The role of long noncoding RNA *SChLAP1* in prostate cancer

by

Anirban Sahu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular and Cellular Pathology)
in the University of Michigan
2015

Doctoral Committee:

Professor Arul M. Chinnaiyan, Chair
Professor David G. Beer
Professor Gregory R. Dressler
Professor David R. Engelke
Assistant Professor Sundeep Kalantry

To my family, thank you for believing in me.

Acknowledgements

Graduate school has been a unique journey filled with exciting moments, difficult challenges, and unforgettable memories. While I have learned a lot about science over the last four years, I have learned even more about myself. I would like to take this opportunity to thank all of the individuals who have helped me throughout this process, providing guidance, mentorship, friendship, and support.

First, I would like to thank my mentor, Arul Chinnaiyan, who has given me every opportunity to succeed as a young scientist. He has pushed me to explore novel scientific areas, taught me how to ask relevant research questions, and equipped me with the resources and tools I need to answer them. Thank you for guiding me through this experience. Additionally, I would like to thank the members of my thesis committee, David Beer, Gregory Dressler, David Engelke, and Sundeep Kalantry, for useful discussions and helpful feedback throughout the course of my research.

I am extremely grateful to all of my colleagues and collaborators. This thesis represents the collective work of many talented individuals, and I would not be here without your help. In fact, this project would not exist without the prior work done by John Prensner and Matthew Iyer. John, thank you for mentoring me when I first joined the lab and teaching me many of the skills that I used

throughout my graduate training. Matthew, thank you for helping me with all aspects of my research. I am lucky to have made such a great friend during my time in the lab.

It has been a pleasure working with all members of the Chinnaiyan lab. I would like to especially thank Rohit Malik, Yasuyuki Hosono, and Yashar Niknafs for creating such a fun and intellectually stimulating environment in the long noncoding RNA group. Additionally, I would like to thank Felix Feng for help with all of my projects as well as discussions about my career and life. I would also like to thank my students, Ben Chandler, Nithin Edara, and Udit Singhal, for helping me complete these experiments and knowing what to do even when I did not. Thank you Xuhong Cao for organizing everything in lab and making this thesis work possible. Finally, thank you to all of my labmates for such a memorable experience: Brendan Veeneman, Alejandro Balbin, Chad Brenner, Scott Deroo, Irfan Asangani, Ram Mani, Qi Cao, Sameek Roychowdhury, Rohit Mehra, Mohan Dhanasekaran, Carl Engelke, Anton Poliakov, June Escara-Wilke, Marcin Cieslik, Yajia Zhang, Lanbo Xiao, Teng Ma, Sumin Han, Xiang Zhang, Sunita Shankar, Sudhanshu Shukla, George Zhao, and many, many others.

I would like to thank everybody who has provided administrative support over the years. Thank you Karen Giles and Christine Betts for all that you do to make things run so smoothly in MCTP. Thank you Dianna Banka and Jyoti Athanikar for helping me write grants and obtain funding. Thank you Zaneta Nikolovska-

Coleska and Laura Labut for helping all of the graduate students in MCP with everything that we need. Thank you Ron Koenig, Ellen Elkin, Hilkka Ketola, and Laurie Koivupalo for all of your kindness and support through MSTP.

None of this would be possible without the love and support of my friends and family. My parents have worked so hard and sacrificed so much to give me this opportunity. Thank you Mom, Dad, and Nev for always pushing me to be my best and encouraging me at every step of the way. Thank you to my in-laws, Amma, Appa, Sunitha, and Madhav, for giving me a second home in Chicago, treating me like one of your own, and supporting me throughout graduate school.

Finally, to my wife, Aneesha, thank you for everything that you have done for me. You are my best friend, my biggest fan, and you are the only reason I have made it this far. Thank you for your unconditional love and unwavering support. Thank you for dealing with me at all times, day or night, even when I am hungry. Thank you for listening and understanding. Thank you for believing in me and helping me achieve my dreams. I love you very much and hope that I have made you proud.

Table of Contents

List of Figures

List of Tables

Abstract

Prostate cancer is the most common malignancy in U.S. men, accounting for nearly 30,000 deaths annually. While the majority of prostate cancers are indolent, a subset of patients has aggressive disease. However, the molecular basis for this clinical heterogeneity remains incompletely understood.

Long noncoding RNAs (lncRNAs) are an emerging class of regulatory molecules implicated in a diverse range of human malignancies. Here, *SChLAP1* is identified as a novel, highly prognostic lncRNA that is expressed in 15-30% of prostate cancers. Functionally, *SChLAP1* coordinates cancer cell invasion *in vitro* and metastatic spread *in vivo*. Mechanistically, *SChLAP1* interacts with and antagonizes the tumor-suppressive SWI/SNF nucleosome-remodeling complex.

While deleterious SWI/SNF mutations occur in 20% of all cancers, they are relatively rare in prostate cancer. Within prostate cancer, SWI/SNF mutations are associated with low *SChLAP1* expression, suggesting that high *SChLAP1* expression may represent a mutation-independent modality of SWI/SNF inhibition. Employing a previously described antagonistic model between SWI/SNF and Polycomb Repressive Complex 2 (PRC2), *SChLAP1* is found to enhance PRC2 function in prostate cancer. Additionally, *SChLAP1*-expressing cells are more sensitive to pharmacologic EZH2 inhibition.

Further characterization of *SChLAP1* reveals a 250bp region near the 3'-end that mediates its invasive phenotype and coordinates its interaction with SWI/SNF. Additionally, *SChLAP1* interacts with BRG1-containing but not BRM-containing SWI/SNF complexes, and knockdown of BRM in *SChLAP1*-expressing cells exposes a synthetic lethal vulnerability in prostate cancer.

Finally, the largest biomarker discovery project to date in prostate cancer identifies *SChLAP1* as one of the best genes for predicting metastatic progression. Characterization of *SChLAP1* expression by *in situ* hybridization shows that *SChLAP1* expression is enriched in metastatic samples. Additionally, *SChLAP1* can be detected in patient urine samples and may be useful as a non-invasive biomarker. Lastly, targeting *SChLAP1* with antisense oligonucleotides (ASO) suggests that directly targeting *SChLAP1* may be an effective therapeutic strategy in prostate cancer.

Taken together, this work defines an essential role for *SChLAP1* in aggressive prostate cancer, uncovers novel aspects of lncRNA biology, and has broad implications for cancer biology.

Chapter 1:

A rationale to study long noncoding RNAs in prostate cancer

Prostate Cancer

Epidemiology

Prostate cancer is the most common non-cutaneous cancer in U.S. men and the second-leading cause of cancer-related death in U.S. men, behind only lung cancer[1]. Approximately 221,000 men will be diagnosed with prostate cancer this year and about 28,000 men will die from prostate cancer this year alone[1]. While 1 in 7 men will be diagnosed with prostate cancer in their lifetime, the majority of men have indolent disease that requires no treatment at all[2]. In fact, nearly 3 million U.S. men are living with prostate cancer today[3], and autopsy studies have found that up to 40% of all men, and over 70% of men over the age of 70, have latent prostate cancer[4,5].

The overall relative survival rates for prostate cancer are very good: 100% at 5 years, 99% at 10 years, and 94% at 15 years[6]. However, as the previous statistics suggest, the public health burden and disease-specific death of prostate cancer are substantial. Therefore, a better understanding of the biological basis behind aggressive, lethal prostate cancer is necessary, and treating those

patients with serious, life-threatening illness at an early stage is a major goal in the clinical management of the disease[7,8].

Risk factors

While one predominant risk factor has not been identified in prostate cancer (akin to smoking and lung cancer), several demographic and lifestyle factors have been associated with an increased risk of disease[9,10]. Age is the most important risk factor: the majority of prostate cancers are found in men above 65 years of age, and very rarely do men under the age of 40 have prostate cancer[11].

Additionally, race is a major factor, as African-American men are more likely to acquire and twice as likely to die from prostate cancer compared to Caucasian men[12-14]. In contrast, Asian-American and Hispanic men are least likely to develop prostate cancer[15,16]. However, the molecular basis behind these racial differences remains unclear[17,18], and differences in access to healthcare do not account for the observed variation[19]. Finally, a family history of prostate cancer increases the odds of developing the disease, with 5-10% of prostate cancers thought to be primarily caused by inherited factors[20-22].

Screening

While digital rectal exams were used for decades to detect prostate cancer, variability in clinician assessment and the predominance of advanced cancer upon detection made it an unsuitable tool for screening[23-25]. The introduction of

the prostate-specific antigen (PSA) serum test in the 1980s led to an increase in detected cases of prostate cancer (**Fig. 1.1**), but also increased risks (*see below*) without clear evidence of decreased mortality[23,26]. Additionally, PSA is not specific to cancer and can be elevated for a variety of reasons, including benign prostatic hyperplasia (BPH) and infection (prostatitis)[23,26]. Furthermore, the cutoffs used for PSA screening were far from perfect, with approximately 15% of men in the normal range (less than 4ng/mL) actually having prostate cancer and approximately 75% of men in the elevated range (4-10ng/mL) not having any sign of prostate cancer upon further testing[29,32]. While the U.S. Preventive Services Task Force (USPSTF) recommends against routine PSA-based prostate cancer screening[27], most clinicians suggest implementing PSA testing on an case-by-case basis[23,26]. In combination with physical exam and overall clinical assessment, an elevated PSA most often requires prostate tissue biopsy for further evaluation.

Precancerous lesions

Prostate cancer develops primarily from the glandular cells in the prostate that secrete fluids and are therefore called adenocarcinomas[28-31]. While cancers can arise from other cells in the prostate, they are extremely rare[32]. In this thesis, the term prostate cancer refers to prostate adenocarcinomas (PRAD).

Several precancerous conditions of the prostate may be detected on initial tissue biopsy (**Fig. 1.2**)[30,33]. Proliferative inflammatory atrophy (PIA) is a condition

where prostate cells look smaller than normal and are surrounded by signs of inflammation. Prostatic intraepithelial neoplasia (PIN) is a condition where prostate cells have an abnormal pattern under the microscope, but do not display other characteristics of cancer. Approximately 20% of patients with high-grade PIN will have a cancerous lesion in another region of the prostate and these patients should be more actively monitored[34].

Gleason score

The Gleason method for grading prostate tissues was developed in the 1960s and is based entirely on histologic patterns of carcinoma cells (**Fig. 1.3**)[35]. The Gleason grade ranges from 1 (normal) to 5 (abnormal), and the Gleason score is calculated as the sum of the two regions that account for the majority of a cancer. A Gleason score of 6 or below is considered low-grade and less aggressive. A Gleason score of 8-10 is considered high-grade, more likely to develop metastatic, lethal tumors, and should be treated aggressively[36,37]. The Gleason grade has stood the test of time as a prognostic indicator of prostate cancer, with direct relationships to histopathologic and clinical end points, including tumor size, margin status, progression to metastatic disease, and survival[35-38].

Molecular aberrations

Although prostate cancer genomes are relatively less mutated compared to other cancers[39], molecular characterization studies have revealed a complex, heterogeneous landscape of somatic mutations, gene deletions, gene

amplifications, and chromosomal rearrangements that define the disease[40-44]. *SPOP* has the highest rate of point mutations in prostate cancers, with 5-15% of cases altered in localized disease[40,43,45,46]. *TP53*, the most commonly mutated tumor suppressor in cancer, is deleted in 25-40% of prostate cancers and harbors point mutations in 5-40% of cases[40-42,44,46,47]. The PI3K pathway, which is also frequently altered in human cancer, is affected in 25-70% of prostate cancers[39]. *PTEN*, a tumor suppressor that regulates the PI3K pathway, is deleted in 40% of prostate cancers and mutated in 5-15% of cases[40-42,44,47].

In 2005, the discovery that approximately half of all prostate cancers harbor recurrent ETS (E26 transformation-specific) gene fusions revolutionized the molecular categorization of prostate cancers[48]. Typically, the 5' region of an androgen-regulated gene is fused to a member of the ETS transcription factor family. Most commonly, *TMPRSS2* is fused to *ERG*. *SPINK1*, a secreted protease, is overexpressed in a subset of ETS-positive prostate cancers and is typically associated with more aggressive disease[49]. Additionally, *EZH2*, the catalytic subunit of PRC2, is overexpressed in prostate cancer and associated with aggressive and metastatic disease[50].

As androgen signaling plays such a crucial role in prostate development and cancer progression[51], the mainstay of therapy targets the androgen signaling axis (*see below*). Following treatment, metastatic, castrate-resistant prostate cancers (CRPC) show genetic abnormalities in the androgen receptor (*AR*) gene[52-54].

While amplifications and point mutations are frequently found in treated, metastatic tumors, they are absent in localized disease[40,41,44], suggesting AR aberrations appear as a mechanism of resistance to these therapies.

Treatment Modalities

In addition to standard chemotherapeutic agents, several options are available for the initial management of prostate cancer. Radical prostatectomy, a surgical procedure to completely remove the prostate, is commonly used as first-line curative treatment for localized prostate cancer, although its benefit versus watchful waiting for localized disease is uncertain[55]. Radiation therapy is also implemented at early stages of the disease, and the combination of radiotherapy with androgen deprivation therapy (ADT, *see below*) has become the standard of care for men receiving either treatment[56]. Immunotherapy has recently emerged as an effective treatment modality for prostate cancer, and in 2011, Sipuleucel-T was the first FDA-approved vaccine for the management of cancer[57]. Additionally, bone targeting therapies, such as bisphosphonates and radium-223, have been extremely helpful as the majority of prostate cancer patients will develop bone metastases, leading to pain, deteriorated quality of life, and other serious complications[58].

The basis of therapy for recurrent, incurable metastatic prostate cancer is focused on targeting the androgen signaling pathway[59,60]. The goal of androgen deprivation therapy (ADT) is to reduce the level of androgen signaling in prostate

cells. This can be achieved using several approaches. Overall levels of androgen in the body can be reduced by surgical castration (orchiectomy), decreasing levels of gonadotropin-releasing hormone (LHRH) with drugs such as leuprolide[61], or inhibiting CYP17, an enzyme necessary for androgen biosynthesis, with abiraterone[62]. Drugs targeting the androgen receptor, such as bicalutamide, prevent androgen binding at the receptor[63]. Finally, the nuclear translocation of androgen receptor within cells can be blocked by enzalutamide[64].

Clinical challenges

While the utility of PSA testing for the detection of prostate cancer remains debatable, abnormal PSA tests lead to tissue biopsies, which may cause adverse effects such as pain, bleeding, or infection[65,66]. Additionally, undergoing biopsy can lead to psychological stress and cause anxiety in men[67].

Additionally, indolent prostate cancer continues to be widely overdiagnosed and overtreated[68-70], and the discrimination of aggressive from indolent prostate cancer remains one of the most important areas of research for this disease[8,71]. In particular, the ability to stratify patient outcomes at the time of prostate cancer screening or diagnosis has proven to be challenging. To distinguish between aggressive and indolent tumors, current clinical paradigms rely mainly on pre-operative PSA levels, tumor stage, and Gleason score in order to estimate patient risk[36,38,72-74]. Yet, these remain imperfect tools that inaccurately classify some patients[75-78]. Aggressive treatment of these indolent cancers is associated

with unnecessary risks of urinary, sexual, and bowel dysfunction, which can adversely affect a patient's quality of life[79].

Finally, although several drugs have recently been FDA-approved for the treatment of prostate cancer, therapies for lethal, metastatic disease continue to target the androgen signaling pathway. While decades of research have improved our understanding of the molecular basis underlying aggressive prostate cancer, there have been minimal changes in overall patient outcomes (**Fig. 1.1**). The majority of efforts to address these clinical challenges have focused on proteins; thus, exploration beyond protein-coding genes warrants further study.

The Genome

Definition of a gene

The central dogma of molecular biology states that DNA is transcribed into messenger RNAs, which serve as templates for protein synthesis[80]. For decades, proteins have been studied as the key players in cellular biology and disease pathogenesis. However, the completion of the Human Genome Project in 2001 showed that a surprisingly small number of approximately 25,000 "genes" coded for proteins in human DNA, accounting for only 1.5% of the entire genome[81]. The purpose of these noncoding regions of the genome remained unclear.

Characterizing the genome

In 2007, the Encyclopedia of DNA Elements (ENCODE) Consortium directed a comprehensive sequencing and annotation analysis to find that 60-70% of the genome was transcribed into RNA[82]. With only a small fraction of transcribed RNAs coding for proteins, uncovering the function of the myriad of noncoding transcripts became a major area of investigation. The development of Next-Generation Sequencing (NGS) allowed global, unbiased RNA sequencing (RNA-seq), and several groups began producing large amounts of RNA-seq data to study the transcriptome[83-85]. These efforts revealed that thousands of previously unannotated, noncoding RNAs existed within cells. While the study of noncoding RNAs has been ongoing for decades, the abundance of novel transcripts transformed our understanding of molecular biology and cellular function.

Long noncoding RNAs

Long noncoding RNAs (lncRNAs) are one class of noncoding transcripts. LncRNAs are RNA species >200bp in length commonly characterized by splicing of multiple exons, H3K4me3 promoter methylation, and transcription by RNA polymerase II[86,87]. Efforts to characterize long noncoding transcripts initially relied on these characteristics to define transcribed regions of the genome[87]. As bioinformatics tools improved and RNA-seq data became more readily available, computational tools were used to identify novel, unannotated transcripts[88].

LncRNA function

LncRNA-mediated biology has been implicated in a wide variety of cellular processes, including pluripotency in stem cells[89] and X chromosome inactivation[90]. While some lncRNAs, such as *Xist*, appear to operate exclusively in the nucleus as regulators of gene expression[90,91], other lncRNAs appear to function predominantly in the cytoplasm where they can regulate the stability of mRNAs[92-94].

Several mechanisms of lncRNA activity have been described (**Fig. 1.4**)[95]. Most prominently, lncRNAs have been shown to collaborate with protein partners to form ribonucleoprotein complexes. For example, *Xist* interacts with the Polycomb repressive complex 2 (PRC2), resulting in PRC2 recruitment and H3 lysine 27 trimethylation (H3K27me3) of the inactive X chromosome[96]. *Air* and *Kcnq1ot1* bind to G9a, a histone H3 lysine 9 methylase, to regulate gene expression[97,98]. *ANRIL* associates with PRC1 to regulate the INK4a locus[99]. *Linc-p21* and *PANDA* are two p53-regulated lncRNAs that interact with hnRNPK and NF-Y to regulate transcription. Given this propensity to engage protein complexes, lncRNAs are emerging as decoys, scaffolds, and guides[100].

Cancer-associated lncRNAs

In cancer, lncRNAs are emerging as a prominent layer of previously underappreciated transcriptional regulation, often by collaborating with epigenetic complexes such as Polycomb Repressive Complex 1[99,101] (PRC1) and Polycomb

Repressive Complex 2 (PRC2)[91,96,101-103], among others. Mechanistically, cancer-associated lncRNAs commonly serve as molecular scaffolds that enable recruitment of these complexes to specific genomic loci, allowing for specificity of function[103,104]. Clinically, overexpression of the HOTAIR lncRNA has been shown to correlate with aggressive breast[102], colon[105], hepatocellular[106], and gastrointestinal stromal tumors[107], suggesting that lncRNA-mediated biology plays a prominent role in cancer progression.

However, lncRNAs frequently display lineage-specific expression patterns[88,108], making generalizations about their functional contributions to a variety of contexts difficult. For example, *HOTAIR*, while an important component of breast cancer[102], is virtually absent in prostate cancer[108]. Additionally, in non-disease states, a muscle-specific lncRNA was recently reported as a lineage-specific regulator of muscle differentiation[109]. Therefore, our group sought to uncover the lncRNAs that mediate prostate cancer.

Prostate Cancer Associated Transcripts (PCATs)

Using RNA-seq on a panel of 102 prostate cancer samples and cell lines, our laboratory identified approximately 1,800 lncRNAs expressed in prostate tissues[108]. The majority of the transcripts corresponded to annotated protein coding genes and noncoding RNAs, but a substantial percentage (19.8%) lacked any overlap with previously discovered genes (**Fig. 1.5a**). Differential expression analysis identified 121 lncRNAs, termed PCATs (Prostate Cancer Associated

Transcripts), whose expression patterns distinguished benign, localized cancer and metastatic cancer samples (**Fig. 1.5b**). In that study, experimental work focused on *PCAT-1*, a 1.9-kb polyadenylated lncRNA comprising two exons and located in the Chr8q24 gene desert. While *PCAT-1* showed tissue-specific expression and was selectively upregulated only in prostate cancer, we wanted to identify those lncRNAs with high expression in the subset of aggressive cancers. The highest-ranked lncRNA candidate identified by outlier analysis was *SChLAP1*.

Goals of this thesis

In this thesis, we aim to elucidate the role of long noncoding RNA *SChLAP1* in prostate cancer. We begin by characterizing the transcript and identifying its function in prostate cancer progression. Then, we uncover its mechanism of action and interrogate aspects of *SChLAP1* biology to reveal several therapeutic vulnerabilities in prostate cancer. Finally, we explore translational opportunities of *SChLAP1* in aggressive prostate cancer, with the goal of addressing some of the clinical challenges in prostate cancer.

Figures

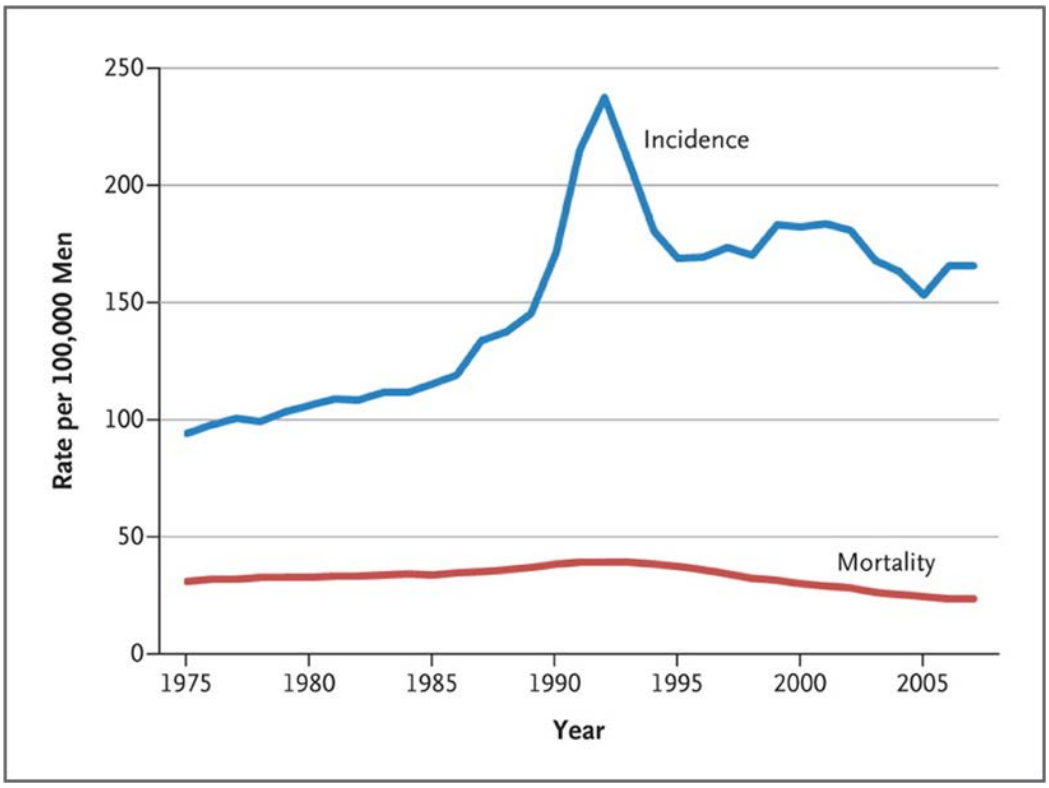


Figure 1.1 Age-Adjusted Incidence of and Mortality from Prostate Cancer in the United States, 1975–2007.

*Reproduced with permission from Hoffman, *N Engl J Med*, 2011. Copyright Massachusetts Medical Society.

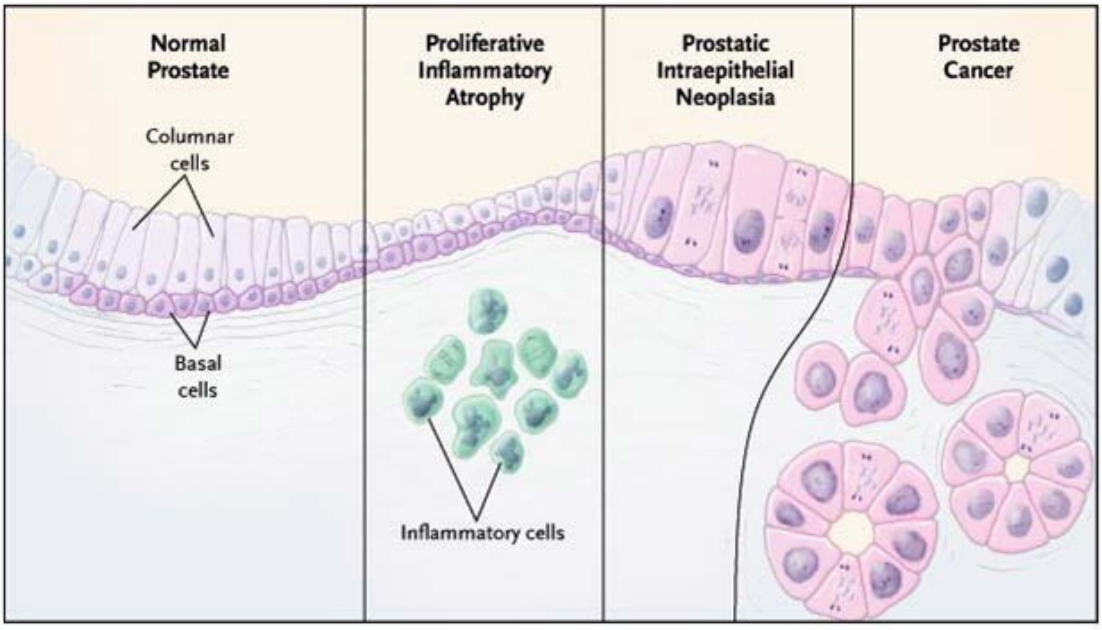


Figure 1.2 Proliferative Inflammatory Atrophy as a Precursor to Prostatic Intraepithelial Neoplasia and Prostate Cancer

*Reproduced with permission from Nelson *et al*., *N Engl J Med*, 2003. Copyright Massachusetts Medical Society.

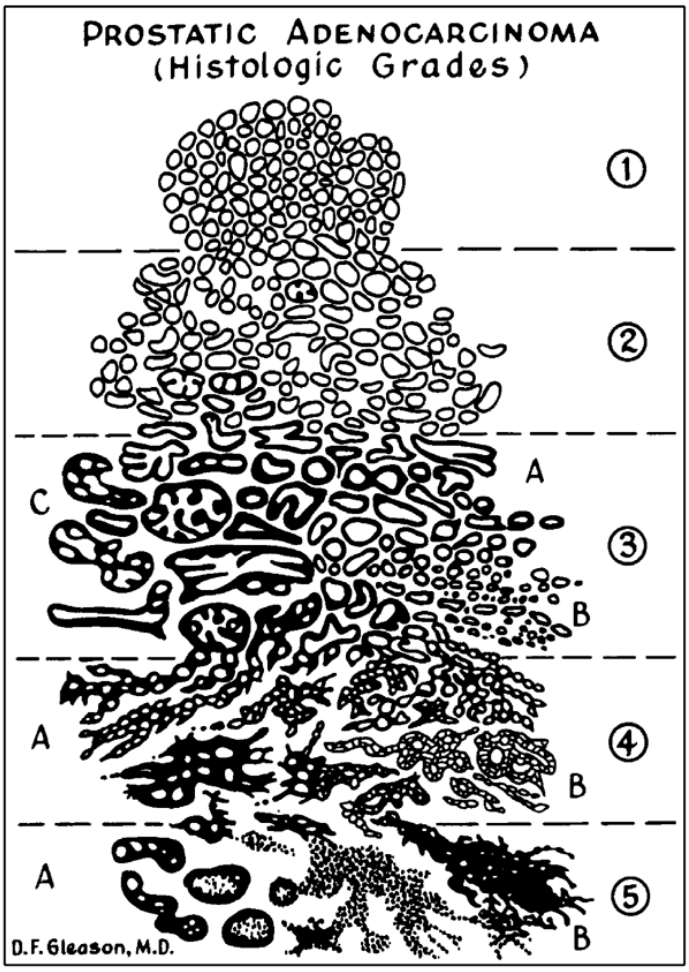


Figure 1.3 Gleason grades: standard drawing

*Reprinted by permission from Macmillan Publishers Ltd: Modern Pathology (Humphrey PA, Gleason grading and prognostic factors in carcinoma of the prostate), copyright 2004

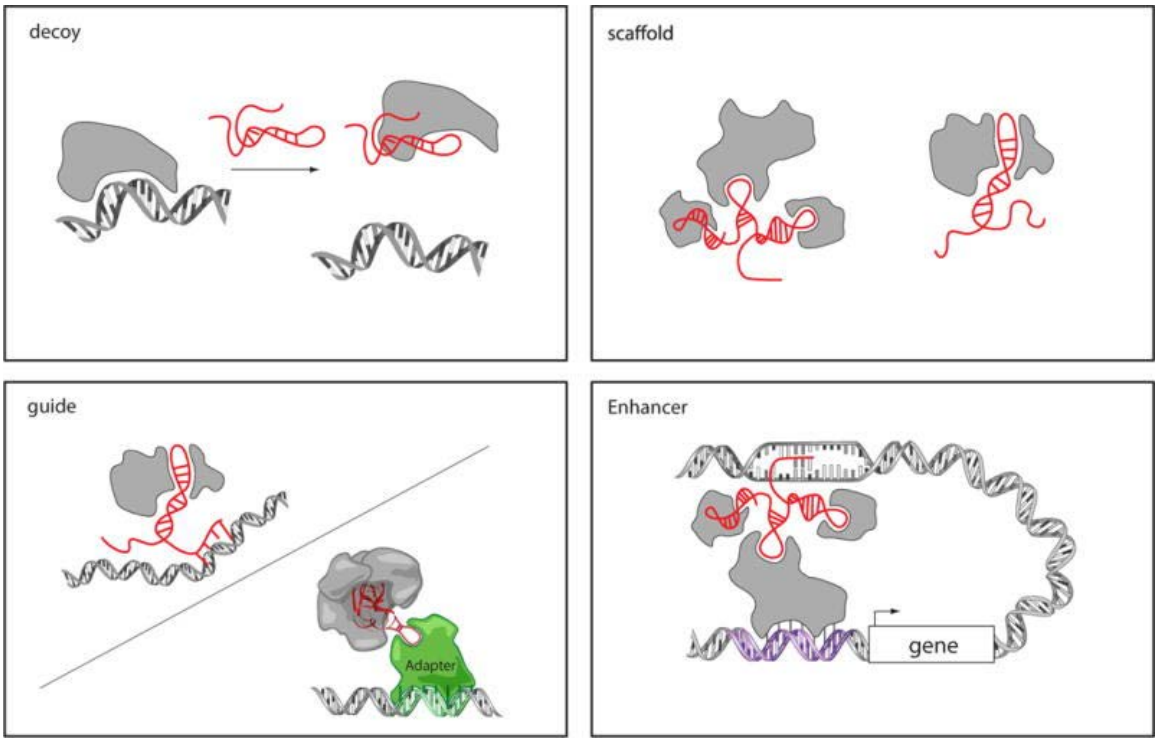


Figure 1.4 Models of lncRNA mechanisms
LncRNAs can act as decoys that titrate away DNA-binding proteins such as transcription factors, scaffolds to bring two or more proteins into a complex or spatial proximity, guides to recruit proteins, such as chromatin modification enzymes, to DNA, or function through chromosome looping in an enhancer-like model.

*From Rinn & Chang, Genome regulation by long noncoding RNAs, *Annu Rev Biochem*, 2012

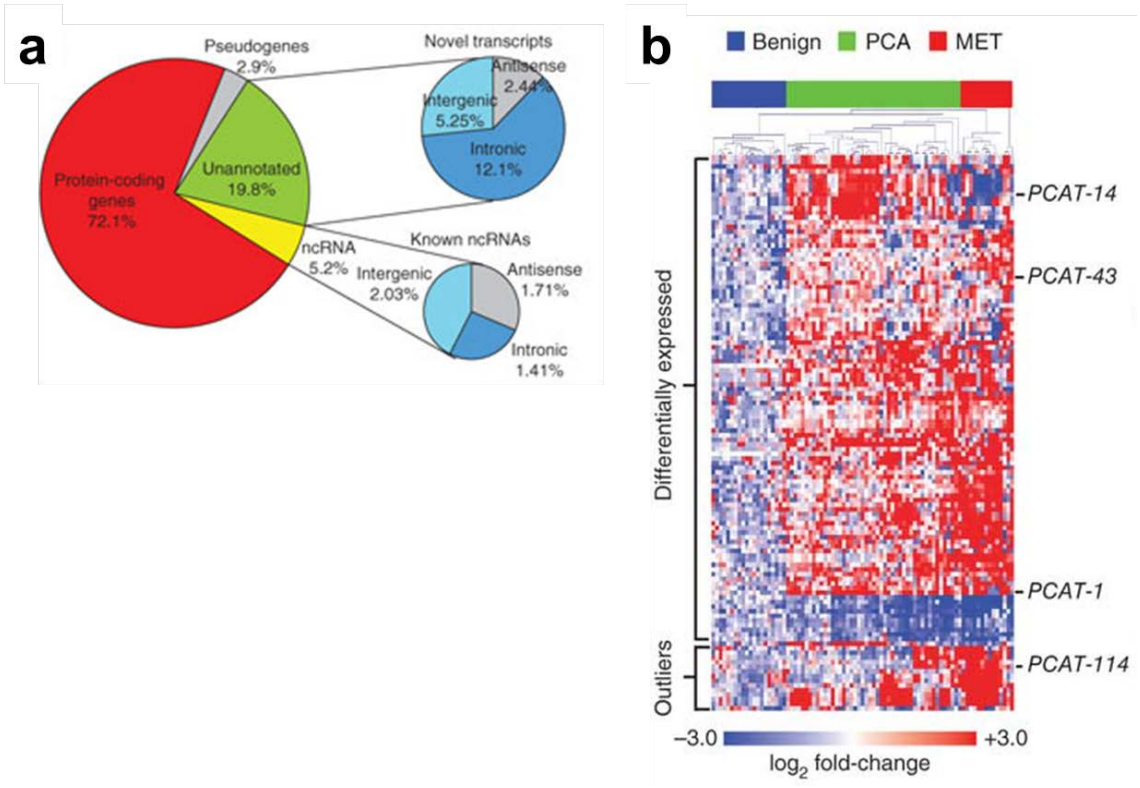


Figure 1.5 Prostate cancer transcriptome sequencing reveals dysregulation of novel transcripts

(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)** 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 the study are indicated on the side.

*Reprinted by permission from Macmillan Publishers Ltd: Nature Biotechnology (Prensner *et al.*, Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression), copyright 2011

References

1. Siegel, R.L., Miller, K.D. & Jemal, A. Cancer statistics, 2015. *CA: a cancer journal for clinicians* **65**, 5-29 (2015).
2. Cooperberg, M.R., Moul, J.W. & Carroll, P.R. The changing face of prostate cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **23**, 8146-8151 (2005).
3. SEER Stat Fact Sheets: Prostate Cancer. Bethesda, MD, National Cancer Institute, 2015.
4. Van der Kwast, T.H. & Roobol, M.J. Defining the threshold for significant versus insignificant prostate cancer. *Nature reviews. Urology* **10**, 473-482 (2013).
5. Sakr, W.A.*, et al.* High grade prostatic intraepithelial neoplasia (HGPIN) and prostatic adenocarcinoma between the ages of 20-69: an autopsy study of 249 cases. *In vivo* **8**, 439-443 (1994).
6. American Cancer Society: Cancer Facts and Figures 2015. Atlanta, Ga: American Cancer Society, 2015. Available online.
7. Shen, M.M. & Abate-Shen, C. Molecular genetics of prostate cancer: new prospects for old challenges. *Genes & development* **24**, 1967-2000 (2010).
8. Prensner, J.R., Rubin, M.A., Wei, J.T. & Chinnaiyan, A.M. Beyond PSA: the next generation of prostate cancer biomarkers. *Science translational medicine* **4**, 127rv123 (2012).
9. Brawley, O.W., Knopf, K. & Thompson, I. The epidemiology of prostate cancer part II: the risk factors. *Seminars in urologic oncology* **16**, 193-201 (1998).
10. Bostwick, D.G.*, et al.* Human prostate cancer risk factors. *Cancer* **101**, 2371-2490 (2004).
11. Howlader N, N.A., Krapcho M, Garshell J, Neyman N, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Cho H, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). SEER Cancer Statistics Review, 1975-2010. (National Cancer Institute. Bethesda, MD,, 2013).
12. Optenberg, S.A.*, et al.* Race, treatment, and long-term survival from prostate cancer in an equal-access medical care delivery system. *Jama* **274**, 1599-1605 (1995).
13. Hoffman, R.M.*, et al.* Racial and ethnic differences in advanced-stage prostate cancer: the Prostate Cancer Outcomes Study. *Journal of the National Cancer Institute* **93**, 388-395 (2001).
14. Kheirandish, P. & Chinegwundoh, F. Ethnic differences in prostate cancer. *British journal of cancer* **105**, 481-485 (2011).
15. Bunker, C.H.*, et al.* High prevalence of screening-detected prostate cancer among Afro-Caribbeans: the Tobago Prostate Cancer Survey. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **11**, 726-729 (2002).

16. Hayat, M.J., Howlader, N., Reichman, M.E. & Edwards, B.K. Cancer statistics, trends, and multiple primary cancer analyses from the Surveillance, Epidemiology, and End Results (SEER) Program. *The oncologist* **12**, 20-37 (2007).
17. Powell, I.J. & Bollig-Fischer, A. Minireview: the molecular and genomic basis for prostate cancer health disparities. *Molecular endocrinology* **27**, 879-891 (2013).
18. Martin, D.N., Starks, A.M. & Ambs, S. Biological determinants of health disparities in prostate cancer. *Current opinion in oncology* **25**, 235-241 (2013).
19. Graham-Steed, T.*, et al.* 'Race' and prostate cancer mortality in equal-access healthcare systems. *The American journal of medicine* **126**, 1084-1088 (2013).
20. Kicinski, M., Vangronsveld, J. & Nawrot, T.S. An epidemiological reappraisal of the familial aggregation of prostate cancer: a meta-analysis. *PloS one* **6**, e27130 (2011).
21. Hjelmborg, J.B.*, et al.* The heritability of prostate cancer in the Nordic Twin Study of Cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **23**, 2303-2310 (2014).
22. Gronberg, H., Damber, L. & Damber, J.E. Familial prostate cancer in Sweden. A nationwide register cohort study. *Cancer* **77**, 138-143 (1996).
23. Hoffman, R.M. Clinical practice. Screening for prostate cancer. *The New England journal of medicine* **365**, 2013-2019 (2011).
24. Smith, D.S. & Catalona, W.J. Interexaminer variability of digital rectal examination in detecting prostate cancer. *Urology* **45**, 70-74 (1995).
25. Chodak, G.W., Keller, P. & Schoenberg, H.W. Assessment of screening for prostate cancer using the digital rectal examination. *The Journal of urology* **141**, 1136-1138 (1989).
26. Roobol, M.J. & Carlsson, S.V. Risk stratification in prostate cancer screening. *Nature reviews. Urology* **10**, 38-48 (2013).
27. Moyer, V.A. & Force, U.S.P.S.T. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Annals of internal medicine* **157**, 120-134 (2012).
28. Boxer, R.J. Adenocarcinoma of the prostate gland. *Urological survey* **27**, 75-94 (1977).
29. Gittes, R.F. Carcinoma of the prostate. *The New England journal of medicine* **324**, 236-245 (1991).
30. Nelson, W.G., De Marzo, A.M. & Isaacs, W.B. Prostate cancer. *The New England journal of medicine* **349**, 366-381 (2003).
31. DeMarzo, A.M., Nelson, W.G., Isaacs, W.B. & Epstein, J.I. Pathological and molecular aspects of prostate cancer. *Lancet* **361**, 955-964 (2003).
32. Nadal, R., Schweizer, M., Kryvenko, O.N., Epstein, J.I. & Eisenberger, M.A. Small cell carcinoma of the prostate. *Nature reviews. Urology* **11**, 213-219 (2014).

33. De Marzo, A.M.*, et al.* Human prostate cancer precursors and pathobiology. *Urology* **62**, 55-62 (2003).
34. Epstein, J.I. Precursor lesions to prostatic adenocarcinoma. *Virchows Archiv : an international journal of pathology* **454**, 1-16 (2009).
35. Gleason, D.F. & Mellinger, G.T. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. *The Journal of urology* **111**, 58-64 (1974).
36. Epstein, J.I. An update of the Gleason grading system. *J Urol* **183**, 433-440 (2010).
37. Humphrey, P.A. Gleason grading and prognostic factors in carcinoma of the prostate. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* **17**, 292-306 (2004).
38. Pierorazio, P.M., Walsh, P.C., Partin, A.W. & Epstein, J.I. Prognostic Gleason grade grouping: data based on the modified Gleason scoring system. *BJU Int* **111**, 753-760 (2013).
39. Barbieri, C.E. & Tomlins, S.A. The prostate cancer genome: perspectives and potential. *Urologic oncology* **32**, 53 e15-22 (2014).
40. Barbieri, C.E.*, et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nature genetics* **44**, 685-689 (2012).
41. Grasso, C.S.*, et al.* The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239-243 (2012).
42. Kumar, A.*, et al.* Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 17087-17092 (2011).
43. Berger, M.F.*, et al.* The genomic complexity of primary human prostate cancer. *Nature* **470**, 214-220 (2011).
44. Taylor, B.S.*, et al.* Integrative genomic profiling of human prostate cancer. *Cancer cell* **18**, 11-22 (2010).
45. Lindberg, J.*, et al.* Exome sequencing of prostate cancer supports the hypothesis of independent tumour origins. *European urology* **63**, 347-353 (2013).
46. Lindberg, J.*, et al.* The mitochondrial and autosomal mutation landscapes of prostate cancer. *European urology* **63**, 702-708 (2013).
47. Beltran, H.*, et al.* Targeted next-generation sequencing of advanced prostate cancer identifies potential therapeutic targets and disease heterogeneity. *European urology* **63**, 920-926 (2013).
48. Tomlins, S.A.*, et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644-648 (2005).
49. Tomlins, S.A.*, et al.* The role of SPINK1 in ETS rearrangement-negative prostate cancers. *Cancer cell* **13**, 519-528 (2008).
50. Varambally, S.*, et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**, 624-629 (2002).
51. Huggins, C. & Hodges, C.V. Studies on prostatic cancer. I. The effect of castration, of estrogen and androgen injection on serum phosphatases in

metastatic carcinoma of the prostate. *CA: a cancer journal for clinicians* **22**, 232-240 (1972).
52. Koivisto, P. Aneuploidy and rapid cell proliferation in recurrent prostate cancers with androgen receptor gene amplification. *Prostate cancer and prostatic diseases* **1**, 21-25 (1997).
53. Linja, M.J. & Visakorpi, T. Alterations of androgen receptor in prostate cancer. *The Journal of steroid biochemistry and molecular biology* **92**, 255-264 (2004).
54. Visakorpi, T.*, et al.* In vivo amplification of the androgen receptor gene and progression of human prostate cancer. *Nature genetics* **9**, 401-406 (1995).
55. Wilt, T.J. The Prostate Cancer Intervention Versus Observation Trial: VA/NCI/AHRQ Cooperative Studies Program #407 (PIVOT): design and baseline results of a randomized controlled trial comparing radical prostatectomy with watchful waiting for men with clinically localized prostate cancer. *Journal of the National Cancer Institute. Monographs* **2012**, 184-190 (2012).
56. Bekelman, J.E.*, et al.* Effectiveness of androgen-deprivation therapy and radiotherapy for older men with locally advanced prostate cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **33**, 716-722 (2015).
57. Cheever, M.A. & Higano, C.S. PROVENGE (Sipuleucel-T) in prostate cancer: the first FDA-approved therapeutic cancer vaccine. *Clinical cancer research : an official journal of the American Association for Cancer Research* **17**, 3520-3526 (2011).
58. Rajpar, S. & Fizazi, K. Bone targeted therapies in metastatic castration-resistant prostate cancer. *Cancer journal* **19**, 66-70 (2013).
59. Sharifi, N., Gulley, J.L. & Dahut, W.L. Androgen deprivation therapy for prostate cancer. *Jama* **294**, 238-244 (2005).
60. Harris, W.P., Mostaghel, E.A., Nelson, P.S. & Montgomery, B. Androgen deprivation therapy: progress in understanding mechanisms of resistance and optimizing androgen depletion. *Nature clinical practice. Urology* **6**, 76-85 (2009).
61. Tolis, G.*, et al.* Tumor growth inhibition in patients with prostatic carcinoma treated with luteinizing hormone-releasing hormone agonists. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 1658-1662 (1982).
62. de Bono, J.S.*, et al.* Abiraterone and increased survival in metastatic prostate cancer. *The New England journal of medicine* **364**, 1995-2005 (2011).
63. Denmeade, S.R. & Isaacs, J.T. A history of prostate cancer treatment. *Nature reviews. Cancer* **2**, 389-396 (2002).
64. Tran, C.*, et al.* Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science* **324**, 787-790 (2009).
65. Wolf, A.M.*, et al.* American Cancer Society guideline for the early detection of prostate cancer: update 2010. *CA: a cancer journal for clinicians* **60**, 70-98 (2010).

66. Loeb, S.*, et al.* Systematic review of complications of prostate biopsy. *European urology* **64**, 876-892 (2013).
67. Fowler, F.J., Jr.*, et al.* The impact of a suspicious prostate biopsy on patients' psychological, socio-behavioral, and medical care outcomes. *Journal of general internal medicine* **21**, 715-721 (2006).
68. Andriole, G.L.*, et al.* Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med* **360**, 1310-1319 (2009).
69. Schroder, F.H.*, et al.* Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med* **360**, 1320-1328 (2009).
70. Draisma, G.*, et al.* Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. *J Natl Cancer Inst* **101**, 374-383 (2009).
71. Etzioni, R., Cha, R., Feuer, E.J. & Davidov, O. Asymptomatic incidence and duration of prostate cancer. *American journal of epidemiology* **148**, 775-785 (1998).
72. Trock, B.J.*, et al.* Tertiary Gleason patterns and biochemical recurrence after prostatectomy: proposal for a modified Gleason scoring system. *J Urol* **182**, 1364-1370 (2009).
73. Eggener, S.E.*, et al.* Predicting 15-year prostate cancer specific mortality after radical prostatectomy. *J Urol* **185**, 869-875 (2011).
74. Epstein, J.I., Feng, Z., Trock, B.J. & Pierorazio, P.M. Upgrading and downgrading of prostate cancer from biopsy to radical prostatectomy: incidence and predictive factors using the modified Gleason grading system and factoring in tertiary grades. *Eur Urol* **61**, 1019-1024 (2012).
75. Mohler, J.L. The 2010 NCCN clinical practice guidelines in oncology on prostate cancer. *J Natl Compr Canc Netw* **8**, 145 (2010).
76. Carter, H.B. Management of low (favourable)-risk prostate cancer. *BJU Int* **108**, 1684-1695 (2011).
77. D'Amico, A.V., Chen, M.H., Roehl, K.A. & Catalona, W.J. Identifying patients at risk for significant versus clinically insignificant postoperative prostate-specific antigen failure. *J Clin Oncol* **23**, 4975-4979 (2005).
78. Freedland, S.J.*, et al.* Risk of prostate cancer-specific mortality following biochemical recurrence after radical prostatectomy. *JAMA* **294**, 433-439 (2005).
79. Wilt, T.J.*, et al.* Systematic review: comparative effectiveness and harms of treatments for clinically localized prostate cancer. *Annals of internal medicine* **148**, 435-448 (2008).
80. Crick, F. Central dogma of molecular biology. *Nature* **227**, 561-563 (1970).
81. Lander, E.S.*, et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
82. Consortium, E.P.*, et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816 (2007).
83. Cancer Genome Atlas Research, N. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519-525 (2012).

84. Ravasi, T.*, et al.* An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* **140**, 744-752 (2010).
85. Harrow, J.*, et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome research* **22**, 1760-1774 (2012).
86. Prensner, J.R. & Chinnaiyan, A.M. The emergence of lncRNAs in cancer biology. *Cancer Discov* **1**, 391-407 (2011).
87. Guttman, M.*, et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223-227 (2009).
88. Cabili, M.N.*, et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* **25**, 1915-1927 (2011).
89. Guttman, M.*, et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* (2011).
90. Lee, J.T. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev* **23**, 1831-1842 (2009).
91. Rinn, J.L.*, et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311-1323 (2007).
92. Gong, C. & Maquat, L.E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* **470**, 284-288 (2011).
93. Poliseno, L.*, et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033-1038 (2010).
94. Salmena, L., Poliseno, L., Tay, Y., Kats, L. & Pandolfi, P.P. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* **146**, 353-358 (2011).
95. Rinn, J.L. & Chang, H.Y. Genome regulation by long noncoding RNAs. *Annual review of biochemistry* **81**, 145-166 (2012).
96. Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J. & Lee, J.T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750-756 (2008).
97. Nagano, T.*, et al.* The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**, 1717-1720 (2008).
98. Pandey, R.R.*, et al.* Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Molecular cell* **32**, 232-246 (2008).
99. Yap, K.L.*, et al.* Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Molecular cell* **38**, 662-674 (2010).
100. Wang, K.C. & Chang, H.Y. Molecular mechanisms of long noncoding RNAs. *Molecular cell* **43**, 904-914 (2011).

101. Kotake, Y.*, et al.* Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* **30**, 1956-1962 (2011).
102. Gupta, R.A.*, et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071-1076 (2010).
103. Tsai, M.C.*, et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689-693 (2010).
104. Chu, C., Qu, K., Zhong, F.L., Artandi, S.E. & Chang, H.Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Molecular cell* **44**, 667-678 (2011).
105. Kogo, R.*, et al.* Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res* **71**, 6320-6326 (2011).
106. Yang, Z.*, et al.* Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation. *Ann Surg Oncol* **18**, 1243-1250 (2011).
107. Niinuma, T.*, et al.* Upregulation of miR-196a and HOTAIR Drive Malignant Character in Gastrointestinal Stromal Tumors. *Cancer Res* **72**, 1126-1136 (2012).
108. Prensner, J.R.*, et al.* Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol* **29**, 742-749 (2011).
109. Cesana, M.*, et al.* A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **147**, 358-369 (2011).

Chapter 2:

The long noncoding RNA *SChLAP1* promotes aggressive prostate cancer and antagonizes the SWI/SNF complex[1]

Abstract

Prostate cancer is a clinically heterogeneous disease in which only a subset of patients has aggressive cancer whereas others have indolent disease[1-3]. However, the molecular basis for this heterogeneity remains incompletely understood[4-6]. Here, we characterize a novel lncRNA termed *SChLAP1 (S*econd *Ch*romosome *L*ocus *A*ssociated with *P*rostate-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

[1] This chapter was previously published as the following manuscript: Prensner JR*, Iyer MK*, Sahu A*, Asangani IA, Cao Q, Patel L, Vergara IA, Davicioni E, Ehro N, Ghadessi M, Jenkins RB, Triche TJ, Malik R, Bedenis R, McGregor N, Ma T, Chen W, Han S, Jing X, Cao X, Wang X, Chandler B, Yan W, Siddiqui J, Kunju LP, Dhanasekaran SM, Pienta K, Feng FY, Chinnaiyan AM. The long noncoding RNA *SChLAP1* promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat Genet*, 45(11):1392-8, 2013. *These authors contributed equally

least in part by antagonizing tumor-suppressive functions of the SWI/SNF complex.

Introduction

Prostate cancer is the most common non-cutaneous cancer in U.S. men, with over 200,000 prostate cancer diagnoses per year[1]. However, while 1 in 6 men are diagnosed with prostate cancer, only 1 in 32 men die from this disease[2,5], and it is estimated that only 20% of prostate cancer patients have a high-risk cancer[3]. Thus, most prostate cancer patients die with their disease, but not from it. While mutational events in key genes characterizes a subset of lethal prostate cancers[4,6,7], the molecular basis for aggressive disease remains poorly understood.

Long non-coding RNAs (lncRNAs) are polyadenylated RNA species >200bp in length commonly characterized by splicing of multiple exons, H3K4me3 promoter methylation, and transcription by RNA polymerase II[8,9]. lncRNA-mediated biology has been implicated in a wide variety of cellular processes, including pluripotency in stem cells[10] and X chromosome inactivation[11]. In cancer, lncRNAs are emerging as a prominent layer of previously underappreciated transcriptional regulation, often by collaborating with epigenetic complexes such as Polycomb Repressive Complex 1[12,13] (PRC1) and Polycomb Repressive Complex 2 (PRC2)[12,14-17], among others.

Here, we hypothesized that prostate cancer aggressiveness was governed by uncharacterized lncRNAs and sought to discover lncRNAs whose expression characterized the subset of prostate cancer patients with aggressive disease.

Results

We previously used RNA-Seq to describe 121 novel lncRNA loci (out of >1,800) that were aberrantly expressed in prostate cancer tissues (**Fig. 2.1a**)[18]. Because only a fraction of prostate cancers present with aggressive clinical features[3], we performed cancer outlier profile analysis[18] (COPA) to nominate intergenic lncRNAs selectively upregulated in a subset of cancers (**Fig. 2.1b**). We observed that only two, *PCAT-109* and *PCAT-114*, which are both located in a "gene desert" on Chromosome 2q31.3, showed striking outlier profiles and ranked among the best outliers in prostate cancer[18] (**Fig. 2.1b**).

Of the two, *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 200kb on Chromosome 2q31.3 (**Fig. 2.1c**). We named this gene *S*econd *Ch*romosome *L*ocus *A*ssociated with *P*rostate-1 (*SChLAP1*) after its genomic location. Published prostate cancer ChIP-Seq data[19] confirmed that the transcriptional start site (TSS) of *SChLAP1* was marked by H3K4 trimethylation (H3K4me3) and its gene body harbored H3K36 trimethylation (H3K36me3) (**Fig. 2.1c**), an epigenetic signature consistent with lncRNAs[9]. We observed numerous *SChLAP1* splicing isoforms of which

three (termed isoforms #1, #2, and #3, respectively) constituted the vast majority (>90%) of transcripts in the cell (**Fig. 2.1d**).

Using quantitative PCR (qPCR), we validated that *SChLAP1* was highly expressed in ~25% of prostate cancers (**Fig. 2.2a**). *SChLAP1* prevalence was more frequent in metastatic compared to localized prostate cancers in this cohort (**Fig. 2.2b**). Next, we found that *SChLAP1* transcripts were located in the nucleus (**Fig. 2.2c**). We confirmed the nuclear localization of *SChLAP1* in human samples using an *in situ* hybridization (ISH) assay in formalin-fixed paraffin-embedded (FFPE) prostate cancers (**Fig. 2.2d**). Additionally, a computational analysis of the *SChLAP1* sequence suggested no coding potential, which was confirmed experimentally by *in vitro* translation assays of three *SChLAP1* isoforms (**Fig. 2.3**).

Next, we performed a network analysis of prostate cancer microarray data in the Oncomine[20] database using signatures of *SChLAP1*-correlated or -anti-correlated genes, given that *SChLAP1* is not measured by expression microarrays. We found a remarkable association with enriched concepts related to prostate cancer progression (**Fig. 2.4a**). For comparison, we next incorporated disease signatures using prostate RNA-seq data as well as additional known prostate cancer genes: *EZH2*, a metastasis gene[21], *PCA3*, a lncRNA biomarker[5], *AMACR*, a tissue biomarker[5], and β-actin (*ACTB*) as a control. A heat-map visualization of significant comparisons confirmed a strong association of

SChLAP1-correlated genes, but not *PCA3*- and *AMACR*-correlated genes, with high-grade and metastatic cancers (**Fig. 2.4b**). Kaplan-Meier analysis similarly showed significant associations between the *SChLAP1* signature and biochemical recurrence[22] and overall survival[23] (**Fig. 2.4c,d**).

To evaluate *SChLAP1* levels with clinical outcomes directly, we next used *SChLAP1* expression to stratify 235 radical prostatectomy localized prostate cancer patients from the Mayo Clinic[24]. Samples were evaluated for three clinical endpoints: biochemical recurrence (BCR), clinical progression to systemic disease (CP), and prostate cancer-specific mortality (PCSM). At the time of this analysis, patients had a median follow-up of 8.1 years.

SChLAP1 was a powerful single-gene predictor of aggressive prostate cancer (**Fig. 2.5a-c**). *SChLAP1* expression was highly significant when distinguishing CP and PCSM ($p = 0.00005$ and $p = 0.002$, respectively) (**Fig. 2.5b,c**). For the BCR endpoint, high *SChLAP1* expression was associated with a rapid median time-to-progression (1.9 vs 5.5 years for *SChLAP1* high and low patients, respectively) (**Fig. 2a**). We further confirmed that this association with rapid BCR using an independent cohort (data not shown). Multivariable and univariable 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 BCR, CP, and PCSM (HR of 3.045, 3.563, and 4.339,

respectively, $p < 0.01$) which were comparable to other clinical factors such as advanced clinical stage and the Gleason histopathological score (**Fid. 2.5d-i**).

To explore the functional role for *SChLAP1*, we performed siRNA knockdowns to compare the impact of *SChLAP1* depletion to that of *EZH2*, which is essential for cancer cell aggressiveness[21]. Remarkably, knockdown of *SChLAP1* (**Fig. 2.6a**) dramatically impaired cell invasion and proliferation *in vitro* at a level comparable to *EZH2* (**Fig. 2.6b**). *SChLAP1* knockdown also impaired cell proliferation (**Fig. 2.6c**). Overexpression of a siRNA-resistant *SChLAP1* isoform rescued the *in vitro* invasive phenotype of 22Rv1 cells treated with siRNA-2 (**Fig. 2.6d,e**). Next, we overexpressed the three most abundant *SChLAP1* isoforms in RWPE benign immortalized prostate cells at physiologic levels similar to the LNCaP cell line (**Fig. 2.7a**). We found that overexpression of *SChLAP1* dramatically increased the ability of RWPE cells to invade *in vitro* but did not impact cell proliferation (**Fig. 2.7b,c**).

To test *SChLAP1 in vivo*, we performed intracardiac injection of 22Rv1 cells stably knocking down *SChLAP1* (**Fig. 2.8a**) and observed that *SChLAP1* depletion impaired metastatic seeding and growth by luciferase signaling at both proximal (lungs) and distal sites (**Fig. 2.8b,c**). Indeed, 22Rv1 sh*SChLAP1* cells displayed both fewer gross metastatic sites overall as well as smaller metastatic tumors when they did form (**Fig. 2.8c,d**). Histopathological analysis of the metastatic 22Rv1 tumors, regardless of *SChLAP1* knockdown, showed uniformly

high-grade epithelial cancer (**Fig. 2.8e**). Interestingly, sh*SChLAP1* subcutaneous xenografts displayed slower tumor progression (**Fig. 2.9a**); however this was due to delayed tumor engraftment rather than decreased tumor growth kinetics with no change in Ki67 staining observed between sh*SChLAP1* and sh*NT* cells (**Fig. 2.9b-g**).

Next, using the chick chorioallantoic membrane (CAM) assay[25], we found that 22Rv1 sh*SChLAP1* #1 cells, which have depleted expression of both isoforms 1 and 2, demonstrated a greatly reduced ability to invade, intravasate and metastasize distant organs (**Fig. 2.10a-c**). Additionally, sh*SChLAP1* cells also showed decreased tumor growth (**Fig. 2.10d**). Importantly, overexpression of RWPE-*SChLAP1* isoform #1 cells partially recapitulated these results, displaying a markedly increased ability to intravasate (**Fig. 2.10e**). RWPE-*SChLAP1* cells did not generate distant metastases or cause altered tumor growth in this model (data not shown). Together, the murine metastasis and CAM data strongly implicate *SChLAP1* in tumor invasion and metastasis through cancer cell intravasation, extravasation, and subsequent tumor cell seeding.

To elucidate mechanisms of *SChLAP1* function, we profiled 22Rv1 and LNCaP *SChLAP1*-knockdown cells, which revealed 165 upregulated and 264 downregulated genes (q-value < 0.001) (**Fig. 2.11a**). After ranking genes according to differential expression[26], we employed Gene Set Enrichment Analysis (GSEA)[27] to search for enrichment across the Molecular Signatures

Database (MSigDB)[28]. Among the highest ranked concepts we noticed genes positively or negatively correlated with the SWI/SNF complex (**Fig. 2.11b**)[29], which was independently confirmed using gene signatures generated from our RNA-Seq data (**Fig. 2.11c-e**). Importantly, *SChLAP1*-regulated genes were inversely correlated with these datasets, suggesting that *SChLAP1* and SWI/SNF function in opposing manners.

The SWI/SNF complex regulates gene transcription as a multi-protein system that physically move nucleosomes at gene promoters[30]. Loss of SWI/SNF functionality promotes cancer progression and multiple SWI/SNF components are somatically inactivated in cancer[30,31]. SWI/SNF mutations do occur in prostate cancer albeit not commonly[4], and down-regulation of SWI/SNF complex members characterizes subsets of prostate cancer[29,32]. Thus, antagonism of SWI/SNF activity by *SChLAP1* is consistent with the oncogenic behavior of *SChLAP1* and the tumor suppressive behavior of the SWI/SNF complex.

To directly test whether *SChLAP1* antagonizes SWI/SNF-mediated regulation, we performed siRNA knockdown of *SNF5* (also known as *SMARCB1)* (**Fig. 2.12a**), an essential subunit that facilitates SWI/SNF binding to histone proteins[30,31,33]. A comparison of genes regulated by knockdown of *SNF5* to genes regulated by *SChLAP1* demonstrated an antagonistic relationship where *SChLAP1* knockdown affected the same genes as *SNF5* but in the opposing direction (**Fig. 2.12b**). We used GSEA to quantify and verify the significance of

these findings (FDR < 0.05) (**Fig. 2.12c-e**). Furthermore, a shared *SNF5-SChLAP1* signature of co-regulated genes was highly enriched for prostate cancer clinical signatures for disease aggressiveness (**Fig. 2.13**).

Mechanistically, although *SChLAP1* and *SNF5* mRNA levels were comparable (**Fig. 2.14a**), *SChLAP1* knockdown or overexpression did not alter SNF5 protein abundance (**Fig. 2.14b**), suggesting that *SChLAP1* regulates SWI/SNF activity post-translationally. To explore this possibility, we performed RNA immunoprecipitation assays (RIP) for SNF5. We found that endogenous *SChLAP1*, but not other cytoplasmic or nuclear lncRNAs[8,34], robustly co-immunoprecipitated with SNF5 in both native (**Fig. 2.14c**) and UV-crosslinked conditions (data not shown) as well as with a second SNF5 antibody (data not shown). In contrast, *SChLAP1* did not co-immunoprecipitate with androgen receptor (**Fig. 2.14c**). Furthermore, both *SChLAP1* isoform #1 and isoform #2 co-immunoprecipitated with SNF5 in RWPE overexpression models (**Fig. 2.14d**). SNRNP70 binding to the *U1* RNA was used as a technical control in all cell lines (data not shown). Finally, pulldown of the *SChLAP1* RNA in RWPE-*SChLAP1* isoform #1 cells robustly recovered SNF5 protein, confirming this interaction (**Fig. 2.14e,f**).

To address whether *SChLAP1* modulated SWI/SNF genomic binding, we performed ChIP-Seq of SNF5 in RWPE-*LacZ* and RWPE-*SChLAP1* cells and called significantly enriched peaks with respect to an IgG control. Western blot

validations confirmed SNF5 pull-down by ChIP. After aggregating called peaks from all samples, we found 6,235 genome-wide binding sites for SNF5 (FDR < 0.05), which were highly enriched for sites near gene promoters (data not shown), supporting previous studies of SWI/SNF binding[35-37].

A comparison of SNF5 binding across these 6,235 genomic sites demonstrated a dramatic decrease in SNF5 genomic binding as a result of *SChLAP1* overexpression (**Fig. 2.15a,b**). Of the 1,299 SNF5 peaks occurring within 1kb of a gene promoter, 390 decreased ≥2-fold in relative SNF5 binding (**Fig. 2.15c**). To verify these findings independently, we performed ChIP for SNF5 in 22Rv1 sh-*SChLAP1* cells, with the hypothesis that knockdown of *SChLAP1* should increase SNF5 genomic binding compared to controls. We found that 9 of 12 target genes showed a substantial increase in SNF5 binding (**Fig. 2.15d**), confirming our predictions.

Finally, we used expression profiling of RWPE-*LacZ* and RWPE-*SChLAP1* cells to characterize the relationship between SNF5 binding and *SChLAP1*-mediated gene expression changes. After identifying a gene signature with highly significant expression changes, we intersected this signature with the ChIP-Seq data. We observed that a significant subset of genes with ≥2-fold relative decrease in SNF5 genomic binding were dysregulated when *SChLAP1* was overexpressed (data not shown). Decreased SNF5 binding was primarily associated with downregulation of target gene expression, although the SWI/SNF

complex is known to regulate expression in either direction[30,31]. An integrative GSEA analysis of the microarray and SNF5 ChIP-Seq data demonstrated a significant enrichment for genes that were repressed when *SChLAP1* was overexpressed (q-value = 0.003, **Fig. 2.15e**). Overall, these data argue that *SChLAP1* overexpression antagonizes SWI/SNF complex function by attenuating the genomic binding of this complex, thereby impairing its ability to regulate gene expression properly.

Discussion

Here, we have discovered *SChLAP1*, a highly prognostic lncRNA that is abundantly expressed in 15-30% of prostate cancers and aided the discrimination of aggressive from indolent forms of this disease. Mechanistically, we find that *SChLAP1* coordinates cancer cell invasion *in vitro* and metastatic spread *in vivo*. Moreover, we characterize an antagonistic *SChLAP1*-SWI/SNF axis in which *SChLAP1* impairs SNF5-mediated gene expression regulation and genomic binding (**Fig. 2.16**). Thus, while other lncRNAs such as *HOTAIR* and *HOTTIP* are known to assist epigenetic complexes such as PRC2 and MLL by facilitating their genomic binding and enhancing their functions[14,15,17,38], *SChLAP1* is the first lncRNA, to our knowledge, that impairs a major epigenetic complex with well-documented tumor suppressor function[29-31,39-41]. Taken together, our discovery of *SChLAP1* has broad implications for cancer biology and provides supporting evidence for the role of lncRNAs in the progression of aggressive cancers.

Materials and Methods

Experimental studies

Cell lines

All cell lines were obtained from the American Type Culture Collection (Manassas, VA). Cell lines were maintained using standard media and conditions. Specifically, VCaP and Du145 cells were maintained in DMEM (Invitrogen) plus 10% fetal bovine serum (FBS) plus 1% penicillin-streptomycin. LNCaP and 22Rv1 were maintained in RPMI 1640 (Invitrogen) plus 10% FBS and 1% penicillin-streptomycin. RWPE cells were maintained in KSF media (Invitrogen) plus 10ng/mL EGF (Sigma) and bovine pituitary extract (BPE) and 1% penicillin-streptomycin. All cell lines were grown at 37°C in a 5% CO_2 cell culture incubator. All cell lines were genotyped for identity at the University of Michigan Sequencing Core and tested routinely for *Mycoplasma* contamination.

SChLAP1 or control-expressing cell lines were generated by cloning *SChLAP1* or control into the pLenti6 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 using blasticidin (Invitrogen) for one week. For LNCAP and 22Rv1 cells with stable knockdown of *SChLAP1*, cells were transfected with *SChLAP1* or non-targeting shRNA lentiviral constructs

for 48 hours. GFP+ cells were selected with 1ug/mL puromycin for 72 hours. 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[42]. These programs are part of the University of Michigan Prostate Cancer Specialized Program Of Research Excellence (S.P.O.R.E.). All tissue samples were collected with informed consent under an Institutional Review Board (IRB) approved protocol at the University of Michigan. (SPORE in Prostate Cancer (Tissue/Serum/Urine) Bank Institutional Review Board # 1994-0481).

RNA isolation and cDNA synthesis

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

Quantitative Real-time PCR

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

Integrated DNA Technologies (Coralville, IA) and are listed in **Table 2.1.** The housekeeping genes, *GAPDH, HMBS,* and *ACTB*, were used as loading controls. Fold changes were calculated relative to housekeeping genes and normalized to the median value of the benign samples.

Reverse-transcription PCR

Reverse-transcription PCR (RT-PCR) was performed for primer pairs using Platinum Taq High Fidelity polymerase (Invitrogen). PCR products were resolved on a 1.0% agarose gel. PCR products were either sequenced directly (if only a single product was observed) or appropriate gel products were extracted using a Gel Extraction kit (Qiagen) and cloned into pcr4-TOPO vectors (Invitrogen). PCR products were bidirectionally sequenced at the University of Michigan Sequencing Core using either gene-specific primers or M13 forward and reverse primers for cloned PCR products. All oligonucleotide primers were obtained from Integrated DNA Technologies (Coralville, IA) and are listed in **Table 2.1**.

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 **Table 2.1**. RACE-PCR products were separated on a 1.5% agarose gels. Gel products

were extracted with a Gel Extraction kit (Qiagen), cloned into pcr4-TOPO vectors (Invitrogen), and sequenced bidirectionally using M13 forward and reverse primers at the University of Michigan Sequencing Core. At least three colonies were sequenced for every gel product that was purified.

siRNA knockdown studies

Cells were plated in 100mM plates at a desired concentration and transfected with 20uM experimental siRNA oligos or non-targeting controls twice, at 8 hours and 24 hours post-plating. Knockdowns were performed with Oligofectamine in OptiMEM media. 72 hours post-transfection, cells were trypsinized, counted with a Coulter counter, and diluted to 1 million cells/mL. Knockdown efficiency was determined by qPCR. siRNA sequences (in sense format) for knockdowns were as follows:

SChLAP1 siRNA 1: CCAAUGAUGAGGAGCGGGA
SChLAP1 siRNA 2: CUGGAGAUGGUGAACCCAA
SNF5 siRNA 5: GUGACGAUCUGGAUUUGAA
SNF5 siRNA 7: GAUGACGCCUGAGAUGUUU

Overexpression studies

SChLAP1 full length transcript was amplified from LNCaP cells and cloned into the pLenti6 vector (Invitrogen) along with LacZ controls. Insert sequences were confirmed by Sanger sequencing at the University of Michigan Sequencing Core. Lentiviruses were generated at the University of Michigan Vector Core. The benign immortalized prostate cell line RWPE was infected with lentiviruses expressing *SChLAP1* or *LacZ* and stable pools and clones were generated by

selection with blasticidin (Invitrogen). Similarly, the immortalized cancer cell line 22Rv1 was infected with lentiviruses expressing *SChLAP1* or *LacZ* and stable pools were generated by selection with blasticidin (Invitrogen).

Cell proliferation assays

72 hours post-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 24-well plates and grown in regular media. 48 and 96 hours post-plating, cells were harvested by trypsinizing and counted using a Coulter counter. All assays were performed in quadruplicate.

Basement Membrane Matrix Invasion Assays

For invasion assays, cells were treated with the indicated siRNAs and 72 hours post-transfection, cells were trypsinized, counted with a Coulter counter, and diluted to 1 million cells/mL. Cells were seeded onto the basement membrane matrix (EC matrix, Chemicon, Temecula, CA) present in the insert of a 24 well culture plate. Fetal bovine serum was added to the lower chamber as a chemo-attractant. After 48 hours, 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, the inserts were treated with 150 µl of 10% acetic acid and the absorbance measured at 560nm using a spectrophotometer (GE Healthcare).

shRNA knockdown

The prostate cancer cell lines LNCaP and 22Rv1 were seeded at 50-60% confluency and allowed to attach overnight. Cells were transfected with *SChLAP1* or non-targeting shRNA lentiviral constructs as described previously for 48 hours. GFP+ cells were drug-selected using 1 ug/mL puromycin for 72 hours. 48 hours post-selection cells were harvested for protein and RNA using RIPA buffer or trizol, respectively. RNA was processed as described above.

Gene expression profiling

Expression profiling was performed using the Agilent Whole Human Genome Oligo Microarray (Santa Clara, CA), according to previously published protocols[43]. All samples were run in technical triplicates comparing knockdown samples treated with *SChLAP1* siRNA compared to treatments with non-targeting control siRNA. Expression data was analyzed using the SAM method as described previously[26].

Murine intracardiac and subcutaneous *in vivo* models

All experimental procedures were approved by the University of Michigan Committee for the Use and Care of Animals (UCUCA). *Intracardiac injection model*: 5 x 10^5 cells from one of three experimental cell lines (22Rv1 shNT, 22Rv1 sh*SChLAP1* #1, sh*SChLAP1* #2, all with luciferase constructs incorporated) were introduced to CB-17 severe combine immunodefiecient mice (CB-17 SCID) at 6 weeks of age. Female mice were used to minimize

endogenous androgen production that may stimulate xenografted prostate cells. 15 mice were used per cell line in order 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 one week post injection, bioluminescent imaging of mice was performed weekly using a CCD IVIS system with a 50-mm lens (Xenogen Corp.) and the results were analyzed using LivingImage software (Xenogen). When the mice reached determined endpoint, whole body region of interest (ROI) of 1 x 10^{10} photons, or became fatally ill, the animal was euthanized and the lung and liver resected. Half of the resected specimen was put in an immunohistochemistry cassette and placed in 10% buffered formalin phosphate (Fisher Scientific) for 24 hours, and then transferred to 70% ethanol until further analysis. The other half of each specimen was snap frozen in liquid nitrogen and stored in -80^{0}C. A specimen was disregarded if the tumor was localized to the heart only. After accounting for these considerations, there were 9 mice analyzed for 22Rv1 shNT cells, 14 mice each analyzed for 22Rv1 sh*SChLAP1* #1 and #2 cells. *Subcutaneous injection model*: 1 x 10^6 cells from one of the three previously described experimental cell lines were introduced to mice (CB-17 SCID), ages 5-7 weeks, with a Matrigel scaffold (BD Matrigel Matrix, BD Biosciences) in the posterior dorsal flank region (*n* = 10 per cell line). Tumors were measured weekly using a digital caliper, and endpoint was determined as a tumor volume of 1000 mm^3. When endpoint was reached, or the animal became

fatally ill, the mouse was euthanized and the primary tumor resected. The resected specimen was divided in half: one half in 10% buffer formalin and the other half snap frozen. For histological analyses, FFPE-fixed mouse livers and lungs were sectioned on a microtome into 5uM sections onto glass slides. Slides were stained with hematoxalyn and eosin using standard methods and analyzed by a board-certified pathologist (LPK).

Immunoblot Analysis

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

Protein lysates were boiled in sample buffer, and 10 µg protein was loaded onto a SDS-PAGE gel and run for separation of proteins. Proteins were transferred onto Polyvinylidene Difluoride membrane (GE Healthcare) and blocked for 90 minutes in blocking buffer (5% milk, 0.1% Tween, Tri-buffered saline (TBS-T)). Membranes were incubated overnight at 4C with primary antibody. Following 3 washes with TBS-T, and one wash with TBS, the blot was incubated with horseradish peroxidase-conjugated secondary antibody and the signals visualized by enhanced chemiluminescence system as described by the manufacturer (GE Healthcare). Primary antibodies used were:

SNF5 (1:1000, Millipore, ABD22, rabbit)
SNF5 (1:1000, Abcam, ab58209, mouse)
ACTB (1:5000, Sigma, rabbit)
AR (1:1000. Millipore, 06-680, rabbit)

RNA immunoprecipitation

RIP assays were performed using a Millipore EZ-Magna RIP RNA-Binding Protein Immunoprecipitation kit (Millipore, #17-701) according to the manufacturer's instructions. RIP-PCR was performed as qPCR, as described above, using total RNA as input controls. 1:150th of RIP RNA product was used per PCR reaction. Antibodies used for RIP were Rabbit polyclonal IgG (Millipore, PP64), SNRNP70 (Millipore, CS203216), SNF5 (Millipore, ABD22), SNF5 (Abcam, ab58209), and AR (Millipore, 06-680, rabbit), using 5 – 7 µg of antibody per RIP reaction. All RIP assays were performed in biological duplicate. For UV-crosslinked RIP experiments, cells were subjected to 400J of 254nM UV light twice and then harvested for RIP experiments as above.

Chromatin immunoprecipitation

ChIP assays were performed as described previously[18,19], using antibodies for SNF5 (Millipore ABD22) and Rabbit IgG (Millipore PP64B). Briefly, approximately 10^6 cells were crosslinked per antibody for 10-15 minutes with 1% formaldehyde and the crosslinking was inactivated by 0.125M glycine for 5 minutes at room temperature. Cells were rinsed with cold PBS three times and cell pellets were resuspended in lysis buffer plus protease inhibitors. Chromatin was sonicated to an average length of 500bp, centrifuged to remove debris, and

supernatants containing chromatin fragments were incubated with protein A/G beads to reduce non-specific binding. Then, beads were removed and supernatants were incubated with 6µg of antibody overnight at 4C. Beads were added and incubated with protein-chromatin-antibody complexes for 2 hours at 4C, washed twice with 1X dialysis buffer and four times with IP wash buffer, and eluted in 150 µl IP elution buffer. 1:10th of the ChIP reaction was taken for protein evaluation for validation of ChIP pull-down. Reverse crosslinking was performed by inclubating the eluted product with 0.3 M NaCl at 65C overnight. ChIP product was cleaned up with the USB PrepEase kit (USB). ChIP experiments were validated for specificity by Western blotting.

ChIP-Seq experiments

Paired-end ChIP-Seq libraries were generated following the Illimuna ChIP-Seq protocol with minor modifications. The ChIP DNA was subjected to end-repair and A base addition before ligating with Illumina adaptors. Samples were purified using Ampure beads (Beckman Coulter Inc., Brea CA) and PCR-enriched with a combination of specific index primers and PE2.0 primer under the following conditions: 98C (30 sec), 65C (30 sec), and 72C (40 sec with a 4 sec increment per cycle). After 14 cycles of amplification a final extension at 72C for 5 minutes was carried out. The barcoded libraries were size-selected using a 3% NuSieve Agarose gele (Lonza, Allendale, NJ) and subjected to an additional PCR enrichment step. The libraries were analyzed and quantitated using Bio-

Analyzer (Agilent Technologies, Santa Clara, CA) before subjecting it to paired-end sequencing using the Illumina Hi-Seq platform.

CAM assays

CAM assays were performed as previously described[44]. Briefly, fertilized eggs were incubated in a rotary humidified incubator at 38°C for 10 days. CAM was released by applying mild amount of low pressure to the hole over the air sac and cutting a 1 cm^2 window encompassing a second hole near the allantoic vein. Approximately 2 million cells in 50µl of media were implanted in each egg, windows were sealed and the eggs were returned to a stationary incubator.

For local invasion and intravasation experiments, the upper and lower CAM were isolated after 72hr. The upper CAM were processed and stained for chicken collagen IV (immunofluorescence) or human cytokeratin (immunohistochemistry) as previously described[44].

For metastasis assay, the embryonic livers were harvested on day 18 of embryonic growth and analyzed for the presence of tumor cells by quantitative human Alu-specific PCR. Genomic DNA from lower CAM and livers were prepared using Puregene DNA purification system (Qiagen) and quantification of human-Alu was performed as described[44]. Fluorogenic TaqMan qPCR probes were generated as described above and used to determine DNA copy number.

For xenograft growth assay with RWPE cells, the embryos were sacrificed on day 18 and the extra-embryonic xenograft were excised and weighed.

In situ hybridization

ISH assays were performed as a commercial service from Advanced Cell Diagnostics, Inc. Briefly, cells in the clinical specimens are fixed and permeablized using xylenes, ethanol, and protease to allow for probe access. Slides are boiled in pretreatment buffer for 15 min and rinsed in water. Next, two independent target probes are hybridized to the *SChLAP1* RNA at 40C for 2 hours, with this pair of probes creating a binding site of a preamplifier. After this, the preamplifier is hybridized to the target probes at 30C and amplified with 6 cycles of hybridization followed by 2 washes. Cells are counter-stained to visualize signal. Finally, slides are H&E stained, dehydrated with 100% ethanol and xylene, and mounted in a xylene-based mounting media.

In vitro translation

Full length *SChLAP1, PCAT-1,* or *GUS* positive control were cloned into the PCR2.1 entry vector (Invitrogen). Insert sequences were confirmed by Sanger sequencing at the University of Michigan Sequencing Core. *In vitro* translation assays were performed with the TnT Quick Coupled Transcription/Translation System (Promega) with 1mM methionine and Transcend Biotin-Lysyl-tRNA (Promega) according to the manufacturer's instructions.

ChIRP Assay

ChIRP assays were performed as previously described[45]. Briefly, antisense DNA probes targeting the *SChLAP1* full-length sequence were designed using the online designer at http://www.singlemoleculefish.com. 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 twenty-four probes were designed against *LacZ* RNA as a negative control. All probes were synthesized with 3' biotinylation (IDT). Sequences of all probes are listed in **Table 2.2**. RWPE cells overexpressing *SChLAP1* isoform 1 were grown to 80% confluency in 100mm cell culture dishes. Two dishes were used for each probe set. Prior to harvesting, the cells were rinsed with 1xPBS and crosslinked with 1% glutaraldehyde (Sigma) for 10 min at room temperature. Crosslinking was quenched with 0.125M glycine for 5 min at room temperature. The cells were rinsed twice with 1xPBS, collected and pelleted at 1500xg for 5 min. Nuclei were isolated using the Pierce NE-PER Nuclear Protein Extraction Kit. The nuclear pellet was resuspended in 100mg/ml cell lysis buffer (50 mM Tris, pH 7.0, 10 mM EDTA, 1% SDS, and added before use: 1 mM dithithreitol (DTT), phenylmethylsulphonyl fluoride (PMSF), protease inhibitor and Superase-In (Invitrogen)). The lysate was placed on ice for 10 min and sonicated using a Bioruptor (Diagenode) at the highest setting with 30 sec on and 45 sec 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 100pmol/ml probes were added to the diluted lysate. Hybridization was carried out by end-over-end rotation at 37°C for 4 hours. Magnetic streptavidin C1 beads were prepared by washing three times in cell lysis buffer and then added to each hybridization reaction at 100µl per 100pmol of probes. The reaction was 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 1mL wash buffer (2×SSC, 0.5% SDS, fresh PMSF added). After the last 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 200µl of RNA proteinase K buffer (100 mM NaCl, 10 mM Tris, pH 7.0, 1 mM EDTA, 0.5% SDS) and 1mg/ml proteinase K (Ambion). The sample was incubated at 50°C for 45 min and then boiled for 10 min. RNA was isolated using 500µl of Trizol reagent using the miRNeasy kit (Qiagen) with on-column DNase digestion (Qiagen). RNA was eluted with 10µl H2O and then analyzed by qRT–PCR for the detection of enriched transcripts. For protein elution, beads were resuspended in 3x the original volume of DNase buffer (100 mM NaCl and 0.1% NP-40), and protein was eluted with a cocktail of 100 ug/ml RNase A (Sigma-Aldrich), 0.1 Units/microliter RNase H (Epicenter), and 100 U/ml DNase I (Invitrogen) at 37°C for 30 min. The eluted protein sample was supplemented with NuPAGE® LDS Sample Buffer (Novex) and NuPAGE® Sample Reducing Agent (Novex) to a final concentration of 1x each and then boiled for 10 min before SDS-PAGE Western blot analysis using a SNF5 antibody (Millipore).

RNA-Seq Library Preparation

Total RNA was extracted from healthy and cancer cell lines and patient tissues, and the quality of the RNA were assessed with the 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 protocols[18,46]. For strand-specific library construction, we employed the dUTP method of second-strand marking as described previously[47].

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 Bioinformatic analyses.

Bioinformatics Analysis

Nomination of *SChLAP1* as an outlier using RNA-Seq data

We nominated *SChLAP1* as a prostate cancer outlier using the methodology detailed in Prensner JR et al., *Nature Biotechnology* 2011. Briefly, a modified COPA analysis was performed on the 81 tissue samples in the cohort. RPKM

expression values were used and shifted by 1.0 in order to avoid division by zero. The COPA analysis had the following steps: 1) gene expression values were median centered, using the median expression value for the gene across the all samples in the cohort. This sets the gene's median to zero. 2) The median absolute deviation (MAD) was calculated for each gene, and then each gene expression value was scaled by its MAD. 3) The 80, 85, 90, 98 percentiles of the transformed expression values were calculated for each gene and the average of those four values was taken. Then, genes were rank ordered according to this "average percentile", which generated a list of outliers genes arranged by importance. 4) Finally, genes showing an outlier profile in the benign samples were discarded.

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[48]. Peak calling was performed using MACS [49] according to the published protocols[50]. Data was visualized using the UCSC Genome Browser [51].

RWPE ChIP-Seq data

Sequencing data from RWPE SNF5 ChIP-Seq samples were mapped to human genome version hg19 using BWA 0.5.9[48]. Although we performed paired-end sequencing, the ChIP-Seq reads were processed as single-end to adhere to our

preexisting analysis protocol. Peak calling was performed respect to an IgG control using the MACS algorithm[49]. 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 reports[52]. We retained peaks with an false discovery rate (FDR) less than 5%. 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 one kilobase of transcription start sites (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 package version 1.6.1[53] 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 Gene Set Enrichment Analysis (GSEA).

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 protein SNF5 (*SMARCB1*) as well as control non-targeting siRNAs. These profiling experiments were run as biological duplicates for a total of 4 arrays (2 cell lines x 2 independent siRNAs x 1 protein). Finally, we profiled of 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 4 arrays (2 from RWPE-*SChLAP1* isoform #1 and 2 from RWPE-*SChLAP1* isoform #2).

Processing to determine ranked gene expression lists

All of the microarray data were represented as base-2 log fold-change between targeting versus control siRNAs. We used the CollapseDataset tool provided by the GSEA package to convert from Agilent Probe IDs to gene symbols. Genes measured by multiple probes were consolidated using the median of 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 was 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 the Molecular Signatures Database (MSigDB) version 3.0[28]. Lists of positively and negatively enriched concepts were interpreted manually.

SNF5 siRNA knockdown microarrays

For each SNF5 protein knockdown we nominated genes that were altered by an average of at least 2-fold. These signatures of putative SNF5 target genes were then used to assess enrichment of *SChLAP1*-regulated genes using the GseaPreRanked tool. Additionally, we nominated genes that changed by an average 2-fold or greater across SNF5 knockdown experiments and quantified the enrichment for *SChLAP1* target genes using GSEA.

RWPE *SChLAP1* expression microarrays

The 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 under-expressed. 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 to the GseaPreRanked tool and compared against SNF5 ChIP-Seq promoter peaks that decreased by >2-fold in RWPE-*SChLAP1* cells. Of the 389 genes in the ChIP-Seq gene set, 250 were profiled by the Agilent HumanGenome microarray chip and present in the GSEA gene symbol database.

RNA-Seq data

We assembled an RNA-Seq cohort from prostate cancer tissues sequenced at multiple institutions. We included data 12 primary tumors and 5 benign tissues published in GEO as GSE22260[54], 16 primary tumors and 3 benign tissues released in dbGAP as study phs000310.v1.p1[55], and 17 benign, 57 primary, 14 metastatic tumors sequenced by our own institution and released as dbGAP study phs000443.v1.p1.

RNA-Seq alignment and gene expression quantification

Sequencing data were aligned using Tophat version 1.3.1[56] against the Ensembl GRCh37 human genome build. Known introns (Ensembl release 63) were provided to Tophat. Gene expression across the Ensembl version 63 genes and the *SChLAP1* transcript was quantified by HT-Seq version 0.5.3p3 using the script *htseq-count* (www-huber.embl.de/users/anders/HTSeq/). 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 DESeq version 1.6.1[53]. Read counts were normalized using the *estimateSizeFactors* function and variance was modeled by the *estimateDispersions* function. Differentially expression statistics were computed by the *nbinomTest* function. We called

differentially expressed genes by imposing adjusted p-value cutoffs for cancer versus benign (padj < 0.05), metastasis versus primary (padj < 0.05), and gleason 8+ versus 6 (padj < 0.10).

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, the raw read count data was converted to pseudo-counts using the *getVarianceStabilizedData* function. Gene expression levels were then mean-centered and standardized using the *scale* function in R. Pearson correlation coefficients were computed between each gene of interest and all other genes. Statistical significance of Pearson correlations was determined by comparison to correlation coefficients achieved by 1,000 random permutations of the expression data. We controlled for multiple hypothesis testing using the *qvalue* package in R. The 253-gene *SChLAP1* correlation signature was determined by imposing a cutoff of $q < 0.05$.

Oncomine Concepts Analysis of *SChLAP1* Signature

We separated the 253 genes with expression levels significantly correlated to *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-value < 10^{-6}. We exported these results as nodes and edges of a concept association network, and visualized the network using Cytoscape version 2.8.2. The 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 on the genes *SChLAP1, EZH2, PCA3, AMACR, ACTB*. For each gene we created signatures from the top 5 percent of positively and negatively correlated genes. We performed a large meta-analysis of these correlation signatures across Oncomine datasets corresponding to disease outcome (Glinsky Prostate, 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 Percent of genes up- and down-regulated. 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 associates with both the over- and under-expression gene sets from a single concept, only the most significant result (as determined by odds ratio) is shown.

Analysis of *SChLAP1-SNF5* expression signatures

The si*SCHLAP1* and si*SNF5* gene signatures were generated from Agilent gene expression microarray datasets. For each cell line we obtained a single vector of per-gene fold changes by averaging technical replicates and then taking the median across biological replicates. We merged the individual cell line results using the median of the changes in 22Rv1 and LNCaP. Venn diagram plots were produced using the BioVenn website (http://www.cmbi.ru.nl/cdd/biovenn/)[57]. We then compared the top 10% up-regulated and down-regulated genes for si*SChLAP1* and si*SNF5* to gene signatures downloaded from the Taylor Prostate 3 dataset 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 comparison using the "heatmap.2" function in the "gplots" R package.

Kaplan-Meier Survival Analysis Based on *SChLAP1* Gene Signature

We downloaded prostate cancer expression profiling data and clinical annotations from GSE8402 published by Setlur et. al.[23]. We intersected the 253-gene *SChLAP1* signature with the genes in this dataset and 80 genes in common. We then assigned SChLAP1 expression scores to each patient sample in the cohort using the un-weighted 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 in order to download the expression data for 167 of the 253 genes profiled by the Glinsky prostate dataset[22]. We assigned *SChLAP1* expression scores in a similar fashion and designated the top 20% of patients as "high" for *SChLAP1*. We performed survival analysis using the time to biochemical PSA recurrence and computed statistical significance as above.

PhyloCSF Analysis

46-way multi-alignment FASTA files for *SChLAP1, HOTAIR, GAPDH,* and *ACTB* were obtained using the "Stitch Gene blocks" tool within the Galaxy bioinformatics framework (usegalaxy.org). We evaluated each gene for its

likelihood to represent a protein-coding region using the PhyloCSF software (version released 2012-10-28). 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 reflect the likelihood that a predicted protein coding sequence is preferred over its non-coding counterpart.

Mayo Clinic Cohort Analyses

Study Design

Patients were selected from a cohort of high-risk radical prostatectomy (RP) patients from the Mayo Clinic. The cohort was defined as 1010 high-risk men that underwent RP between 2000 -2006, of which 73 patients developed clinical progression (defined as patients with systemic disease as evidenced by positive bone or CT scan)[58]. High-risk of recurrence was defined as pre-operative PSA >20 ng/ml, pathological Gleason score 8-10, seminal vesicle invasion (SVI), or GPSM score >=10 [59]. The sub-cohort incorporated all 73 CP progression patients and a 20% random sampling of the entire cohort (202 men including 19 with CP). The total case-cohort study was 256 patients, of which tissue specimens were available for 235 patients. The sub-cohort was previously used to validate a genomic classifier (GC) for predicting Clinical Progression[58].

Tissue Preparation

Formalin-fixed paraffin embedded (FFPE) samples of human prostate adenocarcinoma prostatectomies were collected from patients with informed

consent at the Mayo Clinic according to an institutional review board-approved protocol. Pathological review of H&E tissue sections was used to guide macrodissection of tumour from surrounding stromal tissue from three to four 10 µm sections. The index lesion was considered the dominant lesion by size.

RNA Extraction and Microarray Hybridization

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

Microarray Expression Analysis

The normalization and summarization of the microarray samples was done with the frozen Robust Multiarray Average (fRMA) algorithm using custom frozen vectors. These custom vectors were created using the vector creation methods as described previously[60]. Quantile normalization and robust weighted average

methods were used for normalization and summarization, respectively, as implemented in fRMA.

Statistical Analysis

Given the exon/intron structure of isoform 1 of *SChLAP1*, all probe selection regions (or PSRs) that fall within the genomic span of *SChLAP1* were inspected for overlapping with any of the exons of this gene. One PSR, 2518129, was found fully nested within the third exon 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 of high and low expression of *SChLAP1*.

Statistical analysis on the association of *SChLAP1* with clinical outcomes was done using three endpoints (i) Biochemical Recurrence, defined as two consecutive increases of >=0.2ng/ml after RP, (ii) Clinical Progression, defined as a positive CT or bone scan and (iii) Prostate Cancer Specific Mortality (or PCSM).

For CP end point, all patients with CP were included in the survival analysis, whereas the controls in the sub-cohort were weighted in a 5-fold manner in order to be representative of patients from the original cohort. For PCSM end point, patients from the cases who did not die by PCa were omitted, and weighting was applied in a similar manner. For BCR, since the case-cohort was designed based

on CP endpoint, resampling of BCR patients and sub-cohort was done in order to have a representative of the selected BCR patients from the original cohort.

Tables

Gene	Purpose	Sense	Antisense
SChLAP1	qPCR	TGGACACAATTTCAAGTCCTCA	CATGGTGAAAGTGCCTTATACA
AK093002	qPCR	GGGAATCCCCATATCACAGT	GGAAGCTTACAGTTTTCAAGCA
LOC145837	qPCR	CACCTTGGAGAGACCCAGAA	GGCAGATGGAAGAAGTGGAA
U1	qPCR	GGGAGATACCATGATCACGAAGGT	CCACAAATTATGCAGTCGAGTTTCCC
PCAT-1	qPCR	TGAGAAGAGAAATCTATTGGAACC	GGTTTGTCTCCGCTGCTTTA
SNF5	qPCR	GAGACTCTGACAGACGCTGAGA	GTGTGCTGATGGGCTGGTTAC
MALAT-1	qPCR	GCTCAGTTGCGTAATGGAAAG	CATCACCGGAATTCGATCAC
MEG3	qPCR	TTTTGTGCCCAAGGCTCCTGGA	AGGGACTCAAGGAGCCAGGTTA
ANRIL	qPCR	TCTGATTCAACAGCAGAGATCAA	CAGCACACCTAACAGTGATGC
CLCN5	qPCR	GACTTGTGGGCTTTGTCCTC	CCCATCCTGTTTCTTTCGAG
ANGPTL2	qPCR	CCTGGATGGCTCTGTTAACTTC	GGTTCGTCAGCCAGTAAATGTT
TSGA1	qPCR	GTGCTCACCATCGATCTCCT	TTGACCTTTCTCAGGGATGTG
HOXD9	qPCR	CAGCAACTTGACCCAAACAA	TTTCTCCAGCTCAAGCGTCT
NOS2	qPCR	GAAATTGAATGAGGAGCAGGTC	GCCTCGTAAGGAAATACAGCAC
BRG	qPCR	AAAATCGAGAAGGAGGATGACA	CCAAGCTTGATCTTCACTTTGA
HIF2A	qPCR	GATGTGGAAACGGATGAAGAAC	ATGGGGTTTTGGGTGAACTTAT
FLG	qPCR	TGCCATAATTAATCTTTTCAAGCA	CCATATCTGGGTCATCTGGATT
GLUT1	qPCR	GCAGAGATAGACCAGCAGTTCC	AGAGATCTCGATGCGGCTGTA
GAPDH	qPCR	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGAG
HMBS	qPCR	GATGGGCAACTGTACCTGACTGGA	TGGGGCCCTCGTGGAATGTTA
ACTB	qPCR	AAGGCCAACCGCGAGAAG	ACAGCCTGGATAGCAACGTACA
NFIX promoter	ChIP-PCR	TGTTTTCCCAGACTGCTGTG	CACACATGTCACGTCCAGAA
PTPRU promoter	ChIP-PCR	TGTGTCTGAGCGTAGGATGG	ACAACATTCACCCCCAACAT
TMEM248 promoter	ChIP-PCR	ATGCACACACCCATACCATC	TGGCTTGTAGTGTGCATCGT
KIAA1274 promoter	ChIP-PCR	CTCTCCAGCCTGACAGAGGT	CTGTTACCACCTGCCTCACC
KIAA0841	ChIP-PCR	GATTGAAGCTGTAGGAGTGATGC	AGTGGGTTATCCCAGTCACATC
Chr6 Alu	ChIP-PCR	TCCTGTTCCTTTTGACTGGAAT	ATGGCAGATTTTTGCAAGACAG
Chr2 Alu	ChIP-PCR	GGTTTGTTTTGTCTAGGTTAAGTACG	CATGGAAGGCACCCATTTACTA
ATP8B2 promoter	ChIP-PCR	CTTCTGGCTTTCAGGATTGC	TGCCTGTCTGGCTAAGCTCT
IER2 Promoter	ChIP-PCR	TCTCCTCAGCCATCTTGGTT	CGAATAGGCCCAGACAGAAC
KLK3 Promoter	ChIP-PCR	CTTGGAGTGCAAAGGATCTAGG	AGACTTCATCTAGGGGACAAAGG
KLK5 Promoter	ChIP-PCR	TGGTGAGAAGCCAGATCCTT	TCCGGAATGTCACTCCCTAC
NEFL Promoter	ChIP-PCR	AGCAGAACAAGGTCCTGGAA	TTCTCGTTGGTGGCATCTTC
PDPN Promoter	ChIP-PCR	GCCTGGTATTCGAGGTTGTC	ACTAAGCCGCTGCCTCTCCT
STX10 Promoter	ChIP-PCR	CGAATAGGCCCAGACAGAAC	TCTCCTCAGCCATCTTGGTT
TRPC1 Promoter	ChIP-PCR	GTGCGACAAGGGTGAGAGTT	TGCCCCGATATAAGTTGAGG
SChLAP1 full length construct	Molecular cloning	GCTTTTATGAGCTGTAACACTCA	ATTTATAAGTGAAAGAGGTTTA
SChLAP1	3' RACE	TCAGAGGATAAACATGTGATAT	NA
SChLAP1	3' RACE nested	GATAAACATGTGATATGGTTTG	NA
SChLAP1	3' RACE nested	GTGTCCCCACCCAAATATCATCT	NA
SChLAP1	5' RACE	NA	TTAGCCAGGACTTGATGGAAAT
SChLAP1	5' RACE nested	NA	TTCCAGGTACATGGTGAAAGTG

Table 2.1 Primers used in this study

Target	Probe Name	Sequence of Oligo (5'- to -3')
SChLAP1	SCHLAP_1	TTCTGGACACAATTTCAAGT
	SCHLAP_2	AACCTGTATAAGGCACTTTC
	SCHLAP_3	CACCAGAAATCACAAACATG
	SCHLAP_4	AGCCTCTGGGGATGCAGATA
	SCHLAP_5	CCCTGAAGAAGCTGAATATC
	SCHLAP_6	AGCGGGATGGAGAAAGGAGG
	SCHLAP_7	AATTTCTGGAGATGGTGAAC
	SCHLAP_8	TCCGGTGAATATCAGCACTG
	SCHLAP_9	GCGAAAAGAATGAAGAGTCA
	SCHLAP_10	TCATATAAATTGAGTGTATA
	SCHLAP_11	TCCTGGCTAACCTCATTATG
	SCHLAP_12	CCACTTTACATTAATCTGTA
	SCHLAP_13	TTGTAATCCCATTATCCAAA
	SCHLAP_14	CCAATGTTCACTGTGAAGGA
	SCHLAP_15	GAGGGAAAACCCCTTTCACC
TERC Positive Control	TERC_1	aggcccaccctccgcaaccc
	TERC_2	ggcgcctacgcccttctcaa
	TERC_3	ccgctgaaagtcagcgagaa
	TERC_4	ctctagaatgaacggtggaa
	TERC_5	aggtccccgggaggggcgaa
	TERC_6	cgaccgcggcctccaggcgg
	TERC_7	aactcttcgcggtggcagtg
	TERC_8	tgaacctcgccctcgccccc
	TERC_9	gcgcggggactcgctccgtt
	TERC_10	tgagccgagtcctgggtgca
Non-Targetting Control (Built Against LacZ)	LacZ_1	ccagtgaatccgtaatcatg
	LacZ_2	attaagttgggtaacgccag
	LacZ_3	aatgtgagcgagtaacaacc
	LacZ_4	aataattcgcgtctggcctt
	LacZ_5	aattcagacggcaaacgact
	LacZ_6	atcttccagataactgccgt
	LacZ_7	gctgatttgtgtagtcggtt
	LacZ_8	aactgttacccgtaggtagt
	LacZ_9	tttcgacgttcagacgtagt
	LacZ_10	accattttcaatccgcacct
	LacZ_11	ttcatcagcaggatatcctg
	LacZ_12	tggttcggataatgcgaaca
	LacZ_13	agacgattcattggcaccat
	LacZ_14	atttgatccagcgatacagc
	LacZ_15	tttgatggaccatttcggca
	LacZ_16	aaacggggatactgacgaaa
	LacZ_17	atacagaactggcgatcgtt
	LacZ_18	tattcgctggtcacttcgat
	LacZ_19	tttaccttgtggagcgacat
	LacZ_20	aaatccatttcgctggtggt
	LacZ_21	tgtgaaagaaagcctgactg
	LacZ_22	tacgccaatgtcgttatcca
	LacZ_23	atcaatccggtaggttttcc
	LacZ_24	agttttcttgcggccctaat

Table 2.2 ChIRP probe sequences

Figures

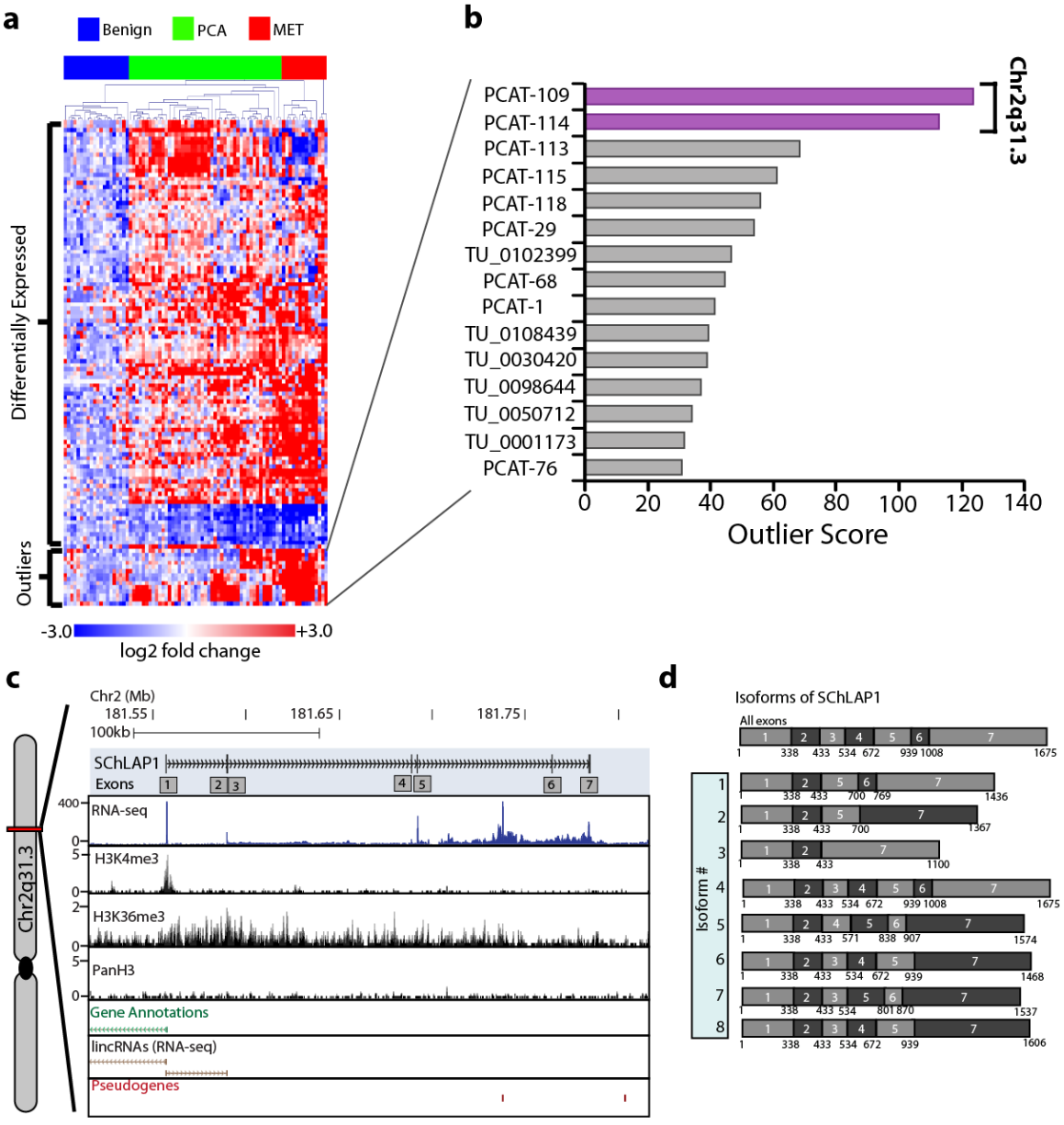


Figure 2.1 Discovery of *SChLAP1* as a prostate cancer lncRNA.
(a) Unsupervised clustering analyses of differentially expressed or outlier unannotated intergenic transcripts clusters benign samples, localized tumors and metastatic cancers. Expression is plotted as log2 fold-change relative to the median of the benign samples. **(b)** Cancer outlier profile analysis (COPA) for intergenic lncRNAs. **(c)** A representation of the *SChLAP1* gene and its annotations in current databases. An aggregated representation of current gene annotations for Ensembl, ENCODE, UCSC, Ref-Seq, and Vega shows no annotation for *SChLAP1*. ChIP-Seq data for H3K4me3 and H3K36me3 show enrichment at the *SChLAP1* gene. Also, RNA-Seq data showing an outlier

sample for *SChLAP1* illustrates its expression. Prominent peak between exons 5 and 6 correspond to a pseudogene. **(d)** A schematic summarizing the observed *SChLAP1* isoforms. A total of 8 isoforms were observed, with isoform #1 and isoform #2 accounting for >90% of transcripts.

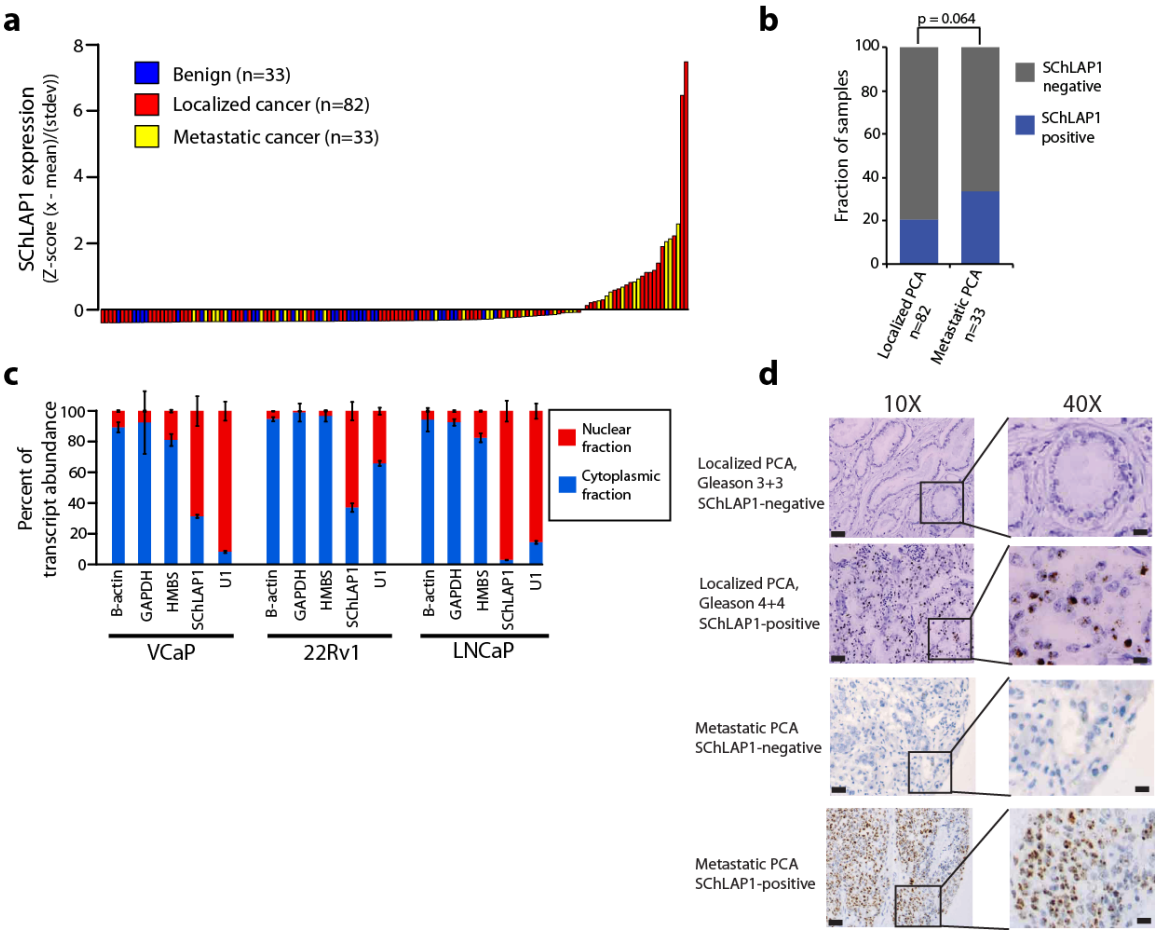


Figure 2.2 Characterization of *SChLAP1* expression.
(a) qPCR for *SChLAP1* on a panel of benign prostate (n=33), localized prostate cancer (n=82), and metastatic prostate cancer (n=33) samples. qPCR data is normalized to the average of (*GAPDH* + *HMBS*) and represented as standardized expression values. **(b)** Prevalence of *SChLAP1* expression in localized prostate cancer tissues and metastatic prostate cancer tissues. P-value was determined by one-sided Fisher's exact test. **(c)** Fractionation of prostate cell lysates demonstrates nuclear expression of *SChLAP1*. *U1* is a positive control for nuclear gene expression. **(d)** *In situ* hybridization of *SChLAP1* in human prostate cancer. *SChLAP1* staining is shown for both localized and metastatic tissues.

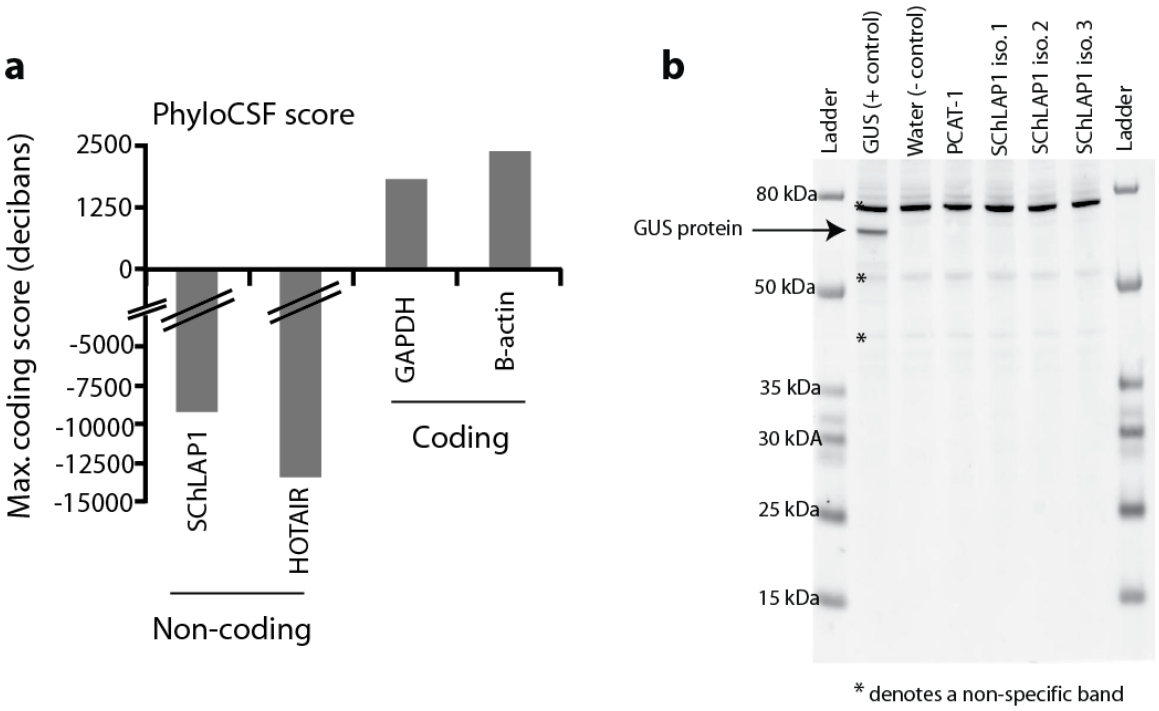


Figure 2.3 *SChLAP1* is a noncoding gene.
(a) Analysis of the coding potential for the *SChLAP1* sequence across 29 mammals in all 3 reading frames using PhyloCSF. *HOTAIR* serves as a control non-coding gene. *GAPDH* and *B-actin* serve as control coding genes. Scores above 0 suggest coding potential whereas scores below 0 suggest no coding potential. **(b)** *In vitro* translation assays for *SChLAP1*. Three isoforms of *SChLAP1* were cloned and tested for protein-coding capacity using an *in vitro* translation assay. GUS is used as a positive control. *PCAT-1* and water serve as negative controls. Non-specific bands are indicated with an asterisk. *SChLAP1* isoforms do not generate a protein in this assay.

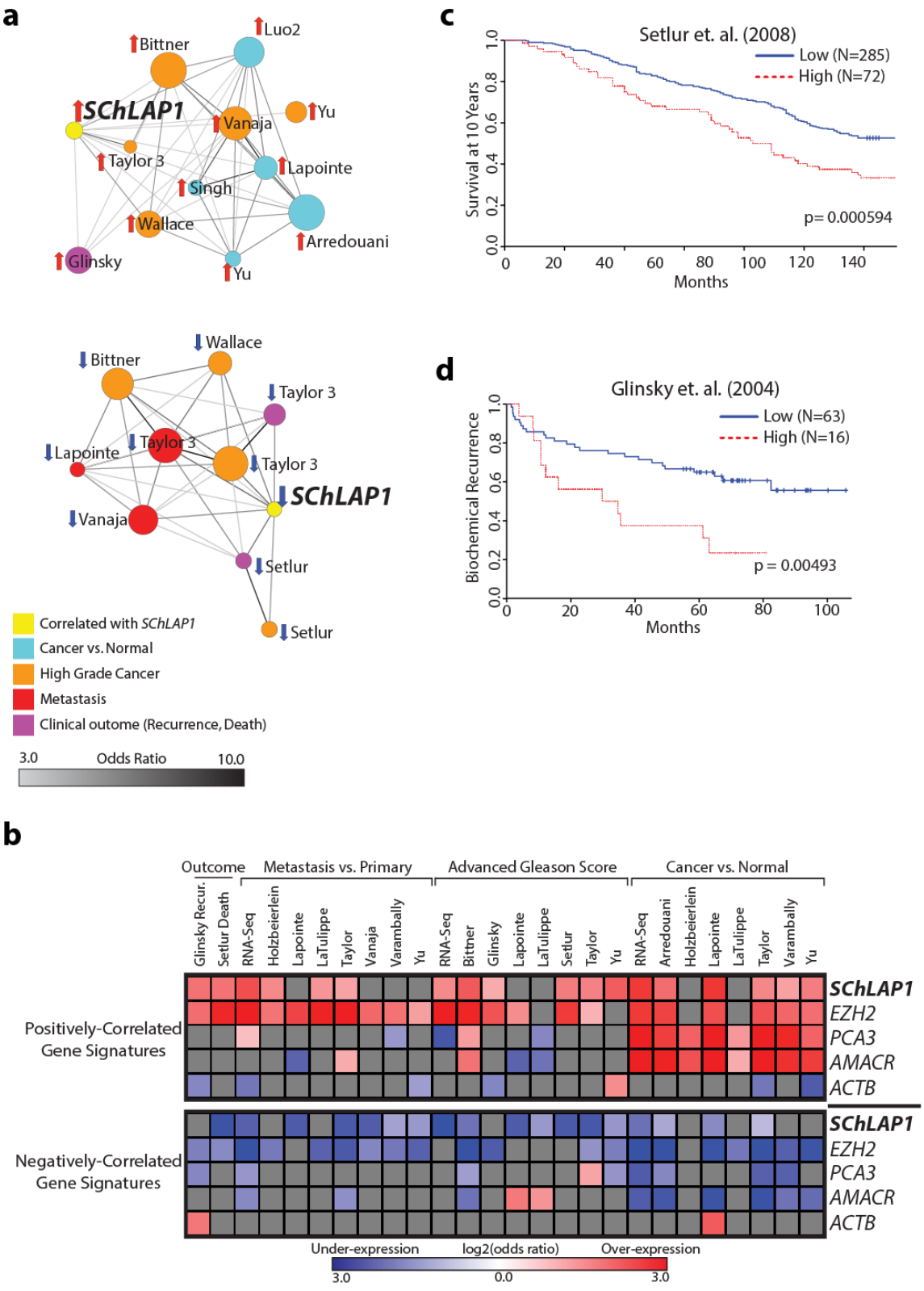

a
Bittner
Luo2
SChLAP1
Vanaja
Yu
Taylor 3
Lapointe
Singh
Wallace
Arredouani
Glinsky
Yu
Wallace
Bittner
Taylor 3
Lapointe
Taylor 3
Taylor 3
SChLAP1
Vanaja
Setlur
Setlur
Correlated with SChLAP1
Cancer vs. Normal
High Grade Cancer
Metastasis
Clinical outome (Recurrence, Death)
3.0
Odds Ratio
10.0
c
Setlur et. al. (2008)
Low (N=285)
High (N=72)
Survival at 10 Years
p= 0.000594
Months
d
Glinsky et. al. (2004)
Low (N=63)
High (N=16)
Biochemical Recurrence
p = 0.00493
Months
b
Outcome
Metastasis vs. Primary
Advanced Gleason Score
Cancer vs. Normal
Glinsky Recur.
Setlur Death
RNA-Seq
Holzbeierlein
Lapointe
LaTulippe
Taylor
Vanaja
Varambally
Yu
RNA-Seq
Bittner
Glinsky
Lapointe
LaTulippe
Setlur
Taylor
Yu
RNA-Seq
Arredouani
Holzbeierlein
Lapointe
LaTulippe
Taylor
Varambally
Yu
Positively-Correlated Gene Signatures
Negatively-Correlated Gene Signatures
SChLAP1
EZH2
PCA3
AMACR
ACTB
Under-expression
log2(odds ratio)
Over-expression
3.0
0.0
3.0

Figure 2.4 *SChLAP1* expression characterizes aggressive prostate cancer. **(a)** Network representation of genes positively or negatively correlated with *SChLAP1* in localized prostate cancers using Oncomine concepts analysis and visualized with the Force Directed Layout algorithm in the Cytoscape tool. Node sizes reflect the number of genes that comprise each molecular concept and node names are labeled according to the author of the primary study. The nodes are colored according to the concept categories indicated in the figure legend. Edges are drawn between nodes with statistically significant enrichment (p-value < 1e-6, odds ratio > 3.0) and darker edge shading implies higher odds ratio. **(b)** Heatmap representation of comparisons between co-expression gene signatures and molecular concepts. Comparisons to positively (top portion) and negatively correlated (bottom portion) gene signatures are shown separately. Comparisons that do not reach statistical significance (q > 0.01 or odds ratio < 2) are shown in grey. Associations with over-expression concepts are colored red, and under-expression concepts blue. **(c)** Kaplan-Meier analysis of prostate cancer outcomes. Patients were stratified according to their *SChLAP1* signature score. Signature scores at or above the 80^{th} percentile were deemed 'High', and the rest 'Low'. Statistical significance was determined by the log rank test. Analysis of the 10-year overall survival probability for prostate cancer patients from the Setlur *et al.* study. **(d)** As in **(c)**, Analysis of the biochemical recurrence probability for prostate cancer patients from the Glinksy *et al.* study.

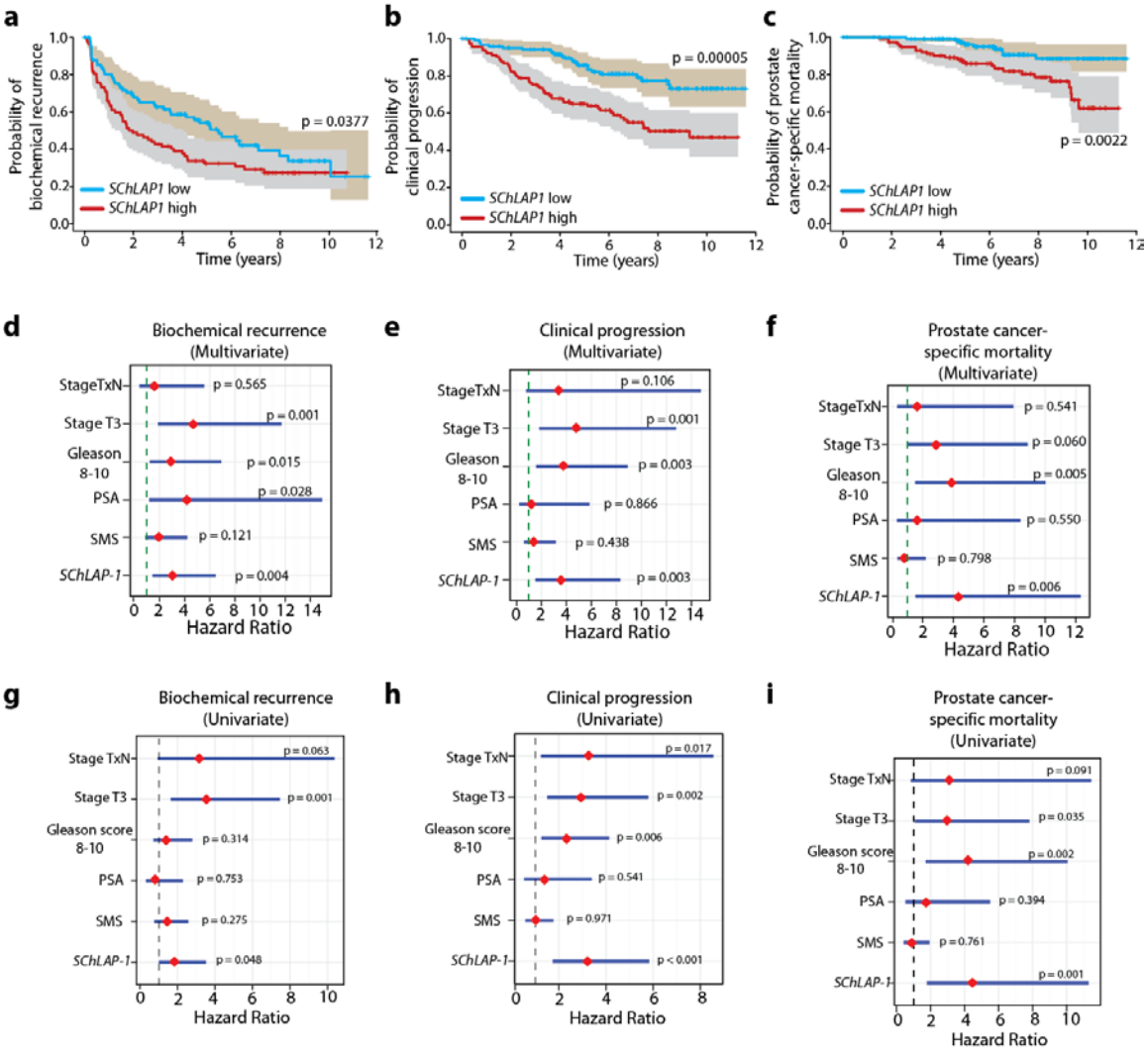


Figure 2.5 *SChLAP1* expression is an independent predictor of patient clinical parameters.

(a-c) 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. **(d-i)** Multivariate and univariate analyses for *SChLAP1* and disease outcomes. **(d-f)** Multivariate survival analyses demonstrate that *SChLAP1* is an independent predictor of prostate cancer biochemical recurrence **(d)**, clinical progression **(e)**, and prostate cancer-specific mortality **(f)** following radical prostatectomy. **(g-i)** Univariate survival analyses for *SChLAP1* for biochemical recurrence **(g)**, clinical progression **(h)**, and prostate cancer-specific mortality **(i)** as in **(d-f)**. For these analyses, clinical significance was adjusted for confounding adjuvant treatment, and Gleason score was dichotomized between those samples ≤7 ≥8. Red diamonds indicate

the median hazard ratio for each factor and blue lines indicate the 95% confidence interval.)

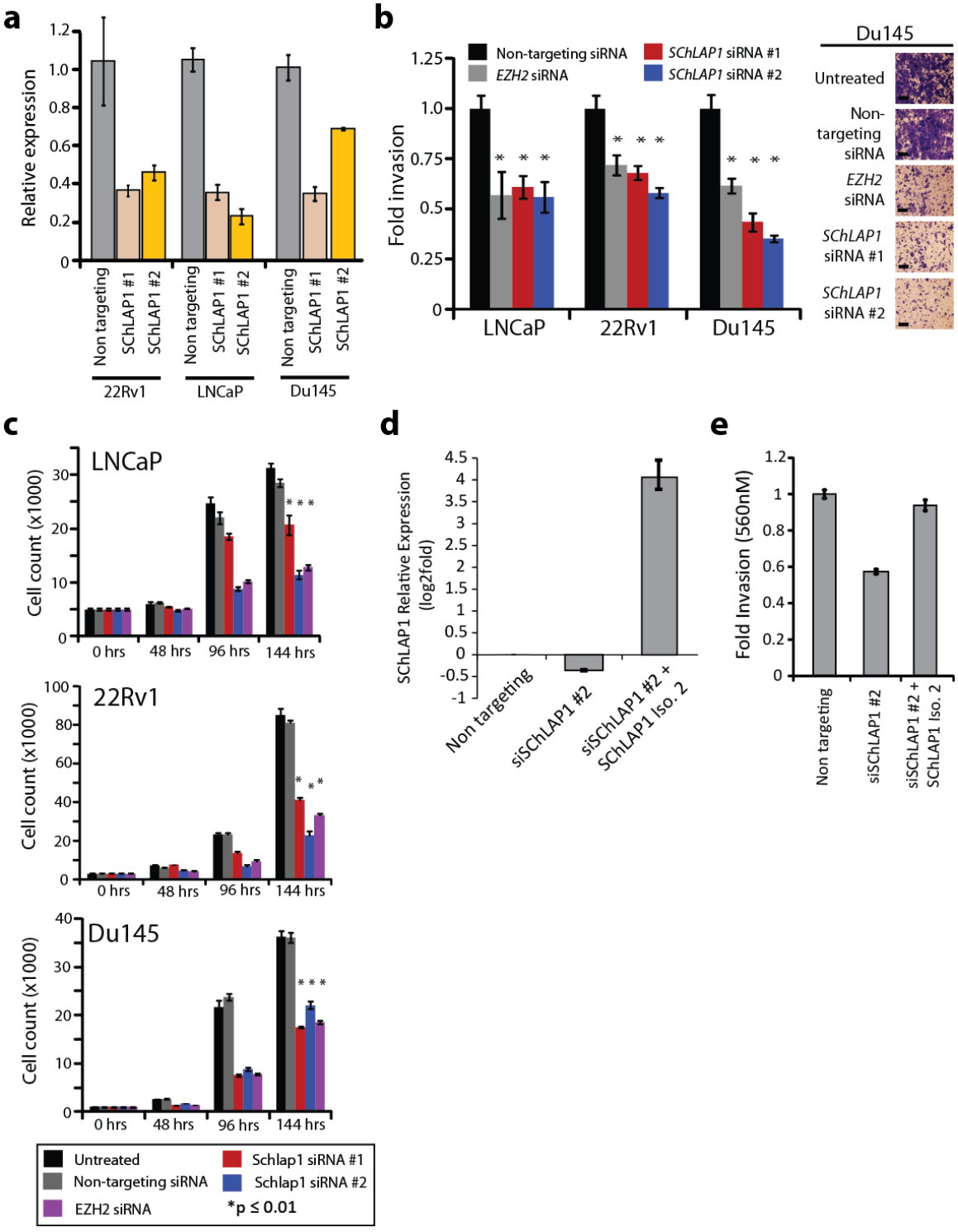


Figure 2.6 *In vitro* knockdown of *SChLAP1* impairs cell invasion and proliferation.
(a) 22Rv1, LNCaP, and Du145 cells were treated with siRNAs against *SChLAP1*. qPCR indicates relative knockdown efficiency in these cell lines. Error bars represent S.E.M. **(b)** siRNA knockdown of *SChLAP1 in vitro* 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. **(c)** Cell proliferation assays for LNCaP, 22Rv1, and Du145 treated with *SChLAP1* siRNAs or non-targeting negative controls. *EZH2* siRNA serves as a positive control. Error bars indicate S.E.M. An asterisk (*) indicates $p < 0.05$ by Student's t-test. Error bars represent S.E.M. **(d)** Expression of *SChLAP1* in 22Rv1 cells treated with non-targeting, siRNA #2 for *SChLAP1*, or siRNA #2 with exogenous overexpression of *SChLAP1* isoform 2. **(e)** Boyden chamber invasion assay data for 22Rv1 cells treated with non-targeting, siRNA #2 for *SChLAP1*, or siRNA #2 with exogenous overexpression of *SChLAP1* isoform 2. Data are represented as absorbance at 560nM. Error bars represent S.E.M.

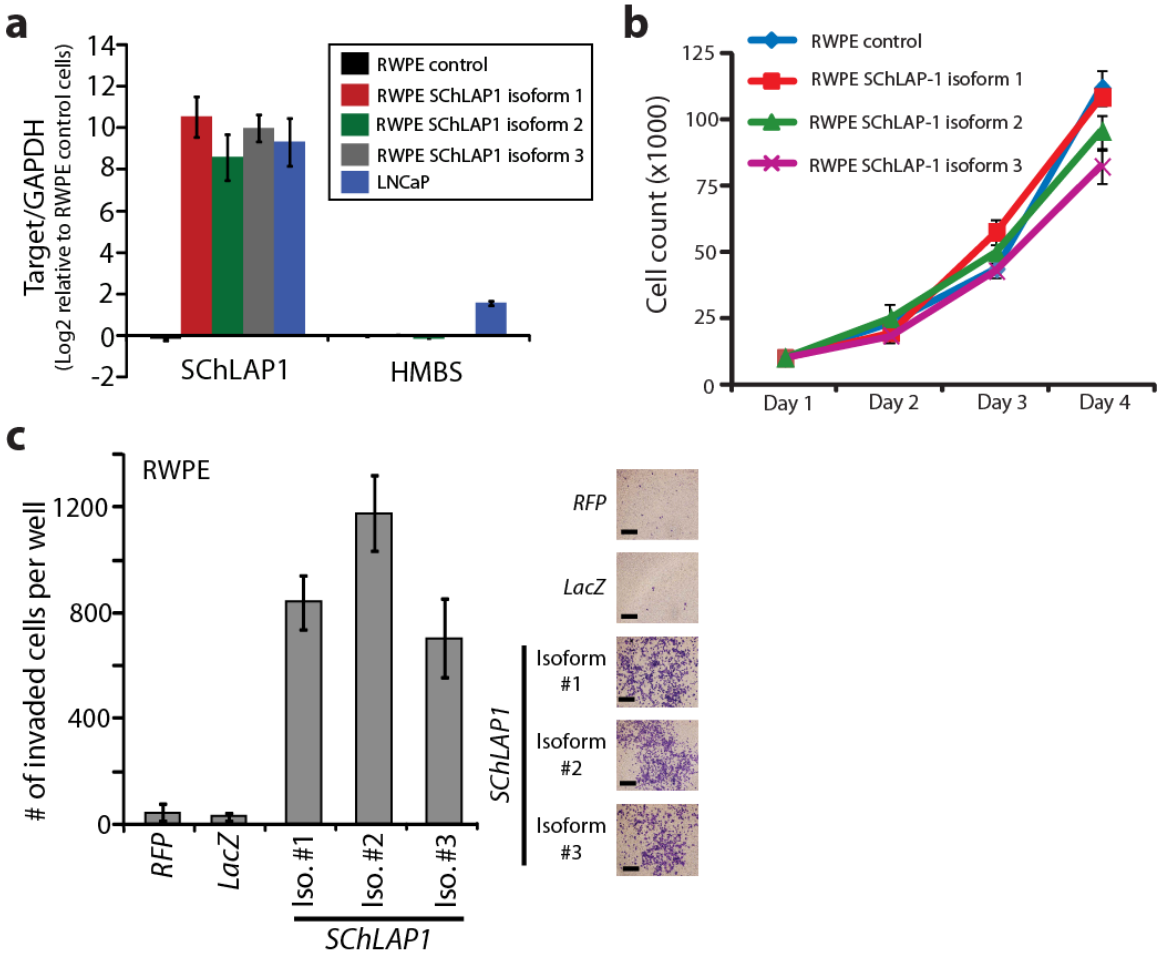


Figure 2.7 Overexpression of *SChLAP1* enhances cell invasion.
(a) Overexpression of *SChLAP1* isoforms 1-3 in RWPE cells was confirmed using qPCR, which demonstrated that the overexpression resulted in comparable levels of *SChLAP1* transcript to LNCaP cells that express this gene endogenously. HMBS serves as a negative control. Error bars represent S.E.M. **(b)** Cell proliferation assays for RWPE cells overexpressing *SChLAP1* isoforms. No significant change in cell proliferation is observed. Error bars represent S.E.M. **(c)** Overexpression of *SChLAP1* in RWPE cells results in increased cellular invasion through Matrigel in Boyden chamber assays.

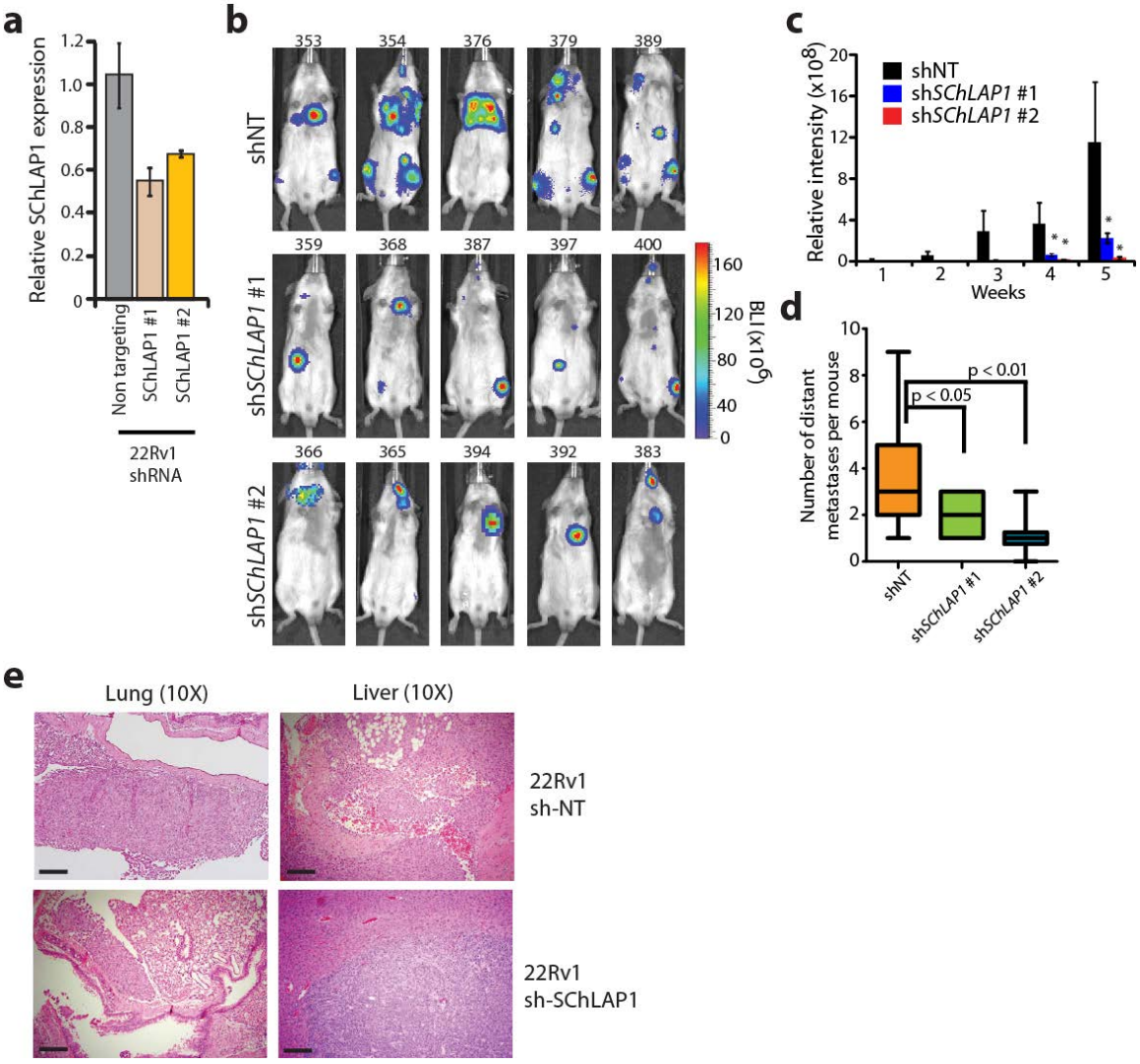


Figure 2.8 *SChLAP1* coordinates metastatic seeding *in vivo*.
(a) Knockdown efficiencies for the shRNA knockdown of *SChLAP1* in LNCAP and 22Rv1 cells. Error bars indicate S.E.M. **(b)** Intracardiac injection of 22Rv1 cells with stable *SChLAP1* knockdown in severe combined immunodeficient (SCID) mice. Example luciferase bioluminescence images from 22Rv1 shNT, sh*SChLAP1* #1, and sh*SChLAP1* #2 mice five weeks following intracardiac injection. Mouse IDs are given above each image. **(c)** The relative intensity of whole-mouse luciferase signal is plotted for 22Rv1 shNT (n=9), sh*SChLAP1* #1 (n=14) and sh*SChLAP1* #2 (n=14) intracardiac injection experiments. **(d)** The number of gross metastatic sites observed by luciferase signal in 22Rv1 sh*SChLAP1* cells or shNT controls. Independent foci of luciferase signal were averaged for shNT (n=9), sh*SChLAP1* #1 (n=14) and sh*SChLAP1* #2 (n=14) mice. **(e)** Histolopathology of murine tumors formed by intracardiac injection of 22Rv1 shNT or 22Rv1 sh-*SChLAP1* cells. Images are taken from the lungs and livers or mice with tumors. Slides are stained with H&E.

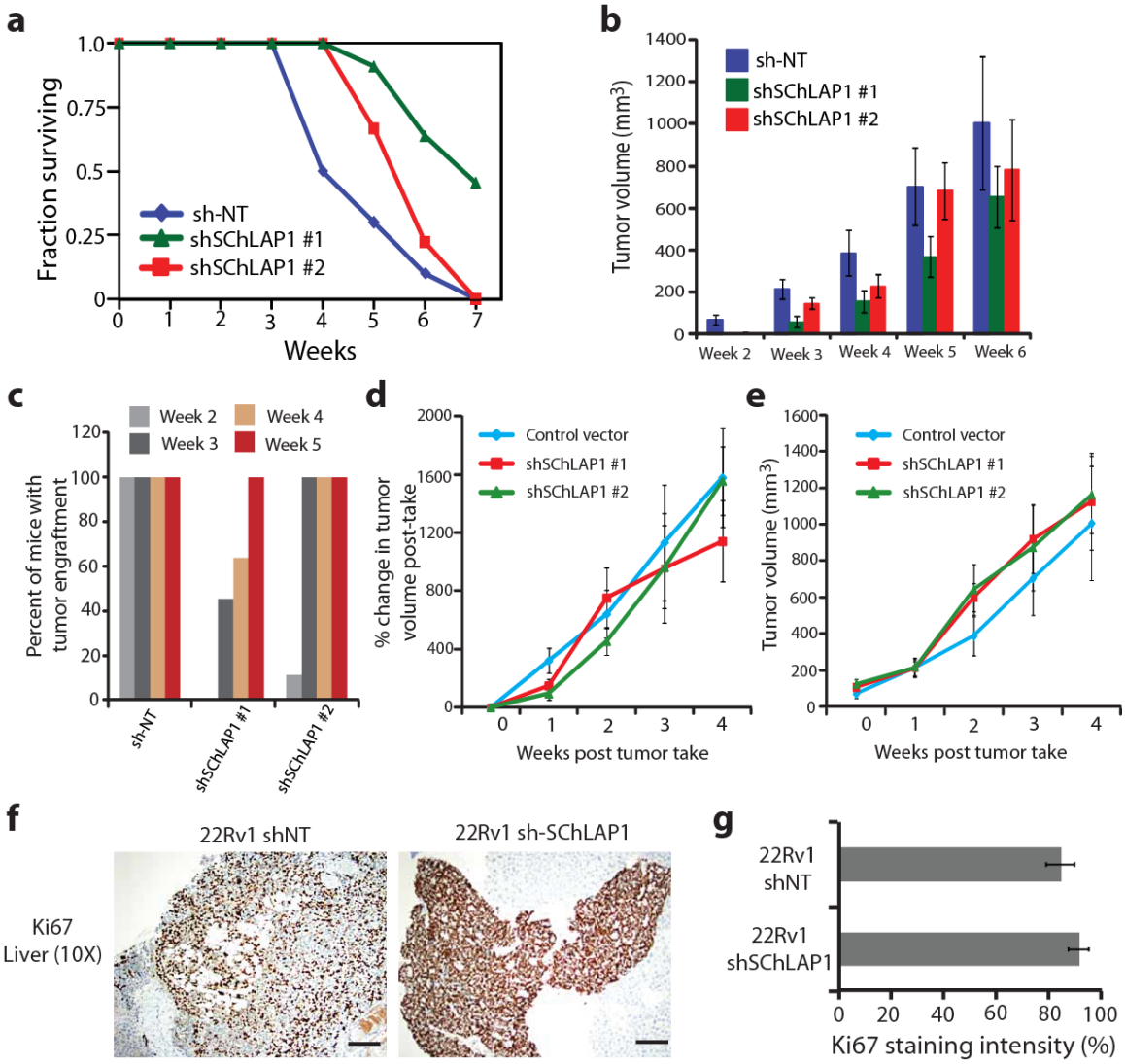


Figure 2.9 Knockdown of *SChLAP1* delays tumor engraftment but not tumor growth kinetics.

(a) The fraction of mice surviving following subcutaneous injection of the 22Rv1 cell lines. This plot represents tumor-specific death of mice sacrificed when the tumor volume reached the maximum allowable volume. **(b)** 22Rv1 cells infected with lentivirus for shNT, sh-*SChLAP1* #1, and sh-*SChLAP1* #2 were injected subcutaneously in mouse flanks and tumor growth was monitored by caliper measurements. N = 10 mice for shNT cells, n = 12 mice for sh-*SChLAP1* #1 cells, n = 9 mice for sh-*SChLAP1* #2 cells. Absolute tumor volume for 22Rv1 shNT, sh-*SChLAP1* #1 and sh-*SChLAP1* #2 cells. Error bars represent S.E.M. **(c)** Percent of mice with tumor engraftment over time. Knockdown of *SChLAP1* delays the onset of tumor engraftment. **(d)** The percent change in tumor volume per cell line normalized to the time of tumor engraftment. Error bars represent S.E.M. **(e)** Tumor volume normalized to the time of tumor engraftment. Error bars represent S.E.M. **(f)** Immunohistochemistry staining for Ki67 in 22Rv1 shNT and sh-*SChLAP1* liver metastases. **(g)** Summary of Ki67 tumor staining for 22Rv1

shNT and sh-*SChLAP1* murine tumors show significant difference in Ki67 staining intensity.

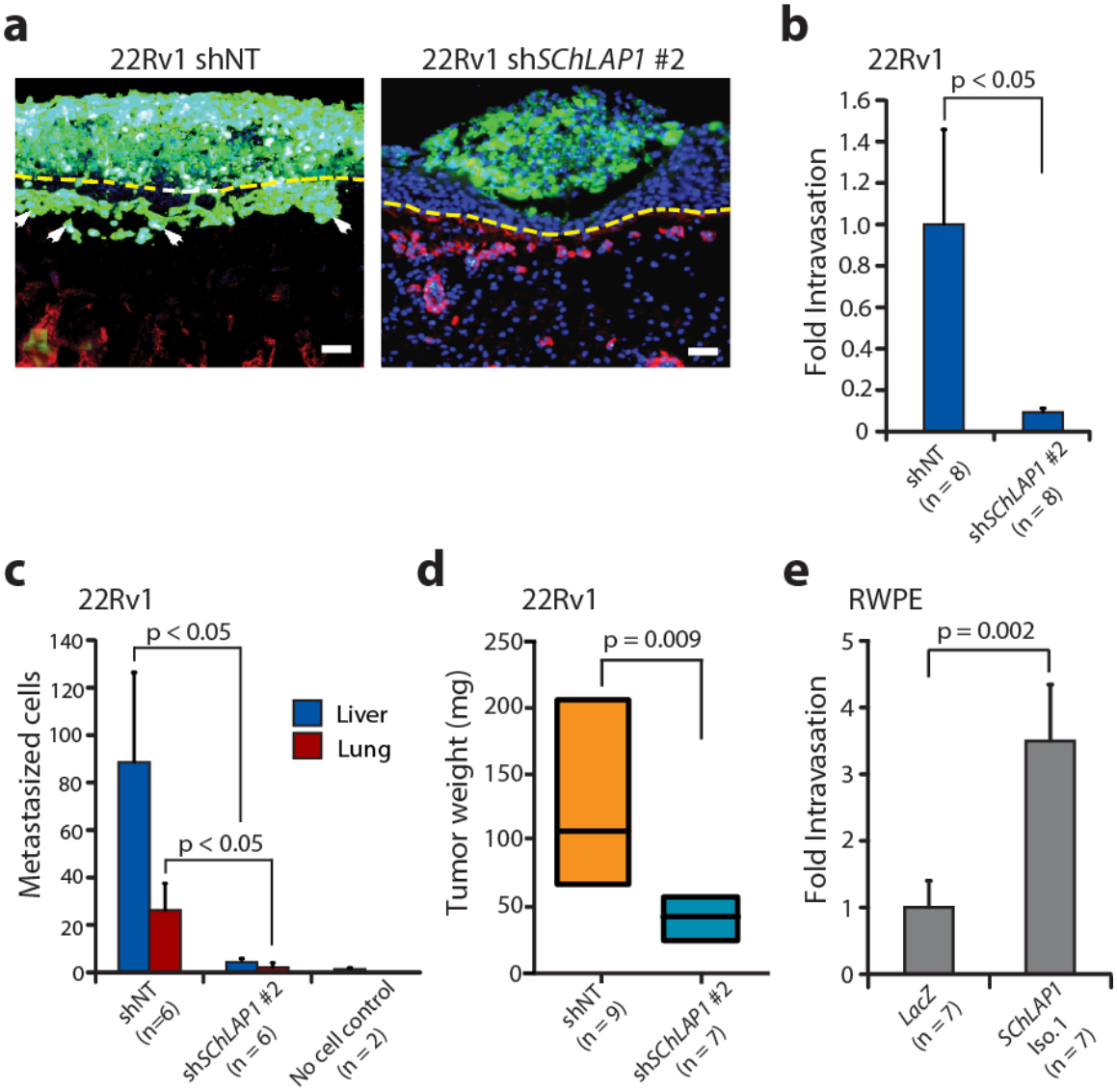


Figure 2.10 Knockdown of *SChLAP1* decreases tumor growth, invasion, and metastasis *in vivo*.
(a) Invasion of 22Rv1-sh*NT* and 22Rv1 sh*SChLAP1* cells in the chick chorioallantoic membrane (CAM) assay. 22Rv1 cells are labeled with GFP. The image is counterstained with chicken collagen IV for vasculature (RFP) and DAPI for nuclei. **(b-e)** Using the CAM assay, 22Rv1 sh*SChLAP1* cells demonstrate decreased intravasation **(b)**, metastatic spread to the liver and lungs **(c)**, and reduced tumor weight **(d)**. **(e)** Quantification of intravasation of RWPE-*LacZ* and RWPE-*SChLAP1* cells in the CAM assay. All data in this figure are represented as mean +/- S.E.M. Statistical significance was determined by a two-tailed Student's *t*-test. An asterisk (*) indicates a p-value < 0.05.

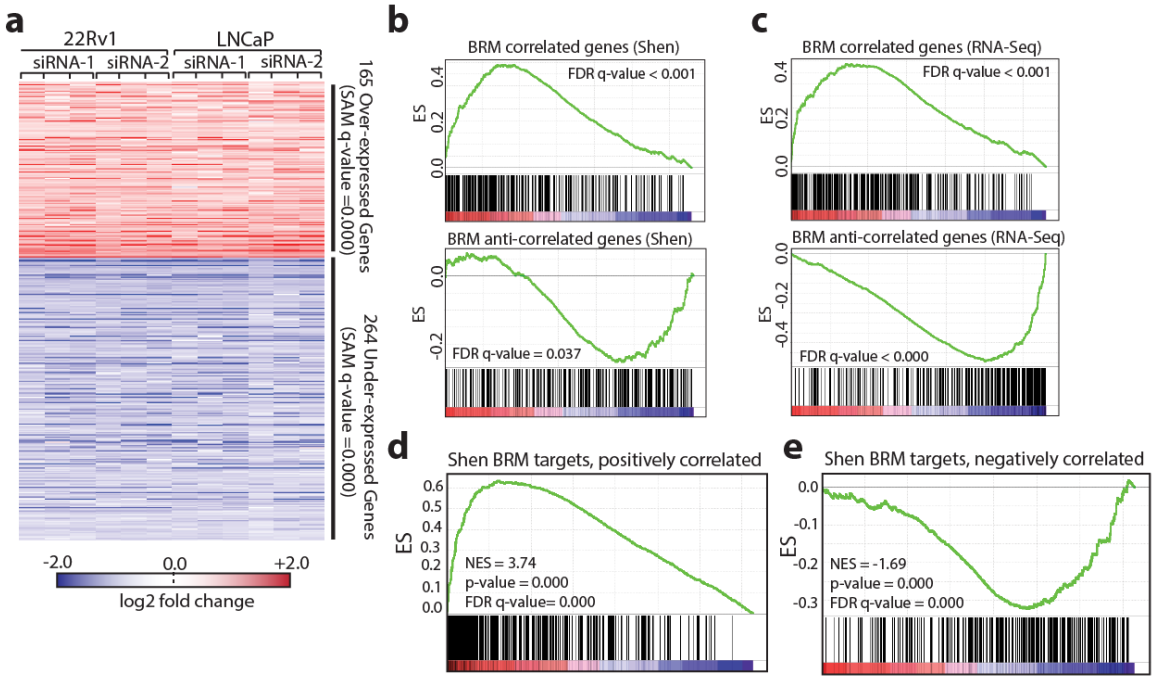


Figure 2.11 Nomination of SWI/SNF concept as a mechanism of *SChLAP1* action.

(a) Transcriptome profiling following *SChLAP1* knockdown *in vitro*. Differentially expressed genes were determined by SAM analysis and represented as a heatmap. **(b-c)** Gene set enrichment analysis (GSEA) of LNCaP and 22Rv1 cells treated with *SChLAP1* siRNAs. GSEA results indicate that *SChLAP1* knockdown results are inversely correlated with SWI/SNF-associated genes using data from Shen *et al.* **(b)** or using RNA-seq data **(c)**. **(d)** Comparison of positively correlated BRM-associated gene signatures in prostate cancer. The BRM-derived signature from RNA-seq samples was compared to the Shen et al. signature by GSEA. A highly significant overlap between the signatures is observed. **(e)** Comparison of negatively correlated BRM-associated gene signatures in prostate cancer. The BRM-derived signature from RNA-seq samples was compared to the Shen et al. signature by GSEA. A highly significant overlap between the signatures is observed.

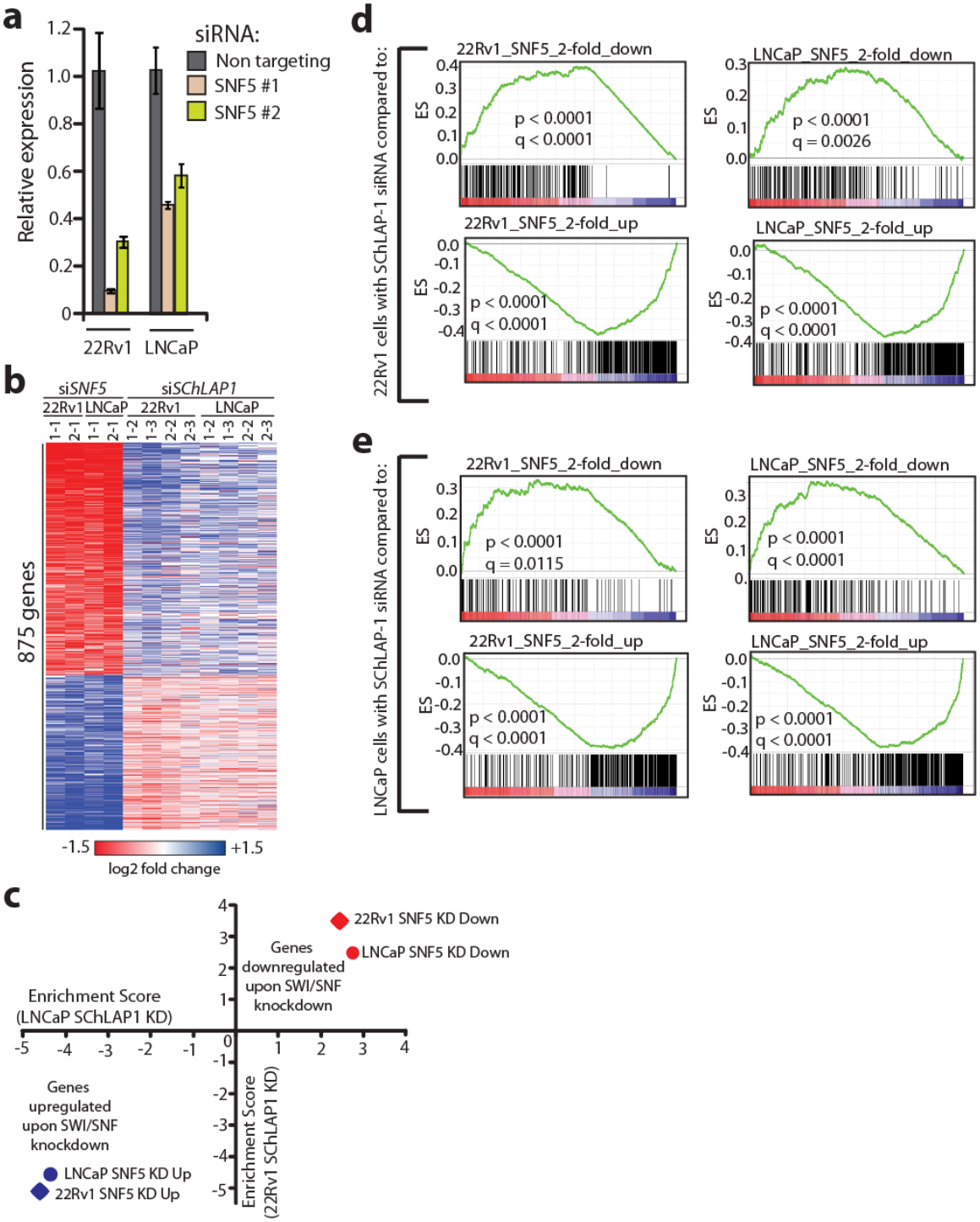


Figure 2.12 *SChLAP1* and the SWI/SNF complex regulate gene expression in an opposing manner.
(a) Knockdown efficiency of *SNF5* siRNAs in 22Rv1 and LNCaP. Error bars represent S.E.M. **(b)** Heatmap results for *SChLAP1* or *SNF5* knockdown in LNCaP and 22Rv1 cells. **(c)** GSEA analysis of *SChLAP1* and *SNF5* knockdowns. Across two cell lines (LNCaP and 22Rv1), *SChLAP1* knockdown had the

opposite effect on gene expression as knockdown of *SNF5*. Here, a positive GSEA normalized enrichment score (NES) indicates genes up-regulated upon *SChLAP1* knockdown, and a negative GSEA NES indicates genes down-regulated upon *SChLAP1* knockdown. **(d)** GSEA results from comparisons of *SChLAP1* and *SNF5* knockdown in 22Rv1 cells. *SChLAP1* was knocked-down using siRNAs in 22Rv1 cells. Gene expression changes were compared using GSEA to expression changes observed using *SNF5* siRNAs in LNCaP or 22Rv1 cells. The enrichment plots of these comparisons are shown. **(e)** GSEA results from comparisons of *SChLAP1* and *SNF5* knockdown in LNCaP cells. *SChLAP1* was knocked-down using siRNAs in LNCaP cells. Gene expression changes were compared using GSEA to expression changes observed using *SNF5* siRNAs in LNCaP or 22Rv1 cells. The enrichment plots of these comparisons are shown.

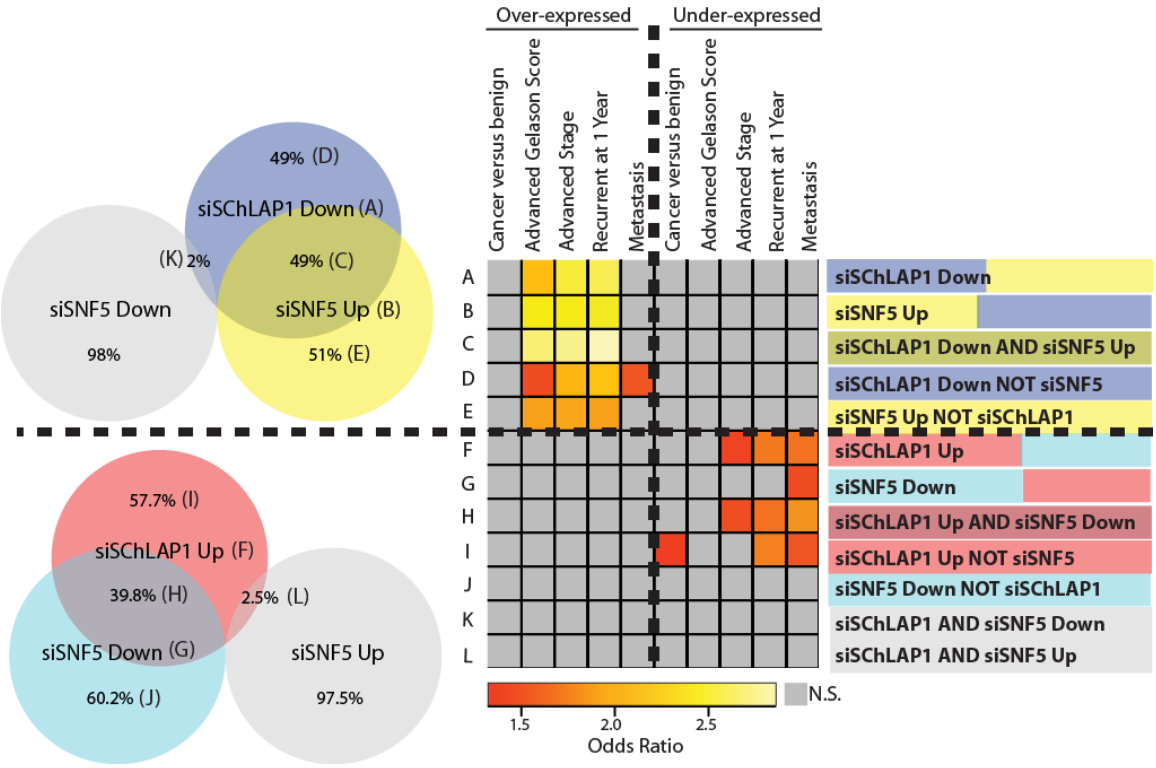


Figure 2.13 *SChLAP1* and *SNF5* co-regulate genes associated with prostate cancer aggressiveness.
The top 10% of up- or down-regulated genes for *SNF5*-knockdown or *SChLAP1*-knockdown microarrays in 22Rv1 and LNCaP were intersected to generate an overlapping gene signature for these knockdown experiments. This signature was analyzed for overlap with the Taylor Prostate 3 Oncomine Concept for disease aggressiveness. Left, Venn diagrams demonstrating overlap of *SChLAP1* and *SNF5*-knockdown experiments. Right, a heatmap visualization showing statistical ($q < 0.05$) overlap of gene signatures from the SNF5 and *SChLAP1* knockdowns with prostate cancer aggressiveness concepts from Oncomine. Odds ratios from the comparisons with q-values <0.05 are shown. One-sided Fisher's exact tests were used for significance.

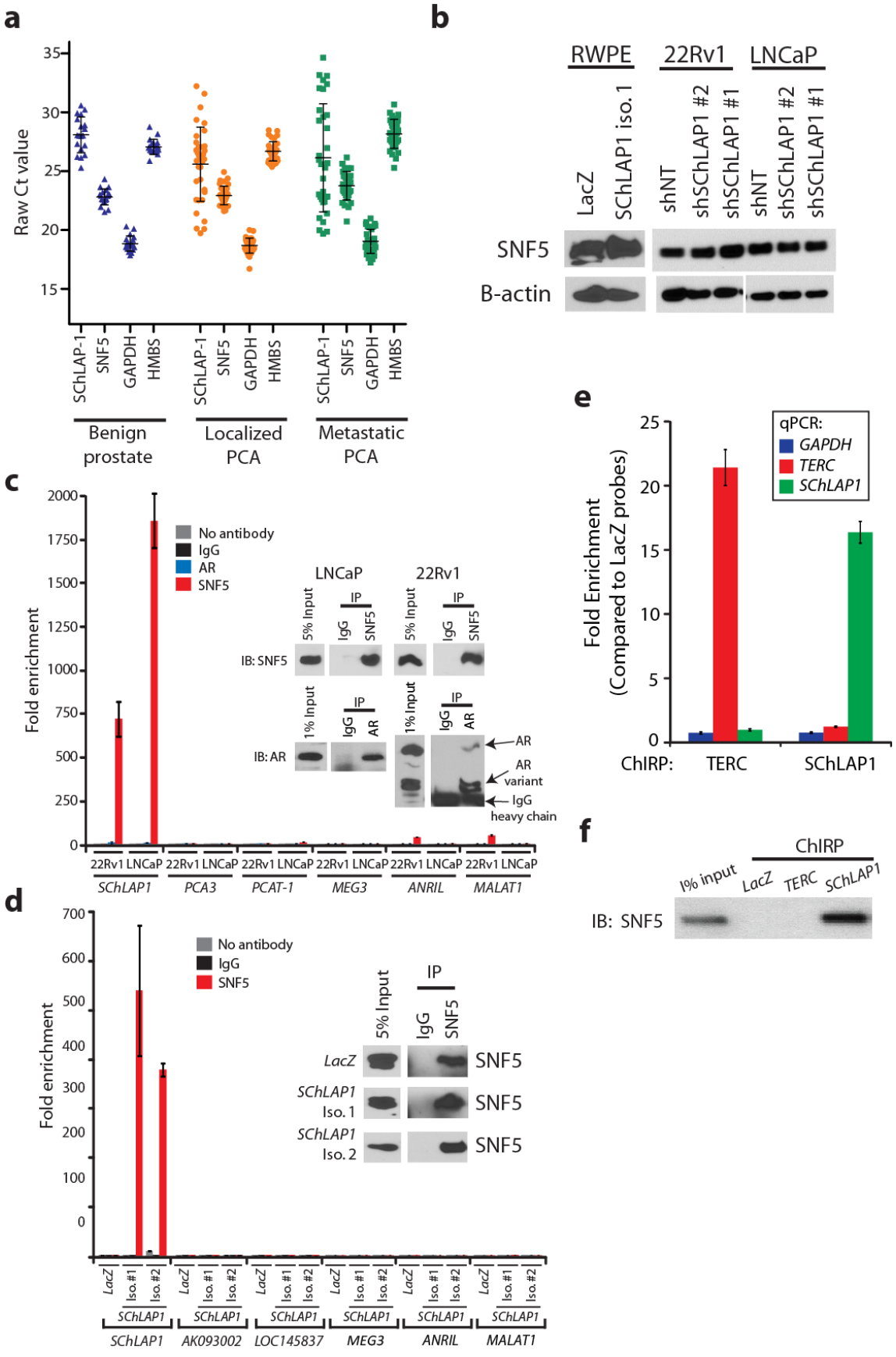

a
Raw Ct value
SChLAP-1
SNF5
GAPDH
HMBS
Benign prostate
Localized PCA
Metastatic PCA
b
RWPE
22Rv1
LNCaP
LacZ
SChLAP1 iso. 1
shNT
shSChLAP1 #2
shSChLAP1 #1
SNF5
B-actin
c
Fold enrichment
No antibody
IgG
AR
SNF5
LNCaP
22Rv1
5% Input
IP
IgG
SNF5
IB: SNF5
1% Input
AR
IB: AR
AR
AR variant
IgG heavy chain
22Rv1 LNCaP
SChLAP1
PCA3
PCAT-1
MEG3
ANRIL
MALAT1
d
Fold enrichment
No antibody
IgG
SNF5
5% Input
IP
LacZ
SChLAP1 Iso. 1
SChLAP1 Iso. 2
Iso.#1
Iso.#2
SChLAP1
AK093002
LOC145837
MEG3
ANRIL
MALAT1
e
Fold Enrichment (Compared to LacZ probes)
qPCR:
GAPDH
TERC
SChLAP1
ChIRP:
TERC
SChLAP1
f
ChIRP
I% input
LacZ
TERC
SChLAP1
IB: SNF5

Figure 2.14 *SChLAP1* interacts with SNF5.
(a) Relative abundance of *SChLAP1* compared to the SWI/SNF complex in human prostate tissues. qPCR cycle threshold (Ct) values for *SChLAP1*, *SNF5*, *GAPDH*, and *HMBS* are shown. *SChLAP1*-positive samples display Ct values in the low 20s, which is consistent with the abundance of *SNF5*. **(b)** Western blot analysis of SNF5 protein abundance in prostate cancer cells either overexpressing *SChLAP1* (RWPE) or with stable knockdown of *SChLAP1* (22Rv1, LNCaP). **(c)** RNA immunoprecipitation (RIP) of SNF5 or AR in 22Rv1 and LNCaP cells. Inset Western blots demonstrate pulldown efficiency. **(d)** RIP analysis of SNF5 in RWPE cells overexpressing *LacZ*, *SChLAP1* isoform #1, or *SChLAP1* isoform #2. Inset Western blots demonstrate pulldown efficiency. **(e)** Pulldown of *SChLAP1* RNA. RWPE-*SChLAP1* isoform #1 cells were treated with biotinylated *SChLAP1*, *TERC* or *LacZ* RNA probes according to the ChIRP protocol. Quantification of RNA pulldown efficiency by qPCR is shown. Error bars indicate S.E.M. **(f)** Pulldown of *SChLAP1* RNA using Chromatin Isolation by RNA Purification (ChIRP) recovers SNF5 protein in RWPE-*SChLAP1* isoform 1 cells. *LacZ* and *TERC* serve as controls.

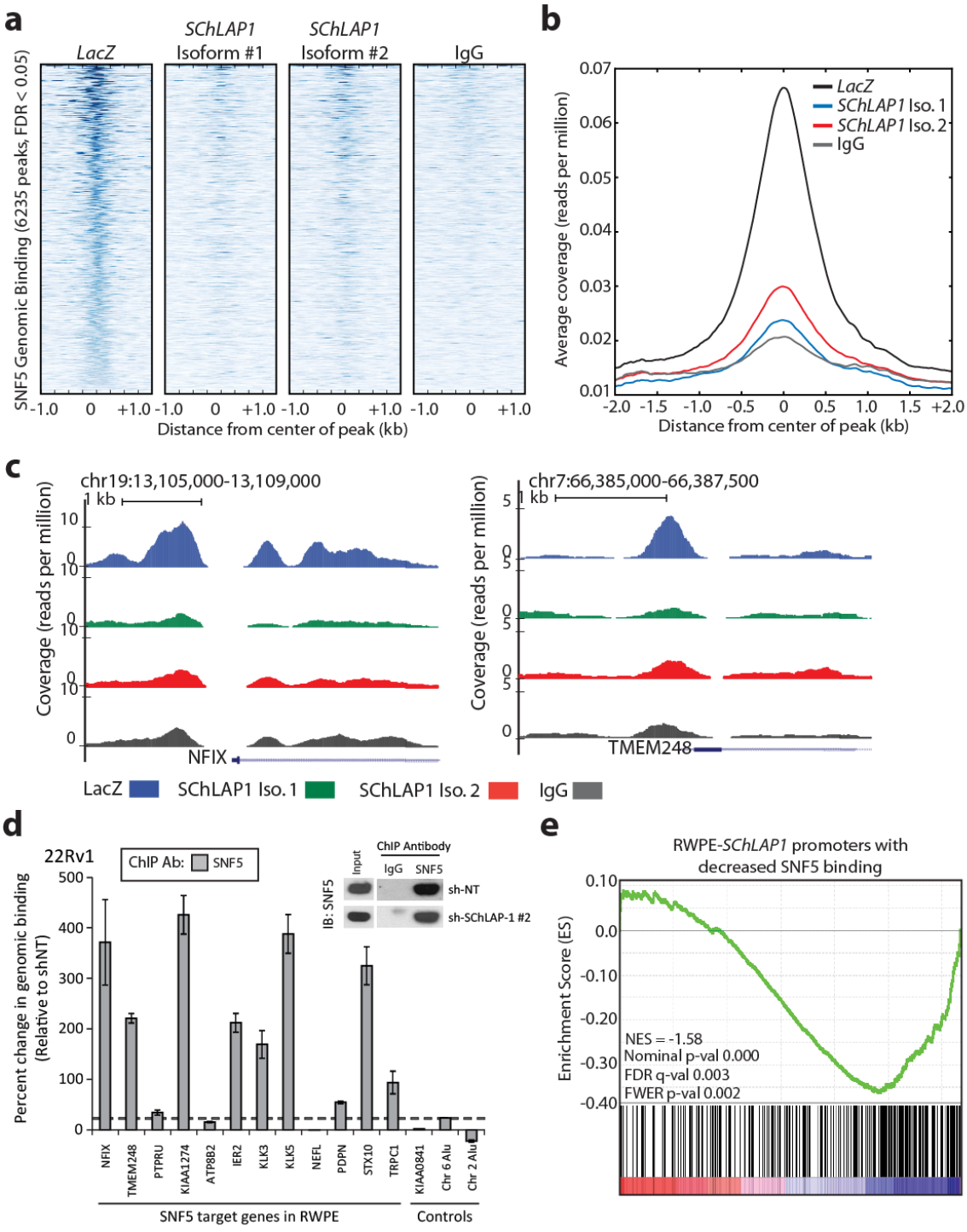


Figure 2.15 *SChLAP1* expression disrupts genomic binding of SNF5. **(a)** A global representation of SNF5 genomic binding over ±2kb window surrounding each SNF5 ChIP-Seq peak in RWPE-*LacZ*, RWPE-*SChLAP1* isoform 1, and RWPE-*SChLAP1* isoform 2 cells. **(b)** A heatmap representation of SNF genomic binding at target sites in RWPE-*LacZ*, RWPE-*SChLAP1* isoform 1, and RWPE-*SChLAP1* isoform 2 cells. A ±1kb interval surrounding the called

SNF5 peak is shown. **(c)** Example ChIP-Seq binding sites for SNF5 on gene promoters. SNF5 binding is higher at gene promoters in RWPE-LacZ cells and decreased upon SChLAP1 overexpression. **(d)** ChIP for SNF5 in 22Rv1 shNT and 22Rv1 sh-SChLAP1 #2. ChIP-PCR for 9 of 12 target genes of SNF5 in RWPE demonstrates an increase in SNF5 binding upon SChLAP1 knockdown. KIAA0841, Chr6 Alu, and Chr 2 Alu serve as negative controls. Data are represented as percent change in genomic binding relative to shNT after being normalized to IgG controls. The inset western blot indicates immunoprecipitation efficiency for SNF5. **(e)** Gene set enrichment analysis results showing significant enrichment of ChIP-Seq promoter peaks with >2-fold loss of SNF5 binding for underexpressed genes in RWPE-*SChLAP1* cells.

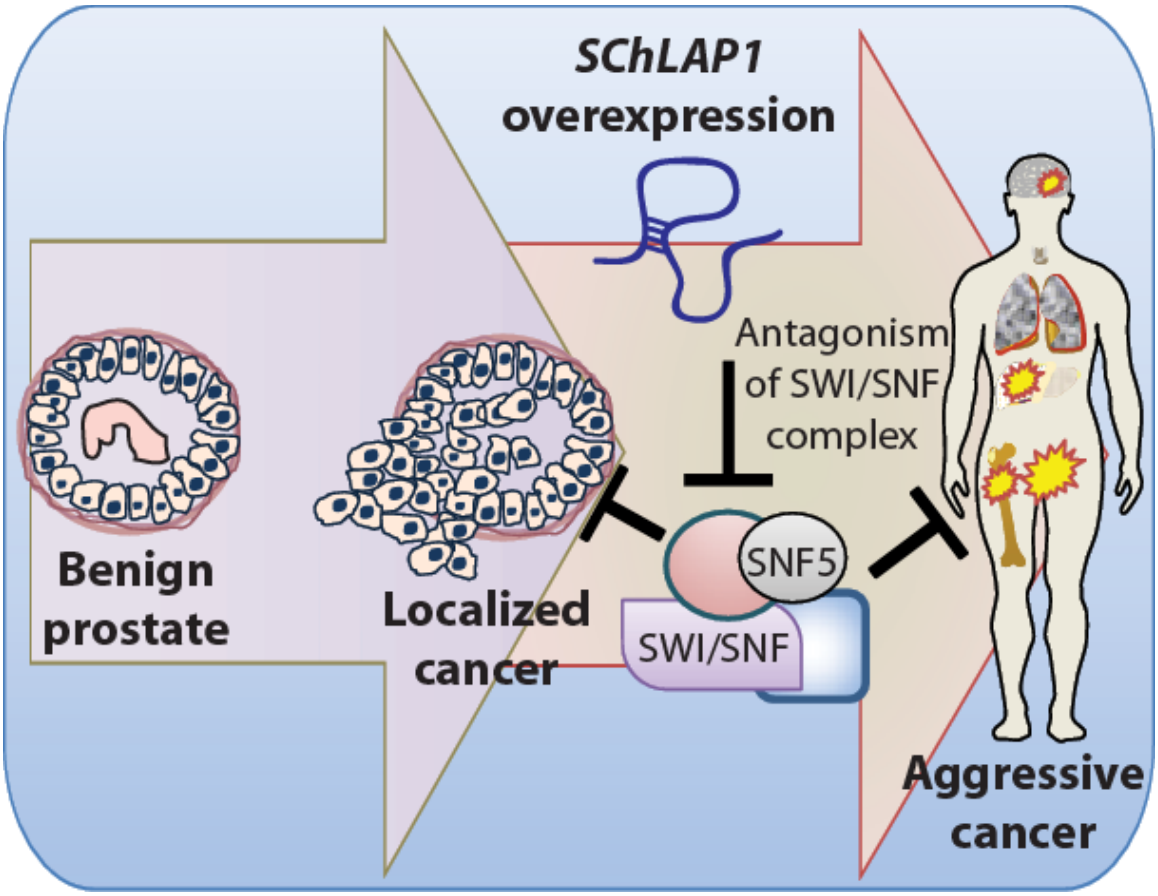


Figure 2.16 A schematic of *SChLAP1* function in prostate cancer.

References

1. Jemal, A., Siegel, R., Xu, J. & Ward, E. Cancer statistics, 2010. *CA Cancer J Clin* **60**, 277-300 (2010).
2. Etzioni, R., Cha, R., Feuer, E.J. & Davidov, O. Asymptomatic incidence and duration of prostate cancer. *American journal of epidemiology* **148**, 775-785 (1998).
3. Cooperberg, M.R., Moul, J.W. & Carroll, P.R. The changing face of prostate cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **23**, 8146-8151 (2005).
4. Grasso, C.S.*, et al.* The mutational landscape of lethal castration-resistant prostate cancer. *Nature* (2012).
5. Prensner, J.R., Rubin, M.A., Wei, J.T. & Chinnaiyan, A.M. Beyond PSA: the next generation of prostate cancer biomarkers. *Science translational medicine* **4**, 127rv123 (2012).
6. Taylor, B.S.*, et al.* Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11-22 (2010).
7. Berger, M.F.*, et al.* The genomic complexity of primary human prostate cancer. *Nature* **470**, 214-220 (2011).
8. Prensner, J.R. & Chinnaiyan, A.M. The emergence of lncRNAs in cancer biology. *Cancer Discov* **1**, 391-407 (2011).
9. Guttman, M.*, et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223-227 (2009).
10. Guttman, M.*, et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* (2011).
11. Lee, J.T. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev* **23**, 1831-1842 (2009).
12. Kotake, Y.*, et al.* Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* **30**, 1956-1962 (2011).
13. Yap, K.L.*, et al.* Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell* **38**, 662-674 (2010).
14. Gupta, R.A.*, et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071-1076 (2010).
15. Tsai, M.C.*, et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689-693 (2010).
16. Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J. & Lee, J.T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750-756 (2008).
17. Rinn, J.L.*, et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311-1323 (2007).

18. Prensner, J.R.*, et al.* Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol* **29**, 742-749 (2011).
19. Yu, J.*, et al.* An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer cell* **17**, 443-454 (2010).
20. Rhodes, D.R.*, et al.* Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* **9**, 166-180 (2007).
21. Varambally, S.*, et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**, 624-629 (2002).
22. Glinsky, G.V., Glinskii, A.B., Stephenson, A.J., Hoffman, R.M. & Gerald, W.L. Gene expression profiling predicts clinical outcome of prostate cancer. *J Clin Invest* **113**, 913-923 (2004).
23. Setlur, S.R.*, et al.* Estrogen-dependent signaling in a molecularly distinct subclass of aggressive prostate cancer. *J Natl Cancer Inst* **100**, 815-825 (2008).
24. Nakagawa, T.*, et al.* A tissue biomarker panel predicting systemic progression after PSA recurrence post-definitive prostate cancer therapy. *PLoS One* **3**, e2318 (2008).
25. Asangani, I.A.*, et al.* Characterization of the EZH2-MMSET histone methyltransferase regulatory axis in cancer. *Mol Cell* **49**, 80-93 (2013).
26. Tusher, V.G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 5116-5121 (2001).
27. Subramanian, A.*, et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545-15550 (2005).
28. Liberzon, A.*, et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740 (2011).
29. Shen, H.*, et al.* The SWI/SNF ATPase Brm is a gatekeeper of proliferative control in prostate cancer. *Cancer research* **68**, 10154-10162 (2008).
30. Roberts, C.W. & Orkin, S.H. The SWI/SNF complex--chromatin and cancer. *Nat Rev Cancer* **4**, 133-142 (2004).
31. Reisman, D., Glaros, S. & Thompson, E.A. The SWI/SNF complex and cancer. *Oncogene* **28**, 1653-1668 (2009).
32. Sun, A.*, et al.* Aberrant expression of SWI/SNF catalytic subunits BRG1/BRM is associated with tumor development and increased invasiveness in prostate cancers. *Prostate* **67**, 203-213 (2007).
33. Dechassa, M.L.*, et al.* Architecture of the SWI/SNF-nucleosome complex. *Molecular and cellular biology* **28**, 6010-6021 (2008).
34. Derrien, T.*, et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* **22**, 1775-1789 (2012).

35. De, S., *et al.* Dynamic BRG1 recruitment during T helper differentiation and activation reveals distal regulatory elements. *Mol Cell Biol* **31**, 1512-1527 (2011).
36. Euskirchen, G.M., *et al.* Diverse roles and interactions of the SWI/SNF chromatin remodeling complex revealed using global approaches. *PLoS Genet* **7**, e1002008 (2011).
37. Yen, K., Vinayachandran, V., Batta, K., Koerber, R.T. & Pugh, B.F. Genome-wide Nucleosome Specificity and Directionality of Chromatin Remodelers. *Cell* **149**, 1461-1473 (2012).
38. Wang, K.C., *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120-124 (2011).
39. Jones, S., *et al.* Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* **330**, 228-231 (2010).
40. Varela, I., *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* **469**, 539-542 (2011).
41. Versteege, I., *et al.* Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. *Nature* **394**, 203-206 (1998).
42. Rubin, M.A., *et al.* Rapid ("warm") autopsy study for procurement of metastatic prostate cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **6**, 1038-1045 (2000).
43. Tomlins, S.A., *et al.* Role of the TMPRSS2-ERG gene fusion in prostate cancer. *Neoplasia* **10**, 177-188 (2008).
44. Asangani, I.A., *et al.* Characterization of the EZH2-MMSET Histone Methyltransferase Regulatory Axis in Cancer. *Mol Cell* (2012).
45. Chu, C., Qu, K., Zhong, F.L., Artandi, S.E. & Chang, H.Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell* **44**, 667-678 (2011).
46. Maher, C.A., *et al.* Chimeric transcript discovery by paired-end transcriptome sequencing. *Proc Natl Acad Sci U S A* **106**, 12353-12358 (2009).
47. Levin, J.Z., *et al.* Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat Methods* **7**, 709-715 (2010).
48. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
49. Zhang, Y., *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137 (2008).
50. Feng, J., Liu, T. & Zhang, Y. Using MACS to identify peaks from ChIP-Seq data. *Curr Protoc Bioinformatics* **Chapter 2**, Unit 2 14 (2011).
51. Kent, W.J., *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002).
52. Shin, H., Liu, T., Manrai, A.K. & Liu, X.S. CEAS: cis-regulatory element annotation system. *Bioinformatics* **25**, 2605-2606 (2009).
53. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106 (2010).

54. Kannan, K.*, et al.* Recurrent chimeric RNAs enriched in human prostate cancer identified by deep sequencing. *Proc Natl Acad Sci U S A* **108**, 9172-9177 (2011).
55. Pflueger, D.*, et al.* Discovery of non-ETS gene fusions in human prostate cancer using next-generation RNA sequencing. *Genome Res* **21**, 56-67 (2011).
56. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111 (2009).
57. Hulsen, T., de Vlieg, J. & Alkema, W. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* **9**, 488 (2008).
58. Buerki, C.*, et al.* Validation of a genomic-clinical classifier for predicting clinical progression in high-risk prostate cancer. *ASCO Annual Meeting, Abstract #4565* (2012).
59. Blute, M.L., Bergstralh, E.J., Iocca, A., Scherer, B. & Zincke, H. Use of Gleason score, prostate specific antigen, seminal vesicle and margin status to predict biochemical failure after radical prostatectomy. *J Urol* **165**, 119-125 (2001).
60. Vergara, I.A.*, et al.* Genomic "Dark Matter" in Prostate Cancer: Exploring the Clinical Utility of ncRNA as Biomarkers. *Front Genet* **3**, 23 (2012).

Chapter 3:

The long noncoding RNA *SChLAP1* enhances PRC2 activity and sensitizes cells to pharmacologic EZH2 inhibition[2]

Abstract

A major challenge in the clinical management of prostate cancer is distinguishing aggressive, lethal tumors from indolent disease[1,2]. Recently, we identified a long noncoding RNA (lncRNA) termed *SChLAP1* that is expressed highly in 15-30% of prostate cancers and significantly associated with metastatic and lethal disease[3-6]. *SChLAP1* enhances cell invasiveness in part by interacting with and abrogating genome-wide binding of the tumor-suppressive SWI/SNF nucleosome-remodeling complex[7]. Approximately 20% of all cancers harbor a mutation in the SWI/SNF complex[8-11], and several recent studies have identified various therapeutic opportunities arising from SWI/SNF inactivation[12-15]. In particular, an antagonistic relationship between the SWI/SNF complex and Polycomb complexes revealed pharmacologic EZH2 inhibition as a promising strategy to target SWI/SNF-mutated cancers[16-18]. Here, we show that *SChLAP1* enhances PRC2 activity and genome-wide binding in prostate cells. Additionally,

[2] This chapter has been prepared as a manuscript for submission with the following authors: Anirban Sahu, Matthew K. Iyer, John R. Prensner, Benjamin Chandler, Xuhong Cao, Saravana M. Dhanasekaran, Yashar S. Niknafs, Nithin Edara, Udit Singhal, Shuang G. Zhao, Yi-Mi Wu, Dan R. Robinson, Rohit Malik, Felix Y. Feng, and Arul M. Chinnaiyan.

SChLAP1-expressing prostate cancer cells are more sensitive to EPZ-6438, a highly specific small molecule inhibitor of EZH2 currently in clinical trials[19]. Furthermore, EPZ-6438 decreased cell growth in *SChLAP1*-overexpressing, but not control, prostate epithelial cells. Our findings indicate that lncRNA-mediated SWI/SNF inhibition may function similarly to SWI/SNF mutation, exposing similar therapeutic opportunities. Moreover, *SChLAP1* expression may identify prostate cancer patients that are more likely to respond to pharmacologic EZH2 inhibition.

Introduction

Prostate cancer is the second-most deadly type of cancer in U.S. men, accounting for nearly 30,000 deaths annually[20]. However, most men diagnosed with prostate cancer have indolent disease, and distinguishing this subset of patients from those with more aggressive, lethal cancer remains a challenge[1,2].

Recently, our group identified a prognostic, prostate cancer-specific long non-coding RNA (lncRNA) termed *SChLAP1* that promotes tumor cell invasion and metastasis to mediate aggressive disease[3]. *SChLAP1* is significantly prognostic for metastatic progression and poor outcomes in prostate cancer[4-6]. Mechanistically, *SChLAP1* functions in part by interacting with the tumor-suppressive SWI/SNF complex to abrogate its genome-wide binding[7].

The multi-subunit SWI/SNF complex acts as a nucleosome-remodeler to influence chromatin architecture and regulate gene expression[9,21]. While

approximately 20% of all cancers harbor a deleterious mutation in the SWI/SNF complex[8-11], SWI/SNF mutations are not common in prostate cancer[22,23]. In fact, among all datasets available on the cBioPortal[24,25], mutations in any of five commonly-mutated SWI/SNF subunits (*SMARCB1* (also known as *SNF5*), *SMARCA2* (also known as *BRM*), *SMARCA4* (also known as *BRG1*), *ARID1A*, and *ARID1B*) are present in less than 5% of prostate cancers (**Fig. 3.1a**, gray bars, and **Table 3.1**). In contrast, *SChLAP1* expression, measured by RNA-seq and obtained from the MiTranscriptome portal[26], is highly specific to prostate cancer, with little to no expression in other types of malignant or benign tissues (**Fig. 3.1a**, red dots). Therefore, we hypothesized that *SChLAP1* expression may mimic SWI/SNF mutations in prostate cancer.

Results

To explore this possibility, we assessed SWI/SNF mutations and *SChLAP1* expression across a cohort of prostate adenocarcinomas obtained and sequenced through the Michigan Oncology Sequencing Project (MI-ONCOSEQ)[27]. SWI/SNF mutations were present in 12 of 86 samples (**Fig. 3.1b**, blue boxes). Using 40 FPKM (**Fig. 3.1b**, dashed gray line) as a cutoff for high (n=28) versus low (n=58) *SChLAP1* expression (**Fig. 3.1b**, barplot), we found that SWI/SNF mutations are significantly associated with low *SChLAP1* expression (p-value<0.05, Fisher's exact, **Fig. 3.1b**). Additionally, when we stratified samples by SWI/SNF mutation status, we found that *SChLAP1* expression was decreased in mutant SWI/SNF samples compared to wild-type

SWI/SNF samples (p-value<0.05, Mann-Whitney *U* test, **Fig. 3.1c**). Taken together, these results suggest that SWI/SNF mutations correlate with low *SChLAP1* expression while high *SChLAP1* expression may represent a mutation-independent, but clinically equivalent, modality of SWI/SNF inhibition.

To investigate this hypothesis, we utilized a previously described model of epigenetic antagonism between the SWI/SNF complex and Polycomb complexes[28-30]. Prior studies have shown enhanced PRC2 (Polycomb Repressive Complex 2) histone methyltransferase activity in cancers with SWI/SNF inactivation[16]. This is particularly intriguing given the well-characterized role of PRC2 in prostate cancer progression[31]. Moreover, using the top 5% of genes negatively correlated to *SChLAP1* in primary prostate tumors, we found that 6 of the top 10 molecular concepts identified by Oncomine Concepts Analysis[32] were associated with PRC2 (odds ratio>4.0 and p-value<$5x10^{-12}$, **Fig. 3.2a**), providing further evidence of a relationship between *SChLAP1* and PRC2.

To determine whether this antagonistic SWI/SNF – PRC2 axis exists in prostate cancer, we performed siRNA knockdown of *SMARCB1*, a core subunit of the SWI/SNF complex that facilitates binding to histone proteins[33], and confirmed a predicted increase in H3K27 trimethylation (H3K27me3) (**Fig. 3.2b**, left), the epigenetic signature of PRC2 methyltransferase activity[34,35]. There was also a subtle increase in H3K9 trimethylation (H3K9me3) upon siRNA knockdown of *SMARCB1*, which may have been due to loading variation or could possibly

suggest that SWI/SNF also regulates other epigenetic complexes. Next, we found that overexpression of two *SChLAP1* isoforms in RWPE benign prostate epithelial cells increased levels of H3K27me3 (**Fig. 3.2b**, right), consistent with an antagonistic effect of *SChLAP1* on SWI/SNF. Furthermore, shRNA knockdown of *SChLAP1* in 22Rv1 prostate cancer cells decreased levels of H3K27me3 (data not shown).

To directly test whether *SChLAP1* enhances PRC2-mediated gene regulation, we performed siRNA knockdown of *EZH2* (**Fig. 3.3a**), the catalytic subunit of PRC2[34], and ranked genes according to differential expression[36]. A comparison of genes regulated by knockdown of *SChLAP1* to genes regulated by *EZH2* demonstrated a cooperative relationship where *EZH2* knockdown affected the same genes as *SChLAP1* and in the same direction (**Fig. 3.3b**). This was in contrast to *SMARCB1* knockdown which demonstrated an antagonistic relationship by regulating the same genes but in the opposite direction (**Fig. 3.3b**). We used Gene Set Enrichment Analysis (GSEA)[37] to quantify and verify the significance of these findings (FDR<0.0001, **Fig 3.3c**). Genes down-regulated by *SChLAP1* knockdown were enriched in genes down-regulated by *EZH2* knockdown and up-regulated by *SMARCB1* knockdown. By contrast, genes up-regulated by *SChLAP1* knockdown were enriched in genes up-regulated by *EZH2* knockdown and down-regulated by *SMARCB1* knockdown.

To address whether *SChLAP1* modulated PRC2 genomic binding, we performed ChIP-seq of SUZ12, a core component of PRC2[34], in RWPE-*LacZ* and RWPE-*SChLAP1* cells and called significantly enriched peaks with respect to an IgG control. A SUZ12 antibody was used for technical reasons, as it had previously been verified for use in ChIP experiments. Western blot validations confirmed SUZ12 pull-down by ChIP (**Fig. 3.4a**). After aggregating called peaks from all samples, we found 30,251 genome-wide binding sites for SUZ12 (FDR < 0.05) which were highly enriched for sites near gene promoters (data not shown).

A comparison of SUZ12 binding across these 30,251 genomic sites demonstrated a dramatic increase in SUZ12 genomic binding as a result of *SChLAP1* overexpression (**Fig. 3.4b,c**). Furthermore, of the 245 most differentially bound SUZ12 peaks occurring within 1kb of a gene promoter, 233 peaks showed increased SUZ12 binding as a result of *SChLAP1* overexpression (**3.4d,e**). Taken together, these results establish an antagonistic role of SWI/SNF on PRC2 in prostate cells, and suggest that *SChLAP1* enhances PRC2 function in part by inhibiting SWI/SNF activity.

EZH2 is overexpressed and functions as an oncogenic driver in several types of cancers[31,38-40], making it an attractive therapeutic target to treat malignancy. Many small molecule inhibitors have been developed to specifically target EZH2 and inhibit its methyltransferase activity[17,41]. However, these inhibitors have been largely ineffective against cells with EZH2 overexpression, but rather have shown

enhanced effectiveness in cells with activating SET domain mutations[19,41,42]. While this is surprising, it suggests that only a certain subset of EZH2-overexpressing tumors become dependent on EZH2 as a driver of the cancer, and therefore respond to EZH2 inhibitors, while EZH2 overexpression functions as a passenger aberration in other cancers. More recently, two studies have shown that EZH2 inhibitors are also effective in SWI/SNF-mutated cancers[17,18]. Therefore, we hypothesized that EZH2 inhibition may be an effective therapeutic strategy in prostate cancer cells with high *SChLAP1* expression.

To test this hypothesis, we treated a panel of prostate cell lines with EPZ-6438, a highly specific small molecule inhibitor of EZH2 currently in clinical trials for patients with advanced solid tumors or with relapsed or refractory B-cell lymphoma[17,19]. After confirming that EPZ-6438 targets EZH2 to decrease H3K27me3 (**Fig. 3.5a**), we performed a cell viability assay (**Fig. 3.5b**) and found that cell lines with high *SChLAP1* expression were more sensitive to EPZ-6438 compared to cell lines with little to no *SChLAP1* expression (**Fig. 3.5c-e**). Importantly, overexpression of *SChLAP1* in RWPE cells increased sensitivity to EPZ-6438 (**Fig. 3.5d**). These data suggest that *SChLAP1*-expressing cell lines are more sensitive to EPZ-6438 (**Fig. 3.5e**). Furthermore, we found that treatment of RWPE-*SChLAP1* cell lines with EPZ-6438 induced strong anti-proliferative effects, whereas control RWPE-*LacZ* cells were minimally affected (**Fig. 3.6a-c**). Overall, these data argue that *SChLAP1* overexpression

antagonizes SWI/SNF complex function, resulting in enhanced PRC2 activity and increased sensitivity to pharmacologic EZH2 inhibition (**Fig. 3.7**).

Discussion

In summary, we have shown that SWI/SNF mutations are associated with low *SChLAP1* expression in prostate cancer, suggesting that high *SChLAP1* expression may represent a mutation-independent, but clinically equivalent, modality of SWI/SNF inhibition. Additionally, we establish an antagonistic relationship between SWI/SNF and PRC2 in prostate cells, and find that *SChLAP1* enhances PRC2 function and genome-wide binding, which may be partly due to SWI/SNF inhibition. Finally, *SChLAP1* expression is one factor that affects EPZ-6438 sensitivity and may be useful in identifying prostate cancer patients that are more likely to respond to pharmacologic EZH2 inhibition. Thus, while mutations in the SWI/SNF complex have been used to reveal specific vulnerabilities in cancer, our work suggests that mechanisms other than mutations, including lncRNA-mediated repression, can inhibit the SWI/SNF complex to expose similar therapeutic opportunities in malignancy. Taken together, our findings have broad implications for cancer biology and provide a rationale for the use of EZH2 inhibitors for therapeutic treatment of prostate cancer.

Materials and Methods

cBioPortal SWI/SNF mutations

Mutation frequency for *SMARCB1*, *SMARCA2*, *SMARCA4*, *ARID1A*, and *ARID1B* in various cancer cohorts were obtained from the cBioPortal (**Table 3.1**)[24,25]. Only those cohorts for which we had corresponding RNA-seq data to assess *SChLAP1* expression were included in this study.

MiTranscriptome *SChLAP1* expression

Expression levels for *SChLAP1* were measured by RNA-seq and obtained from the MiTranscriptome web portal (www.mitranscriptome.org)[26].

MI-ONCOSEQ *SChLAP1* expression and SWI/SNF mutations

SChLAP1 expression and SWI/SNF mutations (*SMARCB1*, *SMARCA2*, *SMARCA4*, *ARID1A*, and *ARID1B*) were assessed in a cohort of prostate adenocarcinomas (86 tissue samples: 5 primary tumors, 79 metastatic samples, 2 unknown) obtained and sequenced through the Michigan Oncology Sequencing Project (MI-ONCOSEQ)[27]. Gene expression was measured by Cufflinks from poly(A)+ transcriptome RNA libraries. Somatic mutations were called by Varscan from exome capture DNA libraries. Because previous studies have found high *SChLAP1* expression in approximately one-third of all prostate cancers[4], we used 40 FPKM as a cutoff for high (n=28) versus low (n=58) *SChLAP1* expression in this cohort. For statistical analysis of SWI/SNF mutation occurrence in *SChLAP1* high and low samples (**Fig. 3.1b**), a one-sided Fisher's

exact test was employed with a null hypothesis that SWI/SNF mutations occur equally in high and low *SChLAP1* samples and an alternative hypothesis that SWI/SNF mutations are enriched in low *SChLAP1* samples. For statistical analysis of *SChLAP1* expression in SWI/SNF wildtype and mutant samples (**Fig. 3.1c**), a Mann-Whitney *U* test was used.

Oncomine Concepts Analysis of *SChLAP1* Signature

A *SChLAP1* gene correlation signature was determined as previously described[3]. We imported the top 5% of genes negatively correlated with *SChLAP1* into Oncomine as a custom concept. We then nominated significantly associated molecular concepts with Odds Ratio > 4.0 and p-value < $5x10^{-12}$ (**Fig. 3.2a**).

Cell lines

All cell lines were obtained from the American Type Culture Collection. Cell lines were maintained using standard media and conditions. Specifically, LNCaP, 22Rv1, C4-2B, and PC3 cells were maintained in RPMI 1640 plus 10% FBS and 1% penicillin-streptomycin. RWPE cells were maintained in KSF media plus 10ng/mL EGF and bovine pituitary extract (BPE) and 1% penicillin-streptomycin. All cell lines were grown at 37C in a 5% CO2 cell culture incubator. All cell lines were genotyped for identity at the University of Michigan Sequencing Core and tested routinely for Mycoplasma contamination.

SChLAP1 or *LacZ* control cell lines were generated as previously described[3]. Briefly, *SChLAP1* full-length transcript was amplified from LNCaP cells and

cloned into the pLenti6 vector along with *LacZ* controls. The benign immortalized prostate cell line RWPE was infected with lentiviruses expressing *SChLAP1* or *LacZ*. Stably-transfected RWPE cells were selected using blasticidin for one week. All lentiviruses were generated by the University of Michigan Vector Core.

siRNA knockdown

Cells were plated in 100mM plates at a desired concentration and transfected with 20nM experimental siRNA oligos or non-targeting controls twice, at 12 hours and 24 hours post-plating. Knockdowns were performed with Lipofectamine RNAiMAX in OPTI-MEM I media. 72 hours post-transfection, cells were harvested. Knockdown efficiency was determined by qPCR. siRNAs used in this study are listed below:

SMARCB1 siRNA #1; ON-TARGETplus SMARCB1; J-010536-05
SMARCB1 siRNA #2; ON-TARGETplus SMARCB1; J-010536-07
EZH2 siRNA #1; Custom ON-TARGET plus; 5′-GAGGUUCAGACGAGCUGAU-3′
EZH2 siRNA #2; Custom ON-TARGET plus; 5′-AGACUCUGAAUGCAGUUGC-3′

Immunoblotting

Cells were lysed in RIPA lysis buffer supplemented with HALT protease inhibitor. Protein lysates were boiled in sample buffer, and 5-10µg protein was loaded onto a SDS-PAGE gel and run for separation of proteins. Proteins were transferred onto Polyvinyl Difluoride membranes and blocked for 30 minutes in blocking buffer (5% milk, 0.1% Tween, Tri-buffered saline (TBS-T)). Membranes were incubated overnight at 4C with primary antibody. After 3 washes with TBS-T, membranes were incubated for 30 minutes at room temperature with horseradish

peroxidase-conjugated secondary antibody. Following 3 washes with TBS-T and one wash with TBS, the signals were visualized by enhanced chemiluminescence system from GE Healthcare. Primary antibodies used in this study are listed below:

SMARCB1; Millipore ABD22; 1:1,000
H3K27me3; Cell Signaling 9733S; 1:1,000
H3K9me3; Abcam ab8898; 1:1,000
Total H3; Abcam ab1791; 1:10,000
B-Actin; Sigma A5316; 1:10,000
SUZ12; Abcam ab12073; 1:1,000

RNA isolation and cDNA synthesis

Total RNA was isolated using Trizol and a miRNeasy Kit with DNase I digestion according to the manufacturer's instructions. RNA integrity was verified on an Agilent Bioanalyzer 2100. cDNA was synthesized from total RNA using Superscript III and random primers.

Quantitative Real-time PCR

Quantitative Real-time PCR (qPCR) was performed using Power SYBR Green Mastermix on an Applied Biosystems 7900HT Real-Time PCR System. All oligonucleotide primers were obtained from Integrated DNA Technologies and are listed below:

SChLAP1 Sense: TGGACACAATTTCAAGTCCTCA
SChLAP1 Antisense: CATGGTGAAAGTGCCTTATACA
EZH2 Sense: TGCAGTTGCTTCAGTACCCATAAT
EZH2 Antisense: ATCCCCGTGTACTTTCCCATCATAAT
ACTB Sense: AAGGCCAACCGCGAGAAG
ACTB Antisense: ACAGCCTGGATAGCAACGTACA
GAPDH Sense: TGCACCACCAACTGCTTAGC
GAPDH Antisense: GGCATGGACTGTGGTCATGAG

Microarray Gene Expression Analysis

Expression profiling was performed using the Agilent Whole Human Genome Oligo Microarray, according to previously published protocols[43]. All samples were run in technical replicates and gene expression was compared between siRNA treatment and non-targeting control. Expression data was analyzed using the SAM method as described previously[36].

Microarray experiments

We performed two-color microarray gene expression profiling of LNCaP and 22Rv1 cells treated with two independent siRNAs targeting *EZH2* as well as control non-targeting siRNAs. These profiling experiments were run in technical duplicate for a total of 8 arrays (4 from LNCaP and 4 from 22Rv1). We also analyzed microarray expression data from *SChLAP1* knockdown in LNCaP and 22Rv1 cells (GSE40383), and *SMARCB1* knockdown in LNCaP and 22Rv1 cells (GSE40384).

Microarray data processing to determine ranked gene expression lists

All of the microarray data were represented as base-2 log fold-change between targeting versus control siRNAs. Genes measured by multiple probes were consolidated using the median of probes. We then ran one-class SAM analysis from the R package ‘samr’ and ranked all genes by the difference between observed versus expected statistics. Additionally, we nominated the top 5% of

changed genes across *SChLAP1* knockdown experiments and quantified the enrichment for *EZH2* and *SMARCB1* target genes using Gene Set Enrichment Analysis (GSEA)[37].

Chromatin Immunoprecipitation

ChIP assays were performed as previously described[3,44,45] using antibodies for SUZ12 (Abcam ab12073) and Rabbit IgG (Millipore PP64B). Briefly, approximately 10^6 cells per antibody were cross-linked for 10-15 minutes with 1% formaldehyde and the cross-linking was inactivated by 0.125M glycine for 5 minutes at room temperature. Cells were rinsed with ice-cold PBS three times and cell pellets were resuspended in lysis buffer plus protease inhibitors. Chromatin was sonicated to an average length of 500bp, centrifuged to remove debris, and supernatants containing chromatin fragments were incubated with protein A/G beads to reduce non-specific binding. Then, beads were removed and supernatants were incubated with 6µg of antibody overnight at 4C. Fresh protein A/G beads were added and incubated with protein-chromatin-antibody complexes for 2 hours at 4C, washed twice with 1X dialysis buffer and four times with IP wash buffer, and eluted in 150µl IP elution buffer. 1:10th of the ChIP reaction was taken for protein evaluation for validation of ChIP pull-down. Reverse crosslinking was performed by incubating the eluted product with 0.3M NaCl at 65C overnight. ChIP product was cleaned up with the USB PrepEase kit. ChIP experiments were validated for specificity by Western blotting.

ChIP-seq experiments

Paired-end ChIP-Seq libraries were generated following the Illumina ChIP-Seq protocol with minor modifications. The ChIP DNA was subjected to end-repair and A base addition before ligating with Illumina adaptors. Samples were purified using Ampure beads and PCR-enriched with a combination of specific index primers and PE2.0 primer under the following conditions: 98C (30 sec), 65C (30 sec), and 72C (40 sec with a 4 sec increment per cycle). After 14 cycles of amplification a final extension at 72ºC for 5 minutes was carried out. The barcoded libraries were size-selected using a 3% NuSieve Agarose gel and subjected to an additional PCR enrichment step. The libraries were analyzed and quantified using Agilent Bioanalyzer 2100 before subjecting to paired-end sequencing using the Illumina Hi-Seq platform.

ChIP-Seq data analysis

Sequencing data from RWPE SUZ12 ChIP-Seq samples were mapped to human genome version hg19 using BWA 0.5.943. Although we performed paired-end sequencing, the ChIP-Seq reads were processed as single-end to adhere to our preexisting analysis protocol. Peak calling was performed respect to an IgG control using the MACS algorithm[46]. For each sample we ran the CEAS program and generated genome-wide reports[47]. We retained peaks with a false discovery rate (FDR) less than 5%. We then aggregated SUZ12 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 8,528 peaks occurred within one kilobase of transcription start sites (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 package version 1.6.3 to compute differential binding between RWPE-*LacZ* and RWPE-*SChLAP1* (both isoforms). We observed that 245 promoter peaks had a significant change in SUZ12 binding (adjusted p-value <0.1), with 233 peaks showing increased SUZ12 binding.

Cell Viability Assays

Cells were plated in a 12-well dish on day 0 and treated the following day with desired concentrations (20μM, 3.2μM, 512nM, 82nM, 13nM, and 2nM) of EPZ-6438 (Selleckchem) or vehicle control (DMSO). During days 2-11, cells were replated as necessary, maintaining cell ratios across various drug concentrations, and treated on the following day. On day 12, cells were plated into a 96-well plate (4 wells per treatment), maintaining cell ratios across various drug concentrations, and treated the following day. On day 14, WST-1 cell proliferation reagent (Roche) was added to each well. Following 2 hours of incubation at 37C, the absorbance of the wells was measured at 450nm. Data was normalized to vehicle control and background signal was removed prior to analysis. GraphPad Prism was used to perform nonlinear regression (log(inhibitor) vs. normalized response) and calculate the IC_{50}.

Cell Proliferation assays

Cells were plated in a 12-well dish on day 0 and treated the following day with either 8μM EPZ-6438 (Selleckchem) or vehicle control (DMSO). During days 2-5, cells were replated as necessary and treated the following day. On day 6, cells were plated into a 96-well plate (4 wells per treatment) at an equal density of 1000 cells per well for each treatment group. On day 7, the cells were treated with EPZ-6438 or vehicle control and the plates were placed in an Incucyte for live-cell imaging to obtain cell confluence data at 12-hour intervals.

Tables

Cancer Type	Altered Cases	Total Cases	Percent
AML	2	200	1.00
Bladder	69	327	21.10
Breast	45	775	5.81
Cervical	7	39	17.95
Colorectal	55	296	18.58
Globlastoma Multiforme	6	382	1.57
Head and Neck	45	385	11.69
Hepatocellular	19	258	7.36
Lower Grade Glioma	30	289	10.38
Lung Adenocarcinoma	84	576	14.58
Lung Squamous Cell	32	178	17.98
Medulloblastoma	15	254	5.91
Melanoma	41	212	19.34
Ovarian	14	316	4.43
Pancreatic	7	99	7.07
Prostate	17	432	3.94
Renal Chromophobe	2	66	3.03
Renal Clear Cell	33	522	6.32
Renal Papillary Cell	25	168	14.88
Stomach	64	220	29.09
Thyroid	6	401	1.50
Uterine Endometrial	100	248	40.32

Table 3.1 SWI/SNF mutation frequency in the cBioPortal
The number of cases in the cBioPortal with an alteration in any of five commonly-mutated SWI/SNF subunits (*SMARCB1, SMARCA2, SMARCA4, ARID1A*, and *ARID1B*) is shown. Only cancer types with corresponding RNA-seq data in the MiTranscriptome Portal are shown.

Figures

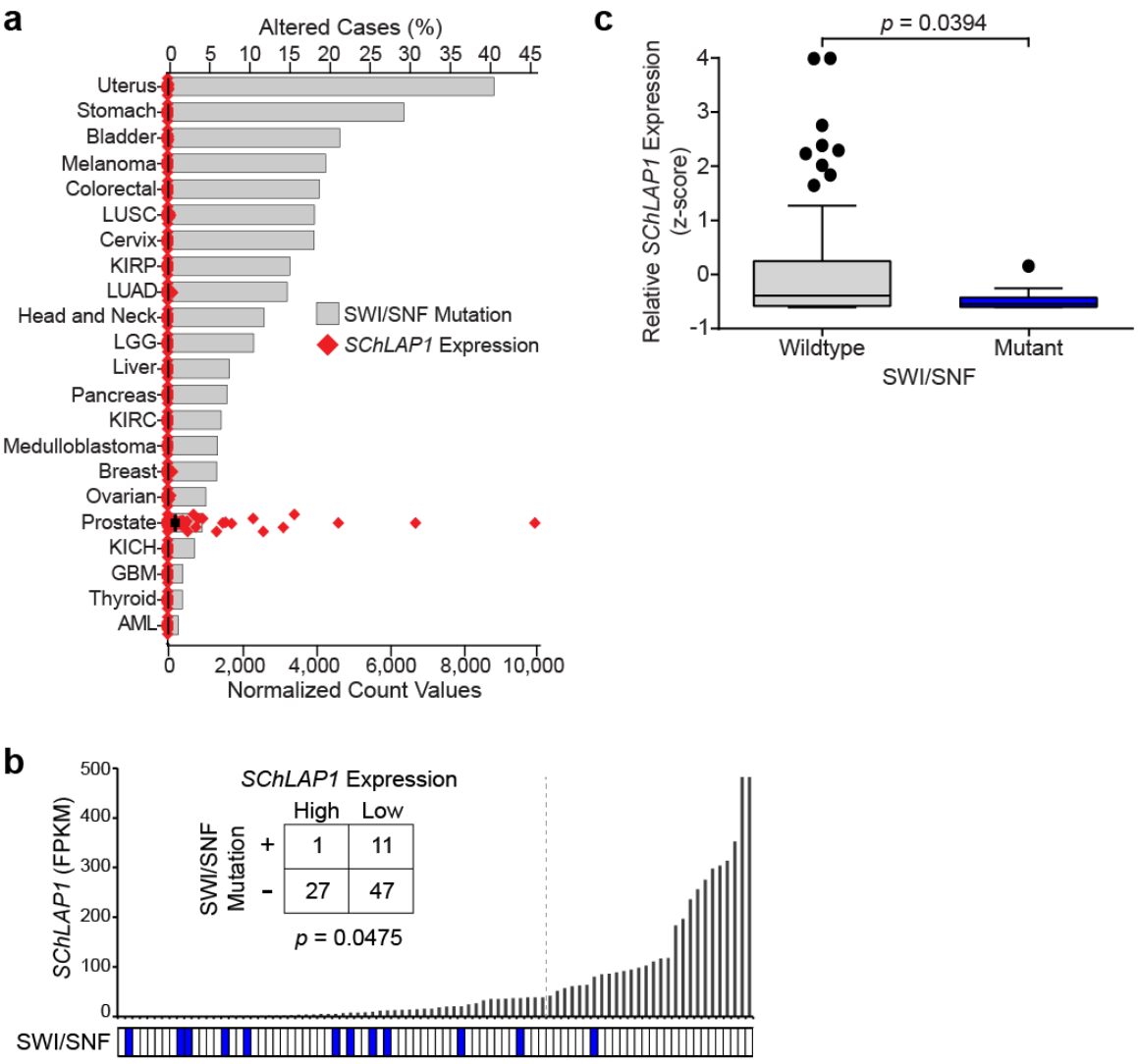


Figure 3.1 Integrated landscape of SWI/SNF mutations and *SChLAP1* expression in cancer.
(a) Gray bars indicate mutation frequency of SWI/SNF across 22 cancer types from the cBioPortal. Bars show the percent of cases with an alteration in at least one of five SWI/SNF subunits (SMARCB1, SMARCA2, SMARCA4, ARID1A or ARID1B). Red dots indicate *SChLAP1* expression across these cancer types from the MiTranscriptome portal. Dots show *SChLAP1* expression in each sample. The black bar represents mean expression within each cancer type. **(b)** Analysis of the MI-ONCOSEQ prostate adenocarcinoma cohort (86 tissue samples: 5 primary tumors, 79 metastatic samples, 2 unknown) for SWI/SNF mutation and *SChLAP1* expression. Blue boxes represent a mutation in one of five SWI/SNF subunits (*SMARCB1, SMARCA2, SMARCA4, ARID1A* or *ARID1B*). *SChLAP1* expression is shown in the barplot. Samples are ordered

according to *SChLAP1* expression. Co-occurence of SWI/SNF mutation and *SChLAP1* expression is shown in the inset. P-value was determined by a Fisher's exact test. **(c)** *SChLAP1* expression in samples wild-type (n=74) and mutant (n=12) for SWI/SNF mutation in the MI-ONCOSEQ prostate adenocarcinoma cohort. Expression is represented as the z-score. P-value was determined by a Mann-Whitney *U* test.

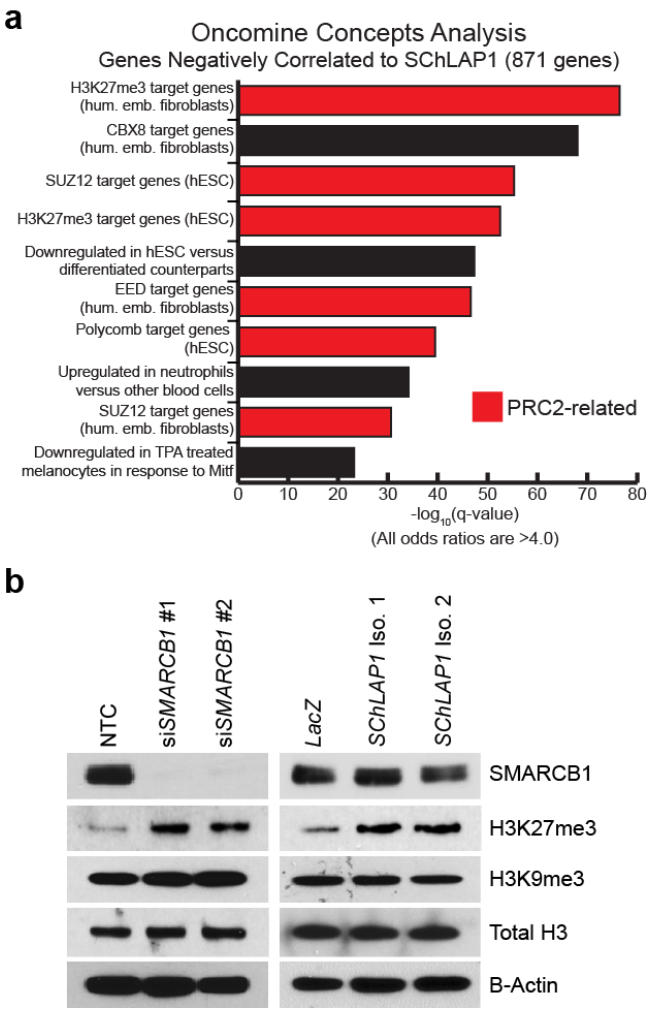


Figure 3.2 *SChLAP1* is associated with PRC2-related concepts and enhances PRC2 histone methyltransferase activity.
(a) Top ten molecular concepts nominated by Oncomine Concepts Analysis using the top 5% of genes negatively correlated to *SChLAP1* in primary prostate tumors. Red bars indicate PRC2-related concepts. **(b)** Knockdown of *SMARCB1* (left) and overexpression of *SChLAP1* (right) in RWPE cells followed by immunoblotting for SMARCB1, H3K27me3, H3K9me3, total H3, and B-Actin.

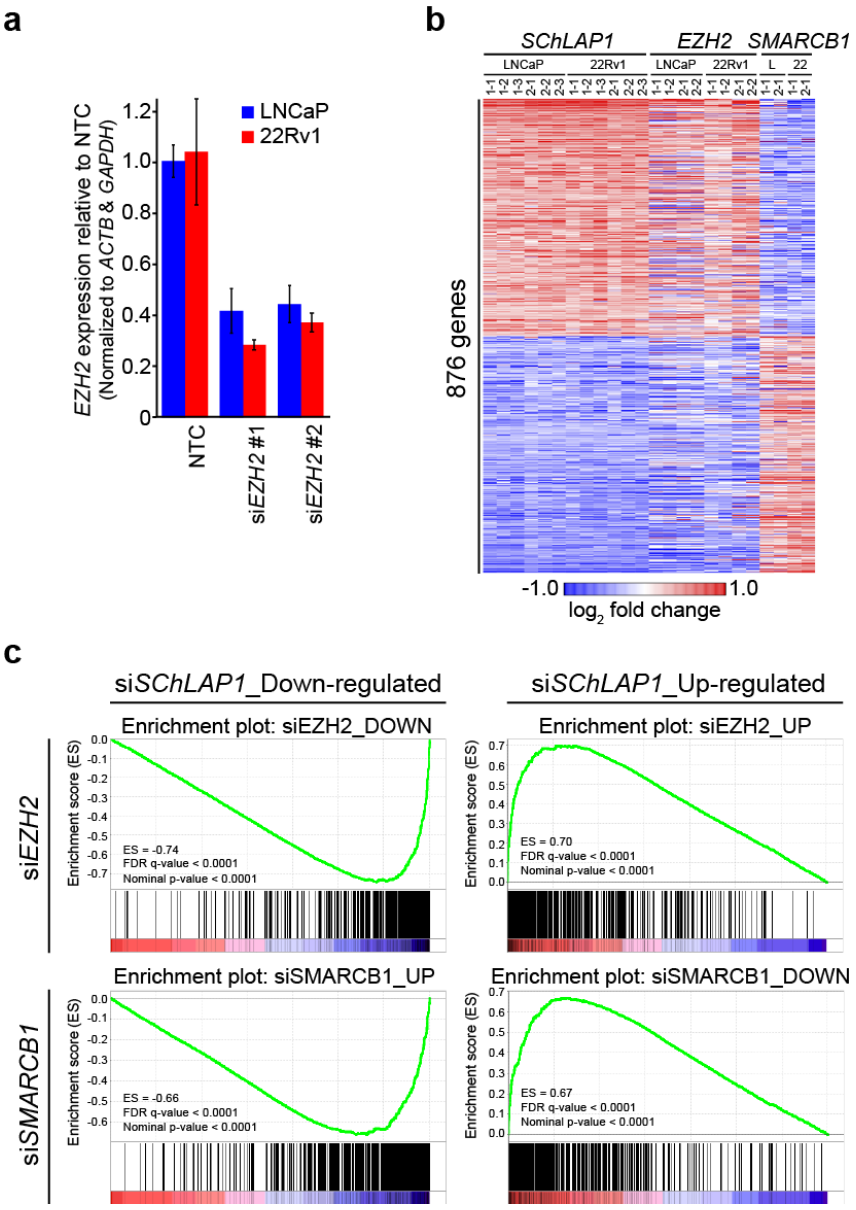


Figure 3.3 *SChLAP1* enhances PRC2-mediated gene regulation and antagonizes SWI/SNF-mediated gene regulation.
(a) LNCaP and 22Rv1 cells were treated with siRNAs against *EZH2*. qPCR indicates relative knockdown efficiency in these cell lines. Error bars represent S.E.M. **(b)** Heatmap results for *SChLAP1*, *EZH2* or *SMARCB1* knockdown in LNCaP and 22Rv1 cells. The numbers above the heatmap indicate the specific siRNA and microarray replicates. **(c)** GSEA results from comparisons of *SChLAP1*, *EZH2* and *SMARCB1* knockdown in LNCaP and 22Rv1 cells. *SChLAP1* was knocked-down using siRNAs in both cells lines. Gene expression changes (down- and up-regulated) across both cell lines were compared using GSEA to expression changes observed using *EZH2* and *SMARCB1* siRNAs. The enrichment plots of these comparisons are shown. ES, enrichment score.

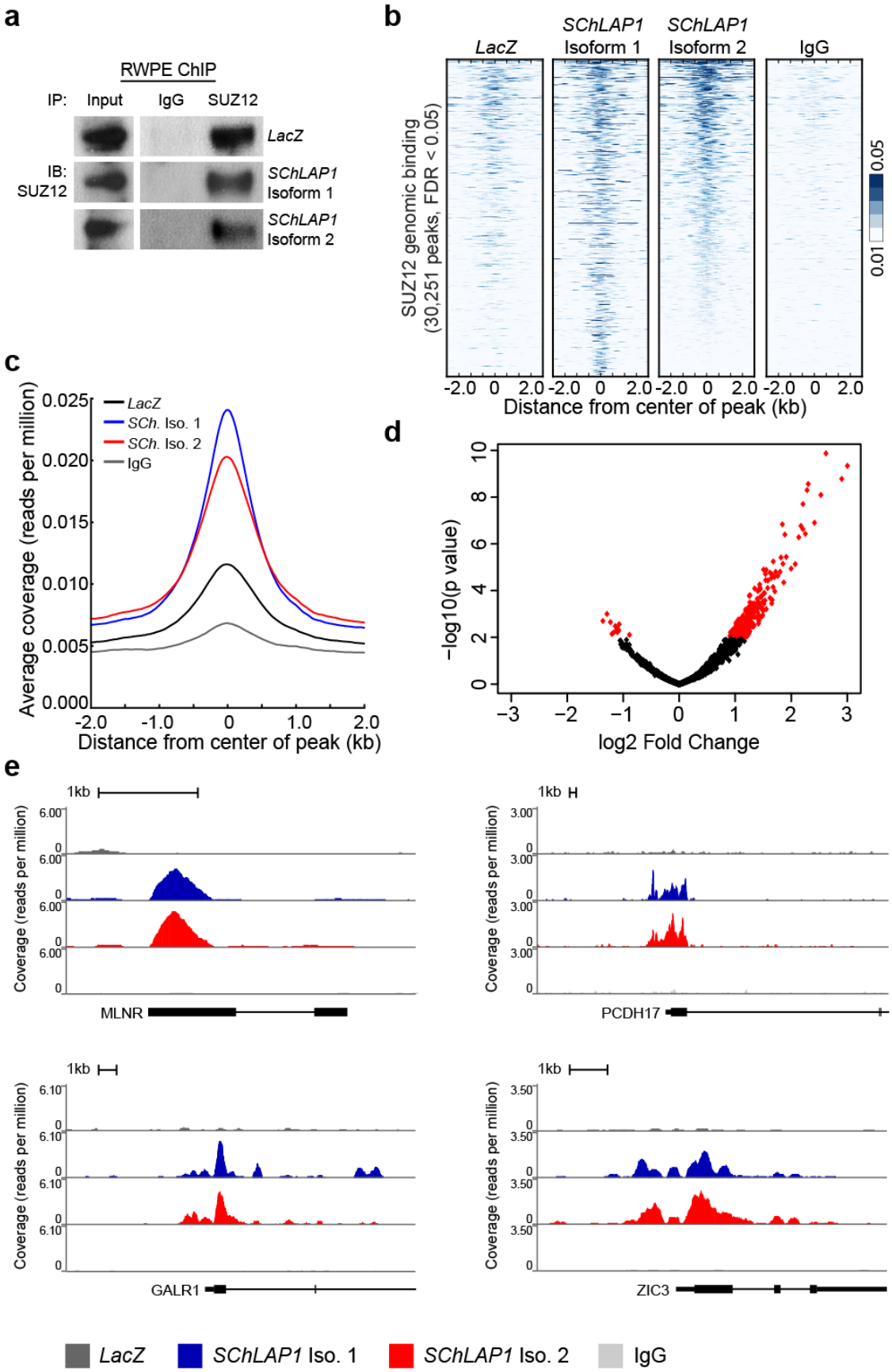
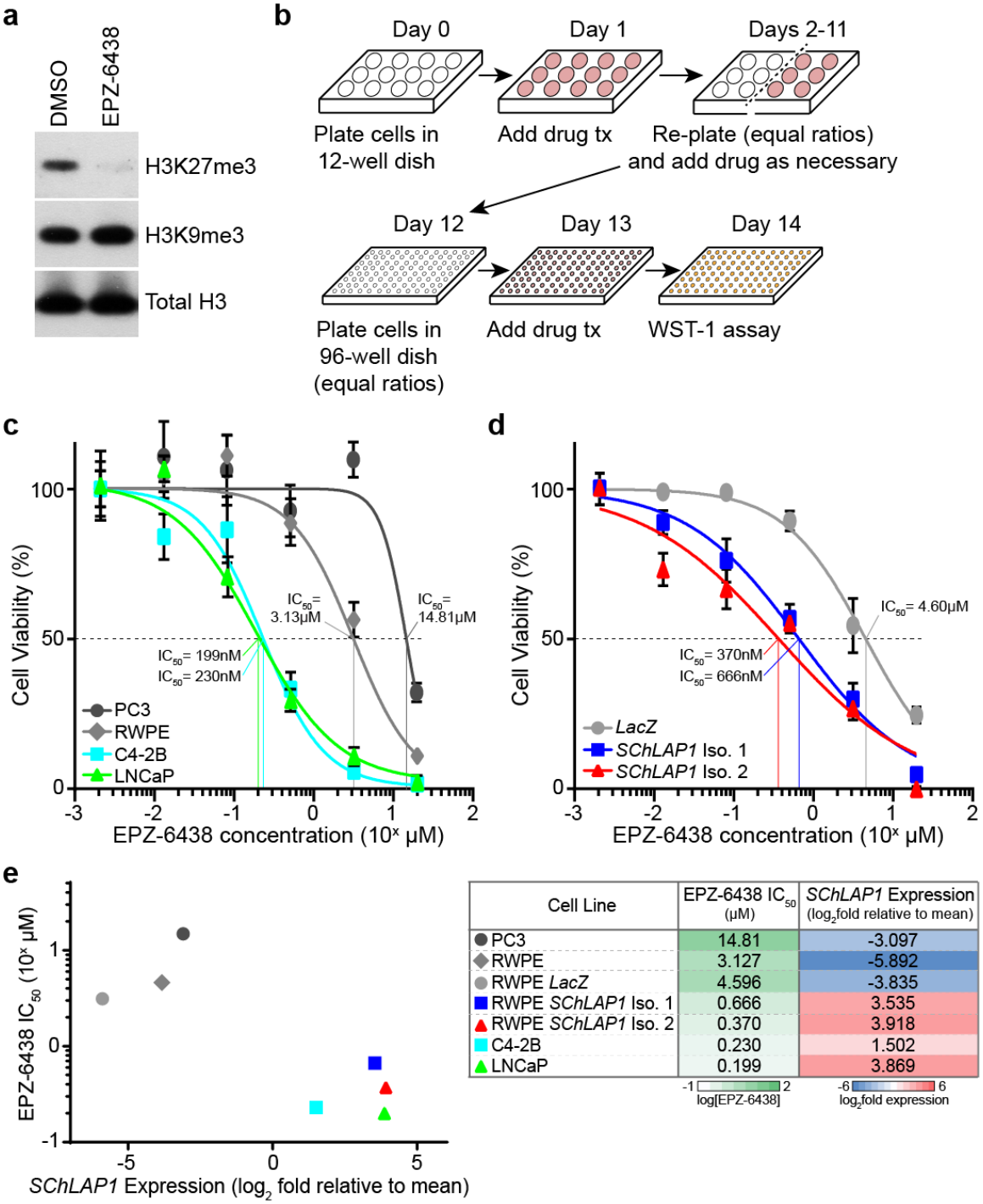

a
RWPE ChIP
IP: Input IgG SUZ12
IB: SUZ12
LacZ
SChLAP1 Isoform 1
SChLAP1 Isoform 2
b
LacZ
SChLAP1 Isoform 1
SChLAP1 Isoform 2
IgG
SUZ12 genomic binding (30,251 peaks, FDR < 0.05)
0.05
0.01
-2.0 0 2.0
Distance from center of peak (kb)
c
Average coverage (reads per million)
LacZ
SCh. Iso. 1
SCh. Iso. 2
IgG
0.025
0.020
0.015
0.010
0.005
0.000
-2.0 -1.0 0 1.0 2.0
Distance from center of peak (kb)
d
−log10(p value)
log2 Fold Change
e
1kb
Coverage (reads per million)
MLNR
PCDH17
GALR1
ZIC3
LacZ
SChLAP1 Iso. 1
SChLAP1 Iso. 2
IgG

Figure 3.4 *SChLAP1* enhances SUZ12 genome-wide binding.
(a) ChIP for SUZ12 protein followed by Western blot. **(b)** A heatmap representation of SUZ12 genomic binding at target sites in RWPE-*LacZ*, RWPE-*SChLAP1* isoform 1 and RWPE-*SChLAP1* isoform 2 cells. A ±2kb interval surrounding the called SUZ12 peak is shown. **(c)** A global representation of SUZ12 genomic binding over ±2kb window surrounding each SUZ12 ChIP-Seq peak in RWPE-*LacZ*, RWPE-*SChLAP1* isoform 1 and RWPE-*SChLAP1* isoform 2 cells. **(d)** Volcano plot showing the relative log2 fold-change in SUZ12 binding between RWPE-LacZ and RWPE-*SChLAP1* (average of both isoforms). Significant peaks are shown in red. **(e)** Example ChIP-Seq binding sites for SUZ12 on gene promoters. SUZ12 binding is lower at gene promoters in RWPE-LacZ cells and increased upon *SChLAP1* overexpression.



Cell Line	EPZ-6438 IC_{50} (µM)	*SChLAP1* Expression (log_2fold relative to mean)
PC3	14.81	-3.097
RWPE	3.127	-5.892
RWPE *LacZ*	4.596	-3.835
RWPE *SChLAP1* Iso. 1	0.666	3.535
RWPE *SChLAP1* Iso. 2	0.370	3.918
C4-2B	0.230	1.502
LNCaP	0.199	3.869

Figure 3.5 *SChLAP1* increases sensitivity to pharmacologic EZH2 inhibition.
(a) LNCaP cells were treated with 1µM EPZ-6438 or vehicle control for 4 days. Western blot for H3K27me3 shows effective drug targeting. **(b)** Experimental design for cell survival assays. Cells were plated at equal density in a 12-well dish and treated with various concentrations of EPZ-6438 or vehicle control. Cells were re-plated as necessary, maintaining cell ratios across conditions. After

12 days of treatment, cells were plated in a 96-well dish, maintaining cell ratios across conditions. On day 14, WST-1 reagent was added to the wells and O.D. at 450nm was measured after 2 hours of incubation at 37C. **(c)** Viability of PC3, RWPE, C4-2B and LNCaP prostate cells at different concentrations of EPZ-6438. Normalized data is represented as the percent of viable cells compared to vehicle control. Cells were treated with EPZ-6438 for 14 days prior to data collection. IC_{50} values of EPZ-6438 in PC3, RWPE, C4-2B and LNCaP cells are indicated on the graph. **(d)** Viability of RWPE-*LacZ*, RWPE-*SChLAP1* isoform 1 and RWPE-*SChLAP1* isoform 2 cells at different concentrations of EPZ-6438. Normalized data is represented as the percent of viable cells compared to vehicle control. Cells were treated with EPZ-6438 for 14 days prior to data collection. IC_{50} values of EPZ-6438 in RWPE-*LacZ*, RWPE-*SChLAP1* isoform 1 and RWPE-*SChLAP1* isoform 2 cells are indicated on the graph. **(e)** IC_{50} values and *SChLAP1* expression for several prostate cell lines. *SChLAP1* expression is represented as log2 fold change relative to the average *SChLAP1* expression in this cell line cohort. Table summarizing EPZ-6438 IC_{50} and *SChLAP1* expression across cell lines is shown to the right.

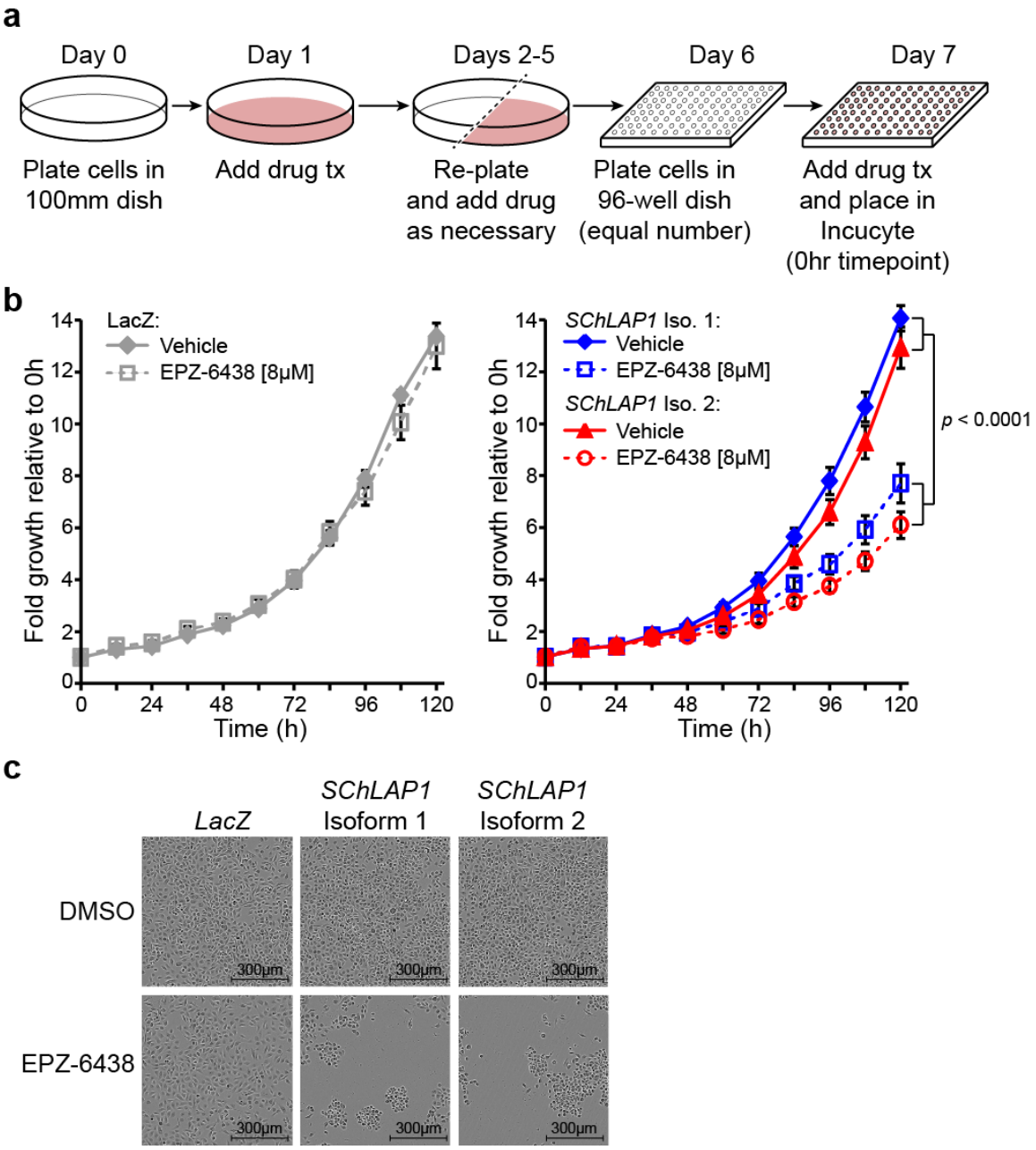


Figure 3.6 EPZ-6438 decreases proliferation in *SChLAP1*-expressing cells. **(a)** Experimental design for cell proliferation assays. Cells were plated at equal density in a 100mm dish and treated with 8μM EPZ-6438 or vehicle control. Cells were re-plated as necessary, maintaining equal densities across conditions. After 6 days of treatment, cells were plated in a 96-well dish at equal densities (1000 cells per well). The following day, drug was added and the plates were placed in an Incucyte machine. **(b)** Cell proliferation of RWPE-*LacZ* (left), RWPE-*SChLAP1* isoform 1 and RWPE-*SChLAP1* isoform 2 (right) cells treated with EPZ-6438 or vehicle control. Cells were pre-treated with EPZ-6438 for 7 days

then replated at equal densities at time 0h on the graph. Cell proliferation assays were performed by Incucyte live-cell imaging. Data shown are fold change in cell confluence vs. time at 12-hour intervals. Each data point is the mean of quadruplicates. P-value was determined by a Student's *t*-test. **(c)** Sample images from cell proliferation assays on day 12 of treatment (day 5 of proliferation assay).

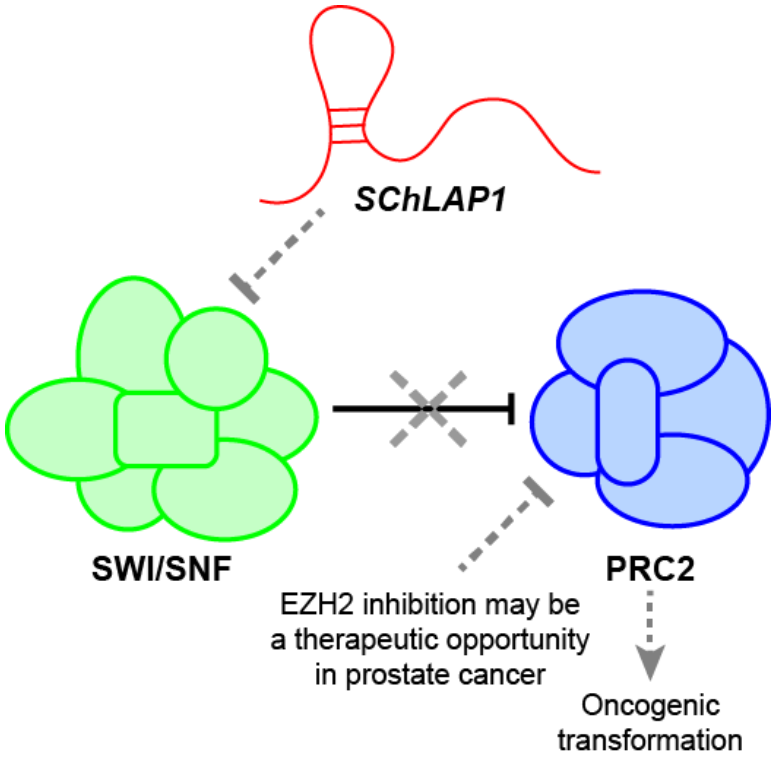


Figure 3.7 Working model of *SChLAP1* in prostate cancer.
In normal cells, SWI/SNF antagonizes PRC2 activity (black line). In the presence of *SChLAP1*, SWI/SNF-mediated repression of PRC2 is lost, leading to oncogenic transformation of cells (gray lines). In this scenario, pharmacologic EZH2 inhibition may be a promising therapeutic strategy. *Figure adapted from Wilson *et al.*, *Cancer Cell*, 2010.

References

1. Cooperberg, M.R., Moul, J.W. & Carroll, P.R. The changing face of prostate cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **23**, 8146-8151 (2005).
2. Shen, M.M. & Abate-Shen, C. Molecular genetics of prostate cancer: new prospects for old challenges. *Genes & development* **24**, 1967-2000 (2010).
3. Prensner, J.R.*, et al.* The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nature genetics* **45**, 1392-1398 (2013).
4. Prensner, J.R.*, et al.* RNA biomarkers associated with metastatic progression in prostate cancer: a multi-institutional high-throughput analysis of SChLAP1. *The Lancet. Oncology* **15**, 1469-1480 (2014).
5. Mehra, R.*, et al.* A novel RNA in situ hybridization assay for the long noncoding RNA SChLAP1 predicts poor clinical outcome after radical prostatectomy in clinically localized prostate cancer. *Neoplasia* **16**, 1121-1127 (2014).
6. Bottcher, R.*, et al.* Novel long non-coding RNAs are specific diagnostic and prognostic markers for prostate cancer. *Oncotarget* (2015).
7. Lee, R.S. & Roberts, C.W. Linking the SWI/SNF complex to prostate cancer. *Nature genetics* **45**, 1268-1269 (2013).
8. Reisman, D., Glaros, S. & Thompson, E.A. The SWI/SNF complex and cancer. *Oncogene* **28**, 1653-1668 (2009).
9. Wilson, B.G. & Roberts, C.W. SWI/SNF nucleosome remodellers and cancer. *Nature reviews. Cancer* **11**, 481-492 (2011).
10. Kadoch, C.*, et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nature genetics* **45**, 592-601 (2013).
11. Shain, A.H. & Pollack, J.R. The spectrum of SWI/SNF mutations, ubiquitous in human cancers. *PloS one* **8**, e55119 (2013).
12. Oike, T.*, et al.* A synthetic lethality-based strategy to treat cancers harboring a genetic deficiency in the chromatin remodeling factor BRG1. *Cancer research* **73**, 5508-5518 (2013).
13. Helming, K.C.*, et al.* ARID1B is a specific vulnerability in ARID1A-mutant cancers. *Nature medicine* **20**, 251-254 (2014).
14. Hoffman, G.R.*, et al.* Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 3128-3133 (2014).
15. Wilson, B.G.*, et al.* Residual complexes containing SMARCA2 (BRM) underlie the oncogenic drive of SMARCA4 (BRG1) mutation. *Molecular and cellular biology* **34**, 1136-1144 (2014).
16. Wilson, B.G.*, et al.* Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. *Cancer cell* **18**, 316-328 (2010).

17. Knutson, S.K.*, et al.* Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 7922-7927 (2013).
18. Bitler, B.G.*, et al.* Synthetic lethality by targeting EZH2 methyltransferase activity in ARID1A-mutated cancers. *Nature medicine* (2015).
19. Knutson, S.K.*, et al.* Selective inhibition of EZH2 by EPZ-6438 leads to potent antitumor activity in EZH2-mutant non-Hodgkin lymphoma. *Molecular cancer therapeutics* **13**, 842-854 (2014).
20. Siegel, R.L., Miller, K.D. & Jemal, A. Cancer statistics, 2015. *CA: a cancer journal for clinicians* **65**, 5-29 (2015).
21. Lu, P. & Roberts, C.W. The SWI/SNF tumor suppressor complex: Regulation of promoter nucleosomes and beyond. *Nucleus* **4**, 374-378 (2013).
22. Grasso, C.S.*, et al.* The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239-243 (2012).
23. Barbieri, C.E.*, et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nature genetics* **44**, 685-689 (2012).
24. Cerami, E.*, et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery* **2**, 401-404 (2012).
25. Gao, J.*, et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling* **6**, pl1 (2013).
26. Iyer, M.K.*, et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nature genetics* (2015).
27. Roychowdhury, S.*, et al.* Personalized oncology through integrative high-throughput sequencing: a pilot study. *Science translational medicine* **3**, 111ra121 (2011).
28. Francis, N.J., Saurin, A.J., Shao, Z. & Kingston, R.E. Reconstitution of a functional core polycomb repressive complex. *Molecular cell* **8**, 545-556 (2001).
29. Shao, Z.*, et al.* Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* **98**, 37-46 (1999).
30. Kia, S.K., Gorski, M.M., Giannakopoulos, S. & Verrijzer, C.P. SWI/SNF mediates polycomb eviction and epigenetic reprogramming of the INK4b-ARF-INK4a locus. *Molecular and cellular biology* **28**, 3457-3464 (2008).
31. Varambally, S.*, et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**, 624-629 (2002).
32. Rhodes, D.R.*, et al.* Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* **9**, 166-180 (2007).
33. Dechassa, M.L.*, et al.* Architecture of the SWI/SNF-nucleosome complex. *Molecular and cellular biology* **28**, 6010-6021 (2008).
34. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343-349 (2011).

35. Cao, R.*, et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039-1043 (2002).
36. Tusher, V.G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 5116-5121 (2001).
37. Subramanian, A.*, et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545-15550 (2005).
38. McCabe, M.T. & Creasy, C.L. EZH2 as a potential target in cancer therapy. *Epigenomics* **6**, 341-351 (2014).
39. Kleer, C.G.*, et al.* EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 11606-11611 (2003).
40. Takawa, M.*, et al.* Validation of the histone methyltransferase EZH2 as a therapeutic target for various types of human cancer and as a prognostic marker. *Cancer science* **102**, 1298-1305 (2011).
41. McCabe, M.T.*, et al.* EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* **492**, 108-112 (2012).
42. Kondo, Y. Targeting histone methyltransferase EZH2 as cancer treatment. *Journal of biochemistry* **156**, 249-257 (2014).
43. Tomlins, S.A.*, et al.* Role of the TMPRSS2-ERG gene fusion in prostate cancer. *Neoplasia* **10**, 177-188 (2008).
44. Prensner, J.R.*, et al.* Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nature biotechnology* **29**, 742-749 (2011).
45. Yu, J.*, et al.* An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer cell* **17**, 443-454 (2010).
46. Zhang, Y.*, et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137 (2008).
47. Shin, H., Liu, T., Manrai, A.K. & Liu, X.S. CEAS: cis-regulatory element annotation system. *Bioinformatics* **25**, 2605-2606 (2009).

Chapter 4:

Characterization of the *SChLAP1* – SWI/SNF interaction reveals a therapeutic opportunity in prostate cancer[3]

Abstract

Prostate cancer is the most common and second-most deadly type of cancer in U.S. men[1]. Identifying those patients with aggressive, lethal disease remains a clinical challenge[2-4]. Recently, we identified a long noncoding RNA (lncRNA) termed *SChLAP1* that is expressed in a subset of prostate cancers[5]. *SChLAP1* mediates cell invasion and metastasis in part by interacting with and abrogating genome-wide binding of the tumor-suppressive SWI/SNF nucleosome-remodeling complex[6]. However, the nature of this relationship remains poorly understood. Also, several recent studies have identified various therapeutic vulnerabilities arising from SWI/SNF subunit mutations[7-13]. Here, we further characterize *SChLAP1* to identify a 250bp region near the 3'-end of the transcript necessary to its function. Additionally, we find that *SChLAP1* interacts with BRG1-, but not BRM-, containing SWI/SNF complexes. Furthermore, BRM knockdown decreased cell invasion and proliferation in *SChLAP1*-overexpressing, but not control, prostate epithelial cells. Our findings indicate that

[3] This chapter has been prepared as a manuscript for submission with the following authors: Anirban Sahu, John R. Prensner, Benjamin Chandler, Qi Cao, Nithin Edara, Udit Singhal, Sumin Han, Matthew K. Iyer, Rohit Malik, Felix Y. Feng, and Arul M. Chinnaiyan.

specific regions of *SChLAP1* are crucial to its function and may warrant further investigation for therapeutic development. These data also provide a rationale for the development and use of BRM-specific small molecule inhibitors. Moreover, lncRNA-mediated SWI/SNF inhibition may expose therapeutic opportunities in residual complexes similar to those arising from subunit mutations.

Introduction

Prostate cancer is the most common type of malignancy in U.S. men, with over 200,000 new diagnoses this year[1]. Furthermore, prostate cancer is the second-most lethal type of malignancy in U.S. men, accounting for nearly 28,000 deaths this year[1]. However, as these statistics suggest, most men diagnosed with prostate cancer die with the disease and from unrelated causes rather than from the disease[4]. Identifying the subset of patients with prostate cancer who will progress to aggressive, lethal disease remains a clinical challenge[3].

Recently, our group identified a prognostic, prostate cancer-specific long non-coding RNA (lncRNA) termed *SChLAP1* that is overexpressed in a subset of aggressive prostate cancers[5]. Subsequent studies found that *SChLAP1* is significantly prognostic for metastatic progression and lethal disease[14-16]. *SChLAP1* mediates cell invasiveness, in part, by interacting with and abrogating genome-wide binding of the tumor suppressive SWI/SNF nucleosome-remodeling complex[6]. However, the nature of this relationship is poorly

understood. Here, we sought to further characterize the regions of *SChLAP1* and components of SWI/SNF most crucial to this collaboration in prostate cancer.

Results

To characterize specific regions of *SChLAP1* essential for its function, we generated deletion constructs tiling every 250bp and overexpressed these in RWPE cells (**Fig. 4.1a,b**). Deletion of a single 250bp region (deletion construct #5, bp 1001 – 1250 for *SChLAP1* isoform 1) shared by all three major isoforms of the RNA abrogated *SChLAP1*-mediated cell invasion in RWPE cells (**Fig. 4.1c**).

To test whether this region is important for SWI/SNF binding, we performed RNA immunoprecipitation (RIP) for SNF5, a core subunit of the SWI/SNF complex that facilitates binding to histone proteins[17], in RWPE cells overexpressing *SChLAP1* isoform 1, *SChLAP1* isoform 2, and *SChLAP1* deletion construct #5, which failed to induce cell invasion. We observed that overexpression of both *SChLAP1* isoform 1 and isoform 2 robustly bound to SNF5, whereas deletion construct #5 failed to bind SNF5 (**Fig. 4.2a**). As controls, we also measured *AK093002* and *LOC145837*, two lncRNAs upregulated in subsets of prostate cancer (**Fig. 4.2b**) that are endogenously expressed in RWPE cells (**Fig. 4.2c**). Control RIP experiments for SNRNP70 protein demonstrated uniformly strong binding to *U1* spliceosomal RNA, which is a well-characterized protein-RNA interaction in the spliceosome ribonucleoprotein complex (**Fig. 4.2d**).

To exclude the possibility that this deletion construct produced an unstable RNA, we performed 5-ethinyl uridine (EU) incorporation assays, which demonstrated that the deletion construct RNA was equally stable when compared to full-length *SChLAP1* RNA (**Fig. 4.3a**). *In silico* modeling with RNAfold[18] of the *SChLAP1* RNA structure suggested the presence of a RNA hairpin in this region that is lost specifically in deletion construct #5 (**Fig. 4.3b**), potentially implicating this secondary structure in the function of the molecule.

The SWI/SNF complex is a multi-subunit epigenetic modifier that regulates gene expression by reorganizing nucleosomes to alter chromatin architecture[19,20]. Each individual SWI/SNF complex contains either BRG1 (also known as SMARCA4) or BRM (also known as SMARCA2) as its enzymatic component, but not both[19,21,22]. While redundancy between these ATPase subunits does exist, there are well-described functional differences between BRG1- and BRM-containing SWI/SNF complexes[19,21,23,24], and this remains an active area of research investigation.

To determine whether *SChLAP1* binds specifically to either type of complex, we performed RIP for BRG1 and BRM in 22Rv1 and LNCaP prostate cancer cells. We found that endogenous *SChLAP1*, but not other prostate-cancer associated lncRNAs [25,26], robustly coimmunoprecipitated with BRG1 but not BRM (**Fig. 4.4a**). To ensure the BRM antibody is able to retrieve intact SWI/SNF complexes, we performed immunoprecipitation followed by immunoblotting for a SMARCD1,

a core component of SWI/SNF[17,19,21] (**Fig. 4.4b**). To extend this analysis further, we assessed the impact of *SChLAP1* expression on BRG1 and BRM genomic binding. Using ChIP-PCR for four target genes of SNF5[5], we found that *SChLAP1* overexpression in RWPE cells preferentially decreased BRG1 binding from SNF5 target promoters by ChIP-PCR, whereas the effect on BRM was markedly more mild (**Fig. 4.4c**). Taken together, these results suggest that SWI/SNF complexes utilizing BRG1 as the enzymatic subunit are the primary target for *SChLAP1*.

The SWI/SNF complex is mutated in approximately 20% of all cancers[19,21,27,28]. Several recent studies have identified various therapeutic opportunities arising from SWI/SNF inactivation[7,9-13]. In particular, targeting the residual relative of a mutated subunit has been shown to result in a synthetic lethal phenotype[8]. For example, BRM has been identified as a synthetic lethal target in BRG1-mutated cancers[11-13], and ARID1B has been identified as a synthetic lethal target in ARID1A-mutated cancers[7]. Furthermore, lncRNAs have been implicated in mediating SWI/SNF biology[5,29,30], and have also been shown to specifically interact with BRG1 to inhibit its chromatin binding[29]. Given *SChLAP1*'s specific interaction with BRG1, we hypothesized that BRM may be a synthetic lethal target in *SChLAP1*-expressing cells.

To test this hypothesis, we performed siRNA knockdown of BRM in RWPE-*LacZ*, RWPE-*SChLAP1* isoform 1, and RWPE-*SChLAP1* isoform 2 cells (**Fig. 4.5a,b**).

We found that BRM knockdown slightly increased cell invasion in RWPE-*LacZ* cells but significantly decreased cell invasion in RWPE-*SChLAP1* cells (**Fig. 4.5c**). Additionally, while BRM knockdown slightly increased cell proliferation in RWPE-*LacZ* cells, consistent with previous studies identifying BRM as a proliferative gatekeeper in prostate cells[31], BRM knockdown significantly decreased cell proliferation in RWPE-*SChLAP1* cells (**Fig. 4.5d**). Interestingly, BRG1 knockdown resulted in decreased cell viability regardless of *SChLAP1* expression (data not shown). While one may hypothesize that BRG1 knockdown should not have any additional effect on *SChLAP1*-overexpressing cells, this could suggest a dose-dependent effect of BRG1 inactivation in either creating a synthetic lethal environment or killing the cell completely. Taken together, these results suggest that BRM can be targeted in *SChLAP1*-expressing cells to decrease cell viability.

To explore the mechanistic basis of this synthetic lethality, we sought to determine whether targeting BRM in *SChLAP1*-expressing cells destabilizes residual SWI/SNF complexes, as previously suggested in SWI/SNF synthetic lethal models[7,11]. While BRM knockdown in RWPE-*LacZ* cells affected the expression of some subunits and their incorporation into the SWI/SNF complex, BRM knockdown in RWPE-*SChLAP1* cells reduced the expression and incorporation of additional components (**Fig. 4.6**). Although further experimentation is necessary to uncover the nature of the SWI/SNF complex under these conditions, these results suggest that SWI/SNF destabilization may

be one of several mechanisms that could explain why targeting BRM in *SChLAP1*-expressing, but not RWPE-*LacZ* control cells, results in decreased cell invasion and proliferation.

Discussion

In summary, we have shown that a 250bp region near the 3'-end of *SChLAP1* mediates its invasive phenotype and coordinates its interaction with the SWI/SNF complex. Additionally, we found that *SChLAP1* interacts with BRG1-containing but not BRM-containing SWI/SNF complexes and preferentially decreases BRG1 genomic binding. Interestingly, our model of *SChLAP1* action shares several similarities to the lncRNA *Mhrt*, a cardioprotective transcript that functions by directly interacting with the helicase domain of BRG1, resulting in decreased chromatin binding by BRG1[29]. Finally, *SChLAP1*'s preference for BRG1 may expose BRM as a therapeutic target in prostate cancer (**Fig. 4.7**). Thus, while mutations in SWI/SNF subunits have been used to reveal specific vulnerabilities in cancer, our work suggests that lncRNA-mediated BRG1-inactivation may uncover similar therapeutic opportunities in malignancy. Taken together, our findings have broad implications for cancer biology and provide additional rationale for the development and use of BRM-specific inhibitors as therapeutic treatment in cancer.

Materials and Methods

Cell lines

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

SChLAP1 or *LacZ* control cell lines were generated as previously described. Briefly, *SChLAP1* full-length transcript was amplified from LNCaP cells and cloned into the pLenti6 vector along with *LacZ* controls. The benign immortalized prostate cell line RWPE was infected with lentiviruses expressing *SChLAP1* or *LacZ*. Stably-transfected RWPE cells were selected using blasticidin for one week. All lentiviruses were generated by the University of Michigan Vector Core.

Basement Membrane Matrix Invasion Assays

For invasion assays, stably-transfected RWPE cells were trypsinized, counted with a Coulter counter, and diluted to 1.5 million cells/mL. 100µL of the cell solution were seeded onto the basement membrane matrix present in the insert of a 24 well culture plate. Fetal bovine serum was added to the lower chamber as

a chemo-attractant. After 48 hours, the non-invading cells and basement membrane matrix were gently removed with a cotton swab. Invasive cells located on the lower side of the chamber were stained with using a Diff-Quik staining kit, air-dried and photographed. The number of invaded cells were counted in 4 different microscopic fields per insert and extrapolated to determine the number of invaded cells per well.

RNA immunoprecipitation

RIP assays were performed using a Millipore EZ-Magna RIP RNA-Binding Protein Immunoprecipitation kit from Millipore according to the manufacturer's instructions. RIP-PCR was performed as qPCR, as described above, using total RNA as input controls. 1:150th of RIP RNA product was used per PCR reaction. Antibodies used for RIP were Rabbit polyclonal IgG (Millipore, PP64), SNRNP70 (Millipore, CS203216), SNF5 (Millipore, ABD22), BRG1 (Abcam ab4081), BRM (Abcam ab15597) and AR (Millipore, 06-680, rabbit), using 5 – 7 µg of antibody per RIP reaction. All RIP assays were performed in biological duplicate.

siRNA knockdown

Cells were plated in 100mM plates at a desired concentration and transfected with 20nM experimental siRNA oligos or non-targeting controls twice, at 12 hours and 24 hours post-plating. Knockdowns were performed with Lipofectamine RNAiMAX in OPTI-MEM I media. 72 hours post-transfection, cells were

harvested. Knockdown efficiency was determined by qPCR. siRNAs used in this study are listed below:

BRM siRNA #1: CAAAGCAGAUGAACGCUAU

BRM siRNA #2: GAAAGGAGGUGCUAAGACA

BRG1 siRNA #1: GCACACCGCUGCAGAACAA

BRG1 siRNA #2: GCGACUCACUGACGGAGAA

Immunoblotting

Cells were lysed in RIPA lysis buffer supplemented with HALT protease inhibitor. Protein lysates were boiled in sample buffer, and 5-10µg protein was loaded onto a SDS-PAGE gel and run for separation of proteins. Proteins were transferred onto Polyvinyl Difluoride membranes and blocked for 30 minutes in blocking buffer (5% milk, 0.1% Tween, Tri-buffered saline (TBS-T)). Membranes were incubated overnight at 4C with primary antibody. Primary antibodies used in this study are listed in **Table 4.1**. After 3 washes with TBS-T, membranes were incubated for 30 minutes at room temperature with horseradish peroxidase-conjugated secondary antibody. Following 3 washes with TBS-T and one wash with TBS, the signals were visualized by enhanced chemiluminescence system from GE Healthcare.

RNA isolation and cDNA synthesis

Total RNA was isolated using Trizol and a miRNeasy Kit with DNase I digestion according to the manufacturer's instructions. RNA integrity was verified on an

Agilent Bioanalyzer 2100. cDNA was synthesized from total RNA using Superscript III and random primers.

Quantitative Real-time PCR

Quantitative Real-time PCR (qPCR) was performed using Power SYBR Green Mastermix on an Applied Biosystems 7900HT Real-Time PCR System. All oligonucleotide primers were obtained from Integrated DNA Technologies and are listed in **Table 4.2**.

Tissue Samples

Prostate tissues were obtained from the radical prostatectomy series and Rapid Autopsy Program at the University of Michigan tissue core[32]. These programs are part of the University of Michigan Prostate Cancer Specialized Program Of Research Excellence (S.P.O.R.E.). All tissue samples were collected with informed consent under an Institutional Review Board (IRB) approved protocol at the University of Michigan. (SPORE in Prostate Cancer (Tissue/Serum/Urine) Bank Institutional Review Board # 1994-0481).

EU pulse chase RNA stability assay

5-ethinyl-uridine RNA stability assays were performed using the Click-iT Nascent RNA Capture Kit (Life Technologies) according to manufacturer instructions. Briefly, RWPE-*SChLAP1* isoform 1 cells or RWPE-*SChLAP1* deletion #5 cells were treated with 5-ethinyl-uridine and RNA was harvested post-chase. Relative

RNA fraction remaining compared to pulse is plotted in the bar graph. Error bars indicate S.E.M.

Chromatin Immunoprecipitation

ChIP assays were performed as previously described[5,26,33] using antibodies for BRG1 (Abcam ab4081), BRM (Abcam ab15597) and Rabbit IgG (Millipore PP64B). Briefly, approximately 10^6 cells per antibody were cross-linked for 10-15 minutes with 1% formaldehyde and the cross-linking was inactivated by 0.125M glycine for 5 minutes at room temperature. Cells were rinsed with ice-cold PBS three times and cell pellets were resuspended in lysis buffer plus protease inhibitors. Chromatin was sonicated to an average length of 500bp, centrifuged to remove debris, and supernatants containing chromatin fragments were incubated with protein A/G beads to reduce non-specific binding. Then, beads were removed and supernatants were incubated with 6µg of antibody overnight at 4C. Fresh protein A/G beads were added and incubated with protein-chromatin-antibody complexes for 2 hours at 4C, washed twice with 1X dialysis buffer and four times with IP wash buffer, and eluted in 150µl IP elution buffer. 1:10th of the ChIP reaction was taken for protein evaluation for validation of ChIP pull-down. Reverse crosslinking was performed by incubating the eluted product with 0.3M NaCl at 65C overnight. ChIP product was cleaned up with the USB PrepEase kit. ChIP experiments were validated for specificity by Western blotting.

Proliferation Assays

72 hours post-transfection with siRNA, cells were trypsinized, counted with a Coulter counter, and plated into a 96-well plate (4 wells per experimental group) at a density of 4000 cells per well. The plates were placed in an Incucyte for live-cell imaging to obtain cell confluence data at 3-hour intervals. Error bars indicate S.E.M.

Immunoprecipitation Assay

Immunoprecipitation assays were performed as previously described[7]. Briefly, 60 million cells per reaction were harvested and nuclear pellets isolated with the NE-PER Nuclear Extraction kit (Pierce). The nuclear pellets were resuspended in 500µl of lysis buffer (20mM Tris pH 7.5, 150mM NaCl, 1% Triton-X, protease inhibitor cocktail), frozen, thawed, then sonicated for 6 cycles of 30 sec on/30sec off on the highset setting of a BioRuptor XL. Lysates were cleared by centrifuging 12,000xg for 10min at 4C. Protein concentration was measured by Bradford reagent. 1mg of protein was used per pulldown. Lysates were precleared with 25µl protein A/G beads. 10% of the precleared lysate was saved as input. 5µg of SMARCC1 antibody (Santa Cruz) was added to each reaction tube and incubated at 4C overnight with end-over-end rotation. Antibody-protein complexes were capture with 45µl of protein A/G beads for 2 hrs at 4C with end-over-end rotation. Bead complexes were washed 4 times with lysis buffer and protein complexes were eluted by boiling beads in 25µl 1xSDS loading buffer

plus 2µl B-mercaptoethanol at 98C for 5 min. Proteins were separated by SDS-PAGE. Antibodies used for immunoblotting are listed in **Table 4.1**.

Tables

Antibody	Vendor	Cat. No.	Concentration
ACTL6A	Bethyl	A301-391A	1 to 3000
ARID1A	Bethyl	A301-041A	1 to 1000
ARID1B	Abcam	ab57461	2.5ug/mL
B-Actin	Sigma	A5316	1 to 5000
BRG1	Abcam	ab4081	1 to 1000
BRM	Abcam	ab15597	1 to 1000
PB1	Bethyl	A301-591A	1 to 1000
SMARCC1	Santa Cruz	sc-9746	1 to 1000
SMARCC2	Bethyl	A301-039A	1 to 1000
SMARCD1	Bethyl	A301-595A	1 to 3000
SMARCE1	Bethyl	A300-810A	1 to 3000
SNF5	Millipore	ABD22	1 to 1000
SNF5	Abcam	ab58209	1 to 1000

Table 4.1 Antibodies used in this study

Gene	Purpose	Sense	Antisense
SChLAP1	qPCR	TGGACACAATTTCAAGTCCTCA	CATGGTGAAAGTGCCTTATACA
AK093002	qPCR	GGGAATCCCCATATCACAGT	GGAAGCTTACAGTTTTCAAGCA
LOC145837	qPCR	CACCTTGGAGAGACCCAGAA	GGCAGATGGAAGAAGTGGAA
U1	qPCR	GGGAGATACCATGATCACGAAGGT	CCACAAATTATGCAGTCGAGTTTCCC
PCAT-1	qPCR	TGAGAAGAGAAATCTATTGGAACC	GGTTTGTCTCCGCTGCTTTA
SNF5	qPCR	GAGACTCTGACAGACGCTGAGA	GTGTGCTGATGGGCTGGTTAC
BRG1	qPCR	AAAATCGAGAAGGAGGATGACA	CCAAGCTTGATCTTCACTTTGA
BRM	qPCR	GGAAGAGGAAGATGAAGAAGAGTC	AAATCGCTCACTACAGGTTTGG
GAPDH	qPCR	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGAG
HMBS	qPCR	GATGGGCAACTGTACCTGACTGGA	TGGGGCCCTCGTGGAATGTTA
B-actin	qPCR	AAGGCCAACCGCGAGAAG	ACAGCCTGGATAGCAACGTACA
NFIX promoter	ChIP-PCR	TGTTTTCCCAGACTGCTGTG	CACACATGTCACGTCCAGAA
PTPRU promoter	ChIP-PCR	TGTGTCTGAGCGTAGGATGG	ACAACATTCACCCCCAACAT
TMEM248 promoter	ChIP-PCR	ATGCACACACCCATACCATC	TGGCTTGTAGTGTGCATCGT
KIAA1274 promoter	ChIP-PCR	CTCTCCAGCCTGACAGAGGT	CTGTTACCACCTGCCTCACC
KIAA0841	ChIP-PCR	GATTGAAGCTGTAGGAGTGATGC	AGTGGGTTATCCCAGTCACATC
Chr6 Alu	ChIP-PCR	TCCTGTTCCTTTTGACTGGAAT	ATGGCAGATTTTTGCAAGACAG

Table 4.2 Primers used in this study

Figures

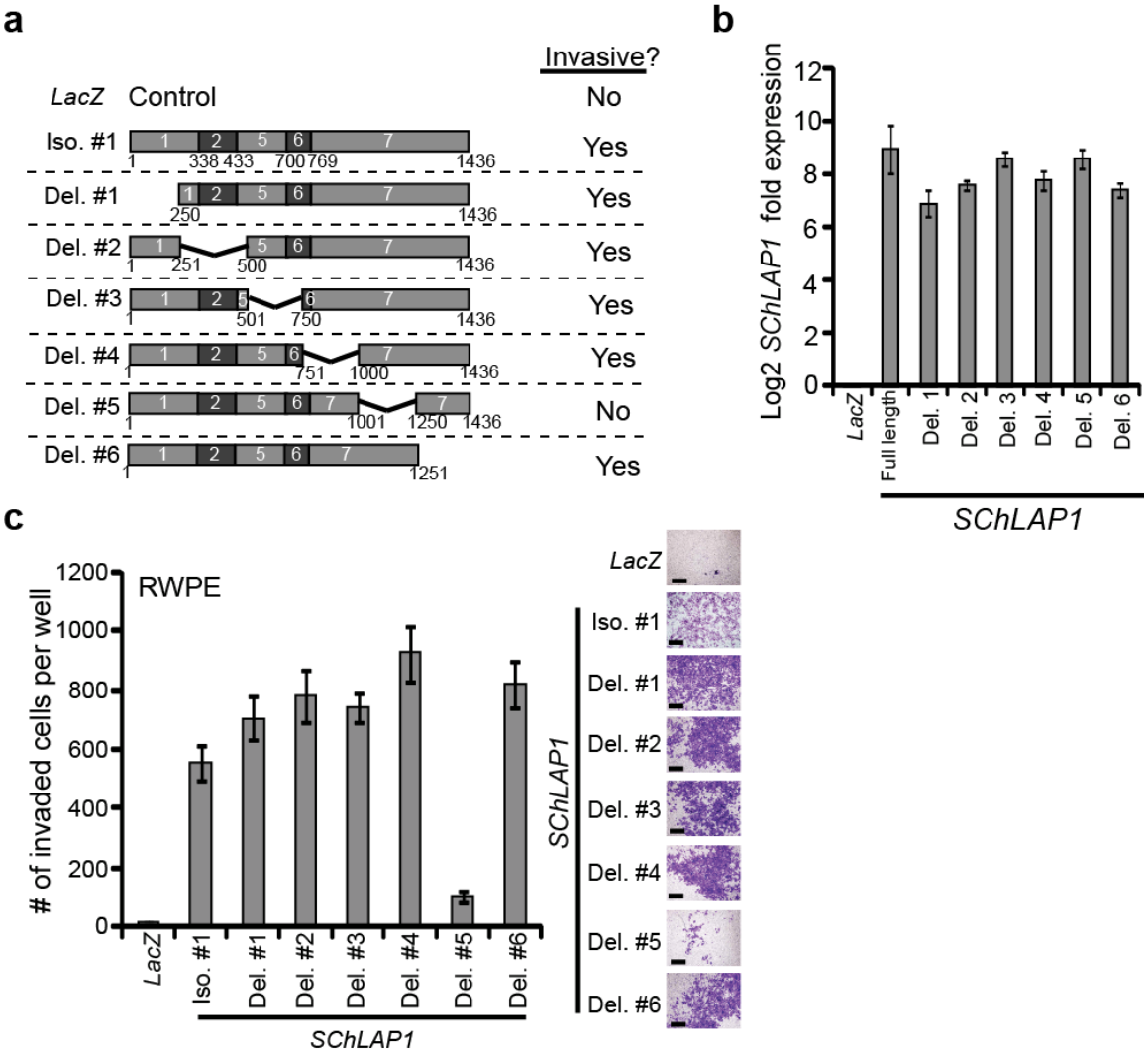


Figure 4.1 A 250bp region of *SChLAP1* is necessary for its invasive phenotype.
(a) Schematic of deletion constructs made for *SChLAP1* and their impact on cell invasion. **(b)** Overexpression of *SChLAP1* deletion constructs in RWPE cells. qPCR for full-length *SChLAP1* and the *SChLAP1* deletion constructs after overexpression in RWPE. **(c)** Deletion constructs of *SChLAP1* were overexpressed in RWPE cells and the resulting cells were assayed for invasion in a Boyden chamber assay. Data are represented as normalized mean +/- S.E.M. Images to the right show representative Boyden chamber membranes following invasion. All images were captured at the same magnification.

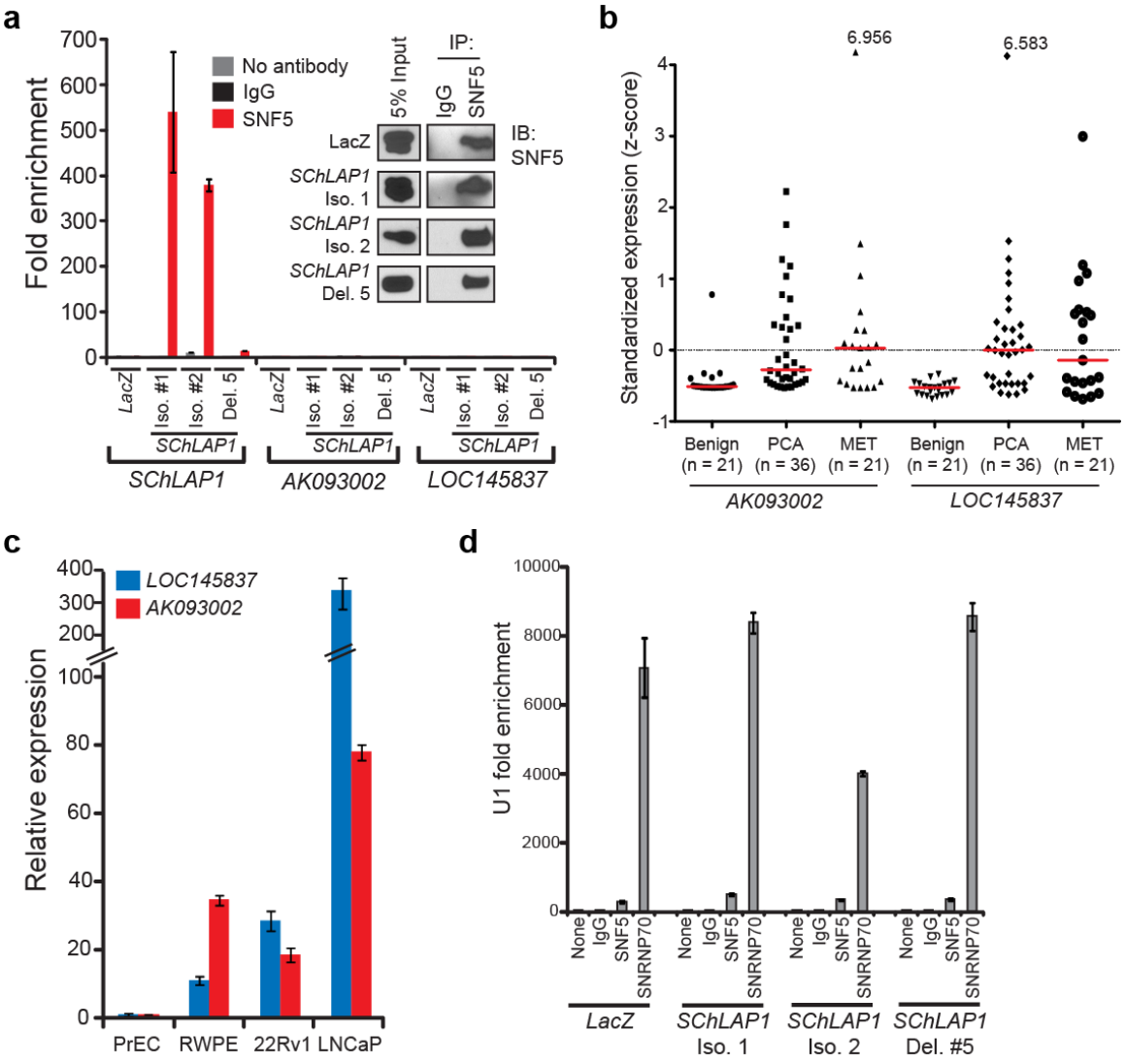


Figure 4.2 A 250bp region of *SChLAP1* is necessary for its interaction with SNF5.
(a) RIP analysis of SNF5 in RWPE cells overexpressing *LacZ*, *SChLAP1* isoform 1, *SChLAP1* isoform 2, or *SChLAP1* deletion construct #5. Inset, protein blots showing pulldown efficiency. *AK093002* and *LOC145837* serve as negative controls. Data are mean +/- S.E.M. **(b)** Expression of *AK093002* and *LOC145837* in a cohort of benign, localized cancer, and metastatic prostate cancers. Expression was measured by qPCR and data are represented as standardized z-score values. Red lines indicate median values for each group. **(c)** Expression of *AK093002* and *LOC145837* in prostate cancer cell lines. qPCR data were normalized to the average of *GAPDH* + *ACTB* and compared to PREC primary non-immortalized prostate cells. Error bars indicate S.E.M. **(d)** Control SNRNP70 experiments in the RWPE-*SChLAP1* overexpression models. Enrichment of *U1* is shown for SNRNP70 IP experiments and serves as a positive technical control for the experiment. Error bars indicate S.E.M.

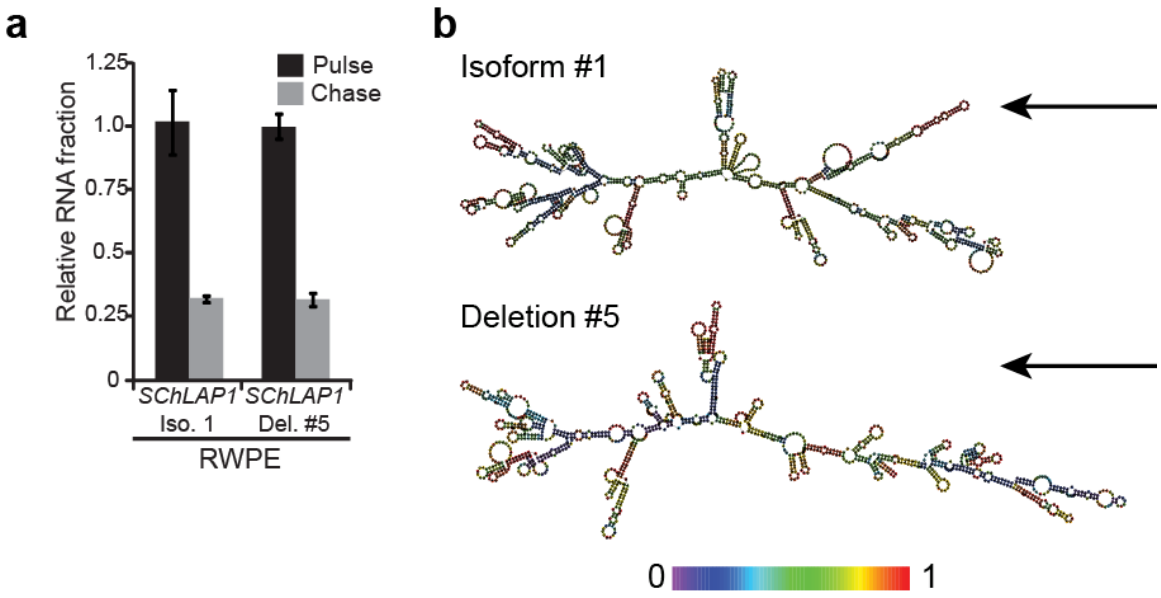


Figure 4.3 *SChLAP1* deletion construct #5 is stable but lacks a hairpin loop. **(a)** RNA stability assays using incorporation of 5-ethinyl-uridine to monitor RNA decay. RWPE-*SChLAP1* cells or RWPE-*SChLAP1* deletion #5 cells were treated with 5-ethinyl-uridine and RNA was harvested post-chase. Relative RNA fraction remaining is plotted. Error bars indicated S.E.M. **(b)** *In silico* structural predictions of *SChLAP1* isoform 1 and deletion construct #5 by RNAfold. Arrows identify a structural hairpin that may be lost in *SChLAP1* deletion construct #5.

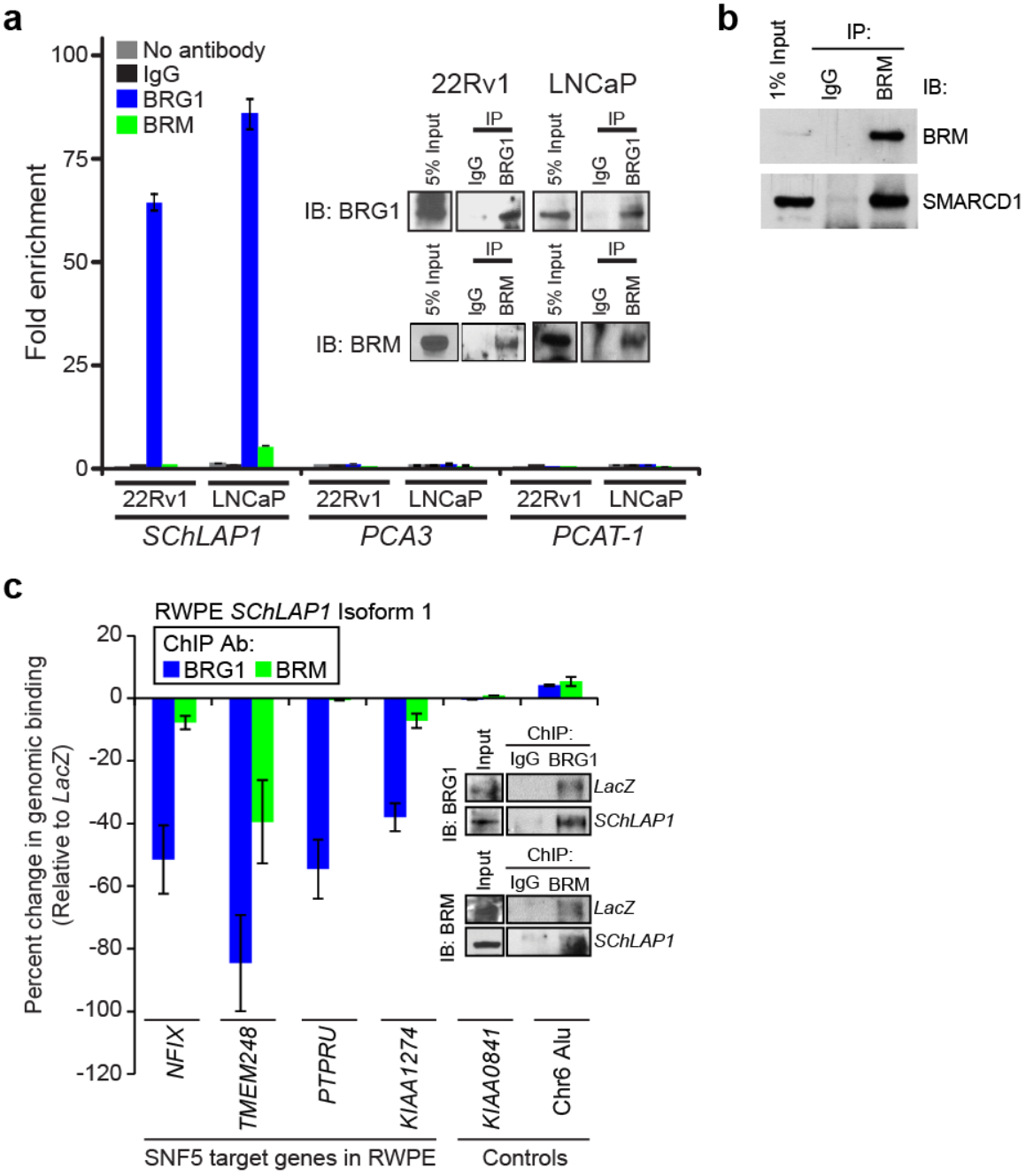


Figure 4.4 *SChLAP1* interacts with and abrogates genomic binding of BRG1.

(a) RIP of BRG1 or BRM in 22Rv1 and LNCaP cells. Inset, protein blots showing pulldown efficiency. *PCA3* and *PCAT-1* serve as negative controls. Data are mean +/- S.E.M. **(b)** Immunoprecipitation for BRM followed by immunoblotting in LNCaP cells. **(c)** ChIP for BRG1 and BRM in RWPE-*SChLAP1* and RWPE-*LacZ* cells. ChIP-PCR for 4 target genes of SNF5. *KIAA0841* and *Chr6 Alu* serve as negative controls. Data are represented as percent change in genomic binding

relative to *LacZ* after being normalized to IgG controls. The inset western blot indicates immunoprecipitation efficiency for BRG1 and BRM.

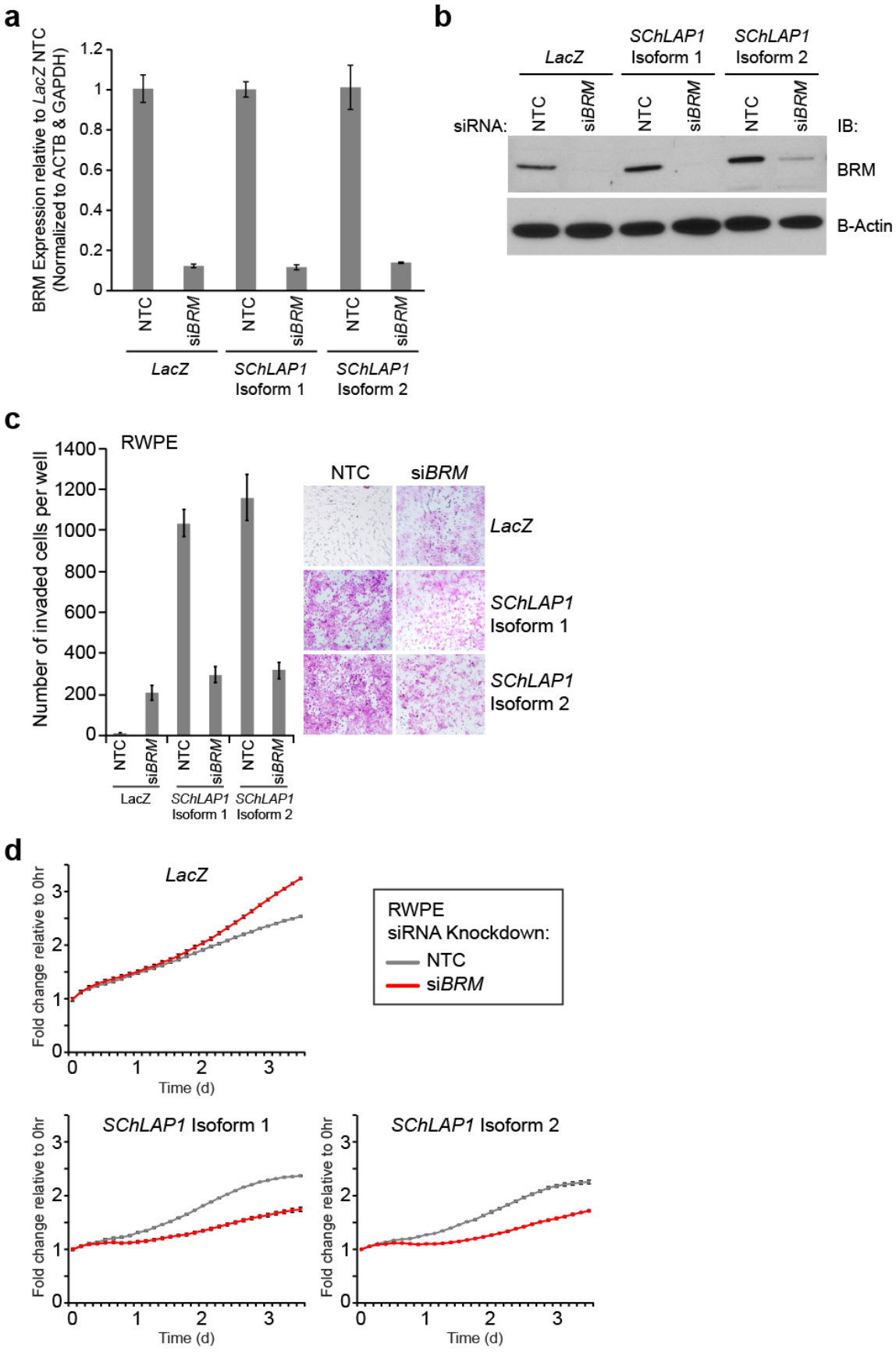

a
BRM Expression relative to LacZ NTC (Normalized to ACTB & GAPDH)
1.2
1
0.8
0.6
0.4
0.2
0
NTC
siBRM
LacZ
SChLAP1 Isoform 1
SChLAP1 Isoform 2
b
LacZ
SChLAP1 Isoform 1
SChLAP1 Isoform 2
siRNA:
IB:
BRM
B-Actin
c
RWPE
Number of invaded cells per well
1400
1200
1000
800
600
400
200
0
NTC
siBRM
LacZ
SChLAP1 Isoform 1
SChLAP1 Isoform 2
d
LacZ
Fold change relative to 0hr
Time (d)
RWPE
siRNA Knockdown:
NTC
siBRM
SChLAP1 Isoform 1
SChLAP1 Isoform 2

Figure 4.5 BRM knockdown in *SChLAP1*-expressing cells decreases cell invasion and proliferation.
(a) Knockdown efficiency of BRM siRNA in RWPE-*LacZ*, RWPE-*SChLAP1* isoform 1, and RWPE-*SChLAP1* isoform 2 cells. qPCR shows BRM knockdown relative to non-targeting control. Error bars represent S.E.M. **(b)** Knockdown efficiency of BRM siRNA in RWPE-*LacZ*, RWPE-*SChLAP1* isoform 1, and RWPE-*SChLAP1* isoform 2 cells. Western blot for BRM is shown. B-Actin serves as a loading control. **(c)** Cell invasion through Matrigel in a Boyden chamber assay of RWPE-*LacZ*, RWPE-*SChLAP1* isoform 1, and RWPE-*SChLAP1* isoform 2 with or without BRM knockdown. Data are represented as normalized mean +/- S.E.M. Images to the right show representative Boyden chamber membranes following invasion. All images were captured at the same magnification. **(d)** Cell proliferation of RWPE-*LacZ*, RWPE-*SChLAP1* isoform 1, and RWPE-*SChLAP1* isoform 2 with or without BRM knockdown. Cell proliferation assays were performed by Incucyte live-cell imaging. Data shown are fold change in cell confluence vs. time at 3-hour intervals. Each data point is the mean of quadruplicates +/- S.E.M.

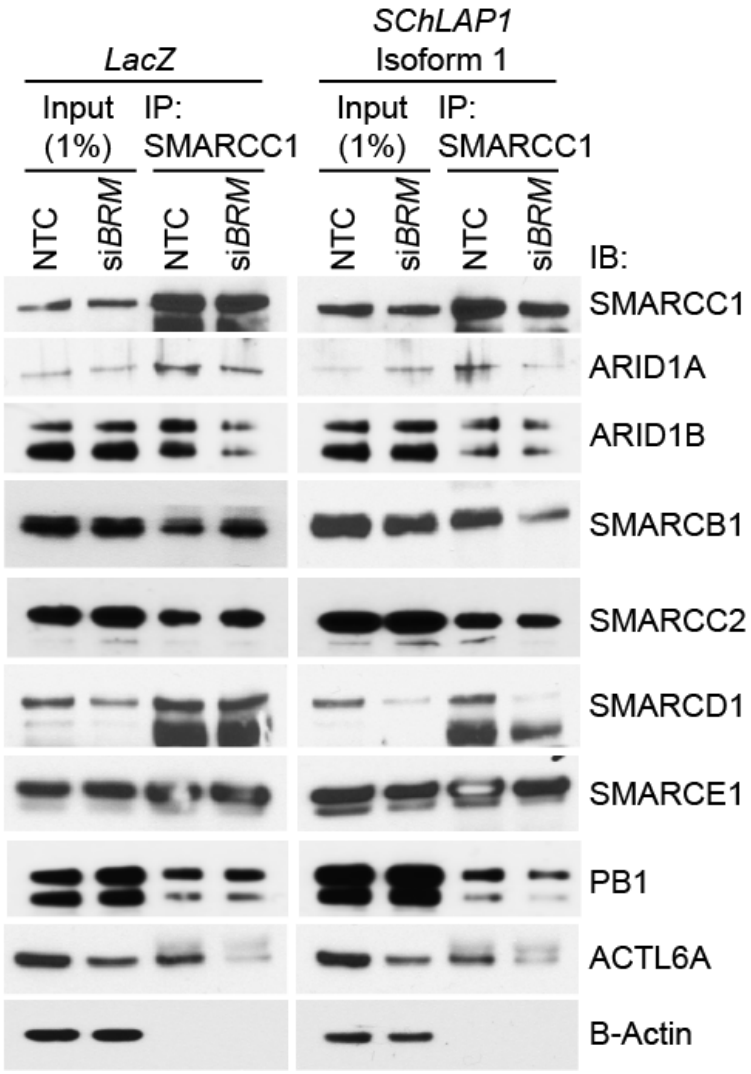


Figure 4.6 BRM knockdown in *SChLAP1*-expressing cells may destabilize the SWI/SNF complex.
Immunoprecipitation for SMARCC1 followed by immunoblotting in RWPE-*LacZ* and RWPE-*SChLAP1* cells treated with non-targeting control siRNA or BRM siRNA.

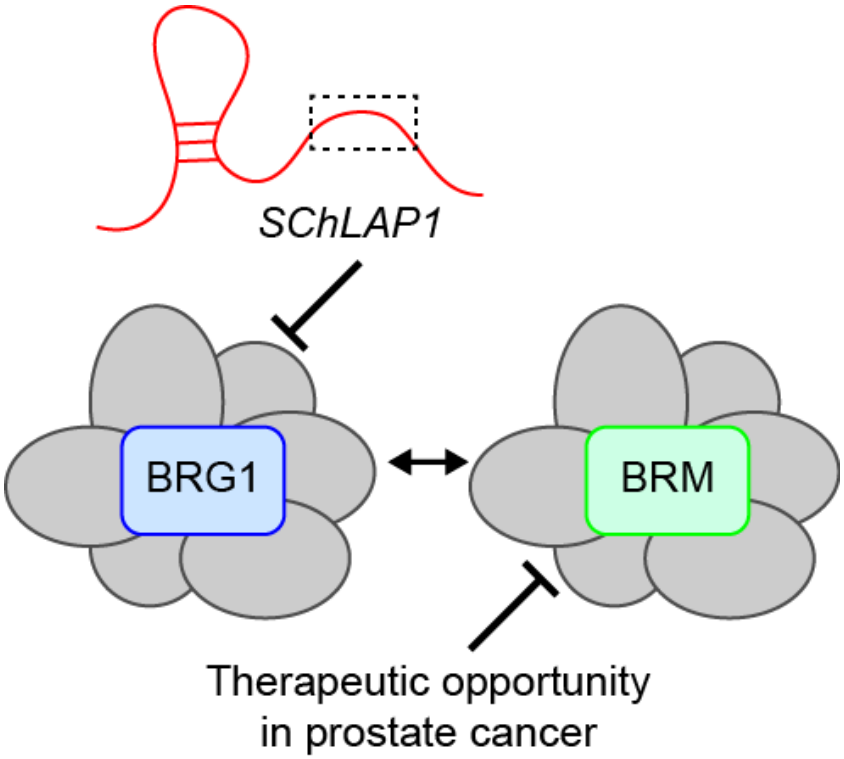


Figure 4.7 Working model of *SChLAP1* in prostate cancer
A 250bp region near the 3'-end of *SChLAP1* mediates its invasive phenotype and interaction with SWI/SNF. *SChLAP1* specifically interacts with and represses BRG1-containing SWI/SNF complexes, exposing BRM as a therapeutic target in prostate cancer.

References

1. Siegel, R.L., Miller, K.D. & Jemal, A. Cancer statistics, 2015. *CA: a cancer journal for clinicians* **65**, 5-29 (2015).
2. Etzioni, R., Cha, R., Feuer, E.J. & Davidov, O. Asymptomatic incidence and duration of prostate cancer. *American journal of epidemiology* **148**, 775-785 (1998).
3. Shen, M.M. & Abate-Shen, C. Molecular genetics of prostate cancer: new prospects for old challenges. *Genes & development* **24**, 1967-2000 (2010).
4. Cooperberg, M.R., Moul, J.W. & Carroll, P.R. The changing face of prostate cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **23**, 8146-8151 (2005).
5. Prensner, J.R.*, et al.* The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nature genetics* **45**, 1392-1398 (2013).
6. Lee, R.S. & Roberts, C.W. Linking the SWI/SNF complex to prostate cancer. *Nature genetics* **45**, 1268-1269 (2013).
7. Helming, K.C.*, et al.* ARID1B is a specific vulnerability in ARID1A-mutant cancers. *Nature medicine* **20**, 251-254 (2014).
8. Helming, K.C., Wang, X. & Roberts, C.W. Vulnerabilities of mutant SWI/SNF complexes in cancer. *Cancer cell* **26**, 309-317 (2014).
9. Knutson, S.K.*, et al.* Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 7922-7927 (2013).
10. Bitler, B.G.*, et al.* Synthetic lethality by targeting EZH2 methyltransferase activity in ARID1A-mutated cancers. *Nature medicine* (2015).
11. Hoffman, G.R.*, et al.* Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 3128-3133 (2014).
12. Wilson, B.G.*, et al.* Residual complexes containing SMARCA2 (BRM) underlie the oncogenic drive of SMARCA4 (BRG1) mutation. *Molecular and cellular biology* **34**, 1136-1144 (2014).
13. Oike, T.*, et al.* A synthetic lethality-based strategy to treat cancers harboring a genetic deficiency in the chromatin remodeling factor BRG1. *Cancer research* **73**, 5508-5518 (2013).
14. Mehra, R.*, et al.* A novel RNA in situ hybridization assay for the long noncoding RNA SChLAP1 predicts poor clinical outcome after radical prostatectomy in clinically localized prostate cancer. *Neoplasia* **16**, 1121-1127 (2014).
15. Prensner, J.R.*, et al.* RNA biomarkers associated with metastatic progression in prostate cancer: a multi-institutional high-throughput analysis of SChLAP1. *The Lancet. Oncology* **15**, 1469-1480 (2014).

16. Bottcher, R.*, et al.* Novel long non-coding RNAs are specific diagnostic and prognostic markers for prostate cancer. *Oncotarget* (2015).
17. Dechassa, M.L.*, et al.* Architecture of the SWI/SNF-nucleosome complex. *Molecular and cellular biology* **28**, 6010-6021 (2008).
18. Gruber, A.R., Lorenz, R., Bernhart, S.H., Neubock, R. & Hofacker, I.L. The Vienna RNA websuite. *Nucleic acids research* **36**, W70-74 (2008).
19. Wilson, B.G. & Roberts, C.W. SWI/SNF nucleosome remodellers and cancer. *Nature reviews. Cancer* **11**, 481-492 (2011).
20. Lu, P. & Roberts, C.W. The SWI/SNF tumor suppressor complex: Regulation of promoter nucleosomes and beyond. *Nucleus* **4**, 374-378 (2013).
21. Reisman, D., Glaros, S. & Thompson, E.A. The SWI/SNF complex and cancer. *Oncogene* **28**, 1653-1668 (2009).
22. Cohen, S.M.*, et al.* BRG1 co-localizes with DNA replication factors and is required for efficient replication fork progression. *Nucleic acids research* **38**, 6906-6919 (2010).
23. Morrison, A.J.*, et al.* INO80 and gamma-H2AX interaction links ATP-dependent chromatin remodeling to DNA damage repair. *Cell* **119**, 767-775 (2004).
24. Wong, A.K.*, et al.* BRG1, a component of the SWI-SNF complex, is mutated in multiple human tumor cell lines. *Cancer research* **60**, 6171-6177 (2000).
25. de Kok, J.B.*, et al.* DD3(PCA3), a very sensitive and specific marker to detect prostate tumors. *Cancer research* **62**, 2695-2698 (2002).
26. Prensner, J.R.*, et al.* Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nature biotechnology* **29**, 742-749 (2011).
27. Kadoch, C.*, et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nature genetics* **45**, 592-601 (2013).
28. Shain, A.H. & Pollack, J.R. The spectrum of SWI/SNF mutations, ubiquitous in human cancers. *PloS one* **8**, e55119 (2013).
29. Han, P.*, et al.* A long noncoding RNA protects the heart from pathological hypertrophy. *Nature* **514**, 102-106 (2014).
30. Zhu, Y., Rowley, M.J., Bohmdorfer, G. & Wierzbicki, A.T. A SWI/SNF chromatin-remodeling complex acts in noncoding RNA-mediated transcriptional silencing. *Molecular cell* **49**, 298-309 (2013).
31. Shen, H.*, et al.* The SWI/SNF ATPase Brm is a gatekeeper of proliferative control in prostate cancer. *Cancer research* **68**, 10154-10162 (2008).
32. Rubin, M.A.*, et al.* Rapid ("warm") autopsy study for procurement of metastatic prostate cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **6**, 1038-1045 (2000).
33. Yu, J.*, et al.* An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer cell* **17**, 443-454 (2010).

Chapter 5:

Clinical utility and translational opportunities of the long noncoding RNA *SChLAP1* in prostate cancer[4]

Abstract

Improved clinical predictors for disease progression are needed for localized prostate cancer, where only a minority of patients experience poor outcomes. Long noncoding RNAs (lncRNAs) are an emerging class of oncogenic molecules implicated in a diverse range of human malignancies. Here, we undertake an unbiased large-scale analysis of genes associated with aggressive clinical course by analyzing gene expression in 1,008 prostate cancer samples using a clinical-grade, high-density Affymetrix GeneChip platform. Among all known genes (coding and noncoding), the lncRNA *SChLAP1* ranked first for elevated expression in patients with metastatic progression. On multivariate modeling,

[4] Parts of this chapter have been previously published in the following manuscripts:

Prensner JR, Zhao S, Erho N, Schipper M, Iyer MK, Dhanasekeran SM, Magi-Galluzzi C, Mehra R, Sahu A, Siddiqui J, Davicioni E, Den RB, Dicker AP, Karnes RJ, Wei JT, Klein EA, Jenkins RB, Chinnaiyan AM, Feng FY. Nomination and validation of *SChLAP1* as an independent risk factor for metastatic prostate cancer progression. *Lancet Oncol,* 15(13):1469-80, 2014.

Mehra R, Shi Y, Udager AM, Prensner JR, Sahu A, Iyer MK, Siddiqui J, Cao X, Wei J, Jiang H, Feng FY, Chinnaiyan AM. A novel RNA in situ hybridization assay for the long noncoding RNA *SChLAP1* predicts poor clinical outcome after radical prostatectomy in clinically localized prostate cancer. *Neoplasia,* 16(12):1121-7, 2014.

Antisense oligonucleotide (ASO) experiments were performed by Lanbo Xiao.

SChLAP1 expression independently predicted metastasis, overall survival, and biochemical recurrence with odds ratios comparable to Gleason score. Next, we sought to develop an RNA in situ hybridization (ISH) assay for *SChLAP1* to investigate the spectrum of *SChLAP1* expression by ISH from benign prostatic tissue to metastatic castration-resistant prostate cancer (mCRPC). Our results demonstrate that *SChLAP1* expression increases with prostate cancer progression. To test the feasibility of a non-invasive test to detect *SChLAP1*, we measured *SChLAP1* expression in 230 urine sediment samples with either biopsy-confirmed cancer or biopsy-negative tissue. Our results show increased incidence and expression of *SChLAP1* RNA in patients at a higher risk for disease progression. Finally, to determine whether directly targeting *SChLAP1* is a viable therapeutic strategy, we evaluate clinical-grade *SChLAP1* anti-sense oligonucleotides (ASOs) and find that two independent ASOs targeting *SChLAP1* robustly reduce *SChLAP1* expression, decrease cell invasion, and reduce PRC2 activity. Taken together, these results establish several translational opportunities for *SChLAP1* in prostate cancer.

Introduction

While a majority of localized prostate cancer patients harbor slow-growing, non-lethal tumors, a smaller fraction of patients experience disease recurrence following definitive first-line therapies, which may lead to metastasis and death[1-6]. Current clinical paradigms rely mainly on pre-operative prostate specific antigen (PSA) levels, tumor stage, and biopsy Gleason score, which assesses cancer

cell histology, in order to estimate patient risk[7-11], yet, these remain imperfect tools that inaccurately classify some patients[12-15]. Thus, characterization of novel prognostic biomarkers, especially ones suitable for non-invasive detection, represents an important research focus for improving patient management.

To date, the majority of biomarker efforts have focused on protein-coding genes, which comprise only 2-3% of all transcribed genes [16,17]. Among the non-coding genes, long non-coding RNAs (lncRNAs) most closely resemble protein-coding genes, in that they are transcribed by RNA polymerase II, polyadenylated, and associated with specific epigenetic signatures (i.e., H3K4me3 at the promoter and H3K36me3 throughout the gene length) [18,19]. Recently, we and others have used next-generation sequencing to define long noncoding RNAs (lncRNAs) as potential biomarkers in prostate cancer[20,21]. Our work led to the analysis of the *SChLAP1* lncRNA as an oncogenic factor in prostate cancer that associates with poor patient outcomes[22]. Furthermore, our preliminary data indicate that *SChLAP1* expression levels can be detected in formalin-fixed, paraffin-embedded (FFPE) tissue sections by RNA in situ hybridization (ISH). Taken together, these results suggest a potential utility of *SChLAP1* as a prostate cancer biomarker.

Here, we undertake an analysis of 1,008 prostate cancer samples using unbiased approaches to define RNA biomarkers associated with metastatic progression. Next, we evaluate *SChLAP1* expression by ISH on FFPE tissue in a

large cohort of patients with clinically localized prostate cancer and lethal metastatic castration-resistant prostate cancer (mCRPC). Then, we measure *SChLAP1* expression in patient urine samples to determine if *SChLAP1* is detectable non-invasively. Finally, we explore the utility of clinical-grade antisense oligonucleotides (ASOs)[23] as a modality to directly target *SChLAP1* as a therapeutic strategy.

Results

We designed a retrospective biomarker discovery analysis according to PROBE criteria[24,25] in which prostate cancer patients who developed metastases were compared to those who did not (**Fig. 5.1a**). We employed 1,008 radical prostatectomy specimens from three academic institutions, comprising four independent patient cohorts (**Fig. 5.1a** and **Fig. 5.2a**). Three cohorts represented case-control study designs; one study was a case-cohort design. Patients were defined as at-risk for recurrence (e.g., pT2 tumor with positive margins or pT3 disease) by current clinical guidelines. Patient characteristics are detailed in **Table 5.1**. All cohorts with available information had mean patient follow-up between seven and fourteen years.

A clinical-grade microarray platform, which contains 5 million probes against 1.4 million unique probeset regions (PSRs), was used to measure global gene expression in an unbiased fashion. Tissue samples from three of four cohorts (not for EMC) were processed in a CLIA-certified laboratory, representing 95%

(960/1,008) of specimens. We analyzed all known protein-coding genes and lncRNAs previously identified in prostate cancer (PCATs).[20] We used metastasis as the primary endpoint. Whereas localized and locally-recurrent disease is potentially curable,[26,27] metastatic disease is incurable, requiring intensive treatment such as next generation anti-androgens and chemotherapy, and frequently progresses to mortality.[6,28-33]

Nomination of SChLAP1 by unbiased expression profiling

Using the MCI cohort (n=545), we performed a global assessment of gene expression differences between tumors from patients who experienced metastasis (n=212) and those who did not (n=333). Mean follow-up was 14 years. We derived median expression values for all genes in each group and compared the relative change in expression between groups. Surprisingly, the top-ranked gene was *SChLAP1*, which was recently characterized as an oncogenic prostate cancer lncRNA (**Fig. 5.1b**).[22] Overall, there were 230 genes whose expression associated with metastasis at a false discovery rate (FDR) or q ≤0·01 (Table S1).

SChLAP1 demonstrated the largest gene expression change between tumors with and without metastatic progression (**Fig. 5.1b**). High *SChLAP1* expression was associated with a markedly higher risk for biochemical recurrence, metastasis, death from prostate cancer, and death from any cause at ten years

post-prostatectomy (p=0·043, p<0·0001, p<0·0001, p<0·0001, respectively, Fisher's exact) (**Fig. 5.1c-f**).

Validation of SChLAP1

For initial validation of *SChLAP1*, we employed a case-cohort MCII set (n=232) of high-risk localized prostate cancer patients (**Table 5.1**) who underwent radical prostatectomy. We observed that *SChLAP1* was again a powerful predictor of time to biochemical recurrence, metastatic progression, and prostate cancer-specific mortality (p=0·002, p=0·0002, p=0·004, respectively, Cox model), with a strong trend for significance in predicting worse overall survival (p=0·066) (**Fig. 5.2b-e**).

We next incorporated data from a third independent cohort of radical prostatectomy tissues from high-risk patients at the CC (n=183, Table 1). Confirming our prior observations, we found a strong association between *SChLAP1* expression and metastatic progression in the CC set (OR=3·1, p=0·021, **Fig. 5.2f**).

Lastly, we searched for additional publicly-available cohorts with clinical annotation that used the Affymetrix HuEx platform and reported a ≥10% metastasis event rate for statistical robustness. We found one cohort from the Erasmus Medical Center (EMC, n=48)[34] and processed these data for *SChLAP1* expression as above (**Table 5.1**). Here, *SChLAP1* expression was again highly

associated with metastases, the only reported outcome in the EMC dataset (p=0·0022, **Fig. 5.2g**), with all metastatic events occurring in patients with high *SChLAP1* expression (data not shown). Together, these four datasets represent 1,008 patients, and all cohorts support a strong association between *SChLAP1* expression and metastasis.

A global comparison of SChLAP1 to other genes

To compare *SChLAP1* to other genes, we measured the receiver-operator-curve (ROC) area-under-the-curve (AUC) metric for metastatic disease progression across all annotated protein-coding genes and PCATs using the MCI and MCII cohorts. These cohorts were most enriched for high-risk patients and adverse outcomes. We plotted the AUC values for both cohorts for the top 1,000 genes (**Fig. 5.3**), of which a small minority displayed substantially higher AUC values in both cohorts (**Fig. 5.3, box**). A focused analysis of the top genes defined *SChLAP1* as the second best single-gene predictor of metastasis (**Fig. 5.3, right**).

SChLAP1 expression by ISH

To determine the utility of *SChLAP1* as a tissue biomarker, we examined *SChLAP1* expression by ISH in large TMA cohorts of patients with clinically localized prostate cancer or lethal mCRPC. Benign prostate glands, clinically localized prostate cancer, and mCRPC demonstrated a spectrum of *SChLAP1* expression by ISH (**Fig. 5.4**). When present, *SChLAP1* staining was

predominantly nuclear (**Fig. 5.4**). Overall, there were significant differences in *SChLAP1* expression between benign prostatic glands, clinically localized prostate cancer, and lethal mCRPC ($P < 0.001$; **Fig. 5.5a**). Benign prostatic glands were available for evaluation in 74 patients with clinically localized prostate cancer and, overall, showed absent to low *SChLAP1* expression (mean ISH product score = 13.8; range = 0-100). Out of a total of 208 patients with clinically localized prostate cancer, tissue from 160 patients (76.9%) was available for *SChLAP1* expression evaluation. Of these, 58 (36.3%) showed no SChLAP1 expression (ISH product score = 0), while the remaining 102 (63.7%) demonstrated a wide spectrum of *SChLAP1* expression (overall mean ISH product score = 44.5; range = 0-337; **Fig. 5.4**).

Relative to benign prostatic glands, *SChLAP1* expression was significantly increased in clinically localized prostate cancer ($P < 0.001$; **Fig. 5.5a**). In addition, high *SChLAP1* expression was associated with an increased proportion of clinically localized tumors with high Gleason score (GS≥8; **Fig. 5.5b**). Rapid autopsy material from a total of 28 patients with lethal mCRPC was available for *SChLAP1* expression evaluation, and a large proportion of the patients (15 cases, 53.6%) demonstrated high *SChLAP1* expression at one or more tissue site. Relative to benign prostatic glands ($P < 0.001$) and clinically localized prostate cancer ($P < 0.001$), *SChLAP1* expression was significantly increased in lethal mCRPC (mean ISH product score = 136.4; range = 0-370; **Fig. 5.5a**).

Overall, these data indicate that *SChLAP1* expression is associated with prostate cancer progression.

Non-invasive detection of SChLAP1 in urine sediments

Next, we sought to evaluate *SChLAP1* expression non-invasively in prostate cancer patients early in their disease course. We employed a University of Michigan cohort of 230 patient urine sediments[35,36] obtained post-digital rectal examination at the time of PSA screening for asymptomatic men. All men subsequently received a diagnostic prostate biopsy to determine whether cancer was present. Although urine sediments also contain bladder cells, *SChLAP1* expression is specific to prostate cells (see **Chapter 3**).

We then measured *SChLAP1* expression in our cohort of urine sediment samples. We observed high *SChLAP1* expression only in a subset of patients (**Fig. 5.6a**), which is consistent with the *SChLAP1* expression profile seen in all previous tissue cohorts[22]. Among patients with biopsy-confirmed cancer, expression of *SChLAP1* was both more frequent and more highly elevated in Gleason 7 patients compared to Gleason 6 patients (p=0·029, Fisher's exact, **Fig. 5.6b**). We were unable to evaluate Gleason ≥8 due to low numbers of patients. Finally, we stratified patients into low, intermediate, and high-risk categories according to standard PSA and Gleason thresholds. We found that *SChLAP1* expression was significantly elevated in intermediate and high risk patients compared to low risk patients (p=0·0022, Mann Whitney U test, **Fig.**

5.6c). These data provide proof-of-principle analyses that *SChLAP1* expression is detectable non-invasively in prostate cancer patient urine samples. However, additional validation is needed to confirm the utility of a urine-based *SChLAP1* test.

Direct targeting of SChLAP1 with antisense oligonucleotides

Finally, we used clinical-grade antisense oligonucleotides (ASOs) designed against *SChLAP1* to determine whether directly targeting *SChLAP1* could be pursued as a therapeutic strategy in prostate cancer. ASOs function by basepair hybridizing to target RNAs, resulting in transcript-specific RNAse H-mediated catalytic degradation[37]. Antisense technology has gained considerable traction over the past few years as several ASOs have been introduced into clinical trials and some have been FDA-approved for clinical use[38-41]. ASOs are a particularly attractive therapeutic modality for several reasons, including predictable human pharmacokinetics, prolonged tissue elimination half-lives, enhanced specificity compared to small molecule inhibitors, and lack of cytochrome P450 enzyme metabolism[23,37,42-44]. These characteristics are thought to make ASOs safer for patients and also more suitable for combination therapies with other drugs.

We tested two independent *SChLAP1* ASOs in 22Rv1 prostate cancer cells, which have moderate to high levels of endogenous *SChLAP1*. Cationic lipid-mediated transfection delivery of the ASOs resulted in significantly decreased *SChLAP1* expression, reduced cell invasion, as well as decreased PRC2 activity

indicated by a reduction in levels of H3K27 trimethylation (H3K27me3) (**Fig. 5.7**). While further experimentation is necessary to determine the *in vivo* effects of *SChLAP1* ASOs, these results suggest that directly targeting *SChLAP1* may be an effective therapeutic modality in the treatment of aggressive prostate cancer.

Discussion

Here, we perform the largest biomarker discovery project to date in prostate cancer, employing over 1,000 patients with one discovery cohort and three validation cohorts. Our study is the first of its kind to: (1) nominate and validate using a high-throughput clinical-grade assay in a CLIA-certified laboratory; (2) use distant metastasis as the primary endpoint; and (3) systematically evaluate non-coding elements in the transcriptome. We find that *SChLAP1* is one of the best genes for predicting metastatic progression. This association between *SChLAP1* and metastasis is robust across multiple independent cohorts.

Next, we characterize *SChLAP1* expression by ISH in a hospital-based cohort of American men treated for clinically localized prostate cancer, as well as a rapid autopsy cohort of patients with lethal mCRPC, using a novel ISH assay. We confirm that *SChLAP1* is predominantly a nuclear RNA transcript, which supports *in vitro* studies of *SChLAP1* in prostate cancer cells and preliminary *in situ* data [22]. We, also find that *SChLAP1* expression is enriched in metastatic samples, suggesting that expression of this lncRNA may be preferentially selected for during prostate cancer progression. Additionally, we find that, in clinically

localized prostate cancer, *SChLAP1* expression is enriched in samples from tumors with high Gleason scores (GS≥8) compared to tumors with lower Gleason scores, which also suggests an association with aggressive disease.

One essential aspect of *SChLAP1* is that its expression is specific to prostate cancer, with minimal expression in all other tumor and tissue types. *SChLAP1*, therefore, is uniquely suitable as a non-invasive biomarker. To this end, we show proof-of-principle data of *SChLAP1* in patient urine samples. Although formalized optimization of urine biomarker assays requires substantial investment and resources,[45] we are encouraged by these data and argue that prioritization of *SChLAP1* during future biomarker development studies may be appropriate.

Finally, we explore the therapeutic efficacy of directly targeting *SChLAP1* using ASOs. Our preliminary *in vitro* data shows effective down-regulation of *SChLAP1* gene expression, decreased cell invasion, and reduced PRC2 activity upon treatment with *SChLAP1* ASOs, suggesting that development of *SChLAP1*-targeted therapies warrants further investigation.

Overall, our work provides compelling evidence that *SChLAP1* expression is highly prognostic and an independent risk factor for metastasis. As such, this study provides insight into the pathogenesis of aggressive prostate cancer and identifies *SChLAP1* as a potential new clinical biomarker for metastatic progression and direct therapeutic target in advanced prostate cancer. Our work

contributes to the growing body of literature suggesting that lncRNAs are major drivers of cancer biology and, therefore, clinically important molecular entities.

Materials and methods

Study Design and Tissue Samples

Banked or archived tumor samples were obtained from three prostatectomy patient cohorts enrolled at the Mayo Clinic or the Cleveland Clinic under informed consent protocols approved by local Institutional Review Board (IRB). Analyses were designed in accordance with PROBE criteria.[24] The Mayo Clinic I (MCI) cohort included 212 patients with metastatic progression and a total of 333 patients without metastatic progression as described.[46] For the Mayo Clinic II (MCII) patients, a case-cohort study design was employed to randomly sample 20% of patients for analysis, in addition to all who developed metastases, from a cohort of 1,010 high-risk men who underwent radical prostatectomy between 2000-2006 as described.[47] The MCII cohort and its outcomes data represent a modified set of patients overlapping with our previously published report of *SChLAP1* with more stringent data quality control filters.[22]

Patients from the Cleveland Clinic (CC) were obtained from a case-control study design sampled from 2,317 conservatively treated radical prostatectomy patients with high-risk features who received no adjuvant or neo-adjuvant therapy from 1987-2008. Patients were sampled to achieve a three:one ratio for non-metastatic (n=134) versus metastatic progression (n=49) patients (Figure S3).

Patient cohorts were designed in accordance with STROBE recommendations for case-control and case-cohort studies.[48,49]

Microarray Hybridization and Gene Expression

RNA extraction from formalin-fixed paraffin embedded (FFPE) samples and microarray hybridization was performed for Mayo Clinic and Cleveland Clinic samples using clinical-grade techniques in a CLIA-certified (Clinical Laboratory Improvement Amendments) laboratory facility. CLIA certification was obtained through the Centers for Medicare and Medicaid Services (CMS) through standard procedures, and laboratory facilities satisfied all criteria required by the CMS for certification. Details regarding the CLIA requirements can be found online at: http://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/index.html. RNA purification, hybridization to Affymetrix Human Exon (HuEx) 1.0 ST GeneChips, and gene expression calculations are detailed in the Supplementary Methods. Partition-Around-Medoids unsupervised clustering was used to define expression subgroups in the MCI cohort. This expression threshold was applied to the other cohorts without additional modification. Microarray data are available on the NCBI Gene Expression Omnibus (GEO) as accession numbers GSE46691 (MCI) and GSE62116 (MCII). The GEO accession number for the CC cohort is pending.

Additional Raw Datasets

We obtained raw Affymetrix HuEx 1.0 ST GeneChip expression data and sample clinical information for Boormans et al.[34] (Erasmus Medical Center; EMC) from the NCBI Gene Expression Omnibus (GSE41408). Gene expression was calculated as above.

Nomination of metastasis-associated genes

We calculated the mean expression of each gene in patients experiencing metastasis versus non-metastatic patients. Fold expression change was calculated with the following formula: ($\log_2$(median_expression_metastatic)–$\log_2$(median_expression_no_metastasis)) / ($\log_2$(median_expression_all_samples) + four). The constant four was used to uniformly increase all expression values above zero to avoid a negative denominator.

Tissue microarray (TMA) construction

TMAs comprised of surgical pathology material from 208 patients with clinically localized prostate cancer were constructed using tumor and benign tissue from radical prostatectomy specimens; all patients had undergone radical prostatectomy at the University of Michigan Health System as primary monotherapy (i.e., no neoadjuvant hormonal or radiation therapy). This radical prostatectomy series is part of the University of Michigan Prostate Cancer Specialized Program of Research Excellence (SPORE) Tissue Core. Three

tissue cores (each 0.6 mm in diameter) were obtained from representative FFPE tissue blocks for each included patient sample. Detailed clinicopathologic data for this cohort (summarized in **Table 1**) are updated and maintained on a secure relational database.

Similarly, TMAs comprised of rapid autopsy material from 60 patients with lethal mCRPC were constructed; this material was obtained as part of the University of Michigan Prostate Cancer SPORE Rapid Autopsy Program, as described previously [50]. All patients received multimodal therapy, including a combination of radical prostatectomy, hormone deprivation, radiation, and/or chemotherapy; detailed clinicopathologic data for a portion of this cohort has been reported previously [51]. As described above, three tissue cores (each 0.6 mm in diameter) were obtained from representative FFPE tissue blocks all metastatic tumor sites, as well as primary tumor within the prostate (when present at the time of autopsy; i.e., no prior radical prostatectomy).

SChLAP1 ISH

SChLAP1 ISH was performed on thin (approximately 4 micron thick) TMA sections (Advanced Cell Diagnostics, Inc., Hayward, CA), as described previously [22]; in parallel, *SChLAP1* ISH was performed on previously identified positive and negative control FFPE tissue sections, and all controls worked adequately (data not shown). All slides were examined for *SChLAP1* ISH signals in morphologically intact cells and scored manually by a study pathologist (R.M.).

Specific *SChLAP1* ISH signal was identified as brown, punctate dots, and expression level was scored as follows: 0 = no staining or less than 1 dot per 10 cells; 1 = 1-3 dots per cell; 2 = 4-9 dots/cell (few or no dot clusters); 3 = 10-14 dots/cell (less than 10% in dot clusters); and, 4 = greater than 15 dots/cell (more than 10% in dot clusters). For each evaluable tissue core, a cumulative ISH product score was calculated as the sum of the individual products of the expression level (0-4) and percentage of cells (0-100) [i.e., (A% x 0) + (B% x 1) + (C% x 2) + (D% x 3) + (E% x 4); total range = 0-400]. For each tissue sample, the ISH product score was averaged across evaluable TMA tissue cores.

SChLAP1 Antisense Oligonucleotides

ASOs were designed by ISIS pharmaceuticals. 22Rv1 cells were plated in a 6-well dish at a desired concentration and transfected with 50nM experimental ASO or non-targeting control. Knockdowns were performed with Lipofectamine RNAiMAX in OPTI-MEM I media. 72 hours post-transfection, cells were harvested for RNA, protein, or plated for invasion assay.

Urine quantitative PCR

Urine samples were collected from 256 patients with informed consent at the time of PSA screening, following a digital rectal exam, and prior to subsequent needle biopsy at the University of Michigan with IRB approval as described.[35] RNA processing and quantitative PCR was performed as described.[52] Data quality control, normalization, and expression calculation were performed

according to standard parameters. A total of 230 urine samples passed quality control metrics and were included for data analysis.

Statistical Analyses

The primary endpoint of metastatic progression was defined as a positive CT scan or bone scan within 10 years of treatment. Fisher's exact test and logistic regression models were used to analyze the relation between each of the three clinical outcomes and clinical factors and biomarkers. To calculate a 10-year event rate, men followed for less than 10 years, and who did not have an event during follow-up, were excluded. In MCI, the original definition of cases and controls were used for fold change and AUC calculation.[53] The association between *SChLAP1* and clinical outcomes was assessed separately for each study and overall in a single logistic regression model stratified by study. Multivariate analyses were performed to assess whether *SChLAP1* was able to increase the predictive ability of standard clinical factors. All clinical covariates were included in the multivariate models. Details on statistical analyses are found in the Supplementary Methods. A p-value <0·05 was considered statistically significant. Kaplan-Meier curves and weighted Cox regression comparing time to metastases between groups defined by *SChLAP1* expression are shown only for the MCII case-cohort study and utilize the weighting method as described previously.[54] The case-control study design of MCI and CC cohorts allows for assessment of relative, but not absolute, incidence of events. Time to event data for EMC was not available. Nonparametric AUC values (equivalent to C

statistics) were calculated separately for each study. Overall values were calculated as the weighted average of the study-specific values with weights proportional to sample size. Testing for improved AUC value between the full model without *SChLAP1* and the full model with *SChLAP1* was done using the likelihood ratio test for *SChLAP1* in the full model with *SChLAP1*.[24] AUC values were analyzed for statistical differences as described previously.[55] For ISH analysis, all statistical analyses were performed using R (version 3.0.2). Mean *SChLAP1* expression for benign prostatic glands, clinically localized prostate cancer, and mCRPC were compared using the Student's t-test and ANOVA. To test for association between *SChLAP1* expression and specific clinicopathologic features in clinically localized prostate cancer, Fisher's exact test and the Student's t-test were used for categorical and continuous data, respectively.

Tables

	Mayo Clinic I (n=545)	Mayo Clinic II (n=232)	Cleveland Clinic (n =183)	Erasmus Medical Center (n = 48)
Age (Years, mean ± SD)	65·3 ± 6·4	63·1 ± 7·4	61·6 ± 6·3	NA
Follow-up (Months, mean ± SD)	160·7 ± 56·2	80·6 ± 30·1	116·6 ± 50·1	NA
Metastatic progression				
No	333 (61%)	157 (68%)	134 (73%)	39 (81%)
Yes	212 (39%)	75 (32%)	49 (27%)	9 (19%)
Pre-operative PSA				
<10	282 (52%)	126 (54%)	127 (69%)	21 (44%)
10 to 20	117 (22%)	62 (27%)	41 (23%)	17 (35%)
>20	131 (24%)	44 (19%)	12 (7%)	8 (17%)
Not available	15 (3%)	0 (0%)	3 (1%)	2 (4%)
Gleason score				
5	0 (0%)	0 (0%)	0 (0%)	0 (0%)
6	60 (11%)	17 (7%)	25 (17%)	23 (48%)
7	271 (49%)	117 (50%)	113 (62%)	16 (33%)
8	68 (13%)	39 (17%)	23 (13%)	8 (17%)
9	134 (24%)	57 (25%)	22 (12%)	1 (2%)
10	9 (2%)	1 (1%)	0 (0%)	0 (0%)
Not available	3 (1%)	1 (1%)	0 (0%)	0 (0%)
Clinical stage				
I	0 (0%)	0 (0%)	0 (0%)	0 (0%)
II	219 (40%)	97 (42%)	0 (0%)	16 (33%)
III	253 (46%)	102 (44%)	0 (0%)	26 (54%)
IV	0 (0%)	0 (0%)	0 (0%)	6 (13%)
Not available	73 (13%)	33 (14%)	183 (100%)	0 (0%)
Extracapsular extension				
Negative	272 (50%)	136 (59%)	51 (28%)	0 (0%)
Positive	273 (50%)	96 (41%)	132 (72%)	0 (0%)
Not available	0 (0%)	0 (0%)	0 (0%)	48 (100%)
Seminal vesicle invasion				
Negative	369 (68%)	149 (64%)	152 (83%)	0 (0%)
Positive	176 (32%)	83 (36%)	31 (17%)	0 (0%)
Not available	0 (0%)	0 (0%)	0 (0%)	48 (100%)
Lymph node invasion				
Negative	472 (87%)	199 (86%)	183 (100%)	0 (0%)
Positive	73 (13%)	33 (14%)	0 (0%)	0 (0%)
Not available	0 (0%)	0 (0%)	0 (0%)	48 (100%)

Surgical margin status				
Negative	279 (51%)	99 (43%)	92 (50%)	0 (0%)
Positive	266 (49%)	133 (57%)	91 (50%)	0 (0%)
Not available	0 (0%)	0 (0%)	0 (0%)	48 (100%)

Table 5.1 Clinical cohort characteristics

Figures

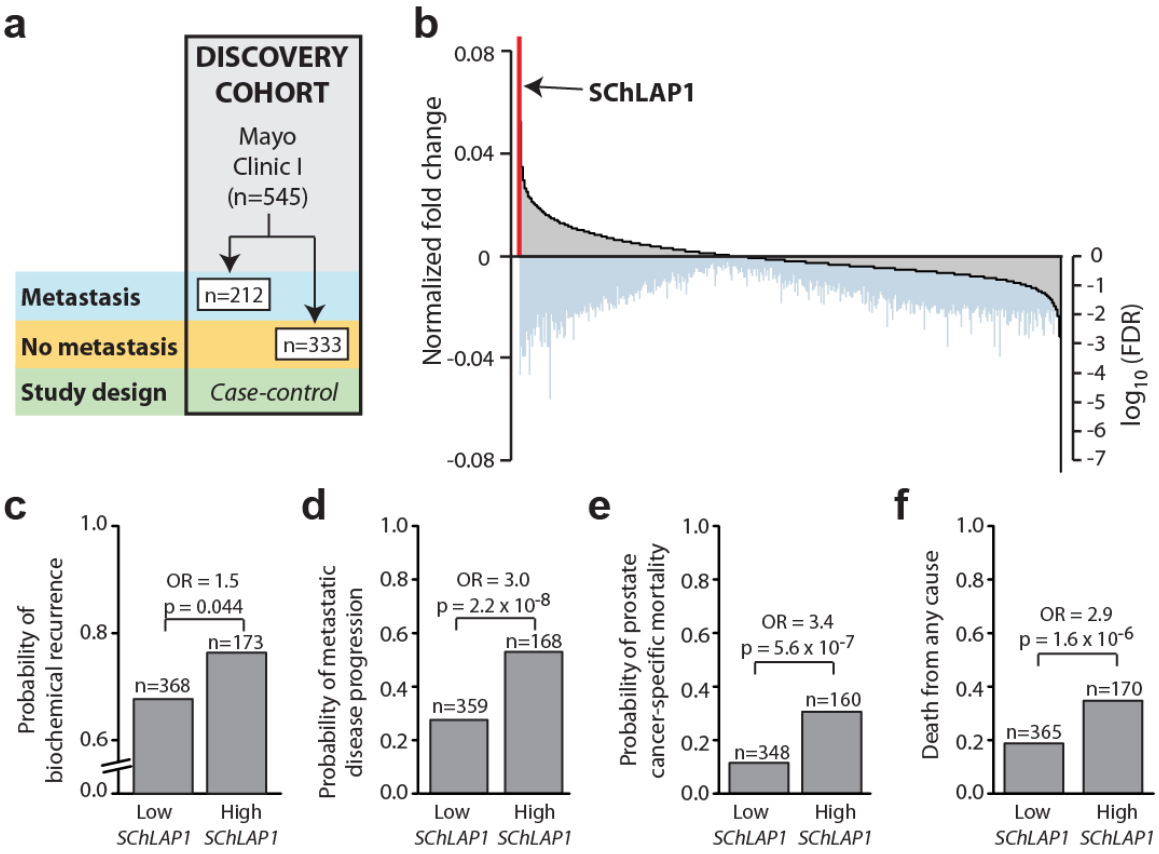


Figure 5.1 Nomination of *SChLAP1* as a top-ranked prognostic gene. **(a)** A schematic overview of the patient specimens in the discovery cohort. **(b)** A global view of gene expression changes associated with metastatic progression. In the Mayo Clinic I cohort (n = 545), gene expression was determined with Affymetrix Exon microarrays and differential expression analysis was performed for patients who experienced metastatic progression compared to those who did not. Ranking all genes according to the fold change of expression between metastatic and non-metastatic samples nominates *SChLAP1* as the top-ranked outlier gene associated with metastatic progression. Log_{10} false discovery rate (FDR) values for each corresponding gene are displayed below. **(c-f)** Patient outcomes in the Mayo Clinic I cohort (n = 545) were stratified by *SChLAP1* expression for biochemical recurrence **(c)**, progression to metastatic disease **(d)**, prostate cancer-specific mortality **(e)**, and overall survival **(f)**. *P* values in **c-f** were determined by a two-tailed Fisher's exact test.

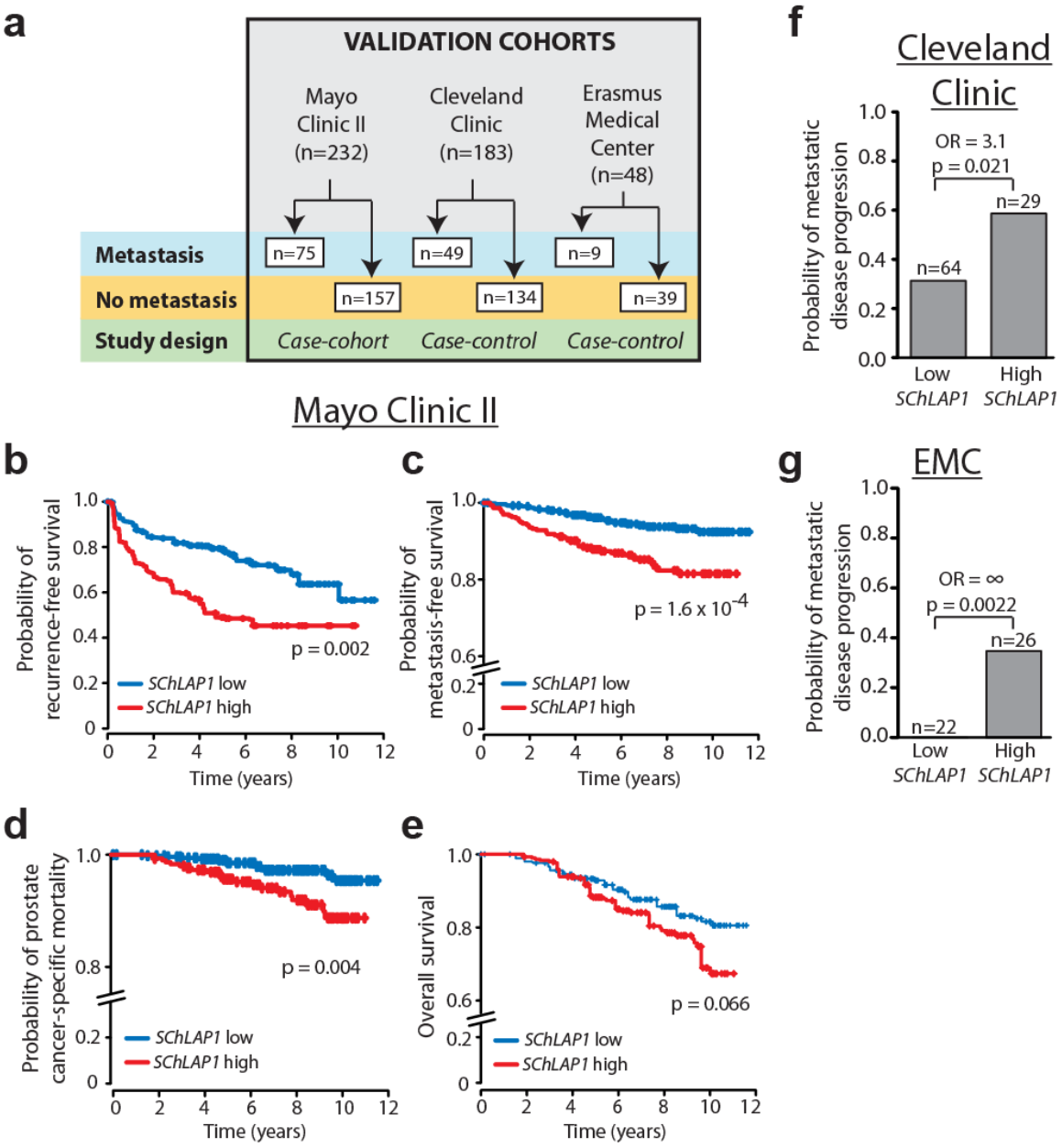


Figure 5.2 Validation of *SChLAP1* as a top-ranked prognostic gene.
(a) A schematic overview of the patient specimens in the validation cohorts. **(b-e)** Patient outcomes in the Mayo Clinic II cohort (n = 232) were stratified by *SChLAP1* expression for biochemical recurrence **(b)**, progression to metastatic disease **(c)**, prostate cancer-specific mortality **(d)**, and overall survival **(e)**. **(f,g)** Patient outcomes for metastasis were stratified by *SChLAP1* expression in the Cleveland Clinic **(f)** and Erasmus Medical Center **(g)** datasets. *P* values in **f-g** were determined by a two-tailed Fisher's exact test.

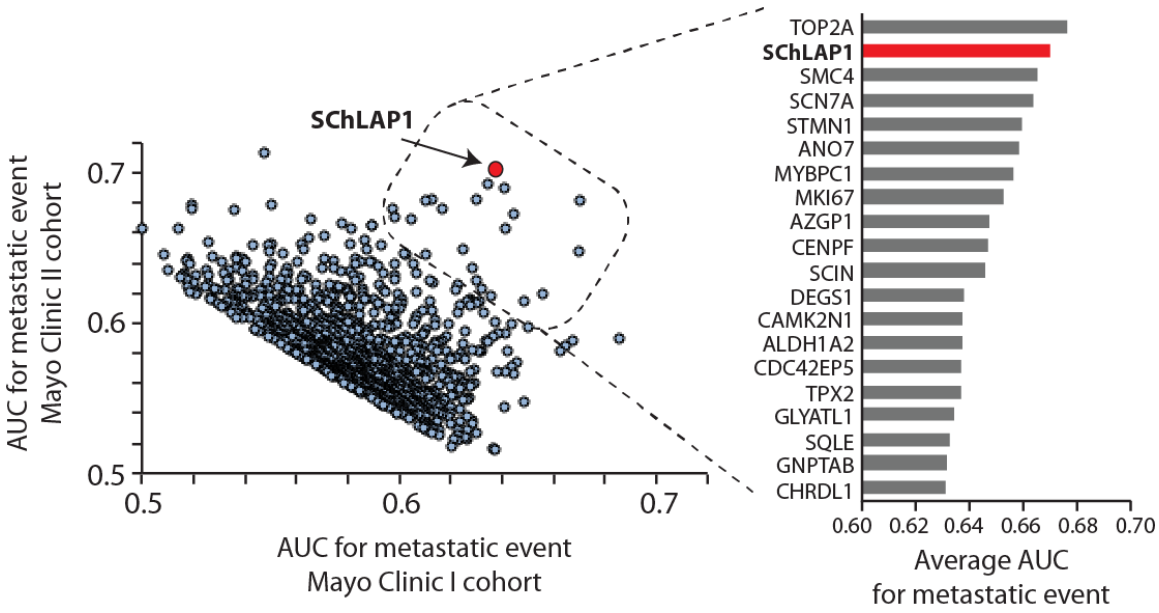


Figure 5.3 A global analysis of *SChLAP1* in the Mayo Clinic I and II cohorts. The Mayo Clinic I and II cohorts were independently analyzed to determine the receiver-operator-curve (ROC) area under the curve (AUC) metric for all genes for the development of metastatic disease. This global analysis generated independent AUC values for each cohort. AUC values for the top 1000 genes in both cohorts were plotted (*left*). A small subset of genes demonstrated superior AUC values in both cohorts (*see outlined box*). A detailed analysis of these top-20 genes is plotted (*right*) using the averaged AUC value between both cohorts. *SChLAP1* ranks #2 overall for prediction of metastatic spread.

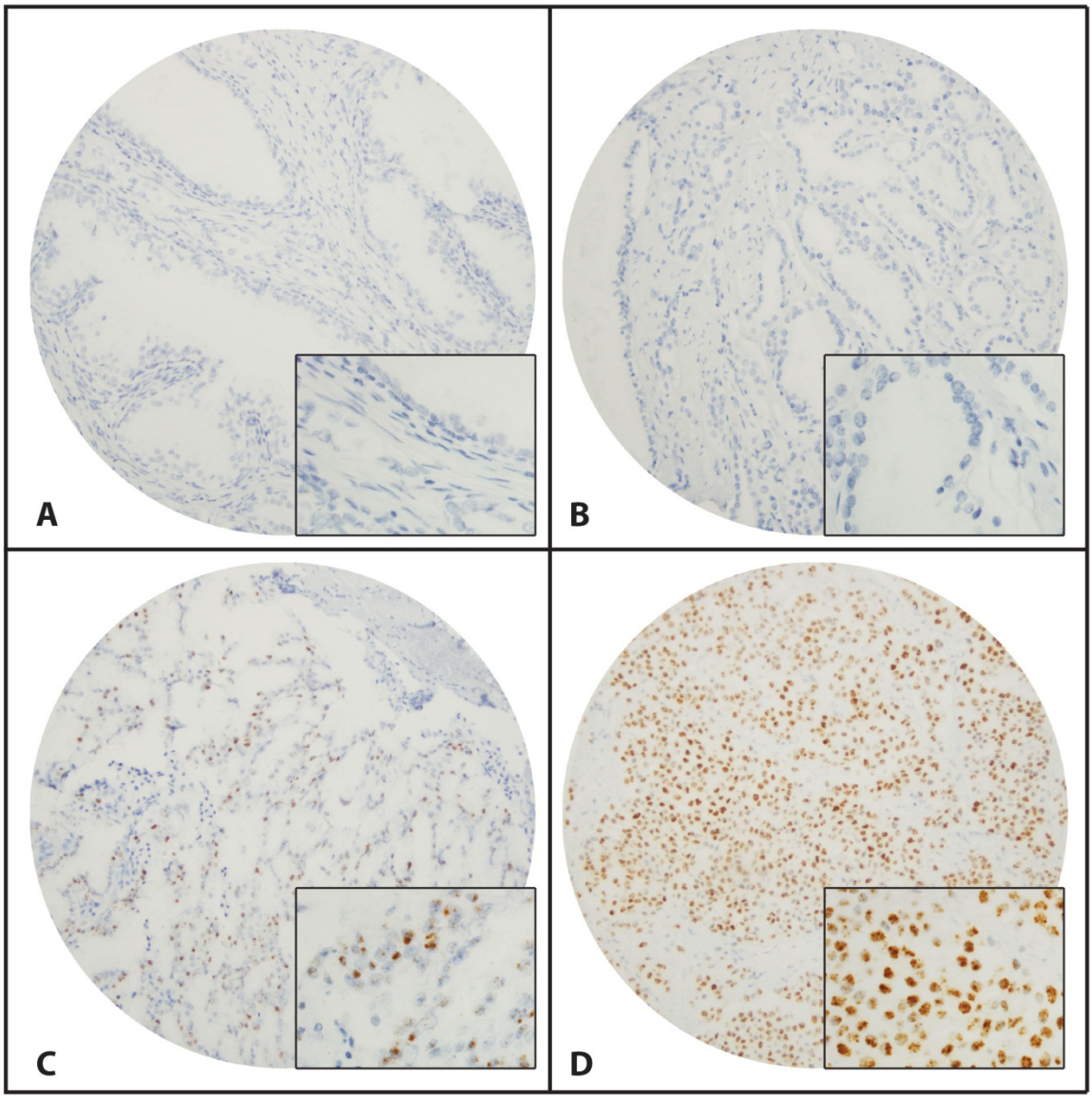


Figure 5.4 Spectrum of *SChLAP1* expression in benign prostatic glands, clinically localized prostate cancer, and lethal mCRPC by ISH.
SChLAP1 expression by chromogenic brown staining in **(a)** benign prostatic glands, **(b)** low and **(c)** high Gleason score clinically localized prostate cancer, and **(d)** lethal mCRPC. *SChLAP1* expression varies from negative to low in benign prostatic glands and low grade clinically localized prostate cancers to high expression in a subset of high grade clinically localized prostate cancers and lethal mCRPC. 100x magnification. Inset in **a**, **b**, **c**, and **d** = 400x magnification.

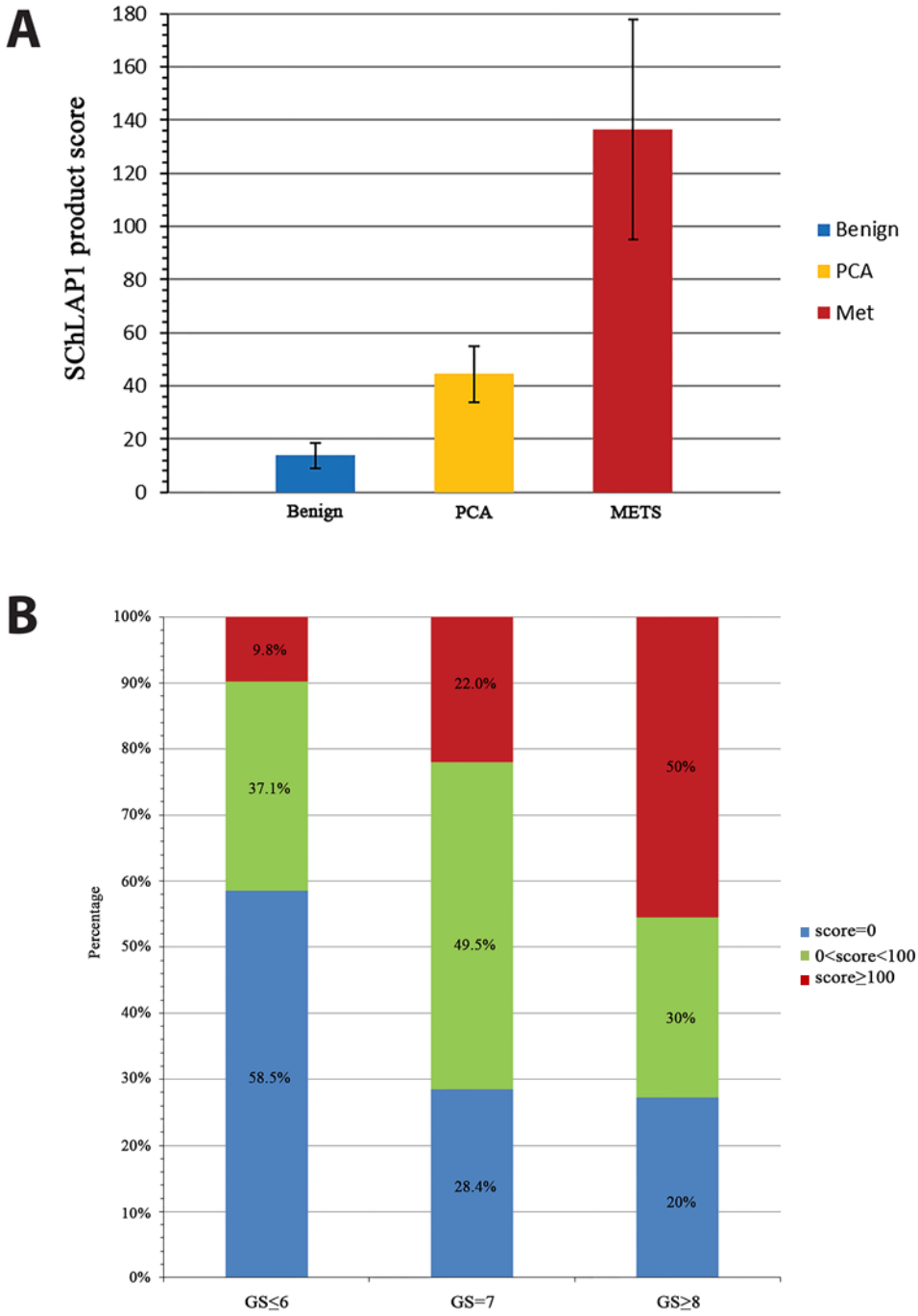


Figure 5.5 *SChLAP1* expression increases with prostate cancer progression.
(a) Histogram representation of mean *SChLAP1* ISH product score for benign prostatic glands (Benign), clinically localized prostate cancer (PCA), and lethal mCRPC (METS) in a large TMA cohort. Error bars represent standard deviation. *SChLAP1* expression is significantly associated with prostate cancer

progression, from benign glands to clinically localized prostate cancer to mCRPC ($P < 0.001$). **(b)** Histogram representation of proportion of clinically localized prostate cancer with negative (ISH product score = 0), low (ISH product score >0 and <100), or high (ISH product score ≥ 100) SChLAP1 expression, stratified by Gleason score (GS). High *SChLAP1* expression is associated with increasing GS in clinically localized prostate cancer.

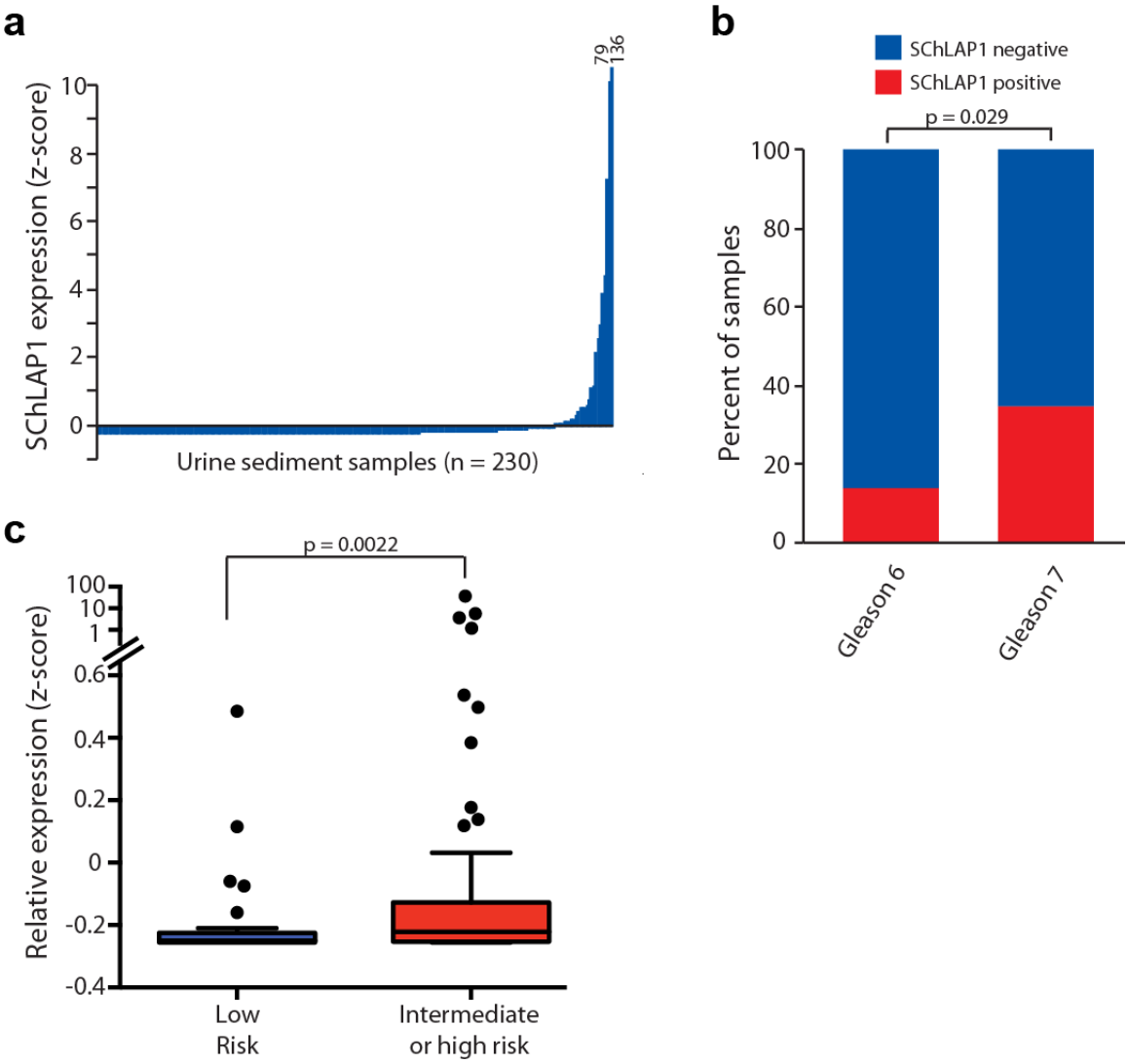


Figure 5.6 Detection of *SChLAP1* in patient urine samples.
(a) Detection of *SChLAP1* RNA in patient urine sediments. Samples are ordered according to *SChLAP1* expression. Expression is represented as the z-score. **(b)** The fraction of Gleason 6 (n = 44) and Gleason 7 (n = 49) urine sediments that demonstrate positive *SChLAP1* expression. *P* value was determined by a two-tailed Fisher's exact test. **(c)** *SChLAP1* expression in urine sediments from low risk (n = 37) and intermediate/high risk patients (n = 68). *P* value was determined by a Mann Whitney *U* test.

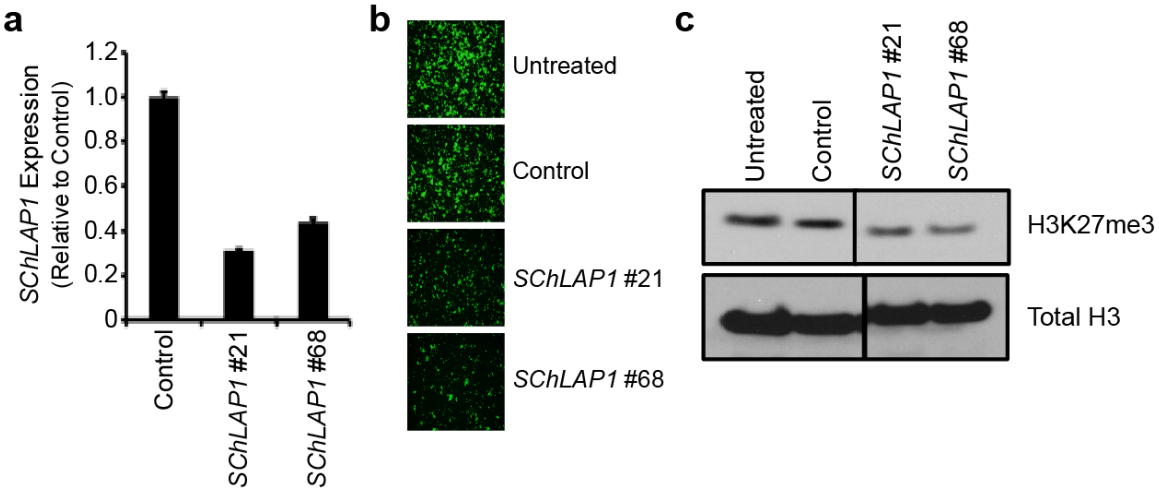


Figure 5.7 The effect of *SChLAP1* ASOs *in vitro*
(a) Knockdown efficiencies for ASO knockdown of *SChLAP1* in 22Rv1 cells. Error bars indicate S.E.M. **(b)** ASO knockdown of *SChLAP1 in vitro* in 22Rv1 prostate cell lines impairs cellular invasion through Matrigel in a Boyden chamber assay. **(c)** ASO knockdown of *SChLAP1* in 22Rv1 cells followed by immunoblotting for H3K27me3 and total H3.

References

1. Wolf, A.M.*, et al.* American Cancer Society guideline for the early detection of prostate cancer: update 2010. *CA Cancer J Clin* **60**, 70-98 (2010).
2. Etzioni, R., Cha, R., Feuer, E.J. & Davidov, O. Asymptomatic incidence and duration of prostate cancer. *American journal of epidemiology* **148**, 775-785 (1998).
3. Makarov, D.V.*, et al.* The natural history of men treated with deferred androgen deprivation therapy in whom metastatic prostate cancer developed following radical prostatectomy. *J Urol* **179**, 156-161; discussion 161-152 (2008).
4. Makarov, D.V., Loeb, S., Getzenberg, R.H. & Partin, A.W. Biomarkers for prostate cancer. *Annu Rev Med* **60**, 139-151 (2009).
5. Prensner, J.R., Rubin, M.A., Wei, J.T. & Chinnaiyan, A.M. Beyond PSA: the next generation of prostate cancer biomarkers. *Science translational medicine* **4**, 127rv123 (2012).
6. Pound, C.R.*, et al.* Natural history of progression after PSA elevation following radical prostatectomy. *JAMA* **281**, 1591-1597 (1999).
7. Epstein, J.I. An update of the Gleason grading system. *J Urol* **183**, 433-440 (2010).
8. Trock, B.J.*, et al.* Tertiary Gleason patterns and biochemical recurrence after prostatectomy: proposal for a modified Gleason scoring system. *J Urol* **182**, 1364-1370 (2009).
9. Pierorazio, P.M., Walsh, P.C., Partin, A.W. & Epstein, J.I. Prognostic Gleason grade grouping: data based on the modified Gleason scoring system. *BJU Int* **111**, 753-760 (2013).
10. Eggener, S.E.*, et al.* Predicting 15-year prostate cancer specific mortality after radical prostatectomy. *J Urol* **185**, 869-875 (2011).
11. Epstein, J.I., Feng, Z., Trock, B.J. & Pierorazio, P.M. Upgrading and downgrading of prostate cancer from biopsy to radical prostatectomy: incidence and predictive factors using the modified Gleason grading system and factoring in tertiary grades. *Eur Urol* **61**, 1019-1024 (2012).
12. Mohler, J.L. The 2010 NCCN clinical practice guidelines in oncology on prostate cancer. *J Natl Compr Canc Netw* **8**, 145 (2010).
13. Carter, H.B. Management of low (favourable)-risk prostate cancer. *BJU Int* **108**, 1684-1695 (2011).
14. D'Amico, A.V., Chen, M.H., Roehl, K.A. & Catalona, W.J. Identifying patients at risk for significant versus clinically insignificant postoperative prostate-specific antigen failure. *J Clin Oncol* **23**, 4975-4979 (2005).
15. Freedland, S.J.*, et al.* Risk of prostate cancer-specific mortality following biochemical recurrence after radical prostatectomy. *JAMA* **294**, 433-439 (2005).
16. Birney, E.*, et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816 (2007).

17. Kapranov, P.*, et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484-1488 (2007).
18. Guttman, M.*, et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223-227 (2009).
19. Prensner, J.R. & Chinnaiyan, A.M. The emergence of lncRNAs in cancer biology. *Cancer Discov* **1**, 391-407 (2011).
20. Prensner, J.R.*, et al.* Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol* **29**, 742-749 (2011).
21. Du, Z.*, et al.* Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nat Struct Mol Biol* **20**, 908-913 (2013).
22. Prensner, J.R.*, et al.* The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nature genetics* **45**, 1392-1398 (2013).
23. Wan, W.B.*, et al.* Synthesis, biophysical properties and biological activity of second generation antisense oligonucleotides containing chiral phosphorothioate linkages. *Nucleic acids research* **42**, 13456-13468 (2014).
24. Pepe, M.S., Feng, Z., Janes, H., Bossuyt, P.M. & Potter, J.D. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J Natl Cancer Inst* **100**, 1432-1438 (2008).
25. Taylor, J.M., Ankerst, D.P. & Andridge, R.R. Validation of biomarker-based risk prediction models. *Clin Cancer Res* **14**, 5977-5983 (2008).
26. Halabi, S.*, et al.* Prognostic model for predicting survival in men with hormone-refractory metastatic prostate cancer. *Journal of Clinical Oncology* **21**, 1232-1237 (2003).
27. Smaletz, O.*, et al.* Nomogram for overall survival of patients with progressive metastatic prostate cancer after castration. *Journal of Clinical Oncology* **20**, 3972-3982 (2002).
28. Basch, E.*, et al.* Abiraterone acetate plus prednisone versus prednisone alone in chemotherapy-naive men with metastatic castration-resistant prostate cancer: patient-reported outcome results of a randomised phase 3 trial. *Lancet Oncol* **14**, 1193-1199 (2013).
29. Ryan, C.J.*, et al.* Abiraterone in metastatic prostate cancer without previous chemotherapy. *New England Journal of Medicine* **368**, 138-148 (2013).
30. Scher, H.I.*, et al.* Increased survival with enzalutamide in prostate cancer after chemotherapy. *New England Journal of Medicine* **367**, 1187-1197 (2012).
31. Kantoff, P.W.*, et al.* Sipuleucel-T immunotherapy for castration-resistant prostate cancer. *New England Journal of Medicine* **363**, 411-422 (2010).

32. de Bono, J.S.*, et al.* Abiraterone and increased survival in metastatic prostate cancer. *New England Journal of Medicine* **364**, 1995-2005 (2011).
33. Catalona, W.J. & Smith, D.S. Cancer recurrence and survival rates after anatomic radical retropubic prostatectomy for prostate cancer: intermediate-term results. *J Urol* **160**, 2428-2434 (1998).
34. Boormans, J.L.*, et al.* Identification of TDRD1 as a direct target gene of ERG in primary prostate cancer. *Int J Cancer* **133**, 335-345 (2013).
35. Laxman, B.*, et al.* A first-generation multiplex biomarker analysis of urine for the early detection of prostate cancer. *Cancer Res* **68**, 645-649 (2008).
36. Tomlins, S.A.*, et al.* Urine TMPRSS2:ERG Fusion Transcript Stratifies Prostate Cancer Risk in Men with Elevated Serum PSA. *Science Translational Medicine* **3**, 94ra72 (2011).
37. Bennett, C.F. & Swayze, E.E. RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. *Annual review of pharmacology and toxicology* **50**, 259-293 (2010).
38. Meng, L.*, et al.* Towards a therapy for Angelman syndrome by targeting a long non-coding RNA. *Nature* **518**, 409-412 (2015).
39. Buller, H.R.*, et al.* Factor XI antisense oligonucleotide for prevention of venous thrombosis. *The New England journal of medicine* **372**, 232-240 (2015).
40. Noveck, R.*, et al.* Effects of an antisense oligonucleotide inhibitor of C-reactive protein synthesis on the endotoxin challenge response in healthy human male volunteers. *Journal of the American Heart Association* **3**(2014).
41. Gaudet, D.*, et al.* Targeting APOC3 in the familial chylomicronemia syndrome. *The New England journal of medicine* **371**, 2200-2206 (2014).
42. Zhang, Y.*, et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137 (2008).
43. Fey, R.A.*, et al.* Local and systemic tolerability of a 2'O-methoxyethyl antisense oligonucleotide targeting interleukin-4 receptor-alpha delivery by inhalation in mouse and monkey. *Inhalation toxicology* **26**, 452-463 (2014).
44. Geary, R.S., Norris, D., Yu, R. & Bennett, C.F. Pharmacokinetics, biodistribution and cell uptake of antisense oligonucleotides. *Advanced drug delivery reviews* (2015).
45. Rittenhouse, H., Blase, A., Shamel, B., Schalken, J. & Groskopf, J. The long and winding road to FDA approval of a novel prostate cancer test: our story. *Clinical Chemistry* **59**, 32-34 (2013).
46. Nakagawa, T.*, et al.* A tissue biomarker panel predicting systemic progression after PSA recurrence post-definitive prostate cancer therapy. *PLoS One* **3**, e2318 (2008).
47. Karnes, R.J.*, et al.* Validation of a genomic classifier that predicts metastasis following radical prostatectomy in an at risk patient population. *J Urol* **190**, 2047-2053 (2013).

48. Vandenbroucke, J.P.*, et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med* **4**, e297 (2007).
49. Sharp, S.J., Poulaliou, M., Thompson, S.G., White, I.R. & Wood, A.M. A review of published analyses of case-cohort studies and recommendations for future reporting. *PLoS One* **9**, e101176 (2014).
50. Rubin, M.A.*, et al.* Rapid ("warm") autopsy study for procurement of metastatic prostate cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **6**, 1038-1045 (2000).
51. Shah, R.B.*, et al.* Androgen-independent prostate cancer is a heterogeneous group of diseases: lessons from a rapid autopsy program. *Cancer Res* **64**, 9209-9216 (2004).
52. Laxman, B.*, et al.* Noninvasive detection of TMPRSS2:ERG fusion transcripts in the urine of men with prostate cancer. *Neoplasia* **8**, 885-888 (2006).
53. Erho, N.*, et al.* Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. *PLoS One* **8**, e66855 (2013).
54. Lin, D.Y. & Ying, Z. Cox regression with incomplete covariate measurements. *Journal of the American Statistical Association* **88**, 1341-1349 (1993).
55. Pepe, M.S., Kerr, K.F., Longton, G. & Wang, Z. Testing for improvement in prediction model performance. *Stat Med* **32**, 1467-1482 (2013).

Chapter 6:

Future directions for investigation

Summary of work

In this thesis, we have discovered *SChLAP1*, a highly prognostic lncRNA that is abundantly expressed in 15-30% of prostate cancers and aided the discrimination of aggressive from indolent forms of this disease. Mechanistically, we find that *SChLAP1* coordinates cancer cell invasion *in vitro* and metastatic spread *in vivo*. Moreover, we characterize an antagonistic *SChLAP1*-SWI/SNF axis in which *SChLAP1* impairs SNF5-mediated gene expression regulation and genomic binding (**Fig. 6.1, upper left**).

Furthermore, we have shown that SWI/SNF mutations are associated with low *SChLAP1* expression in prostate cancer, suggesting that high *SChLAP1* expression may represent a mutation-independent, but clinically equivalent, modality of SWI/SNF inhibition. We establish an antagonistic relationship between SWI/SNF and PRC2 in prostate cells and find that *SChLAP1* enhances PRC2 function in part by inhibiting SWI/SNF activity. Additionally, *SChLAP1* expression is one factor that affects EPZ-6438 sensitivity and may be useful in

identifying prostate cancer patients that are more likely to respond to pharmacologic EZH2 inhibition (**Fig. 6.1, bottom right**).

Further characterization of the *SChLAP1* – SWI/SNF interaction revealed a 250bp region near the 3'-end of *SChLAP1* that mediates its invasive phenotype and coordinates its interaction with the SWI/SNF complex. Additionally, we found that *SChLAP1* interacts with BRG1-, but not BRM-, containing SWI/SNF complexes and preferentially decreases BRG1 genomic binding. Moreover, *SChLAP1*'s preference for BRG1 may expose BRM as a therapeutic target in prostate cancer (**Fig. 6.1, bottom left**).

Finally, we perform the largest biomarker discovery project to date in prostate cancer, finding that *SChLAP1* is one of the best genes for predicting metastatic progression. We characterize *SChLAP1* expression by ISH to confirm that *SChLAP1* is predominantly a nuclear RNA transcript and find that *SChLAP1* expression is enriched in metastatic samples as well as in tumors with high Gleason scores (GS≥8) compared to tumors with lower Gleason scores. As a non-invasive biomarker, we show proof-of-principle data that *SChLAP1* can be detected in patient urine samples. Lastly, we explore the effects of *SChLAP1* ASOs and find effective down-regulation of *SChLAP1* expression, reduced PRC2 activity, and decreased cell invasion, suggesting that direct *SChLAP1* targeting may be an effective therapeutic strategy in prostate cancer (**Fig. 6.1, upper right**).

While we believe this work defines an essential role for *SChLAP1* in aggressive prostate cancer, uncovers novel aspects of lncRNA function, and has broad implications for cancer biology, this study has several limitations that warrant further investigation.

Unexplored areas of study

Characterization of the SChLAP1 – SWI/SNF interaction

Our results describe a 250bp region near the 3'-end of *SChLAP1* that mediates its invasive phenotype as well as interaction with the SWI/SNF complex. It remains unclear whether the 250bp segment alone is sufficient to promote cell invasion and also facilitate an interaction with SWI/SNF. Additionally, it would be useful to determine whether an even smaller region of the *SChLAP1* transcript can recapitulate its full-length function. Several methods have been developed over the last few years to reveal protein-binding sites on RNA transcripts at nucleotide resolution, including PAR-CLIP[1,2], HITS-CLIP[3], including iCLIP[4]. These techniques utilize various cross-linking methods followed by one of several chemistries to modify each protein-bound nucleotide. Sequencing analysis can then be used to identify exactly which nucleotides are participating in an RNA-protein interaction. Uncovering the specific region(s) and nucleotide(s) essential to *SChLAP1* function may reveal areas of the transcript most suitable for direct therapeutic targeting with antisense technology as well as the exact segment that should be utilized for downstream RNA-protein structural analysis.

We have also shown that *SChLAP1* preferentially interacts with BRG1-containing SWI/SNF complexes, resulting in decreased BRG1 genomic binding. Our observations were particularly exciting in light of a recently published manuscript describing a cardioprotective lncRNA termed *Mhrt* that binds to BRG1 to also prevent its interaction with chromatinized DNA[5]. In the paper, the authors assessed the *Mhrt*-BRG1 interaction using *in vitro* biochemical binding assays to show that *Mhrt* directly binds to the BRG1 helicase domain. While it was encouraging to see a shared mechanism of lncRNA function in an independent disease system, the paper revealed several avenues for further interrogation of the *SChLAP1* – SWI/SNF interaction.

First, as our interaction studies were performed using cell-based assays, it is unclear whether *SChLAP1* directly or indirectly binds to SWI/SNF. Additionally, while our results suggest a specificity of *SChLAP1* for BRG1-containing complexes, we are yet to determine if other proteins are involved in this interaction. Based on our findings and the recent manuscript by Han *et al.*, *we* hypothesize that *SChLAP1* binds directly to the SWI/SNF complex, specifically engaging the BRG1 helicase domain. *In vitro* biochemical RNA-protein binding assays are currently underway to test this hypothesis as well as assess if other SWI/SNF proteins are involved in this relationship. These findings may identify novel targets for the development of small molecule inhibitors that could disrupt this interaction. Furthermore, a precise understanding of the RNA regions and

protein domains involved in the *SChLAP1* – SWI/SNF interaction will be helpful for subsequent structural analyses.

Effect of SChLAP1 on SWI/SNF enzymatic function

Our working model suggests that *SChLAP1* functions, in part, by interacting with and abrogating the function of the SWI/SNF complex. *SChLAP1* overexpression leads to decreased SNF5 genome-wide binding. As a nucleosome-remodeling complex, SWI/SNF must engage chromatin to function. Additionally, *SChLAP1*-expressing cells lose SWI/SNF-mediated PRC2 repression[6] and also are more sensitive to EZH2 inhibitors, similar to *SMARCB1*-mutant cells[7]. While this data suggests that *SChLAP1* does indeed inhibit SWI/SNF function, we have not determined if *SChLAP1* affects SWI/SNF enzymatic activity.

One assay to assess SWI/SNF enzymatic activity is micrococcal nuclease sequencing (MNase-seq)[8]. This method uses MNase to digest nucleosome-free regions of DNA, leaving behind a footprint of nucleosome-protected fragments. Sequencing analysis can then be used to reveal nucleosome positions throughout the genome. Dysregulation of SWI/SNF function leads to disorganization of nucleosomes[9-12] and can be detected by MNase-seq[10,13,14]. We performed MNase-seq using RWPE cells with *SChLAP1* overexpression and found that compared to control cells, *SChLAP1*-expressing cells showed differences in nucleosome signal at gene promoters (**Fig. 6.2a**) and also increased fuzziness (a measure of nucleosome phasing and overall organization)

throughout the genome (**Fig. 6.2b**). Although promising, these preliminary results cannot be properly interpreted without comparison to SWI/SNF knockdown as a positive control in our cells. These experiments and subsequent sequencing analysis are currently underway and we hope to soon establish with direct experimental evidence that *SChLAP1* affects SWI/SNF enzymatic activity.

SChLAP1 cellular localization

While cell fractionation assays have shown that *SChLAP1* is predominantly nuclear by qPCR analysis, we do not understand much else about where *SChLAP1* resides within a cell. One technique to directly visualize *SChLAP1* in cells is single-molecule RNA FISH (fluorescence *in situ* hybridization). This method allows simultaneous detection, localization, and quantification of individual RNA molecules at sub-cellular levels in fixed samples using widefield fluorescence microscopy[15]. Pools of antisense complementary oligonucleotide probes with a fluorescent label are used to visualize target transcripts. We collaborated with Biosearch Technologies to perform RNA FISH for *SChLAP1* in LNCaP prostate cancer cells and found that *SChLAP1* is predominantly nuclear, with strong signal coming from three nuclear clouds in each cell (**Fig. 6.3**). While consistent with our previous observations, there are many questions that remain to be answered. Because LNCaP cells have three copies of chromosome 2 (where the *SChLAP1* genomic locus is located), it is likely that these clouds represent bursts of active transcription. The proper experiment with intron- and exon-specific probes to distinguish nascent versus mature RNA needs to be

performed. Additionally, RNA-FISH for *SChLAP1* followed by DNA-FISH for the *SChLAP1* genomic locus will be able to confirm whether these nuclear clouds represent areas of active transcription.

Furthermore, due to the multiplexing capabilities and technical flexibility of RNA FISH, it is possible to simultaneously perform protein immunofluorescence to measure RNA and protein co-localization. For our purposes, it will be useful to perform *SChLAP1* RNA FISH followed by SNF5 immunofluorescence to determine whether *SChLAP1* and SWI/SNF do indeed co-localize in cells. Additionally, we will be able to assess if *SChLAP1* sequesters SWI/SNF to a specific nuclear compartment. Also, these results will show what proportion of total *SChLAP1* engages SWI/SNF and whether determining an alternative mechanism of action for unbound *SChLAP1* transcripts is necessary. Finally, we could perform immunofluorescence with antibodies targeting various SWI/SNF components to discover and verify *SChLAP1* binding specificity.

SChLAP1 interactome analysis

Our mechanistic model has assessed the interaction of *SChLAP1* with the SWI/SNF complex. We have not yet determined whether *SChLAP1* also interacts with and functions through other molecules to cause aggressive prostate cancer. LncRNAs function in a variety of ways, including interacting with multiple protein partners to act as a scaffold, directly engaging DNA to regulate transcription or alter chromatin conformation, and forming complexes with other RNA transcripts

to form a ribonucleoprotein complex[16,17]. Several methods have been recently developed to determine the interactome of lncRNAs in an unbiased manner[18].

Chromatin isolation by RNA purification (ChIRP)[19] is a method that uses biotinylated antisense oligonucleotide probes targeting an RNA transcript to endogenously pulldown the RNA as well as any associated molecules. Following pulldown, RNA is isolated and analyzed by qPCR to ensure efficient target RNA retrieval and probe specificity. Interacting DNA can be analyzed by sequencing (ChIRP-seq) to determine the RNA's genomic binding sites. Proteins can also be isolated and analyzed by western blot to identify RNA-protein interactions.

We performed this technique for *SChLAP1* in RWPE cells with *SChLAP1* overexpression to confirm the interaction between *SChLAP1* and SWI/SNF (**Fig. 2.14e,f**). Additionally, following *SChLAP1* pulldown we isolated a DNA fraction and performed sequencing analysis to find several *SChLAP1* genomic binding sites (**Fig. 6.4a,b**). MEME analysis of these binding sites uncovered two predominant *SChLAP1* DNA binding motifs (**Fig. 6.4c**)[20]. However, several issues remain unresolved. First, it is unclear whether the ChIRP probes themselves are interacting with DNA. Performing this experiment in a cell line with no *SChLAP1* expression will serve as a control for non-specific probe interactors. Second, determining the *SChLAP1* interactome in a cell line with endogenous *SChLAP1* expression rather than in an overexpression cell line model may provide more biologically relevant results. These experiments are

currently underway and sequencing analysis will provide new insights into *SChLAP1* biology.

Additionally, identification of novel protein partners may reveal alternative pathways of *SChLAP1* function. While ChIRP is suitable for protein analysis by western blot, cross-linking methods used in the protocol make it unsuitable for unbiased mass spectrometry analysis. Other methods have been developed to overcome this limitation. RNA antisense purification (RAP) uses longer (~100-150bp) antisense oligonucleotide probes to retrieve endogenous RNA complexes from cell extracts to map RNA interactions with chromatin and also allows mass spectrometry analysis of the isolated protein fraction. We are exploring this technique as well as others to identify novel *SChLAP1* protein partners in an unbiased manner.

Finally, hybrid endogenous and *in vitro* methods can be used to identify RNA-bound proteins. In one technique called Ribotrap (MBL International), the RNA of interest is *in vitro* synthesized and labeled with 5-bromo-UTP (BrUTP) followed by incubation with a cell lysate. After incubation, anti-BrdU monoclonal antibody is used to retrieve the RNA-protein complexes. RNA-bound proteins can then be identified by immunoblotting or mass spectrometry. We have performed Ribotrap for *SChLAP1* with LNCaP cell lysate and analyzed the protein fraction by silver stain and western blot (**Fig. 6.5**). *LacZ* RNA served as a negative control. This

experiment is currently being repeated and biological replicates will be submitted for mass spectrometry analysis.

SChLAP1 regulation

While *SChLAP1* is highly expressed in a subset of aggressive, lethal prostate cancers, it remains unclear what drives *SChLAP1* expression in cells and tissues. Although androgen signaling plays a major role in mediating prostate cancer development and progression, *SChLAP1* is not regulated by androgen (**Fig. 6.6a**). It is unlikely that a single transcription factor could be driving *SChLAP1* expression, as the entire *SChLAP1* locus is transcriptionally active (*see below*). A more global mechanism, such as hypomethylation, acetylation, or chromatin architecture, is more likely to influence transcription at the *SChLAP1* locus and account for increased gene expression. Preliminary results from 5-azacytidine (DNA methylation inhibitor) and Trichostatin A (HDAC inhibitor) treatment suggest that broad regulatory mechanisms may regulate *SChLAP1* expression (**Fig. 6.6b**). More complex experiments such as 3C-based methods[21] may help uncover whether chromatin architecture also plays a role in *SChLAP1* regulation. A thorough analysis of the *SChLAP1* promoter and nearby enhancers may also provide a better understanding of the factor(s) responsible for *SChLAP1* expression and also identify novel therapeutic targets for prostate cancer treatment.

Furthermore, *SChLAP1* expression is specific to human prostate cancer, with little to no expression in other types of malignant or benign tissues or across other species. This suggests that prostate-specific factors are responsible for *SChLAP1* regulation and also warrants the investigation of lncRNAs in other tissues and diseases that may mediate SWI/SNF function. Although lncRNAs are less conserved than protein-coding genes, the lack of conservation of *SChLAP1* across species suggests that other lncRNAs may serve a similar purpose in regulating SWI/SNF and mediating cancer progression, as well.

SChLAP1 transcript specific functions

An emerging area of RNA biology is the effect of transcript specific function[16,22]. While our studies have interrogated *SChLAP1* function and mechanism at the gene level, we have not evaluated isoform-specific differences. We also have not looked for other *SChLAP1* variants such as transcripts with SNPs (single nucleotide polymorphisms) or fused, chimeric, or truncated versions of the RNA. These analyses may reveal novel aspects of *SChLAP1* crucial to prostate cancer biology.

Additionally, transcription around the *SChLAP1* locus warrants further investigation. We have observed high transcriptional activity in this area and a comprehensive de novo assembly of cancer-associated lncRNAs[22] identified several unannotated transcripts in this region (**Fig. 6.7a**). Furthermore, there is evidence of an antisense transcript originating from the *SChLAP1* transcription

start site, suggesting that it is a bidirectional promoter (**Fig. 6.7b**)[23,24]. These transcripts may also play a crucial role in aggressive prostate cancer and open up new translational opportunities.

Development of SChLAP1 as a biomarker and therapeutic target

While our results have defined *SChLAP1* as one of the most powerful prognostic predictors for aggressive and metastatic prostate cancer, our work has limitations. First, we perform extensive retrospective studies, but not prospective studies. We also have not evaluated *SChLAP1* in the context of androgen-deprivation therapy or radiotherapy. For the *SChLAP1* ISH studies, we analyzed a relatively small number of patients in a single cohort from a one institution. Larger, multi-cohort evaluations will be needed to confirm our findings. Also, further study regarding the relationship between *SChLAP1*, serum PSA, and Gleason score will be needed to establish more specific implementation of *SChLAP1* ISH as an assay in the clinical decision-making algorithm for routine patient care. In addition, our urine PCR assay for *SChLAP1* is a preliminary analysis which does not qualify as a clinical-grade test according to established criteria[25]. Finally, our *SChLAP1* ASO studies were performed in cell line models with cationic lipid-mediated delivery. We need to assess the ability of these ASOs to perform similarly by free-uptake cellular delivery and also show efficacy using *in vivo* experimental models.

Concluding remarks

We live in a very exciting era of science and medicine. New tools have enabled the discovery of novel molecular entities and allowed high-resolution views of cellular activity. Improvements in next-generation sequencing technology and analysis have allowed the implementation of personalized medicine in the clinic. Although prostate cancer remains a devastating disease, advances in prognostic markers and therapeutic agents have greatly improved patient outlook over the last few years. Less than a decade ago, lncRNAs were mostly ignored, often attributed to leaky transcription or sequencing artifacts. Now, entire labs are devoted to the study of lncRNAs, and new techniques are being developed specifically to interrogate lncRNA biology. While *SChLAP1* has proven to be an essential player in aggressive prostate cancer, enhanced bioinformatics analyses have identified thousands of cancer-associated lncRNAs that are likely to play equally important roles across all cancers. Given the tissue- and disease-specific nature of these transcripts, their abundance throughout the genome, and the relatively recent discovery of the majority of these transcripts, it is likely that lncRNAs hold the answer to scientific and medical mysteries that have eluded us for decades. I look forward to what the future holds.

Figures

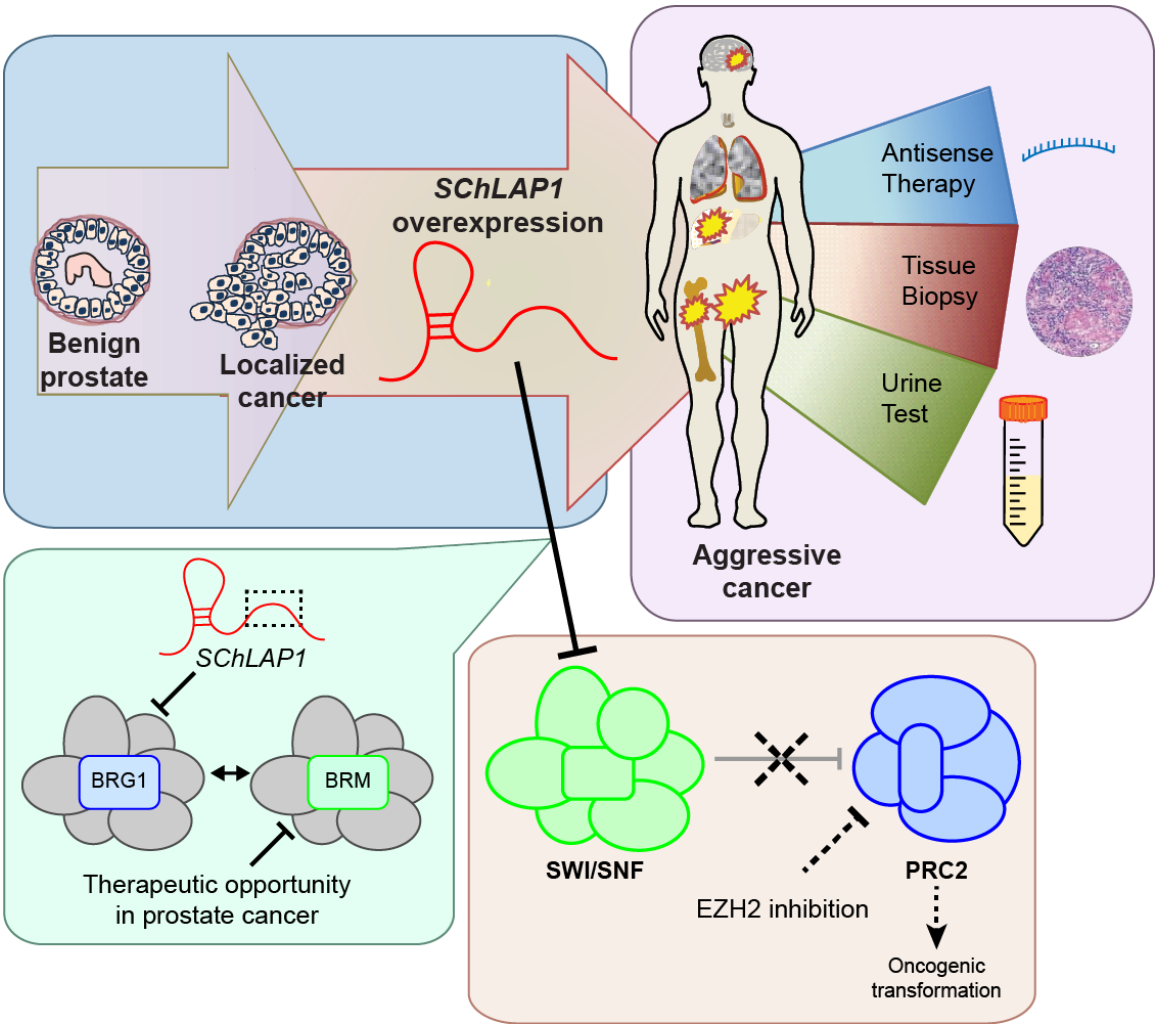


Figure 6.1 Summary of thesis

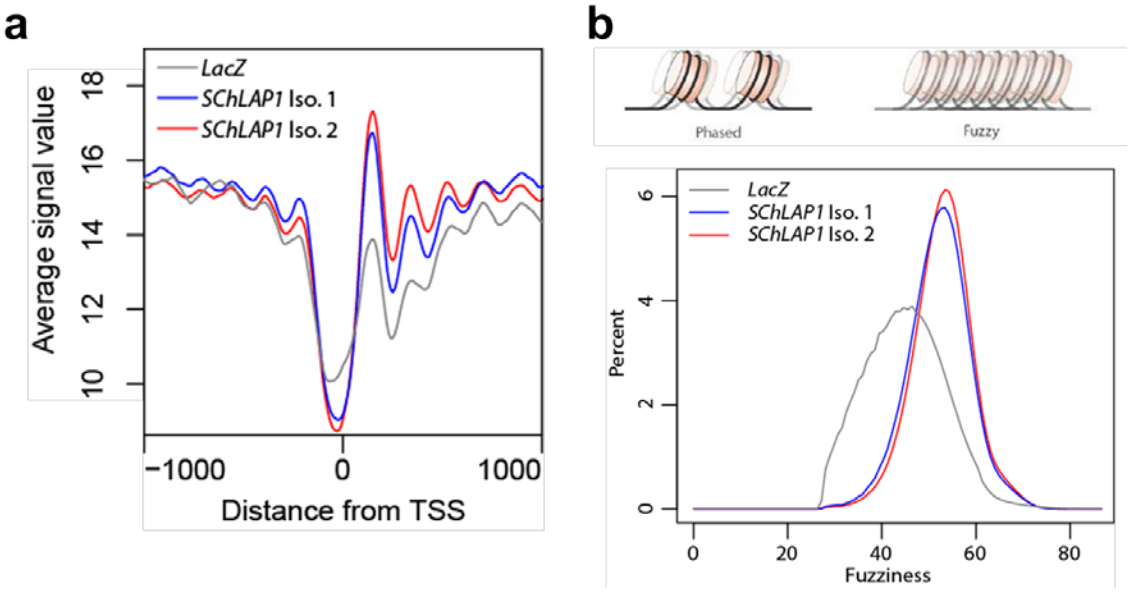


Figure 6.2 MNase-seq shows changes in nucleosome organization in cells with *SChLAP1* overexpression
MNase-seq was performed using RWPE-*LacZ*, RWPE-*SChLAP1* isoform 1, and RWPE-*SChLAP1* isoform 2 cells. Average signal at gene promoters **(a)** and global levels of genome-wide fuzziness **(b)** are shown.

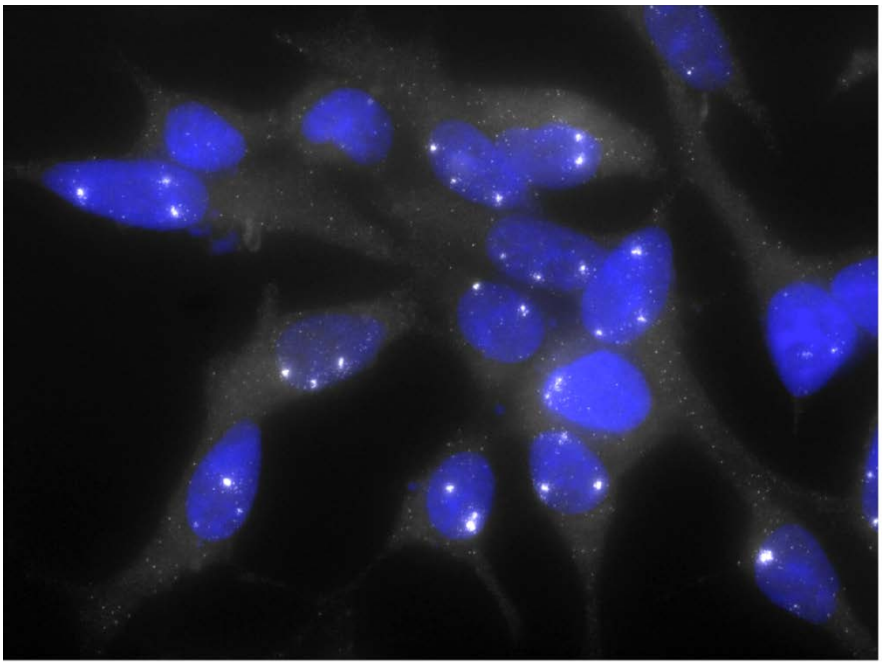


Figure 6.3 *SChLAP1* RNA FISH
Single-molecule RNA FISH (Biosearch) was performed for *SChLAP1* in LNCaP cells using a pool of singly labeled oligonucleotide probes designed to hybridize along the *SChLAP1* transcript. Blue staining (DAPI) represents cell nucleus.

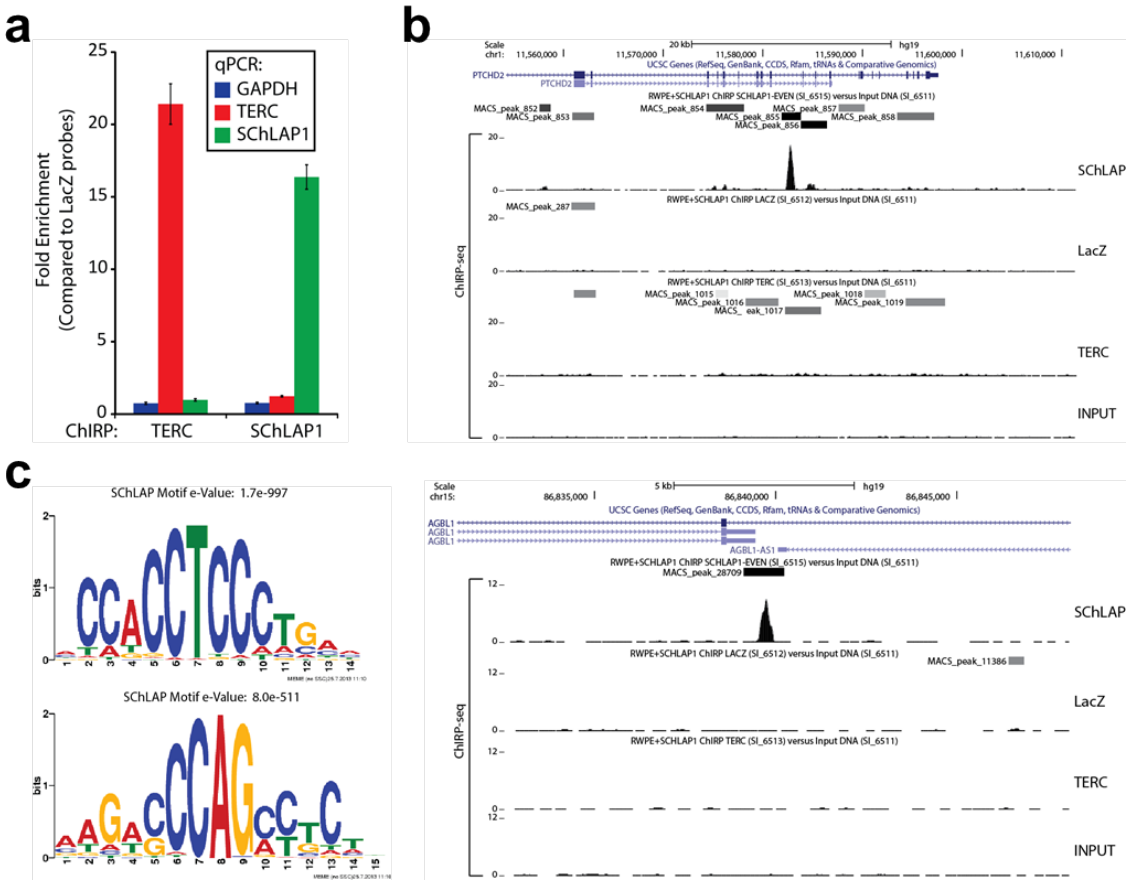


Figure 6.4 ChIRP-seq reveals *SChLAP1* genomic binding sites and a putative DNA-binding motif

(a) Pulldown of *SChLAP1* RNA. RWPE-*SChLAP1* isoform #1 cells were treated with biotinylated *SChLAP1*, *TERC* or *LacZ* RNA probes according to the ChIRP protocol. Quantification of RNA pulldown efficiency by qPCR is shown. Error bars indicate S.E.M. **(b)** Example *SChLAP1* genomic binding sites. **(c)** MEME motif analysis identified two *SChLAP1* DNA-binding motifs.

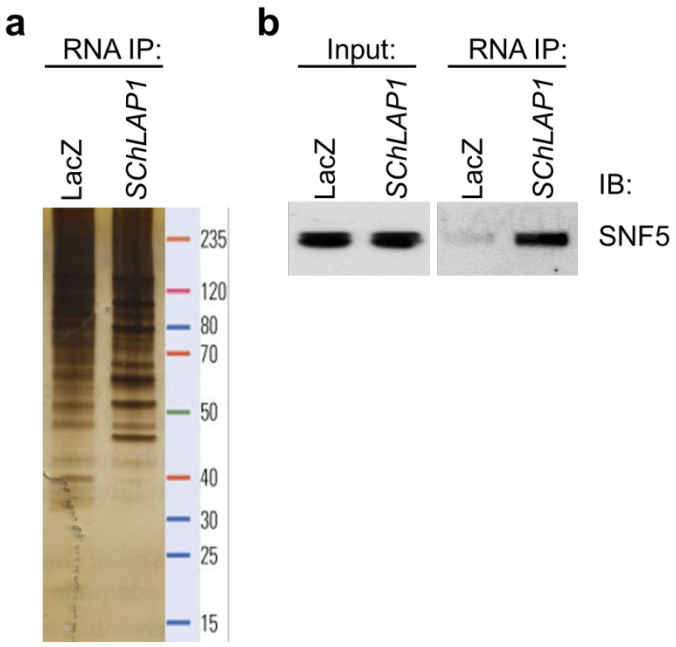


Figure 6.5 Ribotrap for *SChLAP1*
Ribotrap was performed using BrU-labeled *SChLAP1* incubated with LNCaP prostate cancer cell line nuclear lysate followed by silver staining **(a)** and western blot for SNF5 **(b)**.

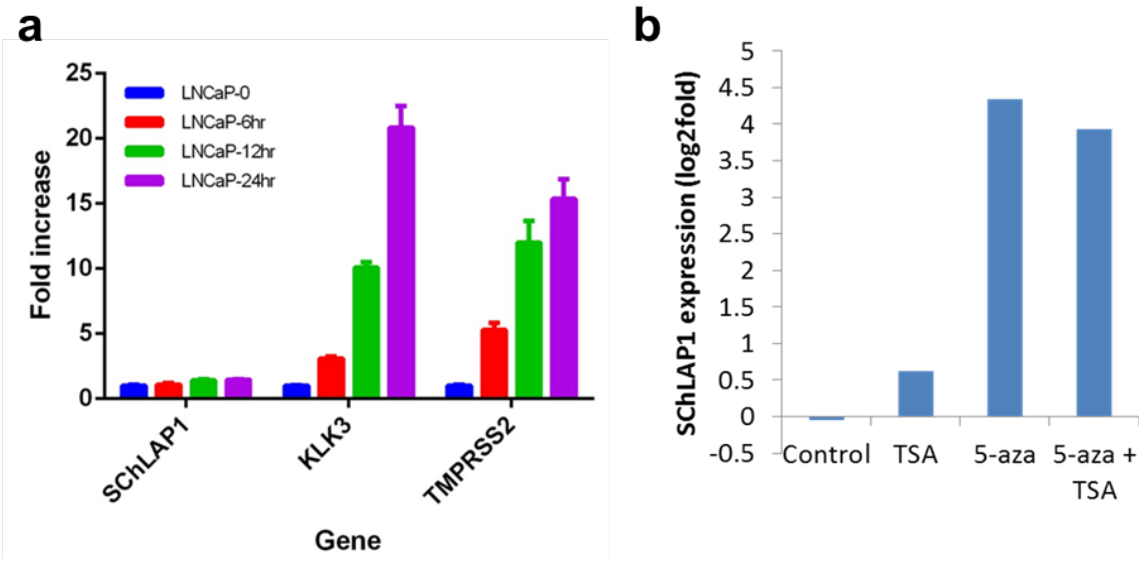


Figure 6.6 *SChLAP1* regulation
(a) Expression of *SChLAP1* in LNCaP prostate cancer cells after DHT (dihydrotestosterone) treatment for 6, 12, and 24 hours. *KLK3* and *TMPRSS2* are androgen-regulated genes that serve as positive controls. **(b)** Expression of *SChLAP1* in PrEC prostate epithelial cells after treatment with Trichostatin A (TSA), 5-azacytidine (5-aza), or both.

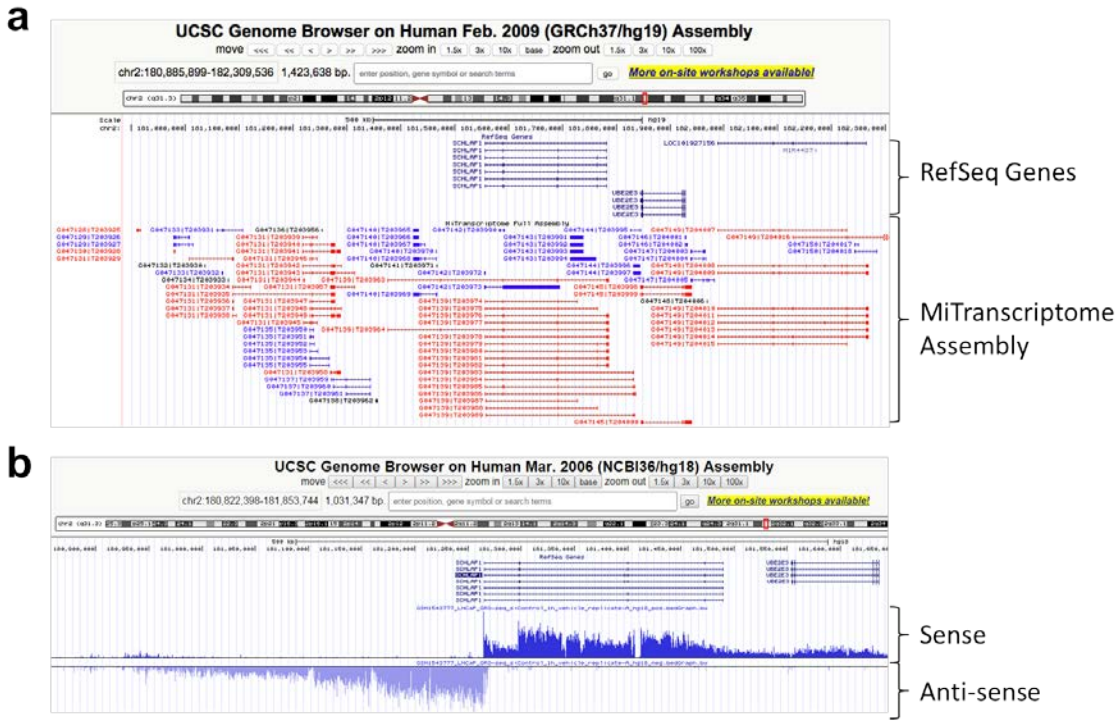


Figure 6.7 Visualization of the *SChLAP1* genomic locus
(a) The UCSC genome browser was used to visualize transcripts from the MiTranscriptome assembly near the *SChLAP1* genomic locus. Top track shows annotated genes from RefSeq. Bottom track shows the myriad of unannotated lncRNA transcripts around *SChLAP1*. **(b)** Strand-specific GRO-seq was performed in LNCaP cells and shows active sense and anti-sense transcription from the *SChLAP1* promoter.

References

1. Hafner, M.*, et al.* PAR-CliP--a method to identify transcriptome-wide the binding sites of RNA binding proteins. *Journal of visualized experiments : JoVE* (2010).
2. Hafner, M.*, et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-141 (2010).
3. Chi, S.W., Zang, J.B., Mele, A. & Darnell, R.B. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460**, 479-486 (2009).
4. Konig, J.*, et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nature structural & molecular biology* **17**, 909-915 (2010).
5. Han, P.*, et al.* A long noncoding RNA protects the heart from pathological hypertrophy. *Nature* **514**, 102-106 (2014).
6. Wilson, B.G.*, et al.* Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. *Cancer cell* **18**, 316-328 (2010).
7. Knutson, S.K.*, et al.* Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 7922-7927 (2013).
8. Cui, K. & Zhao, K. Genome-wide approaches to determining nucleosome occupancy in metazoans using MNase-Seq. *Methods in molecular biology* **833**, 413-419 (2012).
9. Lu, P. & Roberts, C.W. The SWI/SNF tumor suppressor complex: Regulation of promoter nucleosomes and beyond. *Nucleus* **4**, 374-378 (2013).
10. Tolstorukov, M.Y.*, et al.* Swi/Snf chromatin remodeling/tumor suppressor complex establishes nucleosome occupancy at target promoters. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 10165-10170 (2013).
11. Wilson, B.G. & Roberts, C.W. SWI/SNF nucleosome remodellers and cancer. *Nature reviews. Cancer* **11**, 481-492 (2011).
12. Reisman, D., Glaros, S. & Thompson, E.A. The SWI/SNF complex and cancer. *Oncogene* **28**, 1653-1668 (2009).
13. Hu, G.*, et al.* Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1. *Genome research* **21**, 1650-1658 (2011).
14. Jaskelioff, M., Gavin, I.M., Peterson, C.L. & Logie, C. SWI-SNF-mediated nucleosome remodeling: role of histone octamer mobility in the persistence of the remodeled state. *Molecular and cellular biology* **20**, 3058-3068 (2000).
15. Batish, M., Raj, A. & Tyagi, S. Single molecule imaging of RNA in situ. *Methods in molecular biology* **714**, 3-13 (2011).
16. Rinn, J.L. & Chang, H.Y. Genome regulation by long noncoding RNAs. *Annual review of biochemistry* **81**, 145-166 (2012).

17. Wang, K.C. & Chang, H.Y. Molecular mechanisms of long noncoding RNAs. *Molecular cell* **43**, 904-914 (2011).
18. Chu, C., Spitale, R.C. & Chang, H.Y. Technologies to probe functions and mechanisms of long noncoding RNAs. *Nature structural & molecular biology* **22**, 29-35 (2015).
19. Chu, C., Qu, K., Zhong, F.L., Artandi, S.E. & Chang, H.Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Molecular cell* **44**, 667-678 (2011).
20. Bailey, T.L.*, et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic acids research* **37**, W202-208 (2009).
21. de Wit, E. & de Laat, W. A decade of 3C technologies: insights into nuclear organization. *Genes & development* **26**, 11-24 (2012).
22. Iyer, M.K.*, et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nature genetics* (2015).
23. Sigova, A.A.*, et al.* Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 2876-2881 (2013).
24. Puc, J.*, et al.* Ligand-dependent enhancer activation regulated by topoisomerase-I activity. *Cell* **160**, 367-380 (2015).
25. Bustin, S.A.*, et al.* The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* **55**, 611-622 (2009).

Appendix: Author Contributions

Chapter 1

This chapter was written by Anirban Sahu. All figures were reprinted with permission from the following sources:

Figure 1.1: Hoffman, RM. Screening for Prostate Cancer. *NEJM*, 365:2013-2019, 2011.

Figure 1.2: Nelson, WG, De Marzo, AM, Isaacs, WB. Prostate Cancer. *NEJM*, 349:366-381, 2003.

Figure 1.3: Humphrey, PA. Gleason grading and prognostic factors in carcinoma of the prostate. *Mod Path*, 17(3):292-306, 2004.

Figure 1.4: Rinn, JL and Chang, HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem*, 81:145-66, 2012.

Figure 1.5: Prensner, JR, Iyer, MK, *et al.*, Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression.

Chapter 2

This chapter was published in *Nature Genetics*, November 2013, along with the following co-authors: John R. Prensner, Matthew K. Iyer, Irfan A. Asangani, Qi Cao, Lalit Patel, Ismael A. Vergara, Elai Davicioni, Nicholas Erho, Mercedeh Ghadessi, Robert B. Jenkins, Timothy J. Triche, Rohit Malik, Rachel Bedenis,

Natalie McGregor, Teng Ma, Wei Chen, Sumin Han, Xiaojun Jing, Xuhong Cao, Xiaoju Wang, Benjamin Chandler, Wei Yan, Javed Siddiqui, Lakshmi P. Kunju, Saravana M. Dhanasekaran, Kenneth J. Pienta, Felix Y. Feng & Arul M. Chinnaiyan. The article and supplementary information are available online at: http://www.nature.com/ng/journal/v45/n11/full/ng.2771.html.

For this chapter, J.R.P., M.K.I., A.S. and A.M.C. designed the project and directed experimental studies. J.R.P., A.S., Q.C., W.C., S.M.D., B.C., S.H., R.M., L.P., and T.M. performed *in vitro* studies. X.W. performed *in vitro* translation assays. A.S. and I.A.A. performed CAM assays. R.B., N.M. and K.J.P. performed *in vivo* studies. L.P.K. and W.Y. performed histopathological analyses. M.K.I. performed bioinformatics analysis. X.J. and X.C. performed gene expression microarray experiments. J.S. and F.Y.F. facilitated biological sample procurement. F.Y.F. performed clinical analyses. For the Mayo Clinic cohort, R.B.J. provided clinical samples and outcomes data. T.J.T. and E.D. generated and analyzed expression profiles for the Mayo Clinic cohort. E.D., N.E., M.G. and I.A.V. performed statistical analyses of *SChLAP1* expression in the Mayo Clinic cohort. J.R.P., M.K.I., A.S. and A.M.C. interpreted data and wrote the manuscript. All authors discussed results and provided comments. All figures and tables were generated by J.R.P., M.K.I., and A.S.

Chapter 3

This work was performed in collaboration with Matthew K. Iyer, John R. Prensner, Benjamin Chandler, Xuhong Cao, Saravana M. Dhanasekaran, Yashar S. Niknafs, Nithin Edara, Udit Singhal, Shuang G. Zhao, Yi-Mi Wu, Dan R. Robinson, Rohit Malik, Felix Y. Feng, and Arul M. Chinnaiyan. A.S. and A.M.C. designed the project and directed experimental studies. A.S., J.R.P., B.C., S.M.D., N.E., and U.S. performed *in vitro* studies. M.K.I., Y.S.N., and S.G.Z. performed bioinformatics analysis. X.C., Y.M., and D.R.R. performed gene expression microarrays and sequencing. R.M. and F.Y.F. participated in the experimental design. A.S. and A.M.C. interpreted data and wrote the manuscript. All authors discussed results and provided comments. All figures and tables were generated by A.S.

Chapter 4

This work was performed in collaboration with John R. Prensner, Benjamin Chandler, Qi Cao, Nithin Edara, Udit Singhal, Sumin Han, Matthew K. Iyer, Rohit Malik, Felix Y. Feng, and Arul M. Chinnaiyan. A.S., J.R.P. and A.M.C. designed the project and directed experimental studies. A.S., J.R.P., and Q.C. performed deletion construct analysis. S.H. performed RNA stability assays. A.S., B.C., N.E., and U.S. performed BRG1 and BRM experiments. M.K.I., R.M. and F.Y.F. participated in the study design. A.S., J.R.P. and A.M.C. interpreted data and wrote the manuscript. All authors discussed results and provided comments. All figures and tables were generated by A.S. and J.R.P.

Chapter 5

Portions of this chapter were previously published and reproduced with permission from the following manuscripts:

Prensner JR, Zhao S, Erho N, Schipper M, Iyer MK, Dhanasekeran SM, Magi-Galluzzi C, Mehra R, Sahu A, Siddiqui J, Davicioni E, Den RB, Dicker AP, Karnes RJ, Wei JT, Klein EA, Jenkins RB, Chinnaiyan AM, Feng FY. Nomination and validation of *SChLAP1* as an independent risk factor for metastatic prostate cancer progression. *Lancet Oncol,* 15(13):1469-80, 2014.

Mehra R, Shi Y, Udager AM, Prensner JR, Sahu A, Iyer MK, Siddiqui J, Cao X, Wei J, Jiang H, Feng FY, Chinnaiyan AM. A novel RNA in situ hybridization assay for the long noncoding RNA *SChLAP1* predicts poor clinical outcome after radical prostatectomy in clinically localized prostate cancer. *Neoplasia,* 16(12):1121-7, 2014.

For the prognostic studies, J.R.P., S.Z., A.M.C., and F.Y.F. conceived and designed the study. J.R.P., S.Z., N.E., M.K.I., and F.Y.F. developed the methodology. J.R.P., S.Z., N.E., S.M.D., C.M-G., R.M., A.S., A.P.D., R.J.K., J.T.W., and F.Y.F. acquired data. J.R.P., S.Z., M.S., M.K.I., and F.Y.F. analyzed and interpreted data. J.S., E.D., R.B.D., A.P.D., J.T.W., E.A.K., R.B.J., and F.Y.F. provided administrative, technical, or material support. J.R.P., S.Z.,

A.M.C., and F.Y.F. provided study supervision. All authors wrote and provided final approval of the manuscript.

For the *in situ* hybridization studies, R.M. and A.M.C. conceived and designed the study. R.M. developed the methodology. R.M., A.M.U., J.R.P., A.S., M.K.I., and X.C. acquired the data. Y.S. and H.J. performed statistical analyses. F.Y.F. and A.M.C. provided study supervision. All authors wrote and provided final approval of the manuscript.

Antisense oligonucleotide (ASO) experiments were performed by Lanbo Xiao. All figures were generated by J.R.P., S.Z., R.M., and A.S.

Chapter 6

This chapter was written by Anirban Sahu and data was acquired in collaboration with Matthew K. Iyer, Benjamin Chandler, Teng Ma, Biosearch Technologies, Nithin Edara, Rohit Malik, Sudhanshu Shukla, and Yashar S. Niknafs. A.S., M.K.I., B.C., and T.M. performed and analyzed MNase-seq and ChIRP-seq experiments. A.S. and B.T. performed RNA fluorescence *in situ* hybridization assays. A.S. and N.E. performed Ribotrap assay. R.M. and S.S. performed gene regulation experiments. M.K.I. and Y.N. developed the MiTranscriptome portal (www.mitranscriptome.org). GRO-seq data was obtained from Puc, J.*, et al.* Ligand-dependent enhancer activation regulated by topoisomerase-I activity. *Cell*, 160:367-380, 2015. All figures were generated by A.S.