Award Number: W81XWH-11-1-0337

TITLE: Discovery of Novel Gene Elements Associated with Prostate Cancer Progression

PRINCIPAL INVESTIGATOR: Arul M. Chinnaiyan, M.D., Ph.D.

CONTRACTING ORGANIZATION: Regents of the University of Michigan
Ann Arbor, MI 48109

REPORT DATE: December 2014

TYPE OF REPORT: FINAL

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.
Discovery of Novel Gene Elements Associated With Prostate Cancer Progression

Arul Chinnaiyan

E-Mail: chinnaiyangrants@umich.edu

Regents of the University of Michigan
Ann Arbor, Michigan 48109

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

Approved for Public Release; Distribution Unlimited

Long non-coding RNAs (ncRNAs) have gained recent attention as potentially important players in cancer biology. This proposal employs innovative next generation sequencing and ab initio transcriptome assembly methodologies to discover novel ncRNAs and incorporate them into expression signatures that stratify indolent versus aggressive prostate cancer. In this reporting period, we describe the identification and characterization of a prostate cancer-specific ncRNA named PCAT-1. Here, we demonstrated that PCAT-1 is a regulator of cell proliferation and is a target of the Polycomb Repressive Complex 2 (PRC2). We further found that patterns of PCAT-1 and PRC2 expression stratified patient tissues into molecular subtypes distinguished by expression signatures of PCAT-1-repressed target genes. PCAT-1 may potentially serve as a novel biomarker of aggressive disease that can be developed for clinical diagnostic and prognostic utility.
# Table of Contents

1. Introduction ........................................................................... 1

2. Keywords .............................................................................. 1

3. Overall Project Summary ..................................................... 2

4. Key Research Accomplishments ........................................... 8

5. Conclusion ............................................................................ 8


7. Inventions, Patents and Licenses ........................................... 10

8. Reportable Outcomes .......................................................... 10

9. Other Achievements ............................................................ 10

10. References .......................................................................... 10

11. Appendices ......................................................................... 12
1. INTRODUCTION:

Noncoding RNAs (ncRNAs) are emerging as key players in human cancer, with the potential to serve as novel markers of disease and to reveal uncharacterized aspects of tumor biology. ncRNAs are often cell-type specific, have biologically important roles (1, 2) and may interact with known cancer genes such as EZH2 (3). Indeed, several well-described examples, such as HOTAIR (3, 4) and ANRIL (5, 6), indicate that ncRNAs may have essential roles in cancer biology, typically facilitating epigenetic gene repression via chromatin modifying complexes (7, 8). Moreover, ncRNA expression may be correlated with specific clinical phenotype that can be predictive of patient outcomes and thus may have utility in diagnostic tests (4, 9). The characterization of RNA species, their functions, and their clinical applicability is therefore a major area of biological and clinical importance.

Earlier, we reported the discovery of cancer-associated ncRNA transcripts (PCATs) that were identified from a cohort of 102 prostate tissues and cells lines. One ncRNA, PCAT-1, was characterized as a prostate-specific regulator of cell proliferation and we showed that it is a target of the Polycomb Repressive Complex 2 (PRC2). Further, patterns of PCAT-1 and PRC2 expression stratified patient tissues into molecular subtypes distinguished by expression signatures of PCAT-1-repressed target genes (12).

We also identified a long ncRNA spanning ~250kb in a chromosome 2 gene desert that we have designated “Second Chromosome Locus Associated with Prostate (SChLAP1)”. SChLAP1 is highly over-expressed in a subset of prostate cancer (about 20%) where it appears to have an effect on invasiveness and contributes to the development of lethal cancer in part by antagonizing the tumor-suppressive functions of the SWI/SNF complex (14).

Most recently, we demonstrated that PCAT-1 represses the BRCA2 tumor suppressor gene leading to downstream impairment of homologous recombination and PCAT-1-expressing cells exhibit sensitization to PARP1 inhibitors (15).

Over the funding period of this award, we made tremendous progress towards the aim of identifying novel gene elements involved with prostate cancer progress. We discovered and functionally characterized two prostate cancer-specific long non-coding RNAs, PCAT-1 in in vitro and in vivo studies as well as elucidating their molecular mechanism by which they function. We are currently testing their potential to serve as clinical biomarkers for diagnostic/prognostic utility.

2. KEYWORDS

Unannotated RNAs, non-coding RNA (ncRNA), next-generation sequencing, prostate cancer biomarkers, tumor biology.
3. OVERALL PROJECT SUMMARY:

To employ next generation sequencing to comprehensively annotate expressed regions in the prostate cancer transcriptome:

1) lncRNA PCAT-1 implicated in prostate cancer

We applied high-throughput transcriptome sequencing (RNA-Seq) on a large cohort of >100 prostate cancer samples in order to define the RNA landscape in this disease. We found that nearly 20% of polyadenylated RNA species represent unannotated transcripts, including 121 intergenic, unannotated lncRNAs whose aberrant expression in prostate cancer ranked highly as some of the most differentially-expressed RNAs in this disease (Figure 1). These 121 unannotated transcripts were ranked and named as Prostate Cancer Associated Transcripts (PCATs) as defined by their fold change in localized tumor relative to benign tissue. We validated multiple unannotated transcripts; qPCR for four transcripts (PCAT-114, PCAT-14, PCAT-43, PCAT-1) on two independent cohorts of prostate tissues confirmed predicted cancer-specific expression patterns (data for PCAT-1 is shown in Figure 1B).

Figure 1: 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. PCAT-1. qPCR analysis was performed by normalizing to GAPDH and the median expression of the benign samples. *Nat Biotechnol*. 2011;29(8):742-9.
PCAT-1 was strikingly upregulated in a subset of metastatic and high-grade localized (Gleason score ≥7) cancers. To examine 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 2, top panel). Next, we designed siRNA oligos to PCAT-1 and performed knockdown experiments in LNCaP cells that express higher levels of PCAT-1. 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 2, bottom panel).

We hypothesized that PCAT-1 may have coordinated expression with the oncoprotein EZH2, a core PRC2 protein that is also upregulated in solid tumors and contributes to a metastatic phenotype (10, 11). Surprisingly, we found that PCAT-1 and EZH2 expression were nearly mutually exclusive (12). This suggests that outlier PCAT-1 and EZH2 expression may define two subsets of high-grade disease. These findings represent the first comprehensive study of lincRNAs in prostate cancer, provide a computational framework for large-scale RNA-Seq analyses, and describe PCAT-1 as a novel prostate cancer ncRNA functionally implicated in disease progression.

2. **SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex**

Among the 121 PCATs whose expression discriminated benign tissues, localized prostate cancer, and metastatic prostate disease (Figure 1A, 3A), we noted that a large region of transcription on Chromosome 2 contained the two most prominent novel outliers in prostate cancer (Figure 3B). Using in vitro cell line models, we characterized SChLAP1 (Second Chromosome Locus Associated with Prostate-1) as a novel seven-exon gene (Figure 3C). We found that SChLAP1 was upregulated in approximately 20% of localized and metastatic prostate cancers (Figure 3D).

To elucidate the function of SChLAP1 in prostate cancer, we performed siRNA knockdowns of this gene. We found that knockdown of SChLAP1 profoundly inhibited prostate cell line invasiveness in vitro using Boyden Chamber assays (Figure 4A). Furthermore, overexpression of SChLAP1 isoforms in RWPE benign immortalized prostate cells dramatically increased the
invasiveness of these cells in vitro (Figure 4B). SchLAP1 overexpression did not, however, change the cell proliferation rate for these cells (data not shown).

We next investigated the ability for SchLAP1 to coordinate invasion and metastasis in vivo. We used intracardiac injection of 22Rv1 prostate cancer cells with stable knockdown of SchLAP1 via shRNAs. We found that SchLAP1 knockdown impaired the ability for 22Rv1 to seed distant tissues and metastasize in mice (Figure 4C, D). This was due to impaired cell seeding and not decreased tumor growth kinetics, and Ki67 staining showed no change in cellular proliferation rate following SchLAP1 knockdown (data not shown).

We next used the chick chorioallantoic membrane (CAM) assay to further dissect the role of SchLAP1 in disease aggressiveness. We found that SchLAP1 knockdown in 22Rv1 cells impaired cellular invasion (Figure 4E), intravasation (Figure 4F), and metastasis to distant organs (Figure 4G). SchLAP1 knockdown further impaired tumor growth (Figure 4H). Importantly, overexpression of SchLAP1 in RWPE cells partially recapitulated these phenotypes and led to an increase in cellular intravasation in the CAM assay (Figure 4I). Taken together, our data demonstrate a strong role for SchLAP1 in mediating aggressive features of prostate cancer biology.

To evaluate the relationship of SchLAP1 to prostate cancer patient outcomes, we utilized a cohort of 235 high-risk patients from the Mayo Clinic who underwent radical prostatectomy between 2000-2006. These patients were defined by a pre-operative PSA >20ng/ml, Gleason score of 8 – 10, or found to have seminal vesicle involvement in their disease. We measured SchLAP1 expression in these patients and assessed the prognostic role of SchLAP1 to predict biochemical recurrence (BCR), clinical progression to metastatic disease post-prostatectomy (CP), and prostate-cancer-specific mortality (PCSM). At the time of analysis, patients had a mean 8.1 years follow-up.
Using Kaplan-Meier analyses, we found that the SChLAP1 expression was a powerful single-gene predictor of aggressive prostate cancer (Figure 5A-C). SChLAP-1 expression was highly significant when distinguishing CP and PCSM (p = 0.00005 and p = 0.002, respectively); patients with high SChLAP-1 expression had 5-year CP- and PCSM-free survival of only 65% and 85%, respectively, compared to 85% and 95% for patients with low SChLAP-1 expression (Figure 5B,C). For the BCR endpoint, high SChLAP-1 expression in patient primary tumor specimens was associated with a rapid median time-to-progression (1.9 vs 5.5 years for SChLAP-1 high and low patients, respectively) (Figure 5A). These data suggest that SChLAP-1 expression retains its prognostic utility for defining a subgroup of patients more likely to...
experience BCR, CP, and PCSM even in high-risk patients, where most individuals experienced disease recurrence within ten years post-prostatectomy.

Moreover, we characterized an antagonistic SChLAP1-SWI/SNF axis in which SChLAP1 impairs SNF5-mediated gene expression regulation and genomic binding (data not shown). SChLAP1 is the first lncRNA, to our knowledge, that impairs a major epigenetic complex with well-documented tumor suppressor function.

Figure 5: SChLAP1 expression characterizes aggressive prostate cancer. Kaplan-Meier analyses of prostate cancer outcomes in the Mayo Clinic cohort. SChLAP1 expression was measured using Affymetrix exon arrays and patients were stratified according to their SChLAP1 expression. Patient outcomes were analyzed for biochemical recurrence (A), clinical progression to systemic disease (B), and prostate cancer-specific mortality (C). The shaded regions represent the 95% confidence interval. Nat Genet. 2013 Nov;45(11):1392-8.

3. **PCAT-1 regulates BRCA2 and Controls Homologous Recombination in Cancer**

Recent studies have identified lncRNAs that are induced by genotoxic stress as well as being involved in the repair of DNA damage; however, the role of lncRNAs in the regulation of double strand break (DSB) repair remains unclear. Earlier we observed that PCAT-1 expression was a prostate cancer outlier associated with low levels of BRCA2. We therefore hypothesized that PCAT-1 mediated the repression of BRCA2, and thus PCAT-1 may be involved in the dysregulation of homologous recombination (HR) upon genotoxic stress.

We found an inverse relationship between PCAT-1 and BRCA2 in two independent cohorts of human prostate cancer samples. Using 58 prostate cancer tissues and 20 prostate cancer xenografts derived from human specimens, we found that increasing PCAT-1 expression correlated with decreased BRCA2 expression. We observed that PCAT-1 overexpression led to
Chinnaiyan, Arul M.
DoD: PC100171:
an increase in γ-H2AX foci
(a marker of DSBs that have
not been repaired) under
genotoxic stress conditions.

We next tested whether we
investigated whether PARP1
inhibition selectively kills
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 increased cell death in

response to PARP inhibition
(Figure 6A, left). This change
in cell sensitivity to PARP1
inhibitors was striking, with a
five-fold change in the IC50
for LNCaP and Du145 cells
(Figure 6A, right).

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
SCID mice, consistent with
our previous findings that
PCAT-1 accelerates prostate
cell proliferation in vitro
(Figure 6C). Importantly,
Du145 xenografts retained

both PCAT-1 expression and BRCA2 repression (Figure 6D).

Together, these data suggest that PCAT-1 expression antagonizes BRCA2 expression and PCAT-1 is mechanistically linked to increased prostate cell sensitivity to PARP1 inhibitors via its repression of BRCA2 both in vitro and in vivo. This report is the first to demonstrate a role for lncRNAs in the regulation of DSBs in prostate cancer and suggests a new mechanistic basis for impaired HR in this disease (Cancer Res. 2014 Mar 15;74(6):1651-60).
4. **KEY RESEARCH ACCOMPLISHMENTS:**

- We employed transcriptome sequencing on a cohort of 102 prostate tissues and cell lines and found 121 unannotated intergenic ncRNAs as Prostate Cancer Associated Transcripts (PCATs).
- We observed that several PCATs, including PCAT-1, was upregulated markedly in a set of aggressive prostate cancers and
- We found that PCAT-1 upregulation marked a set of aggressive prostate cancers and stratified against EZH2, a prostate cancer oncogene also upregulated in a set of aggressive cancers.
- *In vitro*, PCAT-1 expression was necessary for cancer cell proliferation and overexpression of PCAT-1 was sufficient to increase benign prostate cell proliferation.
- We characterized another long noncoding RNA termed SChLAP1 that is overexpressed in a subset of prostate cancers.
- SChLAP1 levels independently predicted poor outcomes, including metastasis and prostate cancer-specific mortality.
- *In vitro* and *in vivo* gain-of-function and loss-of-function experiments indicated that SChLAP1 is critical for cancer cell invasiveness and metastasis.
- Mechanistically, SChLAP1 antagonizes the genome-wide localization and regulatory functions of the SWI/SNF chromatin-modifying complex.
- The lncRNA PCAT-1 represses BRCA2 in *in vitro* and *in vivo* studies and PCAT-1 may serve as a predictive biomarker for patient response to PARP1 inhibitor therapy.

5. **CONCLUSION:**

Clinically, ncRNAs are proving to be powerful biomarkers for cancer. ncRNA transcripts can be detected non-invasively in prostate cancer patient fluids such as serum or urine. Previous studies have uncovered PCA3 (prostate cancer antigen-3), a prostate-specific ncRNA upregulated in prostate cancer, as a biomarker for prostate cancer diagnosis (13). Clinical-grade assays for urinary detection of PCA3 are currently available for physicians, and research has shown that adding PCA3 to the standard serum PSA test improves prostate cancer prediction. Future studies exploring the biology of ncRNAs and their role in cancer could lead to the development of additional prostate cancer biomarkers, such as PCAT-1 and/or SChLAP1, for the clinical detection and stratification of aggressive prostate cancers. Further, we have also uncovered the molecular mechanisms by which PCAT-1 and SChLAP1 function that could inform clinical response and outcomes.
6. PUBLICATIONS, ABSTRACTS AND PRESENTATIONS:

PUBLICATIONS:


PRESENTATIONS:


ABSTRACTS:

The Emergence of long non-coding RNAs in Cancer

The clinical heterogeneity of prostate cancer, in which only a subset of patients has aggressive cancer whereas others have indolent disease, remains incompletely understood. Here, we characterize a novel lncRNA termed SChLAP1 (Second Chromosome Locus Associated with Prostate-1) overexpressed in a subset of prostate cancers. SChLAP1 levels independently predicted for poor patient outcomes, including metastasis and prostate cancer specific mortality. In vitro and in vivo gain-of-function and loss-of-function experiments confirmed that SChLAP1 is critical for cancer cell invasiveness and metastasis. Mechanistically, SChLAP1 antagonized the genome-wide localization and regulatory functions of the SWI/SNF chromatin-modifying complex. These results suggest that SChLAP1 is a lncRNA that contributes to the development of lethal cancer at least in part by antagonizing tumor-suppressive functions of the SWI/SNF complex.
7. INVENTIONS, PATENTS, LICENSES:

Innovator: Arul Chinnaiyan, Matthew Iyer and Yashar Niknafs
Invention Title: The Landscape of Long Non-Coding RNAs in Cancer
Filing Date: 6/17/2014
Patent Application Number: 6287

8. REPORTABLE OUTCOMES:
N/A

9. OTHER ACHIEVEMENTS
N/A

10. REFERENCES:


LIST OF PERSONNEL:

Arul M. Chinnaiyan
Qi Cao
Matthew Iyer
Shanker Kalyana-Sundaram
Daniel Miller
Mary Pierce-Burlingame
Yuanyuan Qiao
Yi-Mi Wu
Priority Report

**PCAT-1, a Long Noncoding RNA, Regulates BRCA2 and Controls Homologous Recombination in Cancer**

John R. Prensner$^{1,2}$, Wei Chen$^3$, Matthew K. Iyer$^1$, Qi Cao$^{1,2}$, Teng Ma$^4$, Sumin Han$^3$, Anirban Sahu$^1$, Rohit Malik$^1$, Kari Wilder-Romans$^3$, Nora Navone$^6$, Christopher J. Logothetis$^9$, John C. Araujo$^9$, Louis L. Pisters$^3$, Ashutosh K. Tewari$^{10}$, Christine E. Canman$^5$, Karen E. Knudsen$^{12}$, Naoki Kitabayashi$^{11}$, Mark A. Rubin$^{11}$, Francesca Demichelis$^{11,13}$, Theodore S. Lawrence$^9$, Arul M. Chinnaiyan$^{1,3,6,7}$, and Felix Y. Feng$^{1,3,7}$

Abstract

Impairment of double-stranded DNA break (DSB) repair is essential to many cancers. However, although mutations in DSB repair proteins are common in hereditary cancers, mechanisms of impaired DSB repair in sporadic cancers remain incompletely understood. Here, we describe the first role for a long noncoding RNA (lncRNA) in DSB repair in prostate cancer. We identify PCAT-1, a prostate cancer outlier lncRNA, which regulates cell response to genotoxic stress. PCAT-1 expression produces a functional deficiency in homologous recombination through its repression of the BRCA2 tumor suppressor, which, in turn, imparts a high sensitivity to small-molecule inhibitors of PARP1. These effects reflect a postranscriptional repression of the BRCA2 3’UTR by PCAT-1. Our observations thus offer a novel mechanism of “BRCAness” in sporadic cancers. Cancer Res; 74(6); 1651–60. © 2014 AACR.

Introduction

The uncontrolled accumulation of double-stranded DNA breaks (DSB) represents a putative Achilles heel for cancer cells, because these lesions are toxic and their repair requires regulation of disrupted genetic material (1–3). Several mechanisms, such as nonhomologous 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). In particular, the lethality of excess DSBs has been exploited for the therapeutic treatment of hereditary breast and ovarian cancers harboring BRCA1/2 mutations, which leads to defective HR and increased DSBs (5). 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 (6–8).

Recently, long noncoding RNAs (lncRNA) have emerged as new layer of cell biology (9), contributing to diverse biologic processes. In cancer, aberrant expression of lncRNAs is associated with cancer progression (9, 10), and overexpression of oncogenic lncRNAs can promote tumor cell proliferation and metastasis through transcriptional regulation of target genes (11–13). Recent studies have also identified lncRNAs induced by genotoxic stress as well as involved in the repair of DNA damage (14, 15); however, the role of lncRNAs in the regulation of DSB repair remains unclear.

Here, we report the characterization of PCAT-1 as a prostate cancer lncRNA implicated in the regulation of DSB repair. We find that PCAT-1 represses the BRCA2 tumor suppressor gene, leading to downstream impairment of HR. Importantly, PCAT-1-expressing cells exhibit a BRCA-like phenotype, resulting in cell sensitization to PARP1 inhibitors. In human prostate cancer tissues, high PCAT-1 expression predicts for low BRCA2 expression, supporting our observations in model systems. To our knowledge, this report is the first to demonstrate a role for...
lncRNAs in the regulation of DSBs in prostate cancer and suggests a new mechanistic basis for impaired HR in this disease.

Materials and Methods

For full details on methodology, please refer to the Supplementary Information for a complete Materials and Methods section.

Patient samples

For the University of Michigan patient samples, prostate tissues were obtained from the radical prostatectomy series and Rapid Autopsy Program at the University of Michigan tissue core. These programs are part of the University of Michigan Prostate Cancer Specialized Program of Research Excellence (SPORE). 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 IRB # 1994-0481]. For the Weill Cornell Medical College patient samples, prostate tissues were collected as part of an IRB-approved protocol at Weill Cornell Medical College (New York, NY).

Cell lines

All cell lines were obtained from the American Type Culture Collection (ATCC). Cell lines were maintained using standard media and conditions. Du145-derived cell lines were maintained in Dulbecco's Modified Eagle Medium supplemented with 10% FBS (Invitrogen) and 1% penicillin-streptomycin (Invitrogen) supplemented with bovine pituitary extract, EGF, and 1% penicillin-streptomycin in a 5% CO2 cell culture incubator. RWPE-derived cell lines were maintained in keratinocyte serum-free (Invitrogen) supplemented with 10% FBS (Invitrogen) and 1% penicillin-streptomycin in a 5% CO2 cell culture incubator. LNCAP-derived and PC3-derived cell lines were maintained in RPMI-1640 (Invitrogen) supplemented with 10% FBS and 1% penicillin-streptomycin in a 5% CO2 cell culture incubator. PC3 cells containing the GFP HR assay construct were generated as described previously (16, 17).

**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 transduced into 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 **PCAT-1** or nontargeting short hairpin RNA (shRNA) lentiviral constructs for 48 hours. GFP+ cells were drug-selected using 1 μg/mL puromycin. **PCAT-1** shRNAs were custom generated by Systems Biosciences using the following sequences: shRNA 1 GCAGAAACACCAAUGGAUAUU; shRNA 2 AUACAUAAGACCAUGGAAU.

To ensure cell identity, all cell lines were used for less than 6 months after resuscitation and confirmed by genotyping after resuscitation. DNA samples were diluted to 0.10 ng/μL and ten genotyping loci (D3S1358, D5S818, D7S820, D8S1179, D13S317, D18S51, D21S11, FGA, vWA, and the Amelogenin locus) were analyzed by the University of Michigan DNA Sequencing Core using the Profiler Plus PCR Amplification Kit (Applied Biosystems).

Cell line assays

LNCaP, Du145, PC3, and RWPE cell lines were obtained from the ATCC and maintained in standard conditions. Stable overexpression and knockdown cell lines were generated with lentiviral constructs with blasticidin or puromycin selection as appropriate. RNA isolation and cDNA synthesis were performed according to standard protocols. Quantitative PCR was performed with Power SYBR Green Mastermix on an Applied Biosystems 7900HT Real-Time PCR system. Chemosensitivity assays were performed on 5,000 cells plated per well in 96-well plates and treated with a single dose of olaparib or ABT-888 as indicated for 72 hours. WST assays (Roche) were performed according to the manufacturer’s instructions. Immunofluorescence experiments were performed with 1 × 105 cells in 12-well plates according to standard protocols; RAD51 and γ-H2AX staining was performed 6 hours or 24 hours after treatment, respectively.

Luciferase assays

The indicated cell lines were transfected with full-length **BRCA2** luciferase constructs as well as pRL-TK vector as internal control for luciferase activity. After 2 days of incubation, the cells were lysed and luciferase assays conducted using the dual luciferase assay system (Promega). Each experiment was performed in quadruplicate.

Immunoblot analysis

Cells were lysed in radioimmunoprecipitation assay lysis buffer (Sigma) and briefly sonicated for homogenization. Aliquots of each protein extract were boiled in sample buffer, size fractionated by SDS-PAGE at 4°C, and transferred onto polyvinylidene difluoride membrane (GE Healthcare). The membrane was then incubated at room temperature for 1 to 2 hours in blocking buffer [Tris-buffered saline, 0.1% Tween (TBS-T), 5% nonfat dry milk] and incubated at 4°C 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), **BRCA1** (Cell Signaling Technology, #9025S), **XRCC1** (Abcam, ab1838), **XRCC3** (Abcam, ab97390), **XRCC4** (GeneTex, GTX83406), **Ku70** (BD Biosciences, #611892), **Ku80** (Cell Signaling Technology, #2180S), γ-H2AX (Cell Signaling Technology, #9718) and β-actin (Sigma, A5441).
For immunoblot densitometry, the densitometric scan of the immunoblots was performed using ImageJ. Three replicate experiments were quantified for the final analysis.

**Xenograft assays**

Xenograft experiments were performed according to University of Michigan-approved protocols and conform to their relevant regulatory standards. Five-week-old male severe combined immunodeficient (SCID) mice (CB.17, SCID) were purchased from Charles River, Inc. (Charles River Laboratory). A total of 1 × 10^6 Du145-control or Du145-PCAT-I stable cells were resuspended in 100 μL of saline with 50% Matrigel (BD Biosciences) 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, i.p.) and ketamine (10 mg/kg, i.p.) for chemical restraint before tumor implantation. All tumors were staged for 2 weeks before starting the drug treatment. At the beginning of the third week, mice with tumors (10 tumors/treatment group, average size 150–200 mm³) were treated with olaparib (100 mg/kg, i.p. twice daily for 2 weeks) or radiation. As predicted, overexpression of PCAT-1 impaired HR, genotoxic stress of BRCA2 and upregulation of BRCA2 expression (Fig. 1A, right). To ensure that these observations were not restricted to cell line-based studies, we further confirmed an inverse relationship between PCAT-1 and BRCA2 in two independent cohorts of human prostate cancer samples. Using 58 prostate cancer tissues and 20 prostate cancer xenografts derived from human specimens, we found that increasing PCAT-1 expression correlated with decreased BRCA2 expression (Fig. 1B and Supplementary Fig. S1A). Together, these data suggest that PCAT-1 expression antagonizes BRCA2 expression.

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). Accordingly, treatment of our isogenic cell lines with either a PARP1 inhibitor (olaparib or ABT-888) 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 (18). Specifically, PCAT-1 overexpression decreased RAD51 foci formation after therapy and PCAT-1 knockdown increased RAD51 foci formation after therapy in prostate cells (Fig. 1C and Supplementary Fig. S1B–S1D). We further used a well-characterized HR assay, in which cells employ HR to recombine an I-SceI-cut plasmid to produce GFP signaling (16), to evaluate the function of PCAT-1 on HR directly. We found that transient overexpression of PCAT-1 in PC3 prostate cancer cells resulted in a significant inhibition of GFP signaling following I-SceI–induced HR in addition to decreased RAD51 foci (Fig. 1D and Supplementary Fig. S2A–S2D). Of note, PCAT-1 expression does not show substantial change following induction of DNA damage via radiation (Supplementary Fig. S2F).

**PCAT-1 expression impairs DNA damage repair**

Because PCAT-1 impairs HR, genotoxic stress of PCAT-1–expressing cells should lead to an accumulation of DSBs, which can be visualized using γ-H2AX foci, a marker of DSBs 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 γ-H2AX foci under stress conditions (Fig. 2A and B), indicating that PCAT-1 impairs DSB repair in these cells. Similarly, LNCaP cells with PCAT-I knockout displayed decreased levels of γ-H2AX foci (Fig. 2A and B). Immunoblot analysis of γ-H2AX protein abundance in these cells following genotoxic stress confirmed a downregulation of γ-H2AX with knockdown of PCAT-1 and upregulation of γ-H2AX with overexpression of PCAT-1 (Supplementary Fig. S3).

Finally, we also evaluated the ability for 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 (Supplementary Fig. S4). To exclude a regulatory relationship between PCAT-1 and other major actors in DNA damage, we performed analysis of XRCC1 (base excision repair pathway), XRCC3 (HR), XRCC4 (NHEJ), Ku70 (NHEJ), Ku80 (NHEJ), and BRCA1 (multiple pathways) in our in vitro models, which showed no change in protein abundance upon modulation of PCAT-1 (Supplementary Fig. S5A). Together, these data indicate that PCAT-1 expression may impart cell sensitivity to genotoxic stress by decreasing the HR response through downregulation of BRCA2.

**PCAT-1 expression leads to increased cell death following genotoxic stress**

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 increased cell death in response to PARP inhibition (Fig. 3A, left and Supplementary Fig. S5B–S5D). This change in cell sensitivity to PARP1 inhibitors was striking, with a five-fold change in the IC_{50} for LNCaP and Du145 cells (Fig. 3A, right and Supplementary Fig. S6). Similar results were observed in RWPE cells overexpressing PCAT-1 (Supplementary Fig. S7).

To ensure that these effects were dependent on BRCA2, we undertook rescue experiments by performing knockdown of BRCA2 in LNCaP shPCAT-1 cells (which have increased levels of BRCA2). These experiments demonstrated a corresponding increase in the sensitivity of these cells to PARP1 inhibition in a
dose-dependent manner according to the efficiency of the BRCA2 knockdown (Fig. 3B). We further observed reduced RAD51 foci after treatment following BRCA2 knockdown in LNCaP shPCAT-1 cells as well (Supplementary Fig. S8). To exclude a role for altered cell-cycle distributions in these phenotypes, we performed flow cytometry, which demonstrated no change in cell cycle in our model systems (Supplementary Fig. S9).

**PCAT-1 expression leads to decreased in vivo tumor growth following PARP inhibition**

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 SCID mice, consistent with our previous findings that PCAT-1 accelerates prostate cell proliferation in vitro (Fig. 3C; ref. 10). Moreover, Du145-PCAT-1 xenografts showed marked susceptibility and tumor regression following intraperitoneal 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 (19), is small (Fig. 3C).

Mice in all groups of treatment maintained their body weights and showed no evidence of weight loss (Supplementary Fig. S10A).

Importantly, Du145 xenografts retained both PCAT-1 expression and BRCA2 repression (Fig. 3D). To investigate PCAT-1 signaling under control-treated (DMSO) and...
Figure 3. PCAT-1 expression results in prostate cell sensitivity to PARP inhibition in vitro and in vivo. A, left, LNCaP cells with PCAT-1 knockdown exhibit enhanced cell survival 72 hours after treatment with olaparib. Right, Du145 cells with PCAT-1 overexpression exhibit reduced cell survival 72 hours after treatment with olaparib. Cell survival is determined via WST assays. B, BRCA2 knockdown in LNCaP shPCAT-1 cells rescues cell sensitivity to olaparib. Right, Western blot showing efficiency of BRCA2 knockdown. C, tumor growth curves for Du145-control and Du145-PCAT-1 xenografts following initiation of treatment with DMSO control or 25 μmol/L olaparib. Tumor volumes are normalized to 100, and time = 0 represents the start of treatment administration. Treatment was initiated 3 weeks after xenograft engraftment. D, expression level of PCAT-1 and BRCA2 protein in Du145-PCAT-1 xenografts. Error bars, ± SEM.
olaparib-treated conditions, we also observed in vivo upregulation of PCAT-1–induced target genes (TOP2A, E2F8, BIRC5, and KIF15; Supplementary Fig. S10B) defined by previous microarray profiling of LNCaP cells with PCAT-1 siRNAs and confirmed in RWPE-PCAT-1–overexpressing cells (Supplementary Fig. S10C; ref. 10). 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.

**PCAT-1 does not operate via traditional IncRNA-mediated mechanisms**

Although many lncRNAs are noted to regulate gene transcription through epigenetic mechanisms (11, 13, 20), we did not observe evidence for this possibility with PCAT-1. Although PCAT-1 regulated BRCA2 mRNA in vitro (Supplementary Fig. S11A), treatment of RWPE-LacZ and RWPE-PCAT-1 cells with the DNA methylation inhibitor 5-azacytidine (5-aza), the histone deacetylase inhibitor TSA, or briliidine (5-aza), the histone deacetylase inhibitor TSA, or bisulfite sequencing of the BRCA2 promoter in our isogenic LNCaP and RWPE model systems demonstrated minimal CpG island methylation in all cell lines (Supplementary Fig. S11C). These results suggest that epigenetic repression of BRCA2 is not the primary mechanism of PCAT-1. Moreover, IncRNAs containing Alu elements in their transcript sequence may utilize these repetitive sequences to regulate target gene mRNAs via STAU1-dependent degradation (21). 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 (Supplementary Fig. S11D).

**PCAT-1 regulates BRCA2 post-transcriptionally**

To determine whether PCAT-1 may function in a manner more analogous to microRNAs, which regulate mRNA levels post-transcriptionally (22), we generated a luciferase construct of the BRCA2 3′UTR, which is 902 bp in length (Fig. 4A). 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 (Fig. 4A). Supporting these data, we found that PCAT-1 was localized to the cell cytoplasm (Supplementary Fig. S12A) and overexpression of PCAT-1 in Du145 cells significantly reduced the stability of endogenous BRCA2 mRNA, consistent with a posttranscriptional mechanism (Supplementary Fig. S12B and S12C).

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 (Fig. 4B and Supplementary Fig. S13A). We generated these constructs to establish whether the 5′ end of PCAT-1, which contains portions of ancestral transposase and Alu repeat elements (Fig. 4B; ref. 10), or the 3′ end of PCAT-1, which consists of nonrepetitive 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 (Fig. 4B and C), and for this regulation, the first 250 bp of the PCAT-1 gene were required. In contrast, the 3′ end of PCAT-1 was expendable. Importantly, the 5′ end of PCAT-1 was similarly sufficient to sensitize RWPE cells to olaparib treatment in vitro (Fig. 4D). To rule out the possibility that RNA instability was responsible for the inactivity of the PCAT-1 constructs, we performed RNA stability assays, which demonstrated equivalent rates of RNA decay between full-length PCAT-1 and the inactive PCAT-1 deletion constructs in RWPE cells (Supplementary Fig. S13B). 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.

**Discussion**

To our knowledge, this is the first report of an lncRNA being involved in the DSB repair process in prostate cancer (Supplementary Fig. S14). These data are supported by a striking inverse correlation between PCAT-1 and BRCA2 expression in human prostate cancer samples. Our results expand the potential roles for lncRNAs in cancer biology and contrast strikingly with previous reports that lncRNAs operate epigenetically through chromatin-modifying complexes (23, 24). Indeed, epigenetic regulation likely represents only one of numerous mechanisms for IncRNA function (12, 21, 25, 26). Supporting this notion, we do not observe compelling evidence that PCAT-1 functions in an epigenetic manner, but rather it may exhibit posttranscriptional regulation of its target genes.

Importantly, PCAT-1 is also predominantly cytoplasmic, and thus our work describes the first cytoplasmic prostate IncRNA to be associated with therapeutic response. Cytoplasmic lncRNAs are also less well explored than their nuclear counterparts, and our work sheds light onto the complex mechanistic regulation of cellular processes via cytoplasmic lncRNAs. However, PCAT-1 does exhibit a smaller degree of nuclear expression (see Supplementary Fig. S12A), which may account for our previous observation that PCAT-1 may associate with the nuclear Polycomb Repressive Complex 2 (PRC2). Although our data directly support a role for PCAT-1 in the posttranscriptional regulation of BRCA2, we cannot fully exclude the possibility of additional regulation of BRCA2 at the transcriptional level at this time.

In addition, while the mechanism underlying PCAT-1 function remains incompletely understood, we were intrigued that the 5′ portion of the PCAT-1 RNA, which is comprised of fully unique sequences, was critical for its regulation of BRCA2 mRNA whereas the embedded Alu element was not. Although we did not identify a specific microRNA with high-confidence 7-mer complementary base pair matching to both this region of PCAT-1 and BRCA2 (data not shown), we speculate that alternative mechanisms of mRNA-like mismatch base pairing
may contribute to PCAT-1-mediated regulation in a manner similar to the recently described networks of competing endogenous RNAs (27).

Together, our data suggest that lncRNAs may have a more widespread role in mammalian genome maintenance and DNA repair than previously appreciated. In support of this, a role for small RNAs in human DNA damage repair in human cells has been recently reported and shown to be dependent upon the microRNA biogenesis machinery (28). Of note, Adamson and colleagues nominated the RNA-binding protein RBMX as a novel component of the HR pathway (16), 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 "BRCAness"—the clinical observation that many cancers lacking BRCA1/BRCA2 mutations exhibit the clinical features of impaired DSB repair (2, 29, 30). We hypothesize that other cancers with a BRCA-like phenotype may harbor 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.

Disclosure of Potential Conflicts of Interest
J.R. Prensner has ownership interest (including patents) in patent on PCAT1 and lncRNAs in prostate cancer. A.K. Tewari has honoraria from speakers' bureau from Intuitive Surgical, has ownership interest (including patents) in the US Patents US 6004267 A—Method for diagnosing and staging prostate cancer and US 8241310 B2—Urethral catheterless radical prostatectomy. A.M. Chinnaiyan has ownership interest (including patents) in Gen-Probe and Wafergen and is a consultant/advisory board member of Wafergen. No potential conflicts of interests were disclosed by the other authors.

Authors’ Contributions

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J.R. Prensner, W. Chen, Q. Cao, T. Ma, S. Han, A. Sahsu, R. Malik, K. Wilder-Romans, N. Navone, L.L. Pisters, A.K. Tewari, N. Kitabayashi, M.A. Rubin, F. Demichelis, F.Y. Feng


Writing, review, and/or revision of the manuscript: A.M. Chinnaiyan, F.Y. Feng

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): C.J. Logothetis, J.C. Araujo, A.K. Tewari, F.Y. Feng

Study supervision: A.M. Chinnaiyan, F.Y. Feng

Provided experimental design input regarding analyzing homologous recombination in cells: C.E. Canman

Acknowledgments

The authors thank Mats Ljungman, Saravana M. Dhanasekaran, Chad Brenner, Yi-Mi Wu, Daniel Robinson, Sameek Roychowdhury, and Dan Hamstra for helpful discussions. The authors also thank Benjamin Chandler for technical assistance.

References


The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex

John R Prensner1,10, Matthew K Iyer1,12,10, Anirban Sahu1,10, Irfan A Asangani1, Qi Cao1, Lalit Patel1,3, Ismael A Vergara4, Elai Davicioni4, Nicholas Erho4, Mercedeh Ghadessi4, Robert B Jenkins5, Timothy J Triche4, Rohit Malik1, Rachel Bedenis3, Natalie McGregor3, Teng Ma6, Wei Chen6, Sumin Han6, Xiaojun Jing1, Xuhong Cao1, Xiaoju Wang1, Benjamin Chandler1, Wei Yan1, Javed Siddiqui1, Lakshmi P Kunju1,7,8, Saravana M Dhanasekaran1,7, Kenneth J Pienta1,3, Felix Y Feng1,6,8 & Arul M Chinnaiyan1,2,7–9

Prostate cancers remain indolent in the majority of individuals but behave aggressively in a minority1,2. The molecular basis for this clinical heterogeneity remains incompletely understood3–5. Here we characterize a long noncoding RNA termed SChLAP1 (second chromosome locus associated with prostate-1) that is overexpressed in a subset of prostate cancers. SChLAP1 levels independently predict poor outcomes, including metastasis and prostate cancer–specific mortality. In vitro and in vivo gain-of-function and loss-of-function experiments indicate that SChLAP1 is critical for cancer cell invasiveness and metastasis.

Mechanistically, SChLAP1 antagonizes the genome-wide localization and regulatory functions of the SWI/SNF chromatin-modifying complex. These results suggest that SChLAP1 contributes to the development of lethal cancer at least in part by antagonizing the tumor-suppressive functions of the SWI/SNF complex.

With over 200,000 new cases per year, prostate cancer will be diagnosed in 1 in 6 men in the United States during their lifetime, yet only 20% of individuals with prostate cancer have a high-risk cancer that represents potentially lethal disease1,2,4. Whereas mutational events in key genes characterize a subset of lethal prostate cancers3,5,6, the molecular basis for aggressive disease remains poorly understood.

Long noncoding RNAs (IncRNAs) are RNA species >200 bp in length that are frequently polyadenylated and associated with transcription by RNA polymerase II (ref. 7). IncRNA-mediated biology has been implicated in a wide variety of cellular processes, and, in cancer, IncRNAs are emerging as a prominent layer of transcriptional regulation, often by collaborating with epigenetic complexes7–10.

Here we hypothesized that prostate cancer aggressiveness was governed by uncharacterized IncRNAs and sought to identify IncRNAs associated with aggressive disease. We previously used RNA sequencing (RNA-seq) to describe 121 new IncRNA loci (out of >1,800) that were aberrantly expressed in prostate cancer tissues11. Because only a fraction of prostate cancers present with aggressive clinical features2, we performed cancer outlier profile analysis11 (COPA) to nominate intergenic IncRNAs selectively upregulated in a subset of cancers (Supplementary Table 1). We observed that only two, PCAT-109 and PCAT-114, which are both located in a ‘gene desert’ on chromosome 2q31.3 (Supplementary Fig. 1), had striking outlier profiles distinguishing them from the rest of the candidates11 (Fig. 1a).

Of these two IncRNAs, PCAT-114 was expressed at higher levels in prostate cell lines, and, in the PCAT-114 region, we defined a 1.4-kb polyadenylated gene composed of up to seven exons and spanning nearly 200 kb on chromosome 2q31.3 (Fig. 1b and Supplementary Fig. 2a). We named this gene second chromosome locus associated with prostate-1 (SChLAP1) after its genomic location. Published prostate cancer chromatin immunoprecipitation and sequencing (ChIP-seq) data12 confirmed that the transcriptional start site (TSS) of SChLAP1 was marked by trimethylation of histone H3 at lysine 4 (H3K4me3) and that its gene body harbored trimethylation of histone H3 at lysine 36 (H3K36me3). We observed that SChLAP1 splicing isoforms, which are located in a ‘gene desert’ on chromosome 2q31.3 (Supplementary Fig. 2b), were spliced with H3K4me3 and H3K36me3 modifications.

Using quantitative PCR (qPCR), we confirmed that SChLAP1 was highly expressed in ~25% of prostate cancers (Fig. 1c). SChLAP1 was found to be expressed more frequently in metastatic compared to localized prostate cancers, and its expression was associated with ETS gene fusions in this cohort but not with other molecular events (Supplementary Fig. 2c).

Supplementary Table 1. Supplementary Fig. 1. Supplementary Fig. 2a. Supplementary Fig. 2b.
An analysis of SCHLAP1 expression in localized tumors demonstrated a strong correlation with higher Gleason scores, a histopathological measure of aggressiveness (Supplementary Fig. 4c,d and Supplementary Table 2). Next, we performed a network analysis of prostate cancer microarray data in the Oncomine database (Online Methods and Supplementary Table 3a). We found a striking association with enriched concepts related to prostate cancer progression (Fig. 2a and Supplementary Table 3b). For comparison, we next incorporated disease signatures using prostate RNA-seq data and additional known prostate cancer genes, including AMACR (a tissue biomarker), EZH2 (a metastasis gene) and PCA3 (a lncRNA biomarker), as well as ACTB (encoding β-actin) as a control (Supplementary Fig. 5, Supplementary Table 3c–i and Supplementary Note). A heatmap visualization of significant comparisons confirmed a strong association of SCHLAP1-correlated genes but not of

experimentally by \textit{in vitro} translation assays of the three SCHLAP1 isoforms (Supplementary Fig. 3). Additionally, we found that SCHLAP1 transcripts were located in the nucleus (Fig. 1d). We confirmed the nuclear localization of SCHLAP1 transcripts in human samples (Fig. 1e) using an \textit{in situ} hybridization assay in formalin-fixed, paraffin-embedded prostate cancer samples (Supplementary Fig. 4a,b and Supplementary Note).

An analysis of SCHLAP1 expression in localized tumors demonstrated a strong correlation with higher Gleason scores, a histopathological measure of aggressiveness (Supplementary Fig. 4c,d and Supplementary Table 2). Next, we performed a network analysis of prostate cancer microarray data in the Oncomine database (Online Methods and Supplementary Table 3a). We found a striking association with enriched concepts related to prostate cancer progression (Fig. 2a and Supplementary Table 3b). For comparison, we next incorporated disease signatures using prostate RNA-seq data and additional known prostate cancer genes, including AMACR (a tissue biomarker), EZH2 (a metastasis gene) and PCA3 (a lncRNA biomarker), as well as ACTB (encoding β-actin) as a control (Supplementary Fig. 5, Supplementary Table 3c–i and Supplementary Note). A heatmap visualization of significant comparisons confirmed a strong association of SCHLAP1-correlated genes but not of
PCA3- and AMACR-correlated genes with high-grade and metastatic cancers (Fig. 2b). Kaplan-Meier analysis similarly showed significant associations between the SChLAP1 signature and biochemical recurrence and overall survival (Supplementary Fig. 6a,b).

To directly evaluate the relationship between SChLAP1 levels and clinical outcome, we next used SChLAP1 expression to stratify 235 samples from individuals with localized prostate cancer who underwent radical prostatectomy at the Mayo Clinic (Online Methods and Supplementary Fig. 6c). We evaluated samples for three clinical endpoints: biochemical recurrence, clinical progression to systemic disease and prostate cancer-specific mortality (Supplementary Table 4). At the time of this analysis, subjects had a median follow-up time of 8.1 years.

SChLAP1 was a powerful single-gene predictor of aggressive prostate cancer (Fig. 2c–e). SChLAP1 expression was highly significant when distinguishing disease with clinical progression and prostate cancer-specific mortality ($P = 0.00005$ and $0.002$, respectively; Fig. 2d,e). For the biochemical recurrence endpoint, high SChLAP1 expression was associated with a shorter median time to progression (1.9 versus 5.5 years for individuals with high and low expression of SChLAP1, respectively; Fig. 2c). We further confirmed this association with rapid biochemical recurrence using an independent cohort (Supplementary Fig. 6d). Multivariate and univariate regression analyses of the Mayo Clinic data demonstrated that SChLAP1 expression is an independent predictor of prostate cancer aggressiveness, with highly significant hazard ratios for predicting biochemical recurrence, clinical progression and prostate cancer-specific mortality (hazard ratios of $3.045, 3.563$ and $4.339$, respectively; $P < 0.01$), which are comparable to those for other clinical factors such as advanced clinical stage and Gleason histopathological score (Supplementary Fig. 7 and Supplementary Note).

To explore the functional role of SChLAP1, we performed small interfering RNA (siRNA)-mediated knockdowns to compare the impact of SChLAP1 depletion to that of EZH2, which is essential for cancer cell aggressiveness. Notably, knockdown of SChLAP1 dramatically impaired cell invasion and proliferation in vitro to
**Figure 3** SCHLAP1 coordinates cancer cell invasion *in vitro* and metastatic seeding *in vivo*. (a) siRNA knockdown of SCHLAP1 *in vitro* in three prostate cell lines (LNCaP, 22Rv1, Du145) impairs cellular invasion through Matrigel in a Boyden chamber assay. EZH2 siRNA serves as a positive control for impaired invasion. Images to the right show representative Boyden chamber membranes stained with crystal violet following invasion. All images were captured at the same magnification, as indicated by the scale bars. (b) Overexpression of SCHLAP1 in RWPE cells results in increased cellular invasion through Matrigel in Boyden chamber assays. Images to the right show Boyden chambers as in a in indicated RWPE cell lines; scale bars represent equal magnification of images. (c) Intracardiac injection of severe combined immunodeficient (SCID) mice with 22Rv1 cells with stable SCHLAP1 knockdown. Representative luciferase bioluminescence images (BLI) are shown for mice 5 weeks after intracardiac injection with 22Rv1 cells expressing non-targeting control shRNA (n = 9), SCHLAP1 shRNA 1 (n = 14) or SCHLAP1 shRNA 2 (n = 14). Mouse IDs are given above each image. (d) Relative intensity of whole-mouse luciferase signal is plotted for mice with intracardiac injection of 22Rv1 cells expressing control, non-targeting shRNA (n = 9), SCHLAP1 shRNA 1 (n = 14) or SCHLAP1 shRNA 2 (n = 14). (e) Number of gross metastatic sites observed by luciferase signal in mice injected with 22Rv1 cells expressing SCHLAP1 shRNA or control, non-targeting shRNA. Independent foci with luciferase signal were averaged for mice injected with 22Rv1 cells expressing control shRNA (n = 9), SCHLAP1 shRNA 1 (n = 14) or SCHLAP1 shRNA 2 (n = 14). (f) Invasion of 22Rv1 cells expressing non-targeting or SCHLAP1 shRNA in chick CAM assays. 22Rv1 cells are labeled with GFP. Images are counterstained for chicken collagen IV to visualize vasculature (RFP) and for DAPI to visualize nuclei. Scale bars represent arbitrary units indicating equal magnification of images. (g-i) In the CAM assay, 22Rv1 cells expressing SCHLAP1 shRNA demonstrate decreased invasiveness (g), metastatic spread to the liver and lungs (h) and reduced tumor weight (i) relative to 22Rv1 cells expressing control, non-targeting shRNA. (j) Quantification of invasavation of RWPE cells expressing LacZ or SCHLAP1 in the CAM assay. All data in bar plots are represented as mean ± s.e.m. Statistical significance was determined by two-tailed Student’s t test: *P < 0.05. Box plots in e,i display box-and-whisker plots with the midpoint line indicating the median, box boundaries showing 25th and 75th quartile ranges and whiskers displaying the minimum and maximum values.

an extent comparable to that observed with knockdown of EZH2 (Fig. 3a and Supplementary Fig. 8a,b). Overexpression of an siRNA-resistant SCHLAP1 isoform rescued the *in vitro* invasive phenotype of 22Rv1 cells treated with siRNA-2 (Supplementary Fig. 8c,d). Overexpression of the three SCHLAP1 isoforms in benign, immortalized RWPE prostate cells dramatically increased the ability of these cells to invade *in vitro* but did not affect cell proliferation (Fig. 3b and Supplementary Fig. 8e,f).

To test *Schlapi* in vivo, we performed intracardiac injection of CB-17 SCID mice with 22Rv1 cells stably knocking down *Schlapi* (Supplementary Fig. 9a) and observed that *Schlapi* depletion impaired metastatic seeding and growth, as measured by luciferase signaling at both proximal (lungs) and distal sites (Fig. 3c,d). Indeed, compared to mice injected with 22Rv1 cells expressing a non-targeting control, mice injected with 22Rv1 cells stably expressing short hairpin RNA (shRNA) against *Schlapi* had both fewer gross metastatic sites overall as well as smaller metastatic tumors when they did form (Fig. 3d,e). Histopathological analysis of the metastatic 22Rv1 tumors, regardless of *Schlapi* knockdown, showed uniformly high-grade epithelial cancer (Supplementary Fig. 9b). Interestingly, subcutaneous xenografts with stable knockdown of *Schlapi* showed slower tumor progression; however, this was due to delayed tumor engraftment rather than to decreased tumor growth kinetics, with no change in Ki67 staining observed between cells expressing *Schlapi* shRNA and control cells expressing non-targeting shRNA (Supplementary Fig. 9c–i).

Next, using the chick chorioallantoic membrane (CAM) assay19, we found that 22Rv1 cells expressing *Schlapi* shRNA 2, which have depleted expression of both isoforms 1 and 2, had greatly reduced ability to invade, intravasate and metastasize to distant organs (Fig. 3f–h). Additionally, cells with knockdown of *Schlapi* also
resulted in decreased tumor growth (Fig. 3i). Notably, RWPE cells with overexpression of SlcLAP1 isoform 1 partially supported these results, showing a markedly increased ability to intravasate (Fig. 3j). RWPE cells overexpressing SlcLAP1 did not generate distant metastases or cause altered tumor growth in this model (data not shown). Together, the mouse metastasis and CAM data strongly implicate SlcLAP1 in tumor invasion and metastasis through activity in cancer cell inva-
svasion, extravasation and subsequent tumor cell seeding.

To elucidate the mechanisms of SlcLAP1 function, we profiled 22Rv1 and LNCaP cells with SlcLAP1 knockdown, identifying 165 upregulated and 264 downregulated genes ($p$ value < 0.001) (Supplementary Fig. 10a and Supplementary Table 5a). After ranking genes according to differential expression$^{20}$, we employed Gene Set Enrichment Analysis (GSEA)$^{21}$ to search for enrichment across the Molecular Signatures Database (MSigDB)$^{22}$. Among the highest ranked concepts, we noticed genes positively or negatively correlated with the SWI/SNF complex$^{23}$, and this association was independently confirmed using gene signatures generated from our RNA-seq data (Supplementary Fig. 10b–e and Supplementary Table 5b,c). Notably, SlcLAP1-regulated genes were inversely correlated with these data sets, suggesting that SlcLAP1 and the SWI/SNF complex function in opposing manners.

The SWI/SNF complex regulates gene transcription as a multipro-
tein system that physically moves nucleosomes at gene promoters$^{24}$. Loss of SWI/SNF complex functionality promotes cancer progression, and multiple SWI/SNF components are somatically inactivated in can-
cer$^{24,25}$. SWI/SNF complex mutations do occur in prostate cancer, albeit not commonly$^{26}$, and downregulation of SWI/SNF complex members

Figure 4 SlcLAP1 antagonizes SNF5 function and attenuates SNF5 genome-wide localization. (a) Heatmap results for SlcLAP1 or SMARCB1 knockdown in LNCaP and 22Rv1 cells. The numbers above the heatmap indicate the specific siRNA and microarray replicates. (b) RIP of SNF5 or androgen receptor (AR) in 22Rv1 and LNCaP cells. Inset, protein blots showing pulldown efficiency. Error bars, s.e.m. (c) RIP analysis of SNF5 in RWPE cells overexpressing LacZ, SlcLAP1 isoform 1 or SlcLAP1 isoform 2. Inset, protein blots showing pulldown efficiency. Error bars, s.e.m. (d) Pulldown of SlcLAP1 RNA using chromatin isolation by RNA purification (ChIRP) recovers SNF5 protein in RWPE cells expressing SlcLAP1 isoform 1. LacZ and TERC serve as negative and positive controls, respectively. (e) Global representation of SNF5 genomic binding over a 4-kb window centered on each SNF5 ChIP-seq peak in RWPE cells expressing LacZ, SlcLAP1 isoform 1 or SlcLAP1 isoform 2. (f) Heatmap of SNF5 genomic binding at target sites in RWPE cells expressing LacZ, SlcLAP1 isoform 1 or SlcLAP1 isoform 2. A 2-kb interval centered on the called SNF5 peak is shown. (g) GSEA results showing significant enrichment of ChIP-seq promoter peaks with >2-fold loss of SNF5 binding for underexpressed genes in RWPE cells expressing SlcLAP1. NES, normalized enrichment score; FWER, familywise error rate.
characterizes subsets of prostate cancer\textsuperscript{23,26}. Thus, antagonism of SWI/SNF complex activity by \textit{SchLAP1} is consistent with the oncogenic behavior of \textit{SchLAP1} and the tumor suppressive behavior of the SWI/SNF complex.

To directly test whether \textit{SchLAP1} antagonizes SWI/SNF-mediated regulation, we performed siRNA-mediated knockdown of \textit{SMARCB1} (which encodes the SNF5 protein) (\textbf{Supplementary Fig. 10f}), an essential subunit that facilitates SWI/SNF complex binding to histone proteins\textsuperscript{24,25,27}, and confirmed predicted expression changes for several \textit{SchLAP1}- or SNF5-regulated genes (\textbf{Supplementary Fig. 10g,h}). A comparison of genes whose expression was altered by knockdown of \textit{SMARCB1} to those regulated by \textit{SchLAP1} demonstrated an antagonistic relationship in which \textit{SchLAP1} knockdown affected the same genes as \textit{SMARCB1} knockdown but with opposing directions of effect (\textbf{Fig. 4a} and \textbf{Supplementary Table 5d–h}). We used GSEA to quantify and verify the significance of these findings (false discovery rate (FDR) < 0.05) (\textbf{Supplementary Fig. 10i–k}). Furthermore, a shared \textit{SMARCB1}-\textit{SchLAP1} signature of coregulated genes was highly enriched for prostate cancer clinical signatures for disease aggressiveness (\textbf{Supplementary Fig. 11} and \textbf{Supplementary Table 5i}).

Mechanistically, although \textit{SchLAP1} and \textit{SMARCB1} mRNA levels were comparable (\textbf{Supplementary Fig. 12a}), \textit{SchLAP1} knockdown or overexpression did not alter SNF5 protein abundance (\textbf{Supplementary Fig. 12b}), suggesting that \textit{SchLAP1} regulates SWI/SNF activity post-translationally. To explore this possibility, we performed RNA immunoprecipitation assays (RIPs) for SNF5. We found that endogenous \textit{SchLAP1} but not other cytoplasmic or nuclear \textit{lncRNAs}\textsuperscript{28} robustly coimmunoprecipitated with SNF5 under native conditions (\textbf{Fig. 4b}) and with use of UV cross-linking (\textbf{Supplementary Fig. 12c}), and coimmunoprecipitation was also observed with a second antibody to SNF5 (\textbf{Supplementary Fig. 12d}). In contrast, \textit{SchLAP1} did not coimmunoprecipitate with androgen receptor (\textbf{Fig. 4b}). Furthermore, both \textit{SchLAP1} isoform 1 and isoform 2 coimmunoprecipitated with SNF5 in RWPE overexpression models (\textbf{Fig. 4c} and \textbf{Supplementary Fig. 12e}). SNRNP70 binding to U1 RNA was used as a technical control in all cell lines (\textbf{Supplementary Fig. 12f,g}). Finally, pulldown of \textit{SchLAP1} RNA in RWPE cells overexpressing \textit{SchLAP1} isoform 1 robustly recovered SNF5 protein, confirming this interaction (\textbf{Fig. 4d} and \textbf{Supplementary Fig. 12h}).

To address whether \textit{SchLAP1} modulates SWI/SNF genomic binding, we performed ChIP-seq for SNF5 in RWPE cells expressing \textit{LacZ} or \textit{SchLAP1} and called significantly enriched peaks with respect to an IgG control (Online Methods and \textbf{Supplementary Table 6a}). Protein blot validation confirmed SNF5 pulldown by ChIP (\textbf{Supplementary Fig. 13a}). After aggregating called peaks from all samples, we found 6,235 genome-wide binding sites for SNF5 (FDR < 0.05; \textbf{Supplementary Table 6b}), which were highly enriched for sites near gene promoters (\textbf{Supplementary Fig. 13b}), supporting results from previous studies of SWI/SNF binding\textsuperscript{29–31}.

A comparison of SNF5 binding across these 6,235 genomic sites to those regulated by \textit{SchLAP1} identified a signature of \textit{SchLAP1}-mediated changes in gene expression. After identifying a gene signature with highly significant changes in expression (\textbf{Supplementary Table 6d}), we intersected this signature with the ChIP-seq data. We observed that a substantial subset of genes with ≥2-fold relative change in SNF5 genomic binding were disregulated when \textit{SchLAP1} was overexpressed (\textbf{Supplementary Fig. 14b}). Decreased SNF5 binding was primarily associated with the downregulation of target gene expression (\textbf{Supplementary Table 6e}), although the SWI/SNF complex is known to regulate expression in either direction\textsuperscript{24,25}. Integrative GSEA of the microarray and SNF5 ChIP-seq data demonstrated significant enrichment for genes that were repressed when \textit{SchLAP1} was overexpressed (q value = 0.003; \textbf{Fig. 4g}). Overall, these data argue that \textit{SchLAP1} overexpression antagonizes SWI/SNF complex function by attenuating the genomic binding of this complex, thereby impairing its ability to properly regulate gene expression.

Finally, we used expression profiling of RWPE cells expressing \textit{LacZ} or \textit{SchLAP1} to characterize the relationship between SNF5 binding and \textit{SchLAP1}-mediated changes in gene expression. After identifying a gene signature with highly significant changes in expression (\textbf{Supplementary Table 6d}), we intersected this signature with the ChIP-seq data. We observed that a substantial subset of genes with ≥2-fold relative decrease in SNF5 genomic binding were disregulated when \textit{SchLAP1} was overexpressed (\textbf{Supplementary Fig. 14b}). Decreased SNF5 binding was primarily associated with the downregulation of target gene expression (\textbf{Supplementary Table 6e}), although the SWI/SNF complex is known to regulate expression in either direction\textsuperscript{24,25}. Integrative GSEA of the microarray and SNF5 ChIP-seq data demonstrated significant enrichment for genes that were repressed when \textit{SchLAP1} was overexpressed (q value = 0.003; \textbf{Fig. 4g}). Overall, these data argue that \textit{SchLAP1} overexpression antagonizes SWI/SNF complex function by attenuating the genomic binding of this complex, thereby impairing its ability to properly regulate gene expression.

Here we have discovered \textit{SchLAP1}, a highly prognostic \textit{lncRNA} that is abundantly expressed in ~25% of prostate cancers and that aids in the discrimination of aggressive tumors from indolent forms of the disease. Mechanistically, we find that \textit{SchLAP1} coordinates cancer cell invasion \textit{in vitro} and metastatic spread \textit{in vivo}. Moreover, we characterize an antagonistic \textit{SchLAP1}-SWI/SNF axis in which \textit{SchLAP1} impairs SNF5-mediated regulation of gene expression and genomic binding (\textbf{Supplementary Fig. 14c}). Thus, whereas other \textit{lncRNAs} such as \textit{HOTAIR} and \textit{HOTTIP} are known to assist epigenetic complexes such as PRC2 and MLL by facilitating their genomic binding and enhancing their functions\textsuperscript{8,9,32}, \textit{SchLAP1} is the first \textit{lncRNA}, to our knowledge, that impairs a major epigenetic complex with well-documented tumor suppressor function\textsuperscript{23–25,33–35}. Our discovery of \textit{SchLAP1} has broad implications for cancer biology and provides supporting evidence for the role of \textit{lncRNAs} in the progression of aggressive cancers.

\textbf{METHODS}

Methods and any associated references are available in the online version of the paper.

\textbf{Accession codes.} Sequences for \textit{SchLAP1} isoforms 1–7 have been deposited in GenBank under accession numbers JX117418, JX117419, JX117420, JX117421, JX117422, JX117423 and JX117424. Microarray data have been deposited in the Gene Expression Omnibus (GEO) under accession GSE40386.

\textbf{Note:} Any Supplementary Information and Source Data files are available in the online version of the paper.

\textbf{ACKNOWLEDGMENTS}

We thank O.A. Balbin, S.A. Tomlins, C. Brenner, S. Deroo and S. Roychowdhury for helpful discussions. This work was supported in part by US National Institutes of Health (NIH) Prostate Specialized Program of Research Excellence grant P50CA9568, Early Detection Research Network grant U01 CA111275, US NIH grant R01CA132874-01A1 and US Department of Defense grant PC100171 (A.M.C.). A.M.C. is supported by a Doris Duke Charitable Foundation Clinical Scientist Award, by the Prostate Cancer Foundation and by the Howard Hughes Medical Institute. A.M.C. is an American Cancer Society Research Professor. A.M.C. is a Taubman Scholar of the University of Michigan. F.Y.F. was supported by the Prostate Cancer Foundation and by US Department of Defense grant PC094231, Q.C. was supported by US Department of Defense Postdoctoral

Fellowship PC094725. J.R.P. was supported by US Department of Defense Predoctoral Fellowship PC094290. M.K.I. was supported by US Department of Defense Predoctoral Fellowship BC100238. A.S. was supported by NIH Predoctoral Fellowship F1F30CA180376-01, J.R.P., M.K.I. and A.S. are Fellows of the University of Michigan Medical Scientist Training Program.

AUTHOR CONTRIBUTIONS


COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

ONLINE METHODS

Cell lines. All cell lines were obtained from the American Type Culture Collection. Cell lines were maintained using standard media and conditions. Specifically, VCaP and Du145 cells were maintained in DMEM (Invitrogen) supplemented with 10% FBS and 1% penicillin-streptomycin. LNCaP and 22Rv1 cells were maintained in RPMI 1640 (Invitrogen) supplemented with 10% FBS and 1% penicillin-streptomycin. RWPE cells were maintained in KSF medium (Invitrogen) supplemented with 10 ng/mL epidermal growth factor (EGF; Sigma) and bovine pituitary extract (BPE) and with 1% penicillin-streptomycin. All cell lines were grown at 37 °C in a 5% CO2 cell culture incubator. All cell lines were genotyped for identity at the University of Michigan Sequencing Core and were tested routinely for Mycoplasma contamination.

Cell lines expressing SChLAP1 or control constructs were generated by cloning SChLAP1 or control sequence into the pLent6 vector (Invitrogen), using pcR8 non-directional Gateway cloning (Invitrogen) as an initial cloning vector, and shuttling to pLenti6 using LR clonase II (Invitrogen) according to the manufacturer’s instructions. Stably transfected RWPE and 22Rv1 cells were selected with blasticidin (Invitrogen) for 72 h. For LNCaP and 22Rv1 cells with stable knockdown of SChLAP1, cells were transfected with lentiviral constructs encoding SChLAP1 shRNA or with non-targeting shRNA lentiviral constructs for 48 h. GFP-positive cells were selected with 1 µg/mL puromycin for 72 h. All lentiviruses were generated by the University of Michigan Vector Core.

Tissue samples. Prostate tissues were obtained from the radical prostatectomy series and Rapid Autopsy Program at the University of Michigan tissue core.6 These programs are part of the University of Michigan Prostate Cancer Specialized Program of Research Excellence (SPORE). 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 (Invitrogen) and an RNaseasy kit (Qiagen) with DNase I digestion according to the manufacturers’ instructions. RNA integrity was verified on an Agilent Bioanalyzer 2100 (Agilent Technologies). cDNA was synthesized from total RNA using Superscript III (Invitrogen) and random primers (Invitrogen).

Quantitative RT-PCR. Quantitative RT-PCR was performed using Power SYBR Green MasterMix (Applied Biosystems) on an Applied Biosystems 7900HT Real-Time PCR System. All oligonucleotide primers were obtained from Integrated DNA Technologies (IDT), and primer sequences are listed in Supplementary Table 7a. The housekeeping genes GAPDH, HMBS and ACTB were used as loading controls. Fold changes were calculated relative to housekeeping genes and were normalized to the median value in benign samples.

RT-PCR. RT-PCR was performed for primer pairs using Platinum Taq High-Fidelity polymerase (Invitrogen). PCR products were resolved on a 1.0% agarose gel. PCR products were then 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 pcR8-TOPO vector (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 IDT, and primer sequences are listed in Supplementary Table 7a.

RACE. 5’ and 3’ RACE were 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, the supplied GeneRacer primers and the appropriate gene-specific primers indicated in Supplementary Table 7a. RACE PCR products were separated on a 1.5% agarose gel. Gel products were extracted with a Gel Extraction kit, cloned into pcR8-TOPO vectors and sequenced bidirectionally using M13 forward and reverse primers at the University of Michigan Sequencing Core. At least three colonies were sequenced for every RACE PCR product that was gel purified.

siRNA-mediated knockdown. Cells were plated in 100-mm plates at a desired concentration and transfected with 20 µM experimental siRNA oligonucleotides or non-targeting controls twice at 8 h and 24 h after plating. Knockdown was performed with Oligofectamine in OptiMEM medium. Knockdown efficiency was determined by qPCR. siRNA sequences (in sense orientation) for knockdown experiments are listed in Supplementary Table 7b. At 72 h after transfection, cells were trypsinized, counted with a Coulter counter and diluted to 1 million cells/mL.

Overexpression. Full-length SChLAP1 transcript was amplified from LNCaP cells and cloned into the pLent6 vector along with LacZ control sequence. 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 SChLAP1 or LacZ, and stable pools and clones were generated by selection with blasticidin. Similarly, the immortalized cancer cell line 22Rv1 was infected with lentiviruses expressing SChLAP1 or LacZ, and stable pools were generated by selection with blasticidin.

Cell proliferation assays. At 72 h after transfection with siRNA, cells were trypsinized, counted with a Coulter counter and diluted to 1 million cells/mL. For proliferation assays, 10,000 cells were plated in each well of a 24-well plate and grown in regular growth medium. At 48 h and 96 h after plating, cells were collected by trypsinizing and counted using a Coulter counter. All assays were performed in triplicate.

Basement membrane matrix invasion assays. For invasion assays, cells were treated with the indicated siRNAs, and, at 72 h after transfection, cells were trypsinized, counted with a Coulter counter and diluted to 1 million cells/mL. Cells were seeded onto basement membrane matrix (EC matrix, Chemicon) present in the insert of a 24-well culture plate. FBS was added to the lower chamber as a chemoattractant. After 48 h, the non-invading cells and EC matrix were gently removed with a cotton swab. Invasive cells located on the lower side of the chamber were stained with crystal violet, air dried and photographed. For colorimetric assays, inserts were treated with 150 µL of 10% acetic acid, and absorbance was measured at 560 nm using a spectrophotometer (GE Healthcare).

shRNA-mediated knockdown. The prostate cancer cell lines LNCaP and 22Rv1 were seeded at 50–60% confluency and were allowed to attach overnight. Cells were transfected with lentiviral constructs expressing SChLAP1 or non-targeting shRNA as described previously for 48 h. GFP-positive cells were selected with 1 µg/mL puromycin for 72 h. At 48 h after the start of selection, cells were collected for protein and RNA using RIPA buffer or TRIzol, respectively. RNA was processed as described above.

Gene expression profiling. Expression profiling was performed using the Agilent Whole Human Genome Oligo Microarray according to previously published protocols. All samples were run in technical triplicates, comparing knockdown samples treated with SChLAP1 shRNA to samples treated with non-targeting control siRNA. Expression data were analyzed using the SAM method as described previously.

Mouse 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).

For the intracardiac injection model, 5 x 105 cells from 1 of 3 experimental cell lines (22Rv1-shSChLAP1-1 or 22Rv1-shSChLAP1-2) (two cell lines expressing SChLAP1 shRNA) or 22Rv1-shNT (expressing control vector), all with luciferase constructs incorporated) were introduced into CB-17 SCID mice at 6 weeks of age. Female mice were used to minimize endogenous androgen production that might stimulate xenografted prostate cells. We used 15 mice per cell line to ensure adequate statistical power to distinguish phenotypes between groups. Mice used in these studies were randomized by double-blind injection of cell line samples into mice and were monitored for tumor growth by researchers blinded to the study design. Beginning 1 week after injection, bioluminescent imaging of mice was performed weekly using a CCD IVIS.
system with a 50-mm lens (Xenogen), and the results were analyzed using LivingImage software (Xenogen). When the a mouse reached the determined end point, defined as whole-body region of interest (ROI) of 1 x 10^19 photons, or became fatally ill, it was euthanized, and the lung and liver were resected. Half of the resected specimen was placed in an immunohistochemistry cassette, incubated in 10% buffered formalin phosphate (Fisher Scientific) for 24 h and transferred to 70% ethanol until further analysis. The other half of each specimen was snap frozen in liquid nitrogen and stored at -80 °C. A specimen was disregarded if the tumor was localized only in the heart. After accounting for these considerations, there were 9 mice analyzed for 22Rv1-shNT cells and 14 mice each analyzed for 22Rv1-shSChLAP1-1 and 22Rv1-shSChLAP1-2 cells.

For the subcutaneous injection model, 1 x 10^6 cells from 1 of the 3 previously described experimental cell lines were introduced into mice (CB-17 SCID), aged 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 the end point was defined by tumor volume of 1,000 mm^3. When a mouse reached the end point or became fatally ill, it was euthanized, and the primary tumor was resected. The resected specimen was divided in half: one half was placed in 10% buffer formalin, and the other half was snap frozen. For histological analyses, formalin-fixed, paraffin-embedded mouse livers and lungs were sectioned on a microtome into 5-µm sections on glass slides. Slides were stained with hematoxylin and eosin using standard methods and were analyzed by a board-certified pathologist (L.P.K.).

**Immunoblot analysis.** Cells were lysed in RIPA lysis buffer (Sigma) supplemented with HALT protease inhibitor (Fisher). Protein blotting analysis was performed with standard protocols using polyvinylidene difluoride (PVDF) membrane (GE Healthcare), and signals were visualized with an enhanced chemiluminescence system as described by the manufacturer (GE Healthcare).

Protein lysates were boiled in sample buffer, and 10 µg of protein was loaded onto an SDS-PAGE gel and run for separation of proteins. Proteins were transferred onto PVDF membrane and blocked for 90 min in blocking buffer (5% milk in a solution of 0.1% Tween-20 in Tris-buffered saline (TBS-T)). Membranes were incubated overnight at 4 °C with primary antibody. After three washes with TBS-T and one wash with TBS, the blot was incubated with HRP-conjugated secondary antibody, and signal was visualized with an enhanced chemiluminescence system as described by the manufacturer (GE Healthcare).

**RIP assays.** RIP assays were performed using a Millipore EZ-Magna RIP RNA-Binding Protein Immunoprecipitation kit (Millipore, 17-701) according to the manufacturer’s instructions. RIP PCR was performed as qPCR, as described above, using total RNA as input controls. We used 1/150 volume of the RIP RNA product per PCR reaction. Antibodies used for RIP included rabbit polyclonal IgG (Millipore, PP64) and antibodies to SNRNP70 (Millipore, CS203216), SNF5 (Millipore, ABD22, rabbit), SNF5 (Abcam, ab58209, mouse), β-actin (1:5,000 dilution; Sigma, A5316, mouse) and androgen receptor (Millipore, 06-680, rabbit).

**ChIP assays.** ChIP assays were performed as described previously using antibody for SNF5 (Millipore, ABD22, rabbit) and rabbit IgG (Millipore, PP64B). Briefly, approximately 1 million cells were cross-linked per antibody for 10–15 min with 1% formaldehyde, and crosslinking was inactivated by incubation with 0.125 M glycine for 5 min at room temperature. Cells were rinsed with cold PBS three times, and cell pellets were resuspended in lysis buffer supplemented with protease inhibitors. Chromatin was sonicated to an average length of 500 bp and centrifuged to remove debris, and supernatants containing chromatin fragments were incubated with protein A or protein G beads to reduce non-specific binding. Beads were then removed, and supernatants were incubated with 6 µg of antibody overnight at 4 °C. Fresh beads were added and incubated with protein-chromatin-antibody complexes for 2 h at 4 °C, washed twice with 1× dialysis buffer and four times with IP wash buffer, and eluted in 150 µl of IP elution buffer. One-tenth of the ChIP reaction was taken for protein evaluation for validation of pulldown. Cross-linking was reversed by incubating eluted products with 0.3 M NaCl at 65 °C overnight. ChIP products were cleaned with the USB PrepEase kit. ChIP experiments were validated for specificity of the antibody by protein blotting.

**ChIP-seq experiments.** Paired-end ChIP-seq libraries were generated following the Illumina ChIP-seq protocol with minor modifications. DNA isolated by ChIP assay was subjected to end repair and A tailing before ligation with Illumina adaptors. Samples were purified using AMPure beads (Beckman Coulter) and PCR enriched with a combination of specific index primers and PE2.0 primer under the following conditions: 98 °C (30 s), 65 °C (30 s) and 72 °C (40 s, with the addition of 4 s per cycle). After 14 cycles of amplification, a final extension at 72 °C for 5 min was carried out. Barcoded libraries were size selected using 3% NuSieve Agarose gels (Lonza) and subjected to an additional PCR enrichment step. Libraries were analyzed and quantified using a Bioanalyzer instrument (Agilent Technologies) before they were subjected to paired-end sequencing using the Illumina HiSeq platform.

**CAM assays.** CAM assays were performed as previously described. Briefly, fertilized chicken eggs were incubated in a rotary humidified incubator at 38 °C for 10 d. CAM was released by applying a mild amount of pressure to the hole over the air sac and cutting a 1-cm² window encompassing a second hole near the allantoic vein. Approximately 2 million cells in 50 µl of medium were implanted in each egg, windows were sealed, and eggs were returned to a stationary incubator.

For local invasion and intravasation experiments, the upper and lower CAMs were isolated after 72 h. Upper CAMs were processed and stained for chicken collagen IV (immunofluorescence) or human cytokeratin (immunohistochemistry) as previously described.

For metastasis assays, embryonic livers were isolated on day 18 of embryonic growth and analyzed for the presence of tumor cells by quantitative human Alu-specific PCR. Genomic DNA isolates from lower CAMs and livers were prepared using the Puregene DNA purification system (Qiagen), and quantification with human Alu-specific PCR was performed as described.

Fluoresgenic TaqMan qPCR probes were generated as described above and used to determine DNA copy number. For xenograft growth assays with RWPE cells, embryos were sacrificed on day 18, and an extraembryonic xenografts were excised and weighed.

**In situ hybridization.** In situ hybridization assays were performed as a commercial service from Advanced Cell Diagnostics, Inc. Briefly, cells in the clinical specimens were fixed and permeabilized using xylene, ethanol and protease to allow for probe access. Slides were boiled in pretreatment buffer for 15 min and rinsed in water. Next, two independent target probes were hybridized to SCChLAP1 RNA at 40 °C for 2 h, with this pair of probes creating a binding site for a preamplifier. After this incubation, the preamplifier was hybridized to the target probes at 30 °C and amplified with six cycles of hybridization followed by two washes. Cells were counterstained to visualize signal. Finally, slides were stained with hematoxylin and eosin, dehydrated with 100% ethanol and mounted in a xylene-based mounting medium.

**In vitro translation.** Full-length SCChLAP1, PCAT-1 or GUS positive control sequences 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 1 mM methionine and Transcend Biotin-Lysyl-tRNA (Promega) according to the manufacturer’s instructions.

**ChIRP assays.** ChIRP assays were performed as previously described. Briefly, antisense DNA probes targeting the full-length SCChLAP1 sequence were designed to hybridize to the full-length transcript with high specificity and sensitivity. These probes were then transcribed in vitro using T7 RNA polymerase and labeled with biotin using the Quick Coupled Transcription/Translation System (Promega) with 1 mM methionine and Transcend Biotin-Lysyl-tRNA (Promega) according to the manufacturer’s instructions. Barcoded libraries were size selected using 3% NuSieve Agarose gels (Lonza) and subjected to an additional PCR enrichment step. Libraries were analyzed and quantified using a Bioanalyzer instrument (Agilent Technologies) before they were subjected to paired-end sequencing using the Illumina HiSeq platform.
were designed using the online designer at Stellaris (see URLs). Fifteen probes spanning the entire transcript and unique to the SChLAP1 sequence were chosen. Additionally, ten probes were designed against TERC RNA as a positive control, and 24 probes were designed against LacZ RNA as a negative control. All probes were synthesized with 3' biotinylated (IDT). Sequences of all probes are listed in Supplementary Table 8. RWPE cells overexpressing SChLAP1 isoform 1 were grown to 80% confluency in 100-mm cell culture dishes. Two dishes were used for each probe set. Before being collected, cells were rinsed with 1× PBS and cross-linked with 1% glutaraldehyde (Sigma) for 10 min at room temperature. Cross-linking was quenched by incubation with 0.125 M glycine for 5 min at room temperature. Cells were rinsed twice with 1× PBS, collected and pelleted at 1,500g for 5 min. Nuclei were isolated using the Pierce NE-PER Nuclear Protein Extraction kit. Nuclear pellets were resuspended in 100 mg/ml cell lysis buffer (50 mM Tris, pH 7.0, 10 mM EDTA, 1% SDS, and, added before use, 1 mM dithiothreitol (DTT), phenylmethylsulphonyl fluoride (PMSF), protease inhibitor and Superase-In (Invitrogen)). Lysates were placed on ice for 10 min and sonicated using a Bioruptor (Diagenode) at the highest setting with 30-s on and 45-s off cycles until lysates were completely solubilized. Cell lysates were diluted in twice the volume of hybridization buffer (500 mM NaCl, 1% SDS, 100 mM Tris, pH 7.0, 10 mM EDTA, 15% formamide, and, added before use, DTT, PMSF, protease inhibitor and Superase-In), and 100 nM probes were added to the diluted lysates. Hybridization was carried out by end-over-end rotation at 37°C for 4 h. Magnetic streptavidin C1 beads were prepared by washing three times in cell lysis buffer and were then added to each hybridization reaction at a concentration of 100 µl per 100 pmol of probe. Reactions were incubated at 37°C for 30 min with end-over-end rotation. Bead-probe RNA complexes were captured with magnetic racks (Millipore) and washed five times with 1 ml wash buffer (2× SSC, 0.5% SDS, fresh PMSF added). After the final wash, 20% of the sample was used for RNA isolation, and 80% of the sample was used for protein isolation. For RNA elution, beads were resuspended in 200 µl of RNA proteinase K buffer (100 mM NaCl, 10 mM Tris, pH 7.0, 1 mM EDTA, 0.5% SDS) and 1 mg/ml proteinase K (Ambion). Samples were incubated at 50°C for 45 min and then boiled for 10 min. RNA was isolated using 500 µl of TRIzol reagent and the miRNeasy kit (Qiagen) with on-column DNAse digestion (Qiagen). RNA was eluted with 10 µl of water and then analyzed by quantitative RT-PCR for the detection of enriched transcripts. For protein elution, beads were resuspended in three times the original volume of DNAse buffer (100 mM NaCl, 0.1% NP-40), and protein was eluted with a cocktail of 100 µg/ml RNase A (Sigma-Aldrich), 0.1 U/ml RNase H (Epipenter) and 100 U/ml DNAse I (Invitrogen) at 37°C for 30 min. Eluted protein samples were supplemented with NuPAGE LDS Sample Buffer (Novex) and NuPAGE Sample Reducing Agent (Novex) to a final concentration of 1× each and then boiled for 10 min before SDS-PAGE protein blot analysis using an antibody to SNF5 (Millipore).

RNA-seq library preparation. Total RNA was extracted from healthy and cancer cell lines and subject tissues, and RNA quality was assessed via Agilent Bioanalyzer. Transcriptome libraries from the mRNA fractions were generated following the RNA-seq protocol (Illumina). Each sample was sequenced in a single lane with the Illumina Genome Analyzer II (with a 40- to 80-nt read length) or with the Illumina HiSeq 2000 (with a 100-nt read length) according to published protocols11,49. For strand-specific library construction, we employed the dUTP method of second-strand marking as described previously50.

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. For details regarding the statistical methods employed during microarray, RNA-seq and ChIP-seq data analysis, see below.

Nomination of SChLAP1 as an outlier using RNA-seq data. We nominated SChLAP1 as a prostate cancer outlier as described11. Briefly, a modified COPA analysis was performed on the 81 tissue samples in the cohort. Reads per kilobase per million mapped reads (RPKM) expression values were used and shifted by 1.0 to avoid division by zero. COPA analysis included the following steps: (i) gene expression values were median centered, using the median expression value for the gene across all samples in the cohort, which sets the gene’s median to zero; (ii) the median absolute deviation (MAD) was calculated for each gene, and each gene expression value was then scaled by its MAD; (iii) the 80th, 85th, 90th and 95th percentiles of the transformed expression values were calculated for each gene, the average of those four values was taken, and genes were then ranked according to this ‘average percentile’, which generated a list of outlier genes arranged by importance; and (iv) finally, genes showing an outlier profile in the benign samples were discarded.

LNCap ChIP-seq data. Sequencing data from GSE14097 were downloaded from GEO. Reads from the LNCAP H3K4me3 and H3K36me3 ChIP-seq samples were mapped to human genome version hg19 using BWA 0.5.9 (ref. 51). Peak calling was performed using MACS252 according to published protocols53. Data were visualized using the UCSC Genome Browser54.

RWPE ChIP-seq data. Sequencing data from RWPE SNF5 ChIP-seq samples were mapped to human genome version hg19 using the BWA 0.5.9 algorithm52. Although we performed paired-end sequencing, the ChIP-seq reads were processed as single-end reads to adhere to our preexisting analysis protocol. Basic read alignment statistics are listed in Supplementary Table 6a. Peak calling was performed with respect to an IgG control using the MACS algorithm52. We bypassed the model-building step of MACS (using the ‘nomodel’ flag) and specified a shift size equal to half the library fragment size determined by the Agilent Bioanalyzer (using the ‘shiftsize’ option). For each sample, we ran the CEAS program and generated genome-wide reports55. We retained peaks with an FDR less than 5% (peak calling statistics across multiple FDR thresholds are shown in Supplementary Table 6b). We then aggregated SNF5 peaks from the RWPE–LacZ, RWPE–SChLAP1 isoform 1 and RWPE–SChLAP1 isoform 2 samples using the ‘union’ of the genomic peak intervals. We intersected peaks with RefSeq protein-coding genes and found that 1,299 peaks occurred within 1 kb of TSSs. We counted the number of reads overlapping each of these promoter peaks across each sample using a custom Python script and used the DESeq R package56 version 1.6.1 to compute the normalized fold change between RWPE–LacZ and RWPE–SChLAP1 (both isoforms). We observed that 389 of the 1,299 promoter peaks had at least a 2-fold average decrease in SNF5 binding. This set of 389 genes was subsequently used as a gene set for GSEA (Supplementary Table 6c).

Microarray experiments. We performed two-color microarray gene expression profiling of 22Rv1 and LNCaP cells treated with two independent siRNAs targeting SChLAP1 as well as control non-targeting siRNAs. These profiling experiments were run in technical triplicate for a total of 12 arrays (6 from 22Rv1 and 6 from LNCaP). Additionally, we profiled 22Rv1 and LNCaP cells treated with independent siRNAs targeting SWI/SNF component SMARC81 as well as control non-targeting siRNAs. These profiling experiments were run as biological duplicates for a total of four arrays (two cell lines × two independent siRNAs × one protein). Finally, we profiled RWPE cells expressing two different SChLAP1 isoforms as well as the control LacZ gene. These profiling experiments were run in technical duplicate for a total of four arrays (two from RWPE–SChLAP1 isoform 1 and two from RWPE–SChLAP1 isoform 2).

Processing to determine ranked gene expression lists. All of the microarray data were represented as log2 fold change between targeting versus control siRNAs. We used the CollapseDataset tool provided by the GSEA package to convert Agilent Probe IDs to gene symbols. Genes whose expression was measured by multiple probes were consolidated using the median values obtained with these probes. We then ran one-class SAM analysis from the Multi-Experiment Viewer application and ranked all genes by the difference between observed versus expected statistics. These ranked gene lists were imported to GSEA version 2.07.

SChLAP1 siRNA knockdown microarrays. For the 22Rv1 and LNCaP SChLAP1 knockdown experiments, we ran the GseaPreRanked tool to discover enriched gene sets in MSigDB57 version 3.0. Lists of positively and negatively enriched concepts were interpreted manually.
SMARCB1 siRNA knockdown microarrays. For each SMARCB1 knockdown experiment, we nominated genes that were altered by an average of at least two-fold. These signatures of putative SNF5 target genes were then used to assess enrichment of SChLAP1-regulated genes using the GseaPre Ranked tool. Additionally, we nominated genes whose expression changed by an average of two-fold or greater across SMARCB1 knockdown experiments and quantified the enrichment for SChLAP1 target genes using GSEA.

RWPE SChLAP1 expression microarrays. RWPE-SChLAP1 versus RWPE- LacZ expression profiles were ranked using SAM analysis as described above. A total of 1,245 genes were significantly over- or underexpressed and are shown in Supplementary Table 6d. A q value of 0.0 in this SAM analysis signifies that no permutation generated a more significant difference between observed and expected gene expression ratios. The ranked gene expression list was used as input for the GseaPre Ranked tool and compared against SNF5 ChIP-seq promoter peaks that decreased by >2-fold in RWPE cells overexpressing SChLAP1. Of the 389 genes in the ChIP-seq gene set, 250 were profiled by the Agilent HumanGenome microarray chip and were present in the GSEA gene symbol database. The expression profile across these 250 genes is shown in Supplementary Table 6c.

RNA-seq data. We assembled an RNA-seq cohort from prostate cancer tissues sequenced at multiple institutions. We included data from 12 primary tumors and 5 benign tissues published in GEO (GSE21269)27, from 16 primary tumors and 3 benign tissues released in the database of Genotypes and Phenotypes (dbGAP) (phs000310.v1.p1)38 and from 17 benign, 57 primary and 14 metastatic tumors sequenced by our own institution and released in dbGAP (phs000443.v1.p1). Sample information is shown in Supplementary Table 1a, and sequencing library information is shown in Supplementary Table 1b.

RNA-seq alignment and gene expression quantification. Sequencing data were aligned using TopHat29 version 1.3.1 against the Ensembl GRCh37 human genome build. Known introns (Ensembl release 63) were provided to TopHat. Gene expression across genes in Ensembl version 63 and the SChLAP1 transcript was quantified by HT-Seq version 0.5.3p3 using the script ‘htseq-count’. Reads were counted without respect to strand to avoid bias between unstranded and strand-specific library preparation methods. This bias results from the inability to resolve reads in regions where two genes on opposite strands overlap in the genome.

RNA-seq differential expression analysis. Differential expression analysis was performed using R package DESeq50 version 1.6.1. Read counts were normalized using the ‘estimateSizeFactors’ function, and variance was modeled by the ‘estimateDispersions’ function. Statistics on differential expression were computed by the ‘binomTest’ function. We called differentially expressed genes by imposing adjusted P-value cutoffs for cancer versus benign samples (Padj < 0.05), metastasis versus primary samples (Padj < 0.05) and Gleason score of 8+ versus 6 (Padj < 0.10). Heatmap visualizations of these analyses are presented as Supplementary Figure 5.

RNA-seq correlation analysis. Read count data were normalized using functions from the R package DESeq version 1.6.1. Adjustments for library size were made using the ‘estimateSizeFactors’ function, and variance was modeled using the ‘estimateDispersions’ function using the parameters ‘method=blind’ and ‘sharingMode=fit-only’. Next, raw read count data were converted to pseudocounts using the ‘getVarianceStabilizedData’ function. Gene expression levels were then mean centered and standardized using the ‘scale’ function in R. Pearson’s correlation coefficients were computed between each gene of interest and all other genes. Statistical significance of Pearson’s correlations was determined by comparison to correlation coefficients achieved with 1,000 random permutations of the expression data. We controlled for multiple-hypothesis testing using the ‘qvalue’ package in R. The SChLAP1 correlation signature of 253 genes was determined by imposing a cutoff of q < 0.05.

Oncomine concepts analysis of the SChLAP1 signature. We separated the 253 genes with expression levels significantly correlated with SChLAP1 into positively and negatively correlated gene lists. We imported these gene lists into Oncomine as custom concepts. We then nominated significantly associated prostate cancer concepts with odds ratio > 3.0 and P < 1 × 10⁻⁶. We exported these results as the nodes and edges of a concept association network and visualized the network using Cytoscape version 2.8.2. Node positions were computed using the Force-Directed Layout algorithm in Cytoscape using the odds ratio as the edge weight. Node positions were subtly altered manually to enable better visualization of node labels.

Association of correlation signatures with Oncomine concepts. We applied our RNA-seq correlation analysis procedure to the genes SChLAP1, EZH2, PCA3, AMACR and ACTB. For each gene, we created signatures from the top 5% of positively and negatively correlated genes (Supplementary Table 3). We performed a large meta-analysis of these correlation signatures across Oncomine data sets corresponding to disease outcome (Glinsky prostate and Setlur Prostate), metastatic disease (Holzbeierlein Prostate, Lapointe Prostate, LaTulippe Prostate, Taylor Prostate 3, Vanaja Prostate, Varambally Prostate and Yu Prostate), advanced Gleason score (Bittner Prostate, Glinsky Prostate, Lapointe Prostate, LaTulippe Prostate, Setlur Prostate, Taylor Prostate 3 and Yu Prostate) and localized cancer (Arredouani Prostate, Holzbeierlein Prostate, Lapointe Prostate, LaTulippe Prostate, Taylor Prostate 3, Varambally Prostate and Yu Prostate). We also incorporated our own concept signatures for metastasis, advanced Gleason score and localized cancer determined from our RNA-seq data. For each concept, we downloaded the gene signatures corresponding to the top 5% of genes up- and downregulated. Pairwise signature comparisons were performed using a one-sided Fisher’s exact test. We controlled for multiple-hypothesis testing using the ‘qvalue’ package in R. We considered concept pairs with q < 0.01 and odds ratio > 2.0 as significant. In cases where a gene signature associated with both the over- and underexpression gene sets from a single concept, only the most significant result (as determined by odds ratio) is shown.

Analysis of SChLAP1 and SMARCB1 expression signatures. Gene signatures obtained with knockdown of SChLAP1 and SMARCB1 were generated from Agilent gene expression microarray data sets. For each cell line, we obtained a single vector per gene fold changes by averaging technical replicates and then taking the median across biological replicates. We merged the results from individual cell line using the median of the changes in 22Rv1 and LNCaP cells. Venn diagram plots were produced using the BioVenn website60. We then compared the top 10% of upregulated and downregulated genes with knockdown of SChLAP1 and SMARCB1 to gene signatures downloaded from the Taylor Prostate 3 data set in the Oncomine database. We performed signature comparison using one-sided Fisher’s exact tests and controlled for multiple testing using the R package ‘qvalue’. Signature comparisons with q < 0.05 were considered significantly enriched. We plotted the odds ratios from significant comparisons using the ‘heatmap.2’ function in the ‘ggplot2’ R package.

Kaplan-Meier survival analysis based on the SChLAP1 gene signature. We downloaded prostate cancer expression profiling data and clinical annotations from GSE8402, published by Setlur et al.17. We intersected the 253-gene SChLAP1 signature with the genes in this data set and found 80 genes in common. We then assigned SChLAP1 expression scores to each patient sample in the cohort using the unweighted sum of standardized expression levels across the 80 genes. Given that we observed SChLAP1 expression in approximately 20% of prostate cancer samples, we used the 80th percentile of SChLAP1 expression scores as the threshold for ‘high’ versus ‘low’ scores. We then performed 10-year survival analysis using the ‘survival’ package in R and computed statistical significance using the log-rank test.

Additionally, we imported the 253-gene SChLAP1 signature into Oncomine to download the expression data for 167 of the 253 genes profiled by the Glinsky prostate data set56. We assigned SChLAP1 expression scores in a similar fashion and designated the top 20% of patients as having ‘high’ SChLAP1 scores. We performed survival analysis using the time to biochemical prostate-specific antigen (PSA) recurrence and computed statistical significance as described above.

PhyloCSF analysis. We obtained 46-way multi-alignment FASTA files for SChLAP1, HOTAIR, GAPDH and ACTB using the ‘Stitch Gene blocks’ tool.
within the Galaxy bioinformatics framework. We evaluated each gene for the likelihood that it represented a protein-coding region using PhyloCSF software (version released 28 October 2012). Each gene was evaluated using the phylogeny from 29 mammals (available by default within PhyloCSF) in any of the 3 reading frames. Scores are measured in decibans and represent the likelihood ratio that a sequence is protein-coding rather than noncoding.

Mayo Clinic cohort analyses. Subjects were selected from a cohort of individuals from the Mayo Clinic with high-risk prostate cancer who had undergone radical prostatectomy. The cohort was defined as 1,010 men with high-risk prostate cancer who underwent radical prostatectomy between 2000 and 2006, of whom 73 developed clinical progression (defined as individuals with systemic disease as evidenced by positive bone or computed tomography (CT) scan)45. High risk of recurrence was defined by preoperative PSA levels of >20 ng/ml, pathological Gleason score of 8–10, seminal vesicle invasion (SVI) or Gleason, PSA, seminal vesicle and margin (GPSM) score of ≥10 (ref. 62). The subcohort incorporated all 73 subjects with clinical progression to systemic disease and a random sampling of 20% of the entire cohort (202 men, including 19 with clinical progression). The total case-cohort study included 256 subjects, and tissue specimens were available from 235 subjects. The subcohort was previously used to validate a genomic classifier for predicting clinical progression65.

Tissue preparation. Formalin-fixed, paraffin-embedded samples of human prostate adenocarcinoma prostatectomies were collected from subjects with informed consent at the Mayo Clinic according to an IRB-approved protocol. Pathological review of tissue sections stained with hematoxylin and eosin was used to guide macrodissection of the tumor from surrounding stromal tissue in three to four 10-µm sections. The index lesion was considered as the dominant lesion by size.

RNA extraction and microarray hybridization. For the validation cohort, total RNA was extracted and purified using a modified protocol for the commercially available RNasy FFPE nucleic acid extraction kit (Qiagen). RNA concentrations were calculated using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies). Purified total RNA was subjected to whole-transcriptome amplification using the WT-Ovation FFPE system according to the manufacturer's recommendations with minor modifications (NuGen). For the validation, only the Ovation FFPE WTA System was used. Amplified products were fragmented and labeled using the Encore Biotin Module (NuGen) and hybridized to Affymetrix Human Exon (HuEx) 1.0 ST GeneChips following the manufacturer's recommendations.

Microarray expression analysis. Normalization and summarization of the microarray samples was performed with the frozen Robust Multiarray Average (RMA) algorithm using custom frozen vectors. These custom vectors were created using the vector creation methods described previously63. Quantile normalization and robust weighted average methods were used for normalization and summarization, respectively, as implemented in RMA.

Statistical analysis. Given the exon-intron structure of isoform 1 of SChLAP1, all probe selection regions (or PSRs) that fell within the genomic span of SChLAP1 were inspected for overlap with any of the exons of this gene. One PSR, 2518129, was found to be fully nested within exon 3 of SChLAP1 and was used for further analysis as a representative PSR for this gene. The PAM (Partition Around Medoids) unsupervised clustering method was used on the expression values of all clinical samples to define two groups with high and low expression of SChLAP1.

Statistical analysis on the association of SChLAP1 with clinical outcomes was carried out using three end-points: (i) biochemical recurrence, defined as two consecutive increases in serum PSA of ≥0.2 ng/ml after radical prostatectomy; (ii) clinical progression, defined as a positive CT or bone scan; and (iii) prostate cancer–specific mortality.

For the clinical progression end point, all subjects with clinical progression were included in the survival analysis, whereas controls in the subcohort were weighted in a fivefold manner to be representative of individuals from the original cohort. For the prostate cancer–specific mortality end point, cases who did not die from prostate cancer were omitted, and weighting was applied in a similar manner. For biochemical recurrence, because the case cohort was designed on the basis of the clinical progression end point, resampling of subjects with biochemical recurrence and the subcohort was performed to have a representative of the selected individuals with biochemical recurrence from the original cohort.

Transcriptome Sequencing Identifies PCAT-1, a Novel lincRNA Implicated in Prostate Cancer Progression

John R. Prensner1,8, Matthew K. Iyer1,8, O. Alejandro Balbin1, Saravana M. Dhanasekaran1,2, Qi Cao1, J. Chad Brenner1, Bharathai Laxman3, Irfan Asangani1, Catherine Grasso1, Hal D. Kominsky1, Xuhong Cao1, Xiaoqiang Jing1, Xiaoju Wang1, Javed Siddiqui1, John T. Wei4, Daniel Robinson1, Hari K. Iyer5, Nallasivam Palanisamy1,2,6, Christopher A. Maher1,2, and Arul M. Chinnaiyan1,2,4,6,7

1Michigan Center for Translational Pathology, University of Michigan Medical School, Ann Arbor, Michigan 48109
2Department of Pathology, University of Michigan Medical School, Ann Arbor, Michigan 48109
3Department of Medicine, University of Chicago, Chicago, Illinois 60637
4Department of Urology, University of Michigan Medical School, Ann Arbor, Michigan 48109
5Department of Statistics, Colorado State University, Fort Collins, Colorado 80523
6Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, Michigan 48109
7Howard Hughes Medical Institute, University of Michigan Medical School, Ann Arbor, Michigan 48109

Abstract

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. We nominated 121 such Prostate Cancer Associated Transcripts (PCATs) with cancer-specific expression patterns. Among these, we characterized PCAT-1 as a novel prostate-specific regulator of cell proliferation and target of the Polycomb Repressive Complex 2 (PRC2). We further found that high PCAT-1 and PRC2 expression stratified patient tissues into molecular subtypes distinguished by expression signatures of PCAT-1-repressed target genes. Taken together, the findings presented herein identify PCAT-1 as a novel transcriptional repressor.
implicated in subset of prostate cancer patients. These findings establish the utility of RNA-Seq to identify disease-associated ncRNAs that may improve the stratification of cancer subtypes.

**Keywords**

prostate cancer; transcriptome; next generation sequencing; non-coding RNA; EZH2

**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. 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, or *de novo* assembly of read sequences followed by sequence alignment. These methods provide a powerful framework to uncover uncharacterized RNA species, including antisense transcripts, short RNAs <250 bps, or long ncRNAs (lincRNAs) >250 bps.

While still largely unexplored, ncRNAs, particularly lincRNAs, have emerged as a new aspect of biology, with evidence suggesting that they are frequently cell-type specific, contribute important functions to numerous systems, and may interact with known cancer genes such as *EZH2*. Indeed, several well-described examples, such as *HOTAIR* and *ANRIL*, indicate that ncRNAs may be essential actors in cancer biology, typically facilitating epigenetic gene repression via chromatin modifying complexes. Moreover, ncRNA expression may confer clinical information about patient outcomes and have utility as diagnostic tests. 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 lincRNAs 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. Notably, we discover *PCAT-1*, a novel prostate cancer ncRNA alternately demonstrating either repression by PRC2 or an active role in promoting cell proliferation through transcriptional regulation of target genes. Our findings describe the first comprehensive study of lincRNAs in prostate cancer, provide a computational framework for large-scale RNA-Seq analyses, and describe *PCAT-1* as a novel prostate cancer ncRNA functionally implicated in disease progression.

**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* tumor suppressor gene and gain of oncogenic ETS transcription factor gene fusions 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 and recovered 1.41 billion mapped reads, with a median of 14.7 million mapped reads per sample (Supplementary Table 1 for sample information). We used the Cufflinks ab initio assembly approach3 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 (Fig. 1a and Supplementary Figs. 1 and 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 (Fig. 1b left and Supplementary Methods). The classifier demonstrated a sensitivity of 70.8% and specificity of 88.3% when trained using transcripts that overlapped genes in the AceView database20, including 11.7% of unannotated transcripts that were classified as “expressed” (Fig. 1b right). We then clustered the “expressed” transcripts into a consensus transcriptome and applied additional heuristic filters to further refine the assembly (Supplementary Methods). The final ab initio transcriptome assembly yielded 35,415 distinct transcriptional loci (Supplementary Table 2 and Supplementary Methods).

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 (Fig. 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” (Fig. 2a). These included partially intronic antisense (2.44%), totally intronic (12.1%), and intergenic transcripts (5.25%), consistent with previous reports of unannotated transcription.21, 22, 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 (Fig. 2b), had greater overlap with expressed sequence tags (ESTs) (Supplementary Fig. 3), and displayed a clear but subtle increase in conservation over randomly permuted controls (novel intragenic transcripts p = 2.7 × 10^{-4} ± 0.0002 for 0.4 < ω < 0.8; novel intronic transcripts p = 2.6 × 10^{-5} ± 0.0017 for 0 < ω < 0.4, Fisher's exact test, Fig. 2c). 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 (Fig. 2c, 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 (Supplementary Fig. 4).

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 lines25, VCaP and LNCaP (Supplementary Table 3). After filtering our dataset for transcribed repetitive elements known to display alternative patterns of histone modifications26, 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 (Fig. 2d-g) but not H3K4me1, which characterizes
enhancer regions\textsuperscript{27} (Supplementary Figs. 5 and 6). Interestingly, intergenic ncRNAs showed greater enrichment compared to intronic ncRNAs in these analyses (Fig. 2d-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\textsuperscript{28}, HPN\textsuperscript{29}, and PCA3\textsuperscript{14} (Fig. 2h, Supplementary Fig. 2 and Supplementary Table 4). 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\textsuperscript{6-9}, we identified aberrantly expressed uncharacterized ncRNAs in prostate cancer. We found a total of 1,859 unannotated lincRNAs throughout the human genome. Overall, these intergenic RNAs resided approximately half-way between two protein coding genes (Supplementary Fig. 7), and over one-third (34.1%) were ≥10kb from the nearest protein-coding gene, which is consistent with previous reports\textsuperscript{30} and supports the independence of intergenic ncRNAs genes. For example, visualizing the Chr15q arm using the Circos program (http://mkweb.bcgsc.ca/circos) illustrated genomic positions of eighty-nine novel intergenic transcripts, including one differentially-expressed gene centromeric to TLE3 (Supplementary Fig. 8).

A focused analysis of the 1,859 unannotated intergenic RNAs yielded 106 that were differentially expressed in localized tumors (FDR < 0.05, Fig. 3a). A cancer outlier expression analysis (Supplementary Methods) similarly nominated numerous unannotated ncRNA outliers (Fig. 3b) as well as known prostate cancer outliers, such as ERG\textsuperscript{18}, ETV1\textsuperscript{17, 18}, SPINK1\textsuperscript{31} and CRISP3\textsuperscript{32}. Merging these results produced a set of 121 unannotated transcripts that accurately discriminated benign, localized tumor, and metastatic prostate samples by unsupervised clustering (Fig. 3a). Indeed, clustering analyses using novel ncRNA outliers also suggested disease subtypes (Supplementary Fig. 9). 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 (Supplementary Tables 5 and 6).

### 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) (Supplementary Fig. 10). qPCR for four transcripts (PCAT-114, PCAT-14, PCAT-43, PCAT-1) on two independent cohorts of prostate tissues confirmed predicted cancer-specific expression patterns (Fig. 3c-f and Supplementary Fig. 11). 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 (Supplementary Table 8). This is further supported by expression analysis of these transcripts in our RNA-Seq compendium of 13 tumor types, representing 325 samples (Supplementary Fig. 12). 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\textsuperscript{17} (Supplementary Fig. 13). PCAT-1 and PCAT-14 also showed cancer-specific upregulation when tested on a panel of matched tumor-normal samples (Supplementary Fig. 14).
Of note, PCAT-114, which ranks as the #5 best outlier, just ahead of ERG (Fig. 3b and Supplementary Table 7), 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) (Supplementary Fig. 15). 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.

**PCAT-1, a novel prostate cancer lincRNA**

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 (Supplementary Fig. 16), whereas PCAT-1 incorporates portions of a mariner family transposase, an Alu, and a viral long terminal repeat (LTR) promoter region (Fig. 4a and Supplementary Fig. 17). While PCAT-14 was upregulated in localized prostate cancer but largely absent in metastases (Fig. 3c), PCAT-1 was strikingly upregulated in a subset of metastatic and high-grade localized (Gleason score ≥7) cancers (Fig. 3f and Supplementary Fig. 11). 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 phenotype35, 36. Surprisingly, we found that PCAT-1 and EZH2 expression were nearly mutually exclusive (Fig. 4b), 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 (Supplementary Fig. 18). Next, since Chr8q24 is known to harbor prostate cancer-associated single nucleotide polymorphisms (SNPs) and to exhibit frequent chromosomal amplification37-42, 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 amplification39, 42, by qPCR. Neither c-MYC nor NCOA2 levels showed striking expression relationships to PCAT-1, EZH2, or each other (Supplementary Fig. 19). 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 (Supplementary Figs. 20 and 21).

**PCAT-1 Function and Regulation**

Despite reports showing that upregulation of the ncRNA HOTAIR participates in PRC2 function in breast cancer9, we do not observe strong expression of this ncRNA in prostate (Supplementary Fig. 22), 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 (Fig. 4c,d), as did treatment of VCaP cells with the demethylating agent 5’deoxyazacytidine, the histone deacetylase inhibitor SAHA, or both (Fig. 4e). Chromatin immunoprecipitation (ChIP) assays also demonstrated that SUZ12, a core PRC2 protein, directly binds the PCAT-1 promoter approximately 1kb upstream of the TSS (Fig. 4f). Interestingly, RNA immunoprecipitation (RIP) similarly showed binding of PCAT-1 to SUZ12 protein in VCaP cells (Supplementary Fig. 23a). RIP assays followed by RNase A, RNase H, or DNase I treatment either abolished, partially preserved, or totally preserved this
interaction, respectively (Supplementary Fig. 23b). 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 (Fig. 5a and Supplementary Fig. 24). Next, we designed siRNA oligos to PCAT-1 and performed knockdown experiments in LNCaP cells, which express higher levels of PCAT-1 without PRC2-mediated repression (Supplementary Fig. 25). 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 (Fig. 5b), but not control DU145 cells lacking PCAT-1 expression (Supplementary Fig. 26) or VCaP cells, in which PCAT-1 is expressed but repressed by PRC2 (Supplementary Fig. 27).

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) (Supplementary Fig. 28 and Supplementary Table 9). 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 (Fig. 5c and Supplementary Table 10). These results suggest that PCAT-1’s function is predominantly repressive in nature, similar to other lincRNAs. We next validated expression changes in three key PCAT-1 target genes (BRCA2, CENPE and CENPF) whose expression is upregulated upon PCAT-1 knockdown (Fig. 5a) 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 (Fig. 6a), 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 (Fig. 6b).

Moreover, comparing EZH2-outlier and PCAT-1-outlier patients (see Fig. 4b), 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 (Fig. 6c). Network analysis of the top 20 upregulated genes following PCAT-1 knockdown with the HefaLMP tool further suggested that these genes form a coordinated network (Fig. 6d), corroborating our previous observations. Taken together, these results provide initial data into the composition and function of the prostate cancer ncRNA 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 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, and broadly our study confirms the utility of RNA-Seq in defining functionally-important elements of the genome.

Of particular interest is our discovery of the prostate-specific ncRNA gene PCAT-1, which is markedly overexpressed in a subset of prostate cancers, particularly metastases, and may contribute to cell proliferation in these tumors. 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 (Fig. 6e). 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 PCA. 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Kalpana Ramnarayanan and Roger Morey for technical assistance with next generation sequencing. We thank Robert J. Lonigro, Shanker Kaylana-Sundaram, Terrence Barrette, and Mike Quist for help with sequencing data analysis, and Rohit Mehra, Bo Han, and Khalid Suleman for prostate tissue specimens. We thank Cole Trapnell and Geo Pertea for assistance with computational analyses. We thank Scott Tomlins, Yi-Mi Wu, Sameek Roychowdhury and members of the Chinnaiyan lab for advice and discussions. We thank Rameen Beroukhim for guidance.

This work was supported in part by the NIH Prostate Specialized Program of Research Excellence grant P50CA69568, the Early Detection Research Network grant U01 CA1111275 (to A.M.C), the US National Institutes of Health R01 CA132874-01A1 (to A.M.C.), the Department of Defense grant PC100171 and W81XWH-11-1-0337 (to A.M.C.) and the National Center for Functional Genomics supported by the Department of Defense (to A.M.C.). A.M.C. is supported by the Doris Duke Charitable Foundation Clinical Scientist Award, a Burroughs Welcome Foundation Award in Clinical Translational Research and the Prostate Cancer Foundation. A.M.C. is an American
Online Methods

Cell lines, treatments, and tissues

All prostate cell lines were obtained from the American Type Culture Collection (Manassas, VA), except for PrEC (benign non-immortalized prostate epithelial cells) and PrSMC (prostate smooth muscle cells), which were obtained from Lonza (Basel, Switzerland). Cell lines were maintained using standard media and conditions.

For androgen treatment experiments, LNCaP and VCaP cells were grown in androgen-depleted media for 48 hours and subsequently treated with 5nM methyltrienolone (R1881, NEN Life Science Products) or an equivalent volume of ethanol for 48 hours before harvesting the cells. For drug treatments, VCaP cells were treated with 20μM 5’deoxyazacytidine (Sigma), 500 nM 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. For DZNep treatments, DZNep was dissolved in DMSO and VCAP cells were treated with either 0.1μM of DZNep or vehicle control; RNA was harvested at 72 hours and 144 hours.

Prostate tissues were obtained from the radical prostatectomy series and Rapid Autopsy Program at the University of Michigan tissue core as 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.

RNA isolation; cDNA synthesis; and PCR experiments

Total RNA was isolated using Trizol and an RNasy 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. Reverse-transcription PCR (RT-PCR) was performed with Platinum Taq High Fidelity polymerase (Invitrogen). All oligonucleotide primers are listed in Supplementary Table 12.

For PCR product sequencing, PCR products were resolved on a 1.5% agarose gel, and either sequenced directly or 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.

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 Supplementary Table 12.

**RNA-Seq library preparation**

2 μg total RNA was selected for polyA+ RNA using Sera-Mag oligo(dT) beads (Thermo Scientific), and paired-end next-generation sequencing libraries were prepared as previously described using Illumina-supplied universal adaptor oligos and PCR primers (Illumina). Samples were sequenced in a single lane on an Illumina Genome Analyzer I or Genome Analyzer II flowcell using previously described protocols. 36-45mer paired-end reads were according to the protocol provided by Illumina.

**Overexpression studies**

*PCAT-1* full length transcript was cloned into the pLenti6 vector (Invitrogen) along with RFP and LacZ controls. After confirmation of the insert sequence, lentiviruses were generated at the University of Michigan Vector Core and transfected into the benign immortalized prostate cell line RWPE. RWPE cells stably expressing *PCAT-1*, RFP or LacZ were generated by selection with blasticidin (Invitrogen), and 10,000 cells were plated into 12-well plates. Cells were harvested and counted at day 2, day 4, and day 6 post-plating with a Coulter counter.

**siRNA knockdown studies**

Cells were plated 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 UUAAAGAGAUCACAGUUAUU; siRNA 2 GCAGAAACACCAUGGAUAUU; siRNA 3 AUACAUAAGACCAUGGAAUU; siRNA 4 GAACCUAACUGGACUUAAUU. For *EZH2* siRNA, the following sequence was used: GAGGUUCAGACGCUGAUU.

**shRNA knockdown and western blotting**

Cells were seeded at 50-60% confluency, incubated overnight, and transfected with *EZH2* or non-targeting shRNA lentiviral constructs as described in for 48 hours. GFP+ cells were drug-selected using 1 ug/mL puromycin. RNA and protein were harvested for PCR and Western blotting according to standard protocols. For Western blotting, PVDF membranes (GE Healthcare) were incubated overnight at 4C with either *EZH2* mouse monoclonal (1:1000, BD Biosciences, no. 612666), or *B-Actin* (Abcam, ab8226) for equal loading.

**Gene expression profiling**

Agilent Whole Human Genome Oligo Microarray (Santa Clara, CA) was used for cDNA profiling of *PCAT-1* siRNA knockdown samples or non-targeting control according to standard protocols. All samples were run in technical triplicates against non-targeting control siRNA. Expression array data was processed using the SAM method with an FDR ≤ 0.01. Up- and down-regulated probes were separated and analyzed using the DAVID bioinformatics platform.
Chromatin immunoprecipitation

ChIP assays were preformed as previously described\textsuperscript{25}, where 4 – 7 μg of the following antibodies were used: IgG (Millipore, PP64), SUZ12 (Cell Signaling, #3737), and SUZ12 (Abcam, ab12073). ChIP-PCR reactions were performed in triplicate with SYBRGreen using 1:150\textsuperscript{th} of the ChIP product per reaction.

\textit{In vitro} translation

Full length PCAT-1, Halo-tagged ERG or GUS positive control were cloned into the PCR2.1 entry vector (Invitrogen) and \textit{in vitro} translational assays were performed using the TnT Quick Coupled Transcription/Translation System (Promega) with 1mM methionine and Transcend Biotin-Lysyl-tRNA (Promega) according to the manufacturer’s instructions.

Bioinformatic analyses

Sequencing reads were aligned with TopHat\textsuperscript{19}, and \textit{ab initio} assembly was performed with Cufflinks\textsuperscript{3}. Transcriptome libraries were merged and statistical classifiers were developed and employed to filter low confidence transcripts. Nominated transcripts were compared to UCSC, RefSeq, Vega, Ensembl, and ENCODE database, and coding potential was determined with the txCdsPredict program from UCSC. Transcript conservation was determined with the SiPhy package. Differential expression analysis was performed using SAM methodology, and outlier analysis using a modified COPA method. See the Supplementary Methods for details on the bioinformatics methods used.

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. For details regarding the statistical methods employed during RNA-Seq and ChIP-Seq data analysis, see Supplementary Methods.

References


Figure 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. The merged consensus transcriptome was generated by compiling all individual transcriptome libraries and using a decision tree classifier in order to define high confidence “expressed” transcripts and low confidence “background” transcripts, which were discarded. The example decision tree on the left was produced from transcripts on chromosome 1. The graphics on the right provide a fictional example demonstrating the informatics filtration pipeline. (c) Following informatic processing and filtration of the sequencing data, transcripts were categorized in order to identify unannotated ncRNAs. Transcribed pseudogenes were isolated, and the remaining transcripts were categorized based on overlap with an aggregated set of known...
gene annotations into annotated protein coding, non-coding, and unannotated. Both annotated and unannotated ncRNA transcripts were then separated into intronic, intergenic, and antisense categories based on their relationship to protein coding genes.
Figure 2. 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).
Figure 3. 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), see Supplementary Table 8) 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 4. 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 5. *PCAT-1* promotes cell proliferation

(a) 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. (b) Cell proliferation assays in LNCaP using *PCAT-1* siRNAs. An asterisk (*) indicates $p \leq 0.005$ by a two-tailed Students t-test. (c) 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 6. 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.