist biomarker development in cancer types without robust and specific biomarkers, and in the future RNA-based therapies may be a viable option for clinical oncology. Additional studies in the future will likely focus on the mechanism

of lncRNA biology, either via molecular mechanisms (epigenetics, etc), RNA structural motifs, sequence conservation, or the potential for encoding small proteins, as well as the possibility that somatic alteration of lncRNAs drives subsets of human cancer.

FIGURES

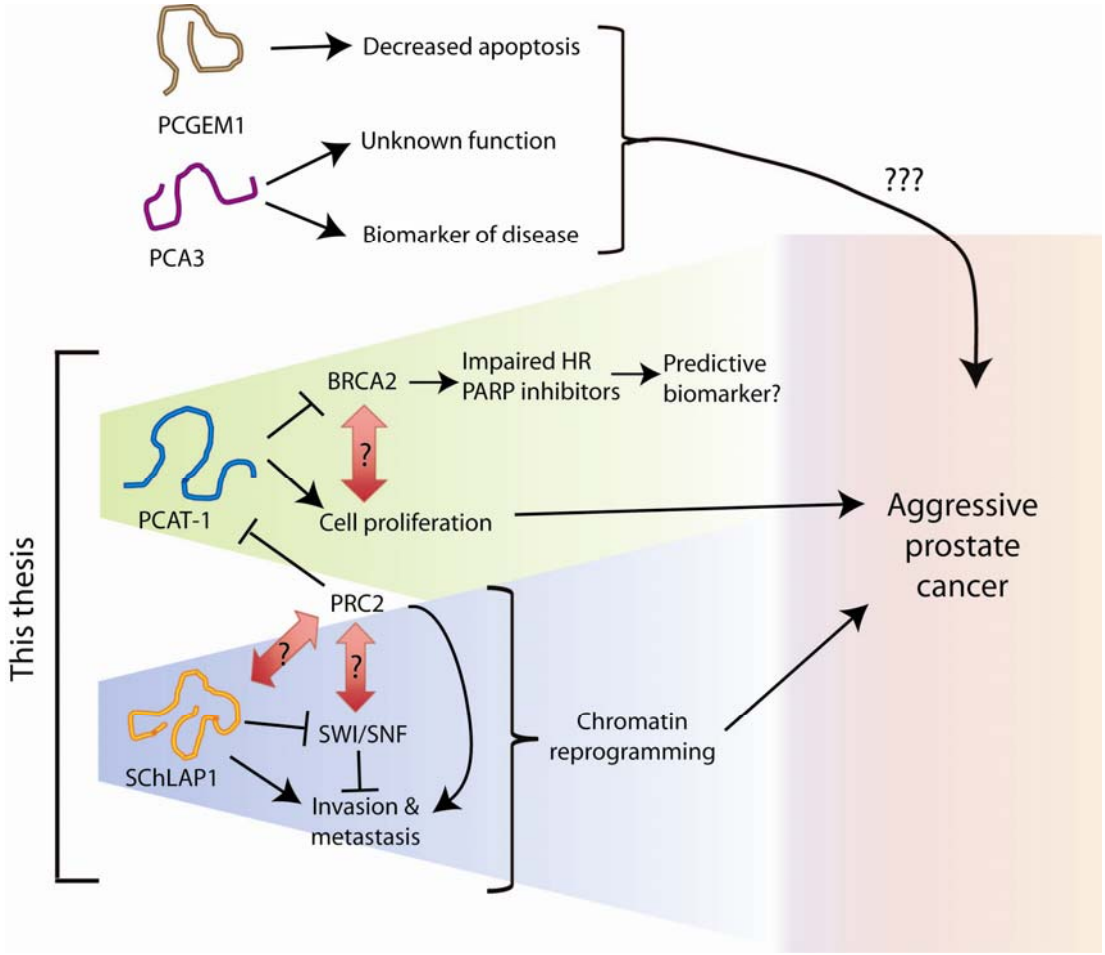


Figure 6.1: The role of lncRNAs in prostate cancer. Four lncRNAs (*PCA3*, *PCGEM1*, *PCAT-1*, and *SChLAP-1*) have been investigated in prostate cancer. Of these, *PCA3* is a clinically-useful biomarker of unknown function. *PCGEM1* decreases apoptosis but its mechanism is unclear. This thesis describes the discovery and characterization of *PCAT-1* and *SChLAP-1*, which contribute to aggressive disease through different mechanisms. *PCAT-1* may also serve as a predictive biomarker through its regulation of homologous recombination. The red arrows with question marks (?) indicate areas of ongoing investigation.

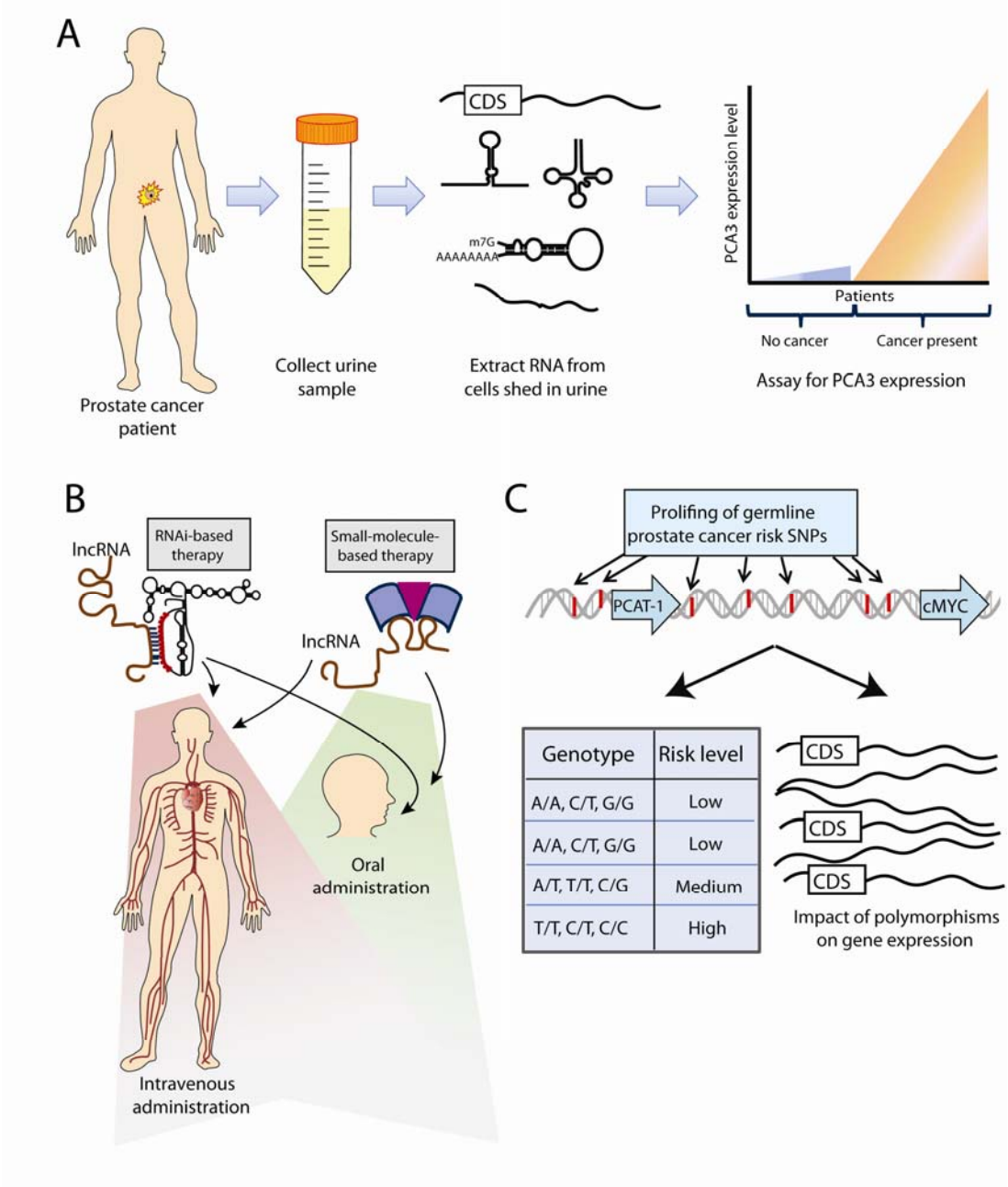


Figure 6.2: Clinical implications of lncRNAs. (**A**) The *PCA3* urine biomarker test for prostate cancer employs a non-invasive approach to disease diagnosis by collecting patient urine samples, isolating nucleic acids from cells in the urine sediment, and quantifying *PCA3* expression. (**B**) lncRNA-based therapies may target the lncRNA by utilizing either RNA interference (RNAi), which uses sequence homology between the lncRNA and the RNAi therapeutic molecule, or a small molecule therapy that interacts with the lncRNA. These therapeutic avenues may be appropriate for systemic therapy by either intravenous or oral administration. (**C**) Genome-wide association studies (GWAS) may provide germline polymorphisms that predict an individual patient's clinical risk for

disease development, response to therapy, or disease aggressiveness, while also providing molecular information through the impact of polymorphisms on gene expression of key genes. Abbreviation: CDS, coding sequence.

REFERENCES

1. R. A. Gupta *et al.*, *Nature* **464**, 1071 (Apr 15, 2010).
2. M. Huarte *et al.*, *Cell* **142**, 409 (Aug 6, 2010).
3. J. L. Rinn *et al.*, *Cell* **129**, 1311 (Jun 29, 2007).
4. M. Guttman *et al.*, *Nature* **458**, 223 (Mar 12, 2009).
5. J. P. Gillet *et al.*, *Proc Natl Acad Sci U S A* **108**, 18708 (Nov 15, 2011).
6. S. P. Jackson, J. Bartek, *Nature* **461**, 1071 (Oct 22, 2009).
7. H. E. Bryant *et al.*, *Nature* **434**, 913 (Apr 14, 2005).
8. C. K. Donawho *et al.*, *Clin Cancer Res* **13**, 2728 (May 1, 2007).
9. H. Farmer *et al.*, *Nature* **434**, 917 (Apr 14, 2005).
10. P. C. Fong *et al.*, *N Engl J Med* **361**, 123 (Jul 9, 2009).
11. M. L. Dechassa *et al.*, *Mol Cell Biol* **28**, 6010 (Oct, 2008).
12. D. Reisman, S. Glaros, E. A. Thompson, *Oncogene* **28**, 1653 (Apr 9, 2009).
13. C. W. Roberts, S. H. Orkin, *Nat Rev Cancer* **4**, 133 (Feb, 2004).
14. H. Shen *et al.*, *Cancer Res* **68**, 10154 (Dec 15, 2008).
15. A. Sun *et al.*, *Prostate* **67**, 203 (Feb 1, 2007).
16. I. Varela *et al.*, *Nature* **469**, 539 (Jan 27, 2011).
17. I. Versteege *et al.*, *Nature* **394**, 203 (Jul 9, 1998).
18. M. N. Cabili *et al.*, *Genes Dev* **25**, 1915 (Sep 15, 2011).
19. M. A. Faghihi *et al.*, *Nat Med* **14**, 723 (Jul, 2008).
20. D. S. Perez *et al.*, *Hum Mol Genet* **17**, 642 (Mar 1, 2008).
21. M. G. Grabherr *et al.*, *Nat Biotechnol* **29**, 644 (2011).
22. J. R. Prensner *et al.*, *Nat Biotechnol*, (Jul 31, 2011).
23. K. C. Wang *et al.*, *Nature* **472**, 120 (Apr 7, 2011).
24. C. Gong, L. E. Maquat, *Nature* **470**, 284 (Feb 10, 2011).
25. M. E. Dinger, K. C. Pang, T. R. Mercer, J. S. Mattick, *PLoS Comput Biol* **4**, e1000176 (Nov, 2008).
26. G. Borsani *et al.*, *Nature* **351**, 325 (May 23, 1991).
27. L. Poliseno *et al.*, *Nature* **465**, 1033 (Jun 24, 2010).
28. M. Guttman *et al.*, *Nat Biotechnol* **28**, 503 (May, 2010).
29. J. R. Prensner *et al.*, *Nat Biotechnol* **29**, 742 (Aug, 2011).
30. C. Trapnell *et al.*, *Nat Biotechnol* **28**, 511 (May, 2010).
31. G. Bejerano *et al.*, *Science* **304**, 1321 (May 28, 2004).
32. C. Braconi *et al.*, *Proc Natl Acad Sci U S A* **108**, 786 (Jan 11, 2011).
33. G. A. Calin *et al.*, *Cancer Cell* **12**, 215 (Sep, 2007).
34. P. Mestdagh *et al.*, *Oncogene* **29**, 3583 (Jun 17, 2010).
35. R. J. Taft, M. Pheasant, J. S. Mattick, *Bioessays* **29**, 288 (Mar, 2007).
36. I. Y. Yotova *et al.*, *Genomics* **92**, 464 (Dec, 2008).
37. P. Schorderet, D. Duboule, *PLoS Genet* **7**, e1002071 (May, 2011).
38. J. R. Prensner, A. M. Chinnaiyan, *Curr Opin Genet Dev* **19**, 82 (Feb, 2009).
39. S. A. Tomlins *et al.*, *Nature* **448**, 595 (Aug 2, 2007).
40. Y. Nakamura *et al.*, *Cancer Genet Cytogenet* **182**, 144 (Apr 15, 2008).
41. R. Garzon, G. A. Calin, C. M. Croce, *Annu Rev Med* **60**, 167 (2009).
42. J. Zhao, B. K. Sun, J. A. Erwin, J. J. Song, J. T. Lee, *Science* **322**, 750 (Oct 31, 2008).

43. Y. Wan, M. Kertesz, R. C. Spitale, E. Segal, H. Y. Chang, *Nat Rev Genet* **12**, 641 (2011).
44. J. G. Underwood *et al.*, *Nat Methods* **7**, 995 (Dec, 2010).
45. M. Kertesz *et al.*, *Nature* **467**, 103 (Sep 2, 2010).
46. D. Wang *et al.*, *Nature* **474**, 390 (Jun 16, 2011).
47. G. L. Lee, A. Dobi, S. Srivastava, *Nat Rev Urol* **8**, 123 (Mar, 2011).
48. S. A. Tomlins *et al.*, *Sci Transl Med* **3**, 94ra72 (Aug 3, 2011).
49. E. A. Gibb, C. J. Brown, W. L. Lam, *Mol Cancer* **10**, 38 (2011).
50. M. E. Davis *et al.*, *Nature* **464**, 1067 (Apr 15, 2010).
51. D. Castanotto, J. J. Rossi, *Nature* **457**, 426 (Jan 22, 2009).
52. M. M. Lee *et al.*, *J Am Chem Soc* **131**, 17464 (Dec 2, 2009).
53. H. Liu, B. Wang, C. Han, *Prostate* **71**, 209 (Feb 1, 2011).
54. L. Jia *et al.*, *PLoS Genet* **5**, e1000597 (Aug, 2009).
55. J. Sotelo *et al.*, *Proc Natl Acad Sci U S A* **107**, 3001 (Feb 16, 2010).
56. N. Ahmadiyeh *et al.*, *Proc Natl Acad Sci U S A* **107**, 9742 (May 25, 2010).
57. E. Pasmant, A. Sabbagh, M. Vidaud, I. Bieche, *FASEB J* **25**, 444 (Feb, 2011).
58. C. C. Chung, S. J. Chanock, *Hum Genet* **130**, 59 (Jul, 2011).
59. K. M. Giacomini *et al.*, *Clin Pharmacol Ther* **81**, 328 (Mar, 2007).
60. M. Guttman *et al.*, *Nature*, (Aug 28, 2011).
61. U. A. Orom *et al.*, *Cell* **143**, 46 (Oct 1, 2010).
62. N. T. Ingolia, S. Ghaemmaghami, J. R. Newman, J. S. Weissman, *Science* **324**, 218 (Apr 10, 2009).
63. E. Oh *et al.*, *Cell* **147**, 1295 (Dec 9, 2011).
64. N. T. Ingolia, L. F. Lareau, J. S. Weissman, *Cell* **147**, 789 (Nov 11, 2011).
65. E. S. Lander *et al.*, *Nature* **409**, 860 (Feb 15, 2001).
66. *Nature* **431**, 931 (Oct 21, 2004).
67. C. B. Burge, S. Karlin, *Curr Opin Struct Biol* **8**, 346 (Jun, 1998).
68. M. Q. Zhang, *Nat Rev Genet* **3**, 698 (Sep, 2002).
69. M. I. Galindo, J. I. Pueyo, S. Fouix, S. A. Bishop, J. P. Couso, *PLoS Biol* **5**, e106 (May, 2007).
70. C. M. Sharma *et al.*, *Nature* **464**, 250 (Mar 11, 2010).
71. T. Kondo *et al.*, *Science* **329**, 336 (Jul 16, 2010).

-

APPENDIX 1

Author contributions

CHAPTER 1

Portions of this chapter were previously published in the following manuscripts:

Prensner JR & Chinnaiyan AM. The emergence of lncRNAs in cancer biology. *Cancer Discovery*. 2011 Oct; 1 (5): 391-407

Prensner JR, Rubin MA, Wei JT, Chinnaiyan AM. Beyond PSA: The next generation of prostate cancer biomarkers. *Science Translational Medicine*. 4, 127rv3 (2012).

Prensner JR & Chinnaiyan AM. Oncogenic gene fusions in epithelial carcinomas. *Current Opinion in Genetics and Development*. 2009 Feb; 19(1): 82-91

All figures and tables in this chapter were made by J.R.P.

CHAPTERS 2 & 3

These chapters were published in *Nature Biotechnology*, August 2011, along with the following co-authors: Matthew K. Iyer, O. Alejandro Balbin, Saravana M. Dhanasekaran, Qi Cao, J. Chad Brenner, Bharathi Laxman, Irfan Asangani, Catherine Grasso, Hal D. Kominsky, Xuhong Cao, Xiaojun Jing, Xiaoju Wang, Javed Siddiqui,

John T. Wei, Daniel Robinson, Hari K. Iyer, Nallasivam Palanisamy, Christopher A. Maher, Arul M. Chinnaiyan. Supplementary information is available online at: http://www.nature.com/nbt/journal/v29/n8/full/nbt.1914.html

For this chapter, Matthew Iyer, John Prensner and Arul Chinnaiyan designed the project and directed experimental studies. Matthew Iyer, O. Alejandro Balbin, Catherine Grasso, and Christopher Maher developed computational platforms and performed sequencing data analysis. Matthew Iyer, O. Alejandro Balbin and Hari Iyer performed statistical analyses. John Prensner, Saravana M. Dhanasekaran, Qi Cao, J. Chad Brenner, Bharathi Laxman, Irfan Asangani, and Hal D. Kominsky performed *in vitro* experimental studies. Xiaojun Jing performed gene expression microarrays. Xiaoju Wang performed *in vitro* translation assays. Xuhong Cao, Daniel Robinson, and Nallasivam Palanisamy performed RNA sequencing. Javed Siddiqui and John T. Wei coordinated biospecimens. John Prensner, Matthew Iyer, and Arul Chinnaiyan wrote the manuscript.

All figures and tables in these chapters were generated by M.K.I. and J.R.P. Figures 2.1, 2.2, 2.4, 2.5, 2.6, 2.7, and 2.10 by M.K.I.; Figures 2.9, 2.11, 2.12, 3.1, 3.2, 3.3, 3.5, 3.6, 3.7, and 3.9 by J.R.P.; Figures 2.3, 2.8, 2.13, 3.4, 3.8, and 3.10 by M.K.I. and J.R.P.

CHAPTER 4

This work was performed in collaboration with Wei Chen, Matthew K. Iyer, Qi Cao, Kari Wilder-Romans, Theodore S. Lawrence, Arul M. Chinnaiyan and Felix Y.

Feng. John Prensner generated molecular constructs and isogenic cell lines. John Prensner, Qi Cao, and Wei Chen performed experimental studies. Wei Chen and Kari Wilder-Romans perform *in vivo* xenograft studies. Theodore S. Lawrence, Arul Chinnaiyan and Felix Feng supervised the experimental work. John Prensner, Wei Chen, Felix Feng, and Arul Chinnaiyan analyzed the data and wrote the manuscript currently in review.

All figures in this chapter were generated by W.C. and J.R.P. Figures 4.2, 4.4, 4.5, 4.8 and 4.9 by W.C.; Figures 4.11, 4.12, 4.13, 4.14 and 4.15 by J.R.P.; Figures 4.1, 4.3, 4.6, 4.7 and 4.10 by W.C. and J.R.P.

CHAPTER 5

This work was performed in collaboration with Rachel Bedenis, Qi Cao, Wei Chen, Arul M. Chinnaiyan, Saravana M. Dhanasekaran, Felix Feng, Matthew K. Iyer, Xiaojun Jing, Natalie McGregor, Lalit Patel, Ken Pienta, Anirban Sahu, and Xiaoju Wang. John Prensner, Qi Cao, Wei Chen, Saravana M. Dhanasekaran, Anirban Sahu and Xiaoju Wang performed *in vitro* experimental studies. Rachel Bedenis, Natalie McGregor and Lalit Patel performed *in vivo* studies. Xiaojun Jing performed gene expression microarrays. Matthew Iyer performed bioinformatics analysis. Felix Feng performed clinical analyses. Ken Pienta and Arul Chinnaiyan supervised the research.

John Prensner, Matthew Iyer, and Arul Chinnaiyan analyzed the data and wrote the manuscript currently in submission.

All figures and tables in these chapters were generated by M.K.I. and J.R.P. Figures 5.4, 5.14, 5.16 and 5.19 by M.K.I.; Figures 5.5, 5.6, 5.7, 5.8, 5.9, 5.10, 5.11, 5.12, 5.15, 5.17 and 5.18 by J.R.P.; Figures 5.1, 5.2, 5.3, 5.13, 5.20, and 5.21 by M.K.I. and J.R.P.

CHAPTER 6

Portions of this chapter were previously published in the following manuscripts:

Prensner JR & Chinnaiyan AM. The emergence of lncRNAs in cancer biology. *Cancer Discovery*. 2011 Oct; 1 (5): 391-407

Prensner JR, Rubin MA, Wei JT, Chinnaiyan AM. Beyond PSA: The next generation of prostate cancer biomarkers. *Science Translational Medicine*. 4, 127rv3 (2012).

Prensner JR & Chinnaiyan AM. Oncogenic gene fusions in epithelial carcinomas. *Current Opinion in Genetics and Development*. 2009 Feb; 19(1): 82-91

All figures in this chapter were made by J.R.P.

APPENDIX 2

List of oligonucleotide primers used

Chapter used in	Application	Gene	Forward Primer	Reverse Primer
Chapter 2, 3	3'RACE	CACNA1D_alus	TTCTCAGGTCTGGATGTAGGG	N/A
Chapter 2, 3	3'RACE	PCAT-1	ACCAGTGGAGAAGAGGCAGA	N/A
Chapter 2, 3	3'RACE	PCAT-14	TTCCACCTGAGGAGAAATGC	N/A
Chapter 2, 3	3'RACE	PCAT-14	GTACCACCCTGCAGGAAAA	N/A
Chapter 2, 3	3'RACE	Schlap1_LINE1	GCTCAGTCCCATCCTCTCAG	N/A
Chapter 2, 3	3'RACE	TU_0085675 region	ACTGCCACCTTGGAGAACAG	N/A
Chapter 2, 3	5'RACE	CACNA1D_alus	N/A	TCCTGCACATTTGGAGAACA
Chapter 2, 3	5'RACE	PCAT-1	N/A	GCTTCAAATGCGAAAAGACC
Chapter 2, 3	5'RACE	PCAT-14	N/A	CGTCACTATGGCCCATTCTT
Chapter 2, 3	5'RACE	PCAT-14	N/A	GGTATGGTGCCGTGCTATTT
Chapter 2, 3	5'RACE	PCAT-26	N/A	ACAGGGGCCCAGCTCCAAAGAGT
Chapter 2, 3	5'RACE	PCAT-6	N/A	TGGCCTGAAGAATCTAACAGCCCCCA
Chapter 2, 3	5'RACE	Schlap1_LINE1	N/A	CCTTGGTGGCTTAGGGTACA
Chapter 2, 3	5'RACE	TU_0085675 region	N/A	AGGAGGTATTCCAGCCCTGT
Chapter 2, 3	CHIP	PCAT1-promoter_1	TGACACCAACTGGTTCCCTATTA	TAAATGCCAGAAGACTCACCTTG
Chapter 2, 3	CHIP	PCAT1-promoter_2	CGAGACTCAGCCAGTGTTTACTC	AGTATGATTGAATGCTGGGACA
Chapter 2, 3	CHIP	PCAT1-promoter_3	TGCTAGTTAATTTCCAGCTCCTTC	CAAGCGAGACTCTGTCTCAAAA
Chapter 2, 3	CHIP	PCAT1-promoter_4	TGAATTCATACCTCGAGTAGAGCA	TCATGTAAGAACTACCAGTGTGTCC
Chapter 2, 3	CHIP	PCAT1-promoter_5	GGGACTGGACTTCACAGAACATA	ATCTTCAGCTTTGGGGTCTACTC
Chapter 2, 3	CHIP	PCAT1-promoter_6	GCACCTAGTCACTTCCCATATCA	CCTCTTGGCTTCAGAAAAGTACA
Chapter 2, 3	CHIP	PCAT1-promoter_7	CCCTTACAATTTGGTGAGACAAG	CCTCAGTTGGCTATGCATCTTAT
Chapter 2, 3	CHIP	PCAT1-promoter_8	ACACTTAACTGGCACTTGTGGAA	CACTGGTGTTCATGGCCTTATTA
Chapter 2, 3	PCR	CACNA1D_Alus_qPCR	ACTGTTTCTCATGCCCCATC	CCACTTTCTCCTCACCCTGA
Chapter 2, 3	PCR	CACNA1D_Alus_qPCR	TTCTCCAAATGTGCAGGATTC	AGGACAGGCAGGCATTAAGA
Chapter 2, 3	PCR	CACNA1D_full length Alus	TTCTCAGGTCTGGATGTAGGG	TCCTGCACATTTGGAGAACA
Chapter 2, 3	PCR	HMBS	TGGGGCCCTCGTGGAATGTTA	GATGGGCAACTGTACCTGACTGGA

Chapter 2, 3	PCR	HOTAIR	GGTAGAAAAAGCAACCACGAAGC	ACATAAACCTCTGTCTGTGAGTGCC
Chapter 2, 3	PCR	PCAT-1	ACCAGTGGAGAAGAGGCAGA	GCTTCAAATGCGAAAAGACC
Chapter 2, 3	PCR	PCAT-1	TGTGCCTCTAAGTGCCAGTG	GGCTGGTCACTATGCTCCTC
Chapter 2, 3	PCR	PCAT-1	TTGTGGAAGCCCCGCAAGGCCTGAA	TGTGGGGCCTGCACTGGCACTT
Chapter 2, 3	PCR	PCAT-1	TGGAGCCGGAGGACCACAGCATCA	TGTGGGGCCTGCACTGGCACTT
Chapter 2, 3	PCR	PCAT-1	AAAAGGTCAGGGTCACTGTTTG	GGATTCAAACTCAACACCAGAA
Chapter 2, 3	PCR	PCAT-1	ATCCACCGCTGCTCTCTGAAT	CCAGAATCTTATGCTTTCTTTGC
Chapter 2, 3	PCR	PCAT-14	GGCGCAGGCCACTCCATCTGGTG	CCTCAATGACCACACTGTAGAG
Chapter 2, 3	PCR	PCAT-14	TTCCACCTGAGGAGAAATGC	GCTGACCGTGGACGTACTTG
Chapter 2, 3	PCR	PCAT-14	GAGATGGAGACACCCCAATC	CGTCACTATGGCCCATTCTT
Chapter 2, 3	PCR	PCAT-14	GCACAAACCCCAAGAGAAAA	GGTATGGTGCCGTGCTATTT
Chapter 2, 3	PCR	PCAT-14_full length	GGCGCAGGCCACTCCATCTGGTG	ATGGAGAATGGCGATGACTT
Chapter 2, 3	PCR	PCAT-26	TGGGGCTTACGGTGCGGGAACG	ACCCTGATGGTAACAGGGGCCCAGC
Chapter 2, 3	PCR	PCAT-26	TGGGGCTTACGGTGCGGGAACGTCT	ACAGGGGCCCAGCTCCAAAGAGT
Chapter 2, 3	PCR	PCAT-26	AAGTGCTGGGGTTACAGGTG	CCCATCAATGGGTGTAGGAG
Chapter 2, 3	PCR	PCAT-29	CCGAGGAATGTGAGGTTCAT	CCTCACCCTTGGAAAATTCA
Chapter 2, 3	PCR	PCAT-29	GGAGCCAATCAACCTGTGCCCCTAGC	TCCCTGGGATGCATATGGTGCCCCT
Chapter 2, 3	PCR	PCAT-29	AACCTGTGCCCCTAGCTGATGTCCCA	GTCCCTGGGATGCATATGGTGCCCCT
Chapter 2, 3	PCR	PCAT-32	TGTCTGACTTCCGACTGACG	TCTCAGTGGGAATACCTGTGC
Chapter 2, 3	PCR	PCAT-32	ACTACCACTCTGGCCACAC	AACCATTTGGGTGCTACTGC
Chapter 2, 3	PCR	PCAT-32	TCCCTTACCAGTGAGCCAAC	GAATACAGCCGCAGACACAA
Chapter 2, 3	PCR	PCAT-43	GGTGGGTGTCTCCAAAGTGCCCACCA	TCCTGGCTCTGCCCTGTCTGTTGGCT
Chapter 2, 3	PCR	PCAT-43	GGAGGTGGGTGTCTCCAAAGTGCCC	AGGTTCCCTCCTGGCTCTGCCCTGT
Chapter 2, 3	PCR	PCAT-43	ATGCCCAGCCCTGCAGCTCAACT	GCAGGCTGACCCTGTGCAGCTCAA
Chapter 2, 3	PCR	PCAT-43	AGATGCCCAGCCCTGCAGCTCAA	TCTGCAGGCTGACCCTGTGCAGCTC
Chapter 2, 3	PCR	PCAT-6	ACTGGTACAGCCCACACTCAGAGCA	TGGCCTGAAGAATCTAACAGCCCCCA
Chapter 2, 3	PCR	PCAT-6	TGGTACAGCCCACACTCAGAGCAGAT	TGGCCTGAAGAATCTAACAGCCCCCA
Chapter 2, 3	PCR	PCAT-6	AAGGTCTTTTGGCAGCTGAA	CCTTGGGCTTTGCACATAAT
Chapter 2, 3	PCR	PCAT-6	CCCAATGTTTTCCCATCACT	CCTTGGGCTTTGCACATAAT
Chapter 2, 3	PCR	SChLAP1 Locus	TTCCCCTCATCTCATCAACC	TCAATTCTGGGCAGGATCTC
Chapter 2, 3	PCR	SChLAP1 Locus	CTCCAGTCATGACCCAGGAT	TGTGTTGACTTACAATTTTCTTATCCA
Chapter 2, 3	PCR	SChLAP1 Locus	TTTTCTCCCAGTTTCCATGC	CAACCAGCAGCCCTTAACTC
Chapter 2, 3	PCR	SChLAP1 Locus	ATCTTGCCTTCTGTGCCACT	GCGTTCCAAAGCTGAGAATC
Chapter 2, 3	PCR	SChLAP1 Locus	GCTTCTCCCTTTGCCTGTAA	TGCAAGGGACATCCACATTA
Chapter 2, 3	PCR	SChLAP1 Locus	GCTCCATCTTCCCCTAGGTC	CAGTCTCACCACCTCCCACT
Chapter 2, 3	PCR	SChLAP1 Locus	CGAGGCACTCAAGCCTTTTA	TCGACTTGTGATACTCTGAAGGA
Chapter 2, 3	PCR	SChLAP1 Locus	CCTAAAATGGCAAAGCGTGT	TAGCTGGCCTGCAGATTGTA
Chapter 2, 3	PCR	SChLAP1 Locus	TCCAGCAACACCTTCAAATG	TGGAGGAGTGAAGGAAACCA
Chapter 2, 3	PCR	SChLAP1 Locus	GTAAACTCCCTCGGTGTTCG	AGTTGCTGAAGGCAGGTGTT
Chapter 2, 3	PCR	SChLAP1 Locus	CTCCCATCTTTCTGCCTCAG	CTGAGAAGCCACCCAAACAT
Chapter 2, 3	PCR	SChLAP1 Locus	CCTTGACACGTGGGGATTAT	TGTCTTTTCCAAAAACCAACTTT
Chapter 2, 3	PCR	SChLAP1 Locus	GAAAAACCTAAGATGCCACCA	TTCAGATTGTTTGCTGTTGGTT

Chapter 2, 3	PCR	SChLAP1 Locus	TAATGCTGGTGAGGACATGG	TCTCCATGGTGGTTGCACTA
Chapter 2, 3	PCR	SChLAP1 Locus	TGCGAAGTCCAAGATCAAGA	TGGGGTCTTTTGGGAGATAA
Chapter 2, 3	PCR	SChLAP1 Locus	ATCTTCCCCCGCTAAACTGT	GGCACACACCATGTATTGGA
Chapter 2, 3	PCR	SChLAP1 Locus	GCAAGCAAAGCTTCAGGTTC	TCTGTGCTCTGGGTATGTCAA
Chapter 2, 3	PCR	SChLAP1 Locus	GAGTCAAGGAACGCTTCCAC	CCCCACCTTAGAGGCATTTT
Chapter 2, 3	PCR	SChLAP1 Locus	AGCAACTCTCCCAGTTCTGC	TGCTGCTACCCCTAGCACTT
Chapter 2, 3	PCR	SChLAP1 Locus	GATGACTTTTGGGTGGAGACA	CATTGGGAGGAATACGGAAA
Chapter 2, 3	PCR	Schlap1_LINE1	GCTCAGTCCCATCCTCTCAG	GACCCTGGGAGAAGTGTTCA
Chapter 2, 3	PCR	Schlap1_LINE1	AAGTGCCAGCATAAGCCAAC	CTGGGGACCAGGATAACAGA
Chapter 2, 3	PCR	Schlap1_LINE1	GCTCAGTCCCATCCTCTCAG	CCTTGGTGGCTTAGGGTACA
Chapter 2, 3	PCR	TMPRSS2	CTGGTGGCTGATAGGGGATA	GGACAAGGGGTTAGGGAGAG
Chapter 2, 3	PCR	TU_0017629 region	ACCCACGGACTCAGGAGCCCTTTCCT	TGCGATGGTCCTGCACCCTCAGTGCT
Chapter 2, 3	PCR	TU_0017629 region	AACAGGCTCCCAGCTCAGCCCT	TGCCGGGACTTGGGGCCAGGTTTT
Chapter 2, 3	PCR	TU_0017629 region	AGCTCAGCCCTGTTCCCCGACACA	TGCCGGGACTTGGGGCCAGGTTTT
Chapter 2, 3	PCR	TU_0017629 region	TGCCCTGAGTCCCGGTTGCATGGGA	AGCCTGGGGACAAGTGGCGACAGCA
Chapter 2, 3	PCR	TU_0017629 region	AGGGCAAGTGGAGAGCAGGTGGCT	AGGCCCGGAGTAACCCAGCCAGTGA
Chapter 2, 3	PCR	TU_0042715 region	TGCCTGTTGTAAACCTATTGAGAC	GCCCTTTGGACATTGTTCTC
Chapter 2, 3	PCR	TU_0042715 region	AAATGCCTGTTGTAAACCTATTGA	ATTCCAGCCCTTTGGACATT
Chapter 2, 3	PCR	TU_0042715 region	ATGGGCAGGGCCAGGCATCAGGAA	TCCCAGGGCCTTGTGCCTCTCTGT
Chapter 2, 3	PCR	TU_0042715 region	GGCCAGGCATCAGGAAGCTCAGCA	AGGGCCTTGTGCCTCTCTGTGCCCT
Chapter 2, 3	PCR	TU_0045472_0	CCCAGGATATGCAAGTCTTCTGCGCC	TGGCCTGAATCCAAAGTCACAGTGGG
Chapter 2, 3	PCR	TU_0045472_0	TCAGTAGCAGAGGCCTGTCA	CCAAAGTCACAGTGGGGTTC
Chapter 2, 3	PCR	TU_0085675 region	ACTGCCACCTTGGAGAACAG	GTTGAAGGGAGGTGGTGAAA
Chapter 2, 3	PCR	TU_0085675 region	TTATGGTTCGCAGCACTGAC	ACCAAACCAGAGCCCCTACT
Chapter 2, 3	PCR	TU_0085675 region	AGCACTGACTGGCTGCTGCTGCT	GGCTATGATCATCTTTGGCCCTGGCA
Chapter 2, 3	PCR	TU_0085675 region	AGTAGGGGCTCTGGTTTGGT	AGGAGGTATTCCAGCCCTGT
Chapter 2, 3	PCR	GSTP1	GACTTGCTGCTGATCCATGA	AGGTTCACGTACTCAGGGGAG
Chapter 2, 3	PCR	FKBP5	AGCCACTGTTGCTGAGCAGGGAGA	TCGGCGTTTCCTCACCATTCCCCACT
Chapter 2, 3	PCR	ERG	CGCAGAGTTATCGTGCCAGCAGAT	CCATATTCTTTCACCGCCCACTCC
Chapter 2, 3	PCR	B-Actin	AAGGCCAACCGCGAGAAG	ACAGCCTGGATAGCAACGTACA
Chapter 2, 3	PCR	c-MYC	GCTCGTCTCAGAGAAGCTGG	GCTCAGATCCTGCAGGTACAA
Chapter 2, 3	PCR	NCOA2	GAGCAGCATGAACCAGATGA	CCCCTCAGAGCAGGATCATT
Chapter 2, 3	PCR, RIP	U1	GGGAGATACCATGATCACGAAGGT	CCACAAATTATGCAGTCGAGTTTCCC
Chapter 2, 3	PCR, Urine	GAPDH	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGAG
Chapter 2, 3	PCR, Urine	KLK3	GAGCACCCCTATCAACCCCCTATT	AGCAACCCTGGACCTCACACCTAA

Chapter 2, 3	PCR, Urine	PCA3	CATGGTGGGAAGGACCTGATGATAC	GATGTGTGGCCTCAGATGGTAAAGTC
Chapter 2, 3	PCR, Urine	PCAT-14	GTACCACCCTGCAGGAAAA	AGGCAGTACCTATGCCGATG
Chapter 2, 3	PCR, Urine	SPINK1	CAAAAATCTGGGCCTTGCTGAGAAC	AGGCCTCGCGGTGACCTGAT
Chapter 2, 3	PCR, Urine, RIP	PCAT-1	TGAGAAGAGAAATCTATTGGAACC	GGTTTGTCTCCGCTGCTTTA
Chapter 4	qPCR	BRCA2	CATACAGTTAGCAGCGACAAAAA	CAAGATGGCTGAAAGTCTGGAT
Chapter 4	qPCR	KIF15	TGCGACTAAAGAAGGAAAATGTC	AAGGTGATGCCTAGGTAGCTGA
Chapter 4	qPCR	BIRC5	CACCGCATCTCTACATTCAAGA	CAAGTCTGGCTCGTTCTCAGT
Chapter 4	qPCR	TOP2A	TTGTTTCGAAAGCAGTCACAAG	GTTTCTTTGCCCGTACAGATTT
Chapter 4	qPCR	E2F8	AAATGGACAATCAGTTGCTGTG	CTAATTGTGACCCTTTGGGTGT
Chapter 4	qPCR	GAPDH	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGAG
Chapter 4	qPCR	B-actin	AAGGCCAACCGCGAGAAG	ACAGCCTGGATAGCAACGTACA
Chapter 4	qPCR	HMBS	GATGGGCAACTGTACCTGACTGGA	TGGGGCCCTCGTGGAATGTTA
Chapter 4	qPCR	PCAT1	TGAGAAGAGAAATCTATTGGAACC	GGTTTGTCTCCGCTGCTTTA
Chapter 4	qPCR	STAU1	TAACATCTCTTCAGGCCACGTA	GTTCTTGTTGTTTTTGGGGAAG
Chapter 4	Cloning	BRCA2_3'UTR	CATTTGCAAAGGCGACAATAAATT	AATCAGTGCCAATTTGAAAGCA
Chapter 4	Cloning	BRCA2_3'UTR_Spe1	GACTAGTGCATTTGCAAAGGCGACAATAAATT	GCAAGCTTAATCAGTGCCAATTTGAAAGCAAG
Chapter 4	Cloning	BRCA2_3'UTR_DEL-1_Spe1	GCACTAGTTGGGAGGAGTGCTTGAGGCCAGGA	GCACTAGTATATGTTGGTTCTGCTATAGTTC
Chapter 4	Cloning	BRCA2_3'UTR_DEL-3_HindIII	GCAAGCTTTTGCATTAAGGAGAACTATTTC	GAAGCTTCCTCAGCTTCTCAAAGTGTTGG
Chapter 4	Cloning	PCAT1_Full_Length	ACACATGGATATTGGATATCTGCT	TAGGCTCAAACACACATTTATTCATC
Chapter 4	Cloning	PCAT1_Deletion_1	GGGAGAAAGGAGATGACGCAAAG	TCATTAGCAGTTCCTGGGCCAA
Chapter 4	Cloning	PCAT1_Deletion_3	GCCAACTGAGGAACCTGAGCAAT	TTGGTGCAAGTTTGGCTTTGGGAAG
Chapter 5	qPCR	SChLAP-1	tggacacaATTTCAAGTCCTCA	CATGGTGAAAGTGCCTTATACA
Chapter 5	qPCR	AK093002	gggaatccccatatcacagt	ggaagcttacagttttcaagca
Chapter 5	qPCR	LOC145837	caccttggagagacccagaa	ggcagatggaagaagtggaa
Chapter 5	qPCR	U1	GGGAGATACCATGATCACGAAGGT	CCACAAATTATGCAGTCGAGTTTCCC
Chapter 5	qPCR	PCAT-1	TGAGAAGAGAAATCTATTGGAACC	GGTTTGTCTCCGCTGCTTTA
Chapter 5	qPCR	SNF5	GAGACTCTGACAGACGCTGAGA	gtgtgctgatgggctggTTAC
Chapter 5	qPCR	BRG1	AAAATCGAGAAGGAGGATGACA	CCAAGCTTGATCTTCACTTTGA
Chapter 5	qPCR	BRM	GAAGAGGAAGATGAAGAAGAGTC	AATCGCTCACTACAGGTTTGG
Chapter 5	qPCR	PCA3	CATGGTGGGAAGGACCTGATGATAC	GATGTGTGGCCTCAGATGGTAAAGTC
Chapter 5	qPCR	GAPDH	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGAG
Chapter 5	qPCR	HMBS	ATGGGCAACTGTACCTGACTGGA	TGGGGCCCTCGTGGAATGTTA
Chapter 5	qPCR	B-actin	AAGGCCAACCGCGAGAAG	CAGCCTGGATAGCAACGTACA
Chapter 5	Cloning	SChLAP-1 full length construct	Gcttttatgagctgtaacactca	ATTTATAAGTGAAAGAGGTTTA
Chapter 5	Cloning	SChLAP-1 Del-1 construct	Cgccgcctttaagaactgtacatgatgagaatggaatgac	GtcattccattctcatcatgtacagttcttaaaggcggcG
Chapter 5	Cloning	SChLAP-1 Del-2	AtgacaccagaaatcacaaagacaacaaaaataatgaagA	TcttcattatttttgttgtctttgtgatttctggtgtcaT

		construct		
Chapter 5	Cloning	SChLAP-1 Del-3 construct	CccaagaggctctgctgggaATGTGGATA TTGAATTATAT	ATATAATTCAATATCCACATtccc agcagagcctcttggG
Chapter 5	Cloning	SChLAP-1 Del-4 construct	TACATTAATCTGTATTGTGACATA GTAGTGAATAAGTCTC	GAGACTTATTCACTACTATGTC ACAATACAGATTAATGTA
Chapter 5	Cloning	SChLAP-1 Del-5 construct	TACATTAATCTGTATTGTGATCCA GTTCAGAGACTATTAG	CTAATAGTCTCTGAACTGGATC ACAATACAGATTAATGTA
Chapter 5	Cloning	SChLAP-1 Del-6 construct	Gcttttatgagctgtaacactca	AGAATAGTATGGGGGAACCAC C
Chapter 5	RACE 3'	SChLAP-1	TCAGAGGATAAACATGtgatat	NA
Chapter 5	RACE nested 3'	SChLAP-1	GATAAACATGtgatatggtttg	NA
Chapter 5	RACE nested 5'	SChLAP-1	gtgtccccacccaaatatcatct	NA
Chapter 5	RACE 5'	SChLAP-1	NA	TTAGCCAGGACTTGATGGAAA T
Chapter 5	RACE nested	SChLAP-1	NA	TTCCAGGTACATGGTGAAAGT G

APPENDIX 3

List of genes differentially regulated by PCAT-1 siRNA in LNCaP cells

Probe UID	Gene Name	Systematic Name	Expected score (dExp)	Observed score(d)	PCAT1 siRNA 1	PCAT1 siRNA 2	PCAT1 siRNA 3
23533	RSRC1	NM_016625	-1.006	4.940	0.521113533	0.334288917	0.633843957
6847	LOC651702	XR_019167	-0.918	4.947	0.347779207	0.440463273	0.652039537
19313	A_24_P255865	A_24_P255865	-0.900	4.954	0.355193733	0.495103133	0.727612313
7250	THC2679528	THC2679528	-0.976	4.959	0.863457823	1.246545633	0.657927357
5893	DBF4	NM_006716	-0.893	4.962	0.300336617	0.399244567	0.391550347
38352	MT1X	NM_005952	-1.337	4.973	0.884343633	0.506372153	1.472981433
30332	TCF25	NM_014972	-1.043	4.973	0.410323063	0.417012783	0.443828227
773	GPAA1	NM_003801	-0.909	4.980	0.39349141	0.392254013	0.374721293
33859	SUV420H2	NM_032701	-0.934	4.982	0.465955283	0.3954617	0.42715913
88	DNTTIP2	NM_014597	-1.022	4.983	0.438897253	0.52095768	0.301800087
9241	ZNF292	ENST00000339907	-0.849	4.984	0.441384457	0.625225883	0.773401753
1344	AKT1	NM_005163	-1.080	4.985	0.51938914	0.41884274	0.851518003
22008	ABCA5	NM_018672	-0.917	4.986	0.389778307	0.410267573	0.634416517
5893	DBF4	NM_006716	-0.923	4.987	0.305547533	0.493902177	0.418649217
38773	FAAH	NM_001441	-0.944	4.988	0.391919553	0.31305381	0.533911253
2475	KIF20A	NM_005733	-0.921	4.998	0.32667465	0.301709357	0.419867413
5768	C14orf106	NM_018353	-1.230	5.002	0.7744003	0.507741253	1.353977367
34338	MGC23270	BC015579	-0.742	5.008	0.486346933	0.509025187	0.8753982
15486	LOC339123	NM_001005920	-0.906	5.009	0.35048762	0.446652143	0.66335048
22189	GAS2L3	NM_174942	-0.943	5.009	0.515002413	0.405029773	0.418236747
1858	ALDH16A1	NM_153329	-0.847	5.017	0.334262947	0.366175573	0.40272684
9809	MGC27348	BC026177	-0.879	5.018	0.31200243	0.387968817	0.449448797
26678	SRFBP1	NM_152546	-0.909	5.026	0.45372356	0.487460857	0.32052505
12393	SAPS1	NM_014931	-1.092	5.030	0.442298353	0.586149067	0.679268087
1344	AKT1	NM_005163	-0.949	5.032	0.381963307	0.456398187	0.71013828
26086	CNP	NM_033133	-0.745	5.035	0.392566453	0.39436468	0.503181743
28385	hCG_21078	XM_371853	-0.821	5.036	0.302173313	0.315085833	0.27650109
5893	DBF4	NM_006716	-0.834	5.038	0.339448313	0.312095023	0.36292373
28071	UNC84B	NM_015374	-1.190	5.040	0.618462433	0.561693517	0.32800544
37861	THC2677659	THC2677659	-1.557	5.047	1.2097785	1.7060216	0.50454075
31827	SGOL2	NM_152524	-1.356	5.050	0.915289353	1.1657253	0.41766023
6473	DNAL4	NM_005740	-0.887	5.050	0.54492778	0.56360382	0.4354317
16923	TUBA1B	NM_006082	-0.937	5.055	0.416320243	0.302067373	0.34673305
2465	KNTC1	NM_014708	-0.853	5.055	0.34466382	0.45486941	0.377695347
39351	CKS1B	NM_001826	-0.820	5.077	0.285845263	0.377309393	0.375315483
18939	SLC3A2	NM_002394	-0.814	5.078	0.303155287	0.31232059	0.350279817
20619	PPM1F	NM_014634	-0.852	5.084	0.366864947	0.596582467	0.554739297
11124	AA971667	AA971667	-0.931	5.088	0.391513137	0.3409588	0.570428247
24622	ARRDC1	NM_152285	-1.001	5.089	0.395349533	0.447496583	0.369562323
10492	THC2679528	THC2679528	-1.594	5.090	0.9883614	0.99597058	1.087137167
27966	SOX7	NM_031439	-0.828	5.097	0.38674495	0.502509607	0.620296883
1344	AKT1	NM_005163	-0.975	5.109	0.507086017	0.400045887	0.7830165
2086	LRRC21	NM_015613	-1.332	5.115	0.999061997	1.066310153	1.142748637
39688	PKN2	NM_006256	-0.976	5.116	0.351516377	0.468048407	0.63927758
40762	SCLT1	NM_144643	-0.924	5.122	0.409959	0.477567233	0.581112767

2969	EIF5B	NM_015904	-1.117	5.124	0.626303633	0.6574326	1.3728249
14042	ENST00000355232	ENST00000355232	-1.171	5.128	0.498998637	0.526135173	0.958924633
7545	BCL7C	NM_004765	-0.913	5.136	0.37449657	0.387230343	0.52944822
2475	KIF20A	NM_005733	-0.892	5.136	0.334010173	0.30140925	0.39019727
37913	SLC44A2	NM_020428	-1.019	5.151	0.33390072	0.393228317	0.42877775
3685	HOXC6	NM_153693	-0.990	5.152	0.44696339	0.90032201	0.64609937
40356	UBE2M	NM_003969	-0.823	5.158	0.316531517	0.422213473	0.471410487
16246	NKTR	NM_001012651	-1.007	5.162	0.41111444	0.559825303	0.337470167
35361	KLHL17	NM_198317	-1.292	5.173	0.699678543	0.40598082	0.459744953
23798	KIAA1641	AK024934	-0.952	5.177	0.349477073	0.49014338	0.631385727
7417	FKSG44	AF334946	-1.704	5.186	1.236010467	0.54856024	1.358906733
26129	BF869497	BF869497	-0.922	5.188	0.37756848	0.460593633	0.68526819
27339	C4orf26	NM_178497	-1.242	5.193	0.62882262	0.801652357	1.113803167
12111	STK33	NM_030906	-0.875	5.198	0.38237799	0.337290137	0.448179493
3968	LOC647768	XR_018202	-0.986	5.213	0.460610143	0.320981443	0.51993272
1344	AKT1	NM_005163	-0.967	5.214	0.47592395	0.44191583	0.754791053
2969	EIF5B	NM_015904	-1.142	5.216	0.659037947	0.662752087	1.395220267
3384	ANKRD26	NM_014915	-1.538	5.223	1.0720811	1.2597755	0.443442473
6272	ONECUT2	NM_004852	-0.833	5.229	0.373997193	0.30682604	0.392072743
20893	MT1G	NM_005950	-1.470	5.234	1.025926667	0.492635607	1.4525373
10063	EPHX1	NM_000120	-0.998	5.239	0.456592407	0.6393029	0.975699467
20405	ADAMTSL5	NM_213604	-1.247	5.243	0.770603783	0.77379168	0.966703587
35061	FLJ20464	AK000471	-1.026	5.254	0.492303003	0.635008183	1.084638367
33073	PIK4CB	NM_002651	-1.035	5.259	0.403256707	0.35605781	0.479554683
11650	C13orf3	BC013418	-0.893	5.262	0.33405941	0.440747533	0.353472893
16024	AAAS	NM_015665	-0.997	5.265	0.40433856	0.474217983	0.72495142
11012	TROAP	NM_005480	-1.000	5.267	0.47867218	0.568151587	0.33242264
39616	MAP2K7	NM_145185	-0.990	5.270	1.156338833	0.96951868	1.3097191
14385	MT1X	NM_005952	-1.394	5.274	0.88888579	0.52054405	1.397178833
35151	ENST00000306565	ENST00000306565	-1.028	5.277	0.468886717	0.621980933	0.8518822
3156	FAM33A	NM_182620	-0.974	5.277	0.387177983	0.325833457	0.360859917
28875	ANKRD12	NM_015208	-2.231	5.277	1.880562533	1.833921833	0.55090053
34058	TCOF1	NM_001008656	-1.084	5.296	0.510207873	0.43407404	0.826990087
36348	ATP2B3	NM_021949	-0.978	5.299	0.491691977	0.714427833	0.518352133
8105	N4BP2	NM_018177	-0.897	5.299	0.495253717	0.55764644	0.352215003
21960	MT1L	X97261	-1.431	5.312	0.992309383	0.498003433	1.412342667
21744	RXRB	NM_021976	-1.035	5.313	0.459925187	0.663733083	0.378826713
36339	MGC11257	NM_032350	-0.924	5.314	0.401290763	0.685803317	0.91626706
2465	KNTC1	NM_014708	-0.921	5.319	0.414101543	0.310730003	0.398239777
38815	KLHDC3	NM_057161	-0.901	5.325	0.5090245	0.38581885	0.62731609
15472	FANCM	NM_020937	-1.393	5.328	0.82116619	0.882122967	1.835790333
14601	SLC25A25	NM_001006641	-1.073	5.351	0.560863153	0.386081937	0.709116947
938	RIOK2	NM_018343	-0.819	5.361	0.36534626	0.41962721	0.46843704
39586	CR749652	CR749652	-0.987	5.364	0.396294413	0.570525597	0.75641686
1360	FLJ10781	NM_018215	-1.079	5.370	0.411075523	0.63895559	0.80072462
16203	HNRPC	NM_031314	-0.998	5.372	0.367765797	0.593727497	0.716934797
18700	LOC731479	ENST00000360524	-1.093	5.374	0.6057479	0.575617207	0.351932263
1344	AKT1	NM_005163	-0.964	5.380	0.598357977	0.41825169	0.743226817
20215	LOC390996	XR_018526	-0.982	5.394	0.405068833	0.545122647	0.783484877
27320	AKAP8L	AL133576	-0.953	5.410	0.418468943	0.63994983	0.85365902
10985	TGM1	NM_000359	-0.905	5.411	0.514855813	0.408505873	0.533184083
21840	A_23_P95125	A_23_P95125	-1.206	5.412	0.651094213	0.613250533	1.248388667
7847	CEP55	NM_018131	-1.093	5.418	0.5886133	0.37273417	0.41119043
30197	CB959193	CB959193	-1.072	5.424	0.495321633	0.79465178	1.193498033
39251	THC2710703	THC2710703	-1.053	5.425	0.428183583	0.8008507	0.970914333
34994	MLLT6	NM_005937	-0.907	5.425	0.39138154	0.38074806	0.31101483
10358	EFNA5	NM_001962	-0.875	5.430	0.44297139	0.932751797	0.692673767
18239	NEK3	NM_002498	-1.027	5.431	0.428340593	0.572266813	0.623114333
5692	A_24_P920715	A_24_P920715	-1.025	5.436	0.534330913	0.423041337	0.483264213
5358	C10orf118	NM_018017	-1.188	5.438	0.750156133	0.783944547	0.40043825
12351	CR749547	CR749547	-0.906	5.447	0.420484167	0.59688365	0.382436553
2930	RAB1B	NM_030981	-1.040	5.452	0.472321977	0.531687413	0.70507138
19692	ZNF136	NM_003437	-0.969	5.454	0.570178987	0.477400177	0.700124533
15540	MT1B	NM_005947	-1.427	5.462	1.054919133	0.51508056	1.4235159
36045	L07392	L07392	-0.935	5.463	0.38727173	0.80135397	0.644476297
34376	LOC554248	BC078169	-0.804	5.467	0.375054717	0.434335617	0.542365383
27300	BC068045	BC068045	-0.992	5.470	0.440173253	0.448673133	0.66106364
16211	PIGU	NM_080476	-0.876	5.490	0.339947603	0.366708197	0.4136233
15210	MT2A	ENST00000245185	-1.387	5.493	0.998746267	0.46660686	1.216663967
26159	AK095583	AK095583	-0.980	5.497	0.474712203	0.385730933	0.621790057

25433	PRKCSH	NM_002743	-0.913	5.503	0.454363187	0.651303267	0.879991203
23778	A_23_P113762	A_23_P113762	-1.641	5.511	0.78288683	1.084225813	1.492523767
31306	SMC3	NM_005445	-1.285	5.525	0.749312787	0.756094753	0.385308323
8894	KRR1	NM_007043	-0.911	5.528	0.35605963	0.328529847	0.33130613
25932	ASNA1	NM_004317	-0.966	5.531	0.441434303	0.36865415	0.53366033
9519	BIRC5	NM_001012271	-0.964	5.533	0.360075717	0.513889463	0.61086785
22360	TUBA1C	NM_032704	-1.056	5.544	0.479116287	0.43058436	0.74580582
22990	MT2A	NM_005953	-1.475	5.554	1.036564667	0.468939107	1.229061433
5893	DBF4	NM_006716	-0.960	5.568	0.39454788	0.438621533	0.438768717
7308	CDIPT	NM_006319	-0.918	5.577	0.446623143	0.376722447	0.401562143
1331	CCDC41	NM_016122	-1.118	5.578	0.614482993	0.447735973	0.872743633
28868	PSIP1	NM_033222	-1.039	5.578	0.484696477	0.557505977	0.34215584
3815	SPINT1	NM_003710	-0.868	5.580	0.312147323	0.4029785	0.356640867
5583	ACTL8	NM_030812	-3.854	5.581	2.942171767	1.879107207	3.912231333
33215	C8orf70	NM_016010	-1.110	5.583	0.561022427	0.448351317	0.814709353
33875	A_24_P3627	A_24_P3627	-2.063	5.590	0.91876261	1.081132867	1.270674253
37791	CENPC1	NM_001812	-1.052	5.591	0.41824091	0.45513327	0.446136253
5651	CENPJ	NM_018451	-1.046	5.593	0.429845383	0.589642533	0.411300947
9519	BIRC5	NM_001012271	-0.938	5.600	0.38230024	0.43965645	0.600935233
8327	C20orf74	AB033098	-1.375	5.604	0.873607833	0.67425798	1.5338235
8287	NME6	NM_005793	-1.020	5.610	0.501350367	0.630588457	1.005878947
10660	PSMD13	NM_175932	-0.848	5.613	0.369453267	0.339212223	0.363020023
28039	THC2611971	THC2611971	-0.933	5.615	0.380078297	0.56931355	0.638574523
25419	SDC3	AB007937	-0.986	5.637	0.414816043	0.568493053	0.74562811
25843	FLJ36166	NM_182634	-2.103	5.645	1.196355917	1.383894067	1.365227817
27886	PADI4	NM_012387	-1.040	5.648	0.46397339	0.775036417	1.039291
5893	DBF4	NM_006716	-0.891	5.648	0.412957697	0.483474507	0.368978977
1344	AKT1	NM_005163	-0.993	5.651	0.538901933	0.460369237	0.812115733
5274	ELK1	NM_005229	-0.960	5.665	0.58595122	0.83713733	0.8490902
37972	SOCS3	NM_003955	-0.971	5.676	0.50843277	0.627204463	0.99061901
5061	ZRF1	NM_014377	-1.064	5.677	0.522003713	0.410465747	0.63694285
9723	TMEM37	NM_183240	-1.045	5.682	0.575322717	1.055485443	1.052750633
18366	CFL1	NM_005507	-0.826	5.693	0.40584895	0.57306705	0.63946067
2475	KIF20A	NM_005733	-0.950	5.695	0.379517643	0.34323663	0.429874857
9175	ARHGAP28	NM_001010000	-1.046	5.708	0.413240087	0.564982223	0.4670017
37270	JRK	NM_003724	-1.173	5.714	0.537261523	0.55554608	0.809393483
5893	DBF4	NM_006716	-0.908	5.717	0.346929533	0.442903727	0.399543373
1369	IGFBP3	NM_001013398	-1.096	5.717	0.615347023	0.632599097	1.00062019
9519	BIRC5	NM_001012271	-0.901	5.717	0.375764173	0.423292463	0.568327037
17504	ADCY6	NM_015270	-0.970	5.735	0.356816677	0.487686517	0.466494907
1344	AKT1	NM_005163	-1.216	5.735	0.55082286	0.555257	0.94155688
6088	NEK3	NM_002498	-0.986	5.736	0.531334133	0.469208713	0.723086047
2465	KNTC1	NM_014708	-0.940	5.738	0.490430607	0.379323967	0.416330273
25730	GSG2	AK056691	-0.869	5.746	0.401150573	0.40580977	0.537110163
20936	A_32_P135469	A_32_P135469	-1.224	5.751	0.55802433	0.7072993	0.47336114
35962	AF100640	AF100640	-3.175	5.753	2.867817033	3.007298167	0.889591033
37306	KIF4A	NM_012310	-1.056	5.758	0.492465713	0.480066237	0.796774763
9519	BIRC5	NM_001012271	-0.965	5.764	0.389337	0.47915467	0.629716953
2465	KNTC1	NM_014708	-0.895	5.774	0.353434943	0.359009393	0.394532187
35194	PHIP	NM_017934	-1.213	5.782	0.616586533	0.62626625	0.38234558
39248	ZNF93	AK096342	-1.134	5.782	0.62031509	0.37851226	0.54668243
9519	BIRC5	NM_001012271	-1.005	5.802	0.397001003	0.513268363	0.658882933
36028	LOC157627	AL832535	-1.201	5.808	0.637521053	0.59587838	1.0501538
29822	A_24_P84711	A_24_P84711	-0.923	5.819	0.425230457	0.37065526	0.360519987
3285	RAD50	NM_005732	-1.086	5.847	0.615432417	0.437646533	0.435127603
1344	AKT1	NM_005163	-1.017	5.852	0.492613403	0.5295083	0.839234777
3285	RAD50	NM_005732	-1.087	5.853	0.5687039	0.38603426	0.44301929
38028	PXMP4	NM_007238	-1.161	5.853	0.49065798	0.872732933	1.025954233
3944	RASL10B	NM_033315	-1.442	5.868	0.8276303	1.083801953	1.3806018
8106	ZNF289	NM_032389	-1.131	5.869	0.592109613	0.87438897	0.49290893
2285	CDKN3	NM_005192	-1.011	5.871	0.423589247	0.391312367	0.57461079
6977	AA045093	AA045093	-0.857	5.904	0.596226227	0.871654783	0.977422967
5893	DBF4	NM_006716	-1.026	5.904	0.37443005	0.454255473	0.45633269
17226	COQ2	NM_015697	-0.951	5.909	0.41728748	0.448028367	0.524524867
21306	BF513730	BF513730	-0.902	5.911	0.443364183	0.376125887	0.45001525
9519	BIRC5	NM_001012271	-0.917	5.924	0.3713759	0.493588077	0.588068383
2828	ZNF146	NM_007145	-0.860	5.925	0.578385017	0.509999627	0.6641796
2465	KNTC1	NM_014708	-0.985	5.927	0.375158947	0.392762413	0.44181522
19235	BC032451	BC032451	-1.061	5.931	0.647958147	0.666673343	1.061447053
7065	PNPLA6	NM_006702	-1.048	5.973	0.420737513	0.463273803	0.502059113

1369	IGFBP3	NM_001013398	-1.370	5.974	0.6840508	0.75719114	1.186419267
1369	IGFBP3	NM_001013398	-1.184	5.977	0.59713945	0.63470889	1.0765522
39787	CRYBA2	NM_005209	-1.001	5.979	0.502473763	0.807285887	1.047568733
3285	RAD50	NM_005732	-1.100	5.990	0.62223484	0.424403617	0.448982877
25404	TRIP11	NM_004239	-2.018	5.992	1.670188267	1.542010033	0.637054323
3827	PTTG2	NM_006607	-0.939	6.000	0.357020577	0.4196713	0.3748663
40463	LOC643401	BC039509	-1.553	6.012	1.107100733	0.65460662	1.4675565
3285	RAD50	NM_005732	-1.034	6.022	0.54367696	0.45766403	0.405674807
34358	FLNA	NM_001456	-1.216	6.029	0.683593313	0.714153417	0.427904197
9519	BIRC5	NM_001012271	-0.934	6.039	0.395572163	0.463541343	0.596245433
35917	MAP9	NM_001039580	-1.191	6.039	0.73893866	0.804490233	0.458019097
5917	PPOX	NM_000309	-1.070	6.048	0.540043097	0.409437883	0.440636673
35830	LOC375748	NM_001010895	-0.952	6.055	0.42159785	0.577351457	0.70910658
1281	AK098270	AK098270	-1.105	6.082	0.51618436	0.7306318	1.00736931
17235	STAG2	NM_006603	-0.971	6.084	0.42422093	0.702816247	0.657899913
1369	IGFBP3	NM_001013398	-1.088	6.099	0.691615633	0.58158993	0.880269103
5683	USP5	NM_003481	-1.023	6.112	0.507237733	0.570293813	0.831264847
25917	A_32_P144634	A_32_P144634	-1.289	6.117	0.799903247	0.769416933	1.337361733
21905	TMTC3	NM_181783	-0.938	6.134	0.471269013	0.699257633	0.59849241
25443	TTK	NM_003318	-1.176	6.142	0.66186526	0.951772133	0.547690443
40405	LOC727992	BX099599	-1.729	6.144	1.320396533	0.947707433	1.4510644
1111	PTK2B	NM_173174	-1.285	6.164	0.639258467	0.5931595	0.966420313
32127	MT2A	NM_005953	-1.548	6.174	1.1225589	0.563942687	1.287562467
34298	RAI2	NM_021785	-1.293	6.179	0.667352407	0.6709583	0.895283753
24194	BAP1	NM_004656	-0.903	6.181	0.426247967	0.48157077	0.51867946
15639	LOC441900	XR_018860	-1.239	6.194	0.687487547	0.64582657	1.0892146
19557	EIF5A	NM_001970	-0.889	6.196	0.357654463	0.42287157	0.444920493
9519	BIRC5	NM_001012271	-0.965	6.198	0.41817245	0.515097547	0.635008633
1111	PTK2B	NM_173174	-1.133	6.209	0.557920643	0.57041109	0.851068867
3285	RAD50	NM_005732	-1.107	6.225	0.606048067	0.416692487	0.478704673
2465	KNTC1	NM_014708	-0.993	6.228	0.390746393	0.417130013	0.463618183
32412	FAM44A	NM_148894	-1.420	6.232	0.99772192	0.8994741	0.510015133
11015	A_32_P138933	A_32_P138933	-1.614	6.235	0.741806227	0.933223433	1.227980167
9519	BIRC5	NM_001012271	-0.898	6.249	0.4018579	0.474494667	0.547166907
1111	PTK2B	NM_173174	-1.010	6.250	0.488387477	0.602771527	0.7455181
27954	THC2760643	THC2760643	-1.363	6.270	0.764793787	1.857968767	1.400994633
39718	KIAA0586	NM_014749	-1.051	6.287	0.48112958	0.3730477	0.425511457
3285	RAD50	NM_005732	-1.136	6.296	0.629687533	0.447152487	0.49195581
23079	SHPRH	NM_173082	-1.139	6.302	0.625485767	0.423523337	0.515336883
7148	KIAA1212	NM_018084	-1.511	6.314	1.048656367	0.863574257	0.79854459
15774	A_32_P62480	A_32_P62480	-1.725	6.324	1.010259983	0.827800917	0.90616938
31360	NAP1L4	NM_005969	-1.072	6.346	0.502304767	0.566436147	0.755094753
36256	PHF19	NM_001009936	-1.063	6.353	0.546339057	0.634155567	0.492962477
5893	DBF4	NM_006716	-1.035	6.357	0.4029716	0.484440597	0.472126817
20599	WDR79	NM_018081	-1.133	6.366	0.64939212	0.45845425	0.70377362
1369	IGFBP3	NM_001013398	-1.105	6.368	0.590328083	0.68390198	1.004822813
21508	PPP5C	NM_006247	-0.925	6.369	0.673417077	0.770083807	0.651410743
26067	IGHMBP2	NM_002180	-1.014	6.371	0.505392033	0.51778365	0.591812193
13005	CGB1	NM_033377	-1.737	6.395	0.937694687	1.3665372	1.3151681
40788	LYPLA2P1	NR_001444	-1.055	6.418	0.56050866	0.493340013	0.659833
3642	TUB	NM_003320	-2.645	6.449	1.533428647	1.055873433	1.6439119
30574	MKI67	NM_002417	-1.171	6.474	0.641194913	0.541955853	0.801107427
40099	USP46	NM_022832	-1.074	6.484	0.459346287	0.687842643	0.678352667
29804	AK021715	AK021715	-1.010	6.485	0.464045273	0.43927597	0.608874213
5479	HLTF	NM_139048	-1.027	6.510	0.423599303	0.490367523	0.543025987
28374	ASH1L	NM_018489	-0.984	6.516	0.538743073	0.493611357	0.478698113
19473	GBA2	NM_020944	-0.894	6.516	0.428892993	0.553983553	0.5866156
4248	SLC30A1	NM_021194	-1.177	6.520	0.764933833	0.6571872	1.179001833
14291	C14orf24	NM_173607	-1.264	6.521	0.641755997	0.68608891	0.467153407
22649	THC2648557	THC2648557	-1.314	6.539	0.783100567	0.892900767	1.218031467
1369	IGFBP3	NM_001013398	-1.294	6.544	0.65774132	0.784666457	1.149141933
1629	ENST00000371189	ENST00000371189	-1.100	6.545	0.559547723	0.465720677	0.578886977
13077	CFL1	NM_005507	-0.985	6.545	0.46893856	0.649925093	0.776248013
27996	BIRC7	NM_022161	-1.065	6.559	0.581351267	0.961036833	1.04183165
1769	A_24_P401521	A_24_P401521	-1.161	6.566	0.5445388	0.847052633	1.035766233
35959	ZNF518	NM_014803	-1.459	6.583	1.0154131	1.387224933	0.718855833
1369	IGFBP3	NM_001013398	-1.249	6.583	0.7376502	0.645824133	1.0946908
9519	BIRC5	NM_001012271	-0.941	6.586	0.420310553	0.548942607	0.606487
31416	SNRP70	NM_003089	-1.119	6.587	0.540011933	0.500633767	0.42571368
37650	FGFR1OP	ENST00000366847	-1.574	6.595	0.868947213	1.0405509	0.77659564

16819	FAM33A	NM_182620	-0.987	6.600	0.406282547	0.385594327	0.465264927
38263	IGKV2-24	BC063599	-1.107	6.602	0.667914563	0.889044167	1.285621467
40224	LOC341378	XR_016562	-1.222	6.631	0.8158969	1.061813767	0.59359718
1369	IGFBP3	NM_001013398	-1.405	6.636	0.753549653	0.746442503	1.231844333
17086	PHF17	NM_199320	-1.131	6.684	0.5612181	0.6884305	0.48410086
11553	A_24_P272503	A_24_P272503	-1.068	6.689	0.442968493	0.541146603	0.57413828
39032	LOC649056	XR_018580	-2.262	6.704	2.5882742	2.499471133	1.478521067
14426	TSR2	NM_058163	-1.302	6.706	0.790769587	0.7495985	1.3135213
40184	SFRS12	AB209694	-0.965	6.713	0.65471357	0.77616179	1.076692967
21366	KIR2DS4	NM_012314	-1.221	6.736	0.669957163	0.994646467	1.189627567
41000	LRRC2	NM_024512	-1.339	6.739	0.671780933	1.057416867	1.081926433
39432	THC2537856	THC2537856	-1.948	6.745	2.1216708	1.9880997	1.0868812
21040	PGS1	NM_024419	-1.128	6.746	0.582391737	0.782717317	1.022129567
21598	ENST00000343253	ENST00000343253	-1.746	6.750	1.233909167	1.384278133	0.7873888
1369	IGFBP3	NM_001013398	-1.162	6.786	0.666790833	0.699803847	1.0568066
24124	A_32_P157671	A_32_P157671	-1.003	6.808	0.4567525	0.506245537	0.44178937
177	BRCA2	NM_000059	-2.132	6.814	1.51479272	1.304196567	1.8578466
35045	RPA4	NM_013347	-1.430	6.817	0.666397837	0.80661291	0.943123237
1999	CCDC69	NM_015621	-1.727	6.844	1.003219847	1.149518467	1.645128733
14858	CD86	NM_006889	-1.364	6.848	0.613720477	0.831415573	0.885769663
32076	LRRCC1	NM_033402	-1.637	6.859	1.3383444	1.070075333	0.674055747
3303	UACA	NM_001008224	-1.456	6.863	1.0628048	0.882884047	0.580912167
3353	A_24_P936067	A_24_P936067	-1.428	6.881	0.9444038	0.991508473	1.121625677
33123	CB162722	CB162722	-1.176	6.952	0.574179317	0.81469171	0.753053833
3008	THC2579650	THC2579650	-1.127	6.953	0.538577063	0.848946497	0.81459915
3285	RAD50	NM_005732	-1.230	6.957	0.670889783	0.49312263	0.526934597
35052	CD619445	CD619445	-1.507	6.961	0.74908649	0.856413467	1.004577987
11850	LOC91431	NM_138698	-1.128	6.981	0.578900337	0.64622077	0.7991376
39629	MIDN	NM_177401	-1.116	6.996	0.533730697	0.558181503	0.464683693
1369	IGFBP3	NM_001013398	-1.262	7.039	0.722907817	0.743149467	1.120351933
3285	RAD50	NM_005732	-1.105	7.056	0.6017554	0.510130527	0.458504667
2465	KNTC1	NM_014708	-0.983	7.091	0.414883687	0.419636013	0.427022257
3079	LOC727768	ENST00000339700	-1.146	7.112	0.557336847	0.754403767	0.931354637
3233	PHF17	NM_199320	-1.123	7.133	0.50628656	0.563930333	0.471989813
35013	AK5	NM_174858	-0.969	7.155	1.044879323	1.377864967	1.5065347
26036	THC2697162	THC2697162	-1.282	7.180	1.5593268	1.566910267	1.388146667
4445	SCLT1	NM_144643	-1.129	7.180	0.643899167	0.490144053	0.65004208
36403	OXTR	NM_000916	-1.203	7.182	0.65636004	0.565799283	0.72835248
1111	PTK2B	NM_173174	-1.124	7.191	0.66676948	0.55215102	0.834798657
1111	PTK2B	NM_173174	-1.122	7.209	0.643002677	0.574078913	0.815721943
13534	SASS6	NM_194292	-1.237	7.231	0.594469133	0.64483456	0.579460933
22374	A_24_P349590	A_24_P349590	-1.070	7.237	0.50525948	0.676091217	0.738450967
1111	PTK2B	NM_173174	-1.188	7.238	0.659862567	0.616939933	0.93864942
22318	MNS1	NM_018365	-1.154	7.241	0.748204967	0.63778115	0.967874253
7567	HNRPCL1	NM_001013631	-1.044	7.245	0.49767689	0.623966953	0.668777027
1111	PTK2B	NM_173174	-1.075	7.290	0.64650101	0.594472533	0.797284517
3285	RAD50	NM_005732	-1.152	7.316	0.591412067	0.466962903	0.495728847
37364	THC2609820	THC2609820	-1.421	7.347	0.772510577	0.840996253	0.653292077
8909	KIAA1212	NM_018084	-1.261	7.353	0.933755987	1.296268	0.921332217
29724	THC2541992	THC2541992	-1.404	7.384	0.720686867	0.9446593	0.832029683
1111	PTK2B	NM_173174	-1.237	7.411	0.703494947	0.626053223	0.958108077
35716	ESF1	NM_016649	-1.440	7.412	1.137021667	0.76417095	0.839484913
3285	RAD50	NM_005732	-1.107	7.413	0.591186447	0.5369251	0.46867373
15	EHMT2	NM_006709	-1.263	7.534	0.720532753	0.722599553	0.584685687
33059	ZNF91	NM_003430	-1.259	7.538	0.734496273	0.825897767	0.562209833
2329	PLK1	NM_005030	-1.000	7.549	0.52634977	0.585007443	0.518939517
13995	KIF11	NM_004523	-1.041	7.562	0.540392967	0.61345032	0.700575577
26524	SETD1A	NM_014712	-1.347	7.591	0.847271053	0.8893106	0.665379177
7597	BX537520	BX537520	-1.590	7.617	0.994465267	1.166917517	1.7125249
1111	PTK2B	NM_173174	-1.163	7.667	0.675232173	0.61131437	0.89205528
22388	THC2672325	THC2672325	-1.144	7.712	0.598596057	0.53114055	0.663180083
9086	FLJ32679	NM_001012452	-1.429	7.727	0.72971708	1.1258943	1.051355313
8520	THC2669975	THC2669975	-1.914	7.730	1.213353143	1.489653933	0.941284133
30402	NCAPG	NM_022346	-1.055	7.752	0.488393787	0.445999163	0.476008677
9830	CASC5	NM_170589	-1.267	7.767	0.857022123	0.90044062	1.359233567
24750	LRRCC1	NM_033402	-1.693	7.812	1.353386833	1.195211567	0.79812034
21761	ZBTB20	BC010934	-1.581	7.891	1.078137813	0.80133861	0.929186493
10879	PHF20L1	NM_032205	-1.228	7.949	0.76263021	0.772459547	0.66079625
20162	PCM1	NM_006197	-1.268	7.977	0.7725081	0.775718483	0.708234167
979	PPIG	NM_004792	-1.136	8.018	0.5626696	0.545135547	0.674137473

38228	KTN1	NM_182926	-1.531	8.039	0.99858982	0.942887673	0.649512647
16734	AKAP9	NM_147171	-2.137	8.068	1.8441631	1.9191385	1.008073767
177	BRCA2	NM_000059	-2.360	8.116	1.833684833	1.277420657	1.922160733
1111	PTK2B	NM_173174	-1.160	8.119	0.610866813	0.72691395	0.8754361
7434	FANCM	NM_020937	-1.253	8.147	0.73361958	0.941641633	0.796186443
12294	LOC150759	AK057596	-1.194	8.183	0.6840992	0.6965006	0.544409453
177	BRCA2	NM_000059	-2.060	8.238	2.008979467	1.4975493	1.7193468
5666	MAF1	NM_032272	-1.137	8.266	0.614623	0.726426197	0.85673299
20611	DYNC2H1	NM_001080463	-1.175	8.308	0.814906827	1.146598567	1.169911767
40329	BRWD1	ENST00000380831	-1.340	8.349	0.815870987	1.229598567	1.280002433
39405	SLITRK6	NM_032229	-1.324	8.446	0.74691398	0.857483533	0.66426292
20557	FLJ39660	NM_001080539	-1.273	8.483	0.6875299	0.7474494	0.86463523
37322	A_32_P11673	A_32_P11673	-1.469	8.514	1.065185867	1.763905833	1.295553467
3367	ASPM	NM_018136	-1.645	8.573	1.2849287	1.312144533	0.813613333
13791	A_32_P222060	A_32_P222060	-1.443	8.617	0.848227383	1.258495633	1.234294367
177	BRCA2	NM_000059	-1.383	8.714	1.229256033	1.493195267	1.148547567
20461	MAP9	NM_001039580	-1.160	8.743	0.718646133	0.738837733	0.682114653
17879	ATRX	BC002521	-1.307	8.817	0.856297917	1.006538533	0.699420147
177	BRCA2	NM_000059	-2.407	8.918	1.613167367	1.654055667	2.218758133
11257	NCAPH2	NM_014551	-1.349	9.101	0.851816193	0.669243023	0.846574533
18995	C1QTNF5	NM_015645	-1.668	9.182	1.3833096	1.4210258	1.155814013
177	BRCA2	NM_000059	-2.033	9.235	1.429109433	1.699343633	1.660988667
9559	RBBP6	NM_006910	-1.317	9.413	0.8733457	0.77431249	0.67727055
3856	FAM44A	NM_148894	-1.346	9.561	0.94367407	0.83779945	0.71739828
9007	C9orf39	NM_017738	-1.356	9.615	0.9222772	0.978756967	1.02163632
19154	TPR	NM_003292	-1.306	9.677	0.8499103	0.90162223	0.683191657
6209	CENPF	NM_016343	-2.283	9.798	1.861923633	1.663239267	1.1828924
2833	FLJ20054	AL831839	-1.642	9.942	1.343169733	1.020266333	1.5220943
26199	GOLGA4	NM_002078	-1.758	10.005	1.3932187	1.432338733	0.963317613
35429	MALAT1	NR_002819	-1.148	10.136	0.64240878	0.62262942	0.708047977
31550	A_32_P167111	A_32_P167111	-3.289	10.191	2.194734433	2.402190167	2.364924667
10950	SLC30A1	ENST00000367000	-1.456	10.431	1.04046762	1.1783135	1.4918507
6839	BDP1	NM_018429	-1.243	11.168	0.71055601	0.72444531	0.703755983
26596	CEP290	NM_025114	-2.084	11.490	1.835236567	1.494201267	1.382930533
36285	GOLGB1	NM_004487	-1.407	11.523	0.885280433	0.90882262	1.064330267
12826	TPR	NM_003292	-1.672	11.800	1.290544267	1.281429067	0.99407887
13760	GOLGA4	NM_002078	-1.709	11.822	1.2912084	1.332447533	1.00790459
5153	NSBP1	NM_030763	-1.414	11.889	0.99900175	1.144781533	0.943524167
5153	NSBP1	NM_030763	-1.506	12.006	0.944791343	1.137348733	1.026634233
177	BRCA2	NM_000059	-1.971	12.083	1.905555633	1.4945863	1.487126333
5153	NSBP1	NM_030763	-1.437	12.215	1.0286332	1.081336033	0.957489913
177	BRCA2	NM_000059	-1.944	12.266	1.804701433	1.488850133	1.380489533
32450	ASPM	NM_018136	-1.571	12.640	1.061606767	1.060196717	0.997615553
37979	ESF1	NM_016649	-1.589	12.644	1.141323667	1.009812867	1.1480279
5153	NSBP1	NM_030763	-1.370	12.775	0.9431407	1.030663177	0.890224167
5153	NSBP1	NM_030763	-1.404	12.987	0.956047347	1.076132167	0.91721711
177	BRCA2	NM_000059	-1.991	13.109	1.648784767	1.360648833	1.566621333
5153	NSBP1	NM_030763	-1.463	13.506	1.058007	0.984533533	0.968643367
5153	NSBP1	NM_030763	-1.452	13.705	1.031717833	1.120189433	0.964975123
7171	FANCM	NM_020937	-1.438	13.711	0.988120267	0.985151513	1.111268667
5153	NSBP1	NM_030763	-1.479	13.861	1.011338267	1.126205	0.974511767
5153	NSBP1	NM_030763	-1.487	14.304	1.02291508	1.046319933	0.9767779
39026	GOLGB1	NM_004487	-1.588	14.454	1.158623267	1.213636767	1.037703967
5153	NSBP1	NM_030763	-1.496	15.014	1.0095847	1.0738679	0.996408
17589	CEP290	NM_025114	-3.085	15.766	2.433021867	2.3530326	1.897406867
177	BRCA2	NM_000059	-2.095	15.853	1.751690633	1.5872352	1.738965433
8003	ANKRD12	NM_015208	-2.243	15.986	1.894270967	1.819415033	1.527371967
16091	NSBP1	NM_030763	-1.607	16.576	1.219613033	1.327510767	1.390599533
34297	ANKRD12	NM_015208	-2.015	17.894	1.822984667	1.764787967	1.727613833
36127	MPHOSPH1	NM_016195	-2.824	18.713	2.1652628	2.436810133	2.223331867
966	CENPE	NM_001813	-2.860	25.769	2.236115633	2.434618067	2.382718733
966	CENPE	NM_001813	-2.566	26.158	2.3450358	2.4236443	2.1953522
966	CENPE	NM_001813	-2.690	26.519	2.274578267	2.493755367	2.316120067
966	CENPE	NM_001813	-2.745	26.785	2.2597778	2.383257133	2.3263037
966	CENPE	NM_001813	-2.534	27.526	2.202902867	2.374423033	2.183361767
966	CENPE	NM_001813	-2.603	28.753	2.373398167	2.371493567	2.219390933
966	CENPE	NM_001813	-2.668	28.921	2.2899048	2.4357717	2.290984567
966	CENPE	NM_001813	-2.623	29.979	2.279469567	2.4181044	2.270165267
966	CENPE	NM_001813	-2.776	30.171	2.3647857	2.471678167	2.3492171
966	CENPE	NM_001813	-2.952	30.271	2.342841133	2.455456233	2.435027933

3122	PLAT	NM_000930	3.085	-13.137	-2.340478267	-2.026714733	-2.671139967
3122	PLAT	NM_000930	2.860	-12.507	-2.1033524	-2.028532167	-2.640381633
3122	PLAT	NM_000930	2.718	-12.403	-1.954678333	-2.289306633	-2.457411767
28729	PCSK9	NM_174936	2.378	-11.839	-1.844288967	-1.946876433	-2.597768133
3122	PLAT	NM_000930	3.175	-11.802	-2.005144867	-2.2210379	-2.72335
3122	PLAT	NM_000930	2.824	-11.471	-2.091276333	-1.9414295	-2.563950333
16285	THRSP	NM_003251	1.503	-10.792	-1.001892567	-1.353822233	-1.390107833
3122	PLAT	NM_000930	2.429	-10.077	-2.096323933	-1.939678033	-2.416908067
29947	BC092452	BC092452	1.272	-9.652	-0.70470064	-0.632113287	-0.75194113
4874	PTGS2	NM_000963	1.906	-9.182	-1.344396267	-1.0644167	-1.602369367
29984	TNFAIP2	NM_006291	1.358	-9.009	-0.836702917	-0.66639271	-0.906576047
16181	A_23_P158380	A_23_P158380	1.390	-8.989	-0.8601257	-0.89270486	-0.8823348
3122	PLAT	NM_000930	1.780	-8.989	-1.950646333	-1.654256333	-2.2400057
20806	MICALCL	NM_032867	1.266	-8.709	-0.663635113	-0.93141427	-0.9377677
15877	TRIM7	NM_033342	1.257	-8.707	-0.741735467	-0.620950323	-0.83477832
17204	CA314936	CA314936	1.358	-8.199	-0.767004867	-0.831700933	-0.591780793
6201	GIPR	NM_000164	1.953	-8.158	-1.804223367	-1.746254433	-1.0566193
38671	KCNK5	NM_003740	1.204	-8.118	-0.604492907	-0.618442513	-0.756560563
29188	THC2644803	THC2644803	1.714	-7.954	-1.26182429	-1.165831	-1.711233533
1679	STAT1	NM_139266	1.073	-7.851	-0.50780306	-0.5694386	-0.658531
4968	RBP5	NM_031491	1.203	-7.726	-0.591627873	-0.590485747	-0.614274423
22503	ARHGAP20	NM_020809	1.812	-7.627	-1.01966633	-1.3101279	-0.905578033
4874	PTGS2	NM_000963	1.816	-7.564	-1.467643233	-1.147307933	-1.5689084
27777	AA854957	AA854957	2.283	-7.480	-1.472720567	-1.4402429	-2.321127767
21454	ISGF3G	NM_006084	1.277	-7.414	-0.759572553	-0.917330233	-1.326219533
1679	STAT1	NM_139266	1.083	-7.323	-0.473540337	-0.6225738	-0.660685347
7483	LAIR1	NM_021706	1.330	-7.298	-0.702031433	-0.886803153	-1.007378067
4193	THC2656519	THC2656519	1.233	-7.245	-0.68373928	-0.849298617	-1.135062433
7432	AATK	NM_001080395	1.093	-7.244	-0.570042307	-0.54732772	-0.773487233
7410	SLC39A10	NM_020342	1.727	-7.190	-1.182429333	-0.696937927	-1.300920633
15020	AK098478	AK098478	1.889	-7.147	-1.14753615	-0.785391933	-1.114865733
21657	RHO	NM_000539	1.035	-7.094	-0.67730392	-0.603532777	-0.613235513
38347	DHTKD1	NM_018706	1.136	-7.079	-0.538328167	-0.621389337	-0.764034233
26487	RAB11FIP4	NM_032932	2.024	-7.070	-1.5672953	-1.2021582	-0.772028633
5310	CV800748	CV800748	1.235	-7.060	-0.665014247	-0.65054795	-0.499491453
15185	FNDC3A	NM_014923	1.198	-6.987	-0.575411163	-0.5936698	-0.848950613
4874	PTGS2	NM_000963	1.986	-6.981	-1.7415197	-1.009077947	-1.7641105
39636	GGTA1	NR_003191	1.506	-6.974	-0.972879833	-0.978762567	-0.563704367
22382	ACSS2	NM_018677	1.020	-6.946	-0.467846453	-0.70541827	-0.677347163
20311	PRKCH	NM_006255	1.284	-6.934	-0.614962163	-0.80861181	-0.5616557
32516	SLC16A3	NM_004207	1.274	-6.906	-0.71810998	-1.2199086	-1.4702735
1194	LOXL2	NM_002318	0.980	-6.893	-0.465528853	-0.61158034	-0.572988537
3122	PLAT	NM_000930	1.624	-6.863	-1.598028433	-1.649845133	-1.8902052
39779	JMJD1A	NM_018433	1.125	-6.858	-0.544273133	-0.515954377	-0.6290375
30172	THC2727548	THC2727548	3.013	-6.831	-1.57174358	-1.495704033	-2.141288067
30066	RGS9BP	NM_207391	1.024	-6.817	-0.438772233	-0.546024937	-0.62252312
14619	DLGAP4	NM_014902	1.007	-6.753	-0.486158857	-0.589980647	-0.68890402
35907	AQP5	BC032946	1.899	-6.701	-1.146254867	-1.0118404	-0.77202461
1679	STAT1	NM_139266	0.978	-6.662	-0.440554857	-0.539309153	-0.605579023
1679	STAT1	NM_139266	1.085	-6.616	-0.447193077	-0.552272707	-0.670089053
20413	PIGR	NM_002644	1.068	-6.604	-0.531717127	-0.952244533	-0.929526513
15979	THC2528836	THC2528836	1.237	-6.579	-0.620215433	-0.654062567	-1.00645251
15191	FAM114A1	NM_138389	1.031	-6.552	-0.51933855	-0.48077252	-0.505364447
15985	PHLDA2	NM_003311	1.047	-6.506	-0.48567673	-0.85902754	-0.81389132
38178	CFB	NM_001710	1.205	-6.480	-0.609473247	-1.306936133	-1.1029469
7414	IFITM1	NM_003641	1.227	-6.445	-0.65705865	-1.1763182	-0.733485597
21113	C8orf4	NM_020130	1.392	-6.390	-0.928227193	-1.090246133	-1.105645157
1679	STAT1	NM_139266	1.054	-6.363	-0.437865333	-0.506963193	-0.644507987
18536	UQCRC2	NM_003366	1.031	-6.362	-0.454220093	-0.5004366	-0.469320307
18166	ENST00000372045	ENST00000372045	2.269	-6.356	-1.24716358	-0.818261367	-1.637104067
37683	AI263083	AI263083	1.622	-6.347	-1.211870467	-1.142831767	-0.598803333
3191	KAZALD1	ENST00000224809	1.019	-6.311	-0.467684917	-0.754526733	-0.548066653
36814	KIAA1446	NM_020836	1.294	-6.309	-0.707675717	-0.50107563	-0.80290935
26130	AP2B1	NM_001282	1.045	-6.292	-0.44220838	-0.409955203	-0.473319857
13330	ACSS2	NM_018677	1.063	-6.290	-0.47101459	-0.741283013	-0.822097243
39521	AA573434	AA573434	1.948	-6.254	-0.80939348	-1.405408433	-1.406230733
29855	TTTY14	NR_001543	0.979	-6.237	-0.501440753	-0.438080143	-0.604608467
1679	STAT1	NM_139266	0.924	-6.230	-0.42827529	-0.484838647	-0.55202351
9926	ANXA9	NM_003568	1.051	-6.216	-0.447571523	-0.791058947	-0.778667223
28754	P2RY2	NM_176072	1.274	-6.200	-0.622990947	-1.1232551	-0.83217842

13081	ZNF226	NM_015919	1.010	-6.185	-0.425115263	-0.640717713	-0.51760771
10338	PLEKHA2	AK098787	1.532	-6.180	-0.988053043	-0.54755066	-1.1418981
40095	LPIN1	AF147446	1.123	-6.152	-1.1019879	-1.405928953	-1.9759118
19789	MORF4L2	NM_012286	0.982	-6.138	-0.462598667	-0.423519463	-0.495940317
5485	GUSBL2	BC065547	1.009	-6.130	-0.460201847	-0.817949553	-0.60525911
3122	PLAT	NM_000930	2.027	-6.109	-2.016946033	-1.18573686	-2.295623
39778	LMAN1	NM_005570	0.913	-6.090	-0.388900693	-0.491598853	-0.440477677
1679	STAT1	NM_139266	0.987	-6.057	-0.40078041	-0.520509533	-0.62760689
11473	CXCL10	NM_001565	1.322	-6.056	-0.74504858	-1.8006046	-1.203655633
39018	TBC1D24	NM_020705	1.160	-6.039	-0.5083218	-0.499618133	-0.67593672
1679	STAT1	NM_139266	1.121	-6.029	-0.42952224	-0.607455713	-0.686771333
25288	EHD4	NM_139265	0.994	-6.021	-0.42301759	-0.367750437	-0.376609233
22818	TBXA2R	NM_001060	1.101	-6.009	-0.590554657	-0.584686533	-0.407919817
37812	NAP1L5	NM_153757	1.006	-6.004	-0.38303683	-0.563317283	-0.612914333
28638	ECGF1	NM_001953	0.904	-5.999	-0.34716604	-0.420867013	-0.4133207
29208	SPRR2G	NM_001014291	1.511	-5.998	-0.897242567	-0.84823631	-0.524488863
15603	THC2585656	THC2585656	0.976	-5.996	-0.39185481	-0.56994412	-0.64962875
27294	ATP8B3	NM_138813	1.704	-5.983	-0.965753843	-0.883361	-1.5825865
1194	LOXL2	NM_002318	1.153	-5.949	-0.55721845	-0.461396557	-0.6771499
25086	THC2524477	THC2524477	1.366	-5.948	-0.709317707	-0.685945063	-0.617770563
428	TNFSF10	NM_003810	1.188	-5.946	-0.605115613	-1.561742133	-1.4817489
304	CAV1	NM_001753	2.550	-5.928	-1.310676643	-1.320399133	-1.246139833
28045	HCP5	L06175	1.063	-5.928	-0.433519643	-0.666887393	-0.814911487
24161	JMJD1A	NM_018433	1.229	-5.918	-0.54888245	-0.566548287	-0.9242438
14359	DIO3	NM_001362	1.192	-5.888	-0.728894363	-1.964018233	-2.156623367
13037	S100A10	NM_002966	1.139	-5.881	-0.502858483	-0.788590417	-0.507900437
18094	ZNF554	NM_152303	0.995	-5.867	-0.371312037	-0.434138917	-0.443653677
19826	AF075063	AF075063	1.038	-5.859	-0.4193786	-0.488466087	-0.57874784
1337	DQ168992	DQ168992	1.507	-5.849	-1.099717433	-0.841722523	-0.74513289
31949	ST3GAL4	AK021929	1.470	-5.840	-0.897867257	-0.97820588	-1.922555133
1679	STAT1	NM_139266	0.893	-5.837	-0.40610774	-0.471493773	-0.540889453
3122	PLAT	NM_000930	3.854	-5.817	-1.723109533	-1.52871322	-2.754939
8534	FCGBP	NM_003890	1.163	-5.802	-0.584425627	-0.493120387	-0.90010621
1679	STAT1	NM_139266	0.935	-5.781	-0.360812767	-0.52107687	-0.57605729
11995	STMN3	NM_015894	1.114	-5.739	-0.5513997	-0.75211232	-1.185465367
17918	KIAA1913	BC044246	1.324	-5.716	-1.2799507	-0.908144267	-1.247602733
7749	CCDC78	NM_173476	0.889	-5.702	-0.35366142	-0.588982633	-0.520895863
27212	LRRC1	NM_018214	1.172	-5.697	-0.548923173	-0.367040983	-0.5135225
25794	JMJD1A	NM_018433	1.037	-5.692	-0.462893907	-0.404190307	-0.50302585
9255	IL6	NM_000600	1.238	-5.692	-0.63877709	-0.950256223	-0.912880973
1194	LOXL2	NM_002318	1.189	-5.689	-0.50370099	-0.501074107	-0.719080057
2546	ZNF143	NM_003442	0.952	-5.679	-0.452416983	-0.34543331	-0.45091503
35567	GDF15	NM_004864	1.024	-5.662	-0.37247048	-0.651989723	-0.55763635
16250	IL27RA	NM_004843	1.230	-5.657	-0.638516037	-0.592530947	-1.0635951
28592	NFKBIZ	NM_031419	0.838	-5.656	-0.399276763	-0.680252967	-0.5927032
26609	BTBD14B	NM_052876	0.867	-5.647	-0.33442763	-0.337814103	-0.399638613
14164	TLL1	NM_012464	0.901	-5.647	-1.036844167	-1.574607867	-2.007416733
23903	ZNF135	AL157426	1.718	-5.628	-1.082710887	-1.1724996	-0.55194276
40233	SLC6A20	NM_020208	1.656	-5.619	-0.7845902	-0.817978057	-1.318159147
26973	SARDH	AF095737	1.012	-5.616	-0.5262521	-0.46970009	-0.36188557
17701	HFE	NM_139006	1.478	-5.604	-1.057104637	-1.2436246	-0.927780383
428	TNFSF10	NM_003810	1.179	-5.598	-0.552905147	-1.6183674	-1.3932727
428	TNFSF10	NM_003810	1.114	-5.589	-0.54777098	-1.528908167	-1.354601533
4685	BQ933774	BQ933774	0.953	-5.587	-0.498210983	-0.913987767	-1.2186037
24261	BRP44L	NM_016098	0.915	-5.583	-0.34748953	-0.51741003	-0.563602347
26184	SPON2	NM_012445	1.607	-5.583	-0.74982117	-0.55882658	-0.709754973
11848	USP18	NM_017414	0.947	-5.583	-0.350507593	-0.577703833	-0.553424653
32079	SMURF1	NM_020429	1.302	-5.536	-0.80861452	-0.492267317	-1.085724233
3101	DCT	NM_001922	1.015	-5.529	-0.45967827	-0.372466253	-0.431676983
8850	AK021467	AK021467	2.645	-5.521	-1.370444067	-0.99604464	-1.832363467
23403	LOC283152	NM_001033658	1.089	-5.511	-0.420166723	-0.402091063	-0.514176857
40790	CCNL2	NM_001039577	0.967	-5.503	-0.45281468	-0.491834283	-0.334235203
12222	CXCL11	NM_005409	1.280	-5.496	-0.697059053	-2.217573367	-1.5537052
32449	ANGPT2	NM_001147	1.955	-5.489	-1.326668333	-0.967197253	-1.18942161
4467	PDGFRB	NM_002609	1.288	-5.480	-1.2410021	-0.923295633	-1.286995283
29897	ARP11	AB039791	1.141	-5.452	-0.575639843	-0.41502278	-0.807079533
511	AK027383	AK027383	1.140	-5.442	-0.58401715	-0.4991183	-0.9423341
31274	C19orf31	NM_001014373	1.561	-5.431	-1.0191593	-0.823552207	-1.545054533
428	TNFSF10	NM_003810	1.045	-5.426	-0.537792677	-1.568648	-1.273448467
2693	C9orf21	NM_153698	0.958	-5.405	-0.43896087	-0.411266743	-0.63277822

428	TNFSF10	NM_003810	1.110	-5.382	-0.51798715	-1.600457933	-1.337337867
36450	THC2669063	THC2669063	0.640	-5.381	-0.455537077	-0.70895041	-0.637471933
39568	SPEG	AK055387	1.371	-5.331	-0.81259646	-1.110026933	-2.1280302
428	TNFSF10	NM_003810	0.961	-5.329	-0.486004013	-1.461768667	-1.238813267
7848	THC2589620	THC2589620	1.560	-5.328	-0.861425253	-1.048731	-0.445495297
428	TNFSF10	NM_003810	1.104	-5.324	-0.511457367	-1.577113467	-1.3176588
19075	A_24_P630916	A_24_P630916	1.136	-5.321	-0.501575477	-0.53755618	-0.317769267
4985	PDXK	NM_003681	0.844	-5.320	-0.312371203	-0.47346731	-0.49974734
1326	KCNK6	NM_004823	0.889	-5.319	-0.350867107	-0.673042287	-0.53419706
428	TNFSF10	NM_003810	1.046	-5.311	-0.479290667	-1.448819733	-1.302678033
6588	CACNA1B	ENST00000371372	1.323	-5.302	-1.1743241	-0.8529725	-1.170360383
5285	APOL6	NM_030641	2.417	-5.297	-1.754723567	-2.460166133	-0.707609757
30362	CXCL11	NM_005409	1.212	-5.282	-0.594433043	-2.067847733	-1.5501496
17787	TMEM173	NM_198282	0.809	-5.280	-0.3642342	-0.600631587	-0.75335053
28635	DMKN	NM_033317	2.256	-5.270	-1.3613215	-1.009538667	-1.183411727
15265	PLA2G3	NM_015715	1.226	-5.268	-0.698538867	-0.885502533	-1.711716633
34693	CLN3	BC111068	0.902	-5.266	-0.312308133	-0.35800294	-0.378894863
428	TNFSF10	NM_003810	1.001	-5.265	-0.492812863	-1.536260633	-1.2612175
4413	THC2590490	THC2590490	1.410	-5.261	-0.693449963	-0.945177547	-0.453792683
15932	PDLIM2	AK074031	1.162	-5.260	-0.883031367	-0.45855397	-0.739803267
15830	RSAD2	NM_080657	1.060	-5.259	-0.522487177	-1.648276867	-1.705192467
33137	SHC3	NM_016848	1.650	-5.250	-1.080824567	-1.643162367	-0.638431073
5497	HRASLS5	BC034222	0.821	-5.246	-0.449225453	-0.39539804	-0.434927643
4940	ACLY	U18197	0.945	-5.225	-0.378183017	-0.541948567	-0.779965267
3882	COL13A1	NM_005203	1.179	-5.220	-0.8208332	-0.698674687	-1.455585967
24047	TMEM45A	NM_018004	1.164	-5.216	-0.5148523	-0.9053135	-1.3825128
29759	LGALS1	NM_002305	1.306	-5.209	-0.652095797	-0.4727589	-1.058351