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

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

Prostate cancers remain indolent in the majority of individuals but behave aggressively in a minority1,2. The molecular basis for this clinical heterogeneity remains incompletely understood3–5. Here we characterize a long noncoding RNA termed SChLAP1 (second chromosome locus associated with prostate-1; also called LUNC00913) that is overexpressed in a subset of prostate cancers. SChLAP1 levels independently predict poor outcomes, including metastasis and prostate cancer-specific mortality. In vitro and in vivo gain-of-function and loss-of-function experiments indicate that SChLAP1 is critical for cancer cell invasiveness and metastasis. Mechanistically, SChLAP1 antagonizes the genome-wide localization and regulatory functions of the SWI/SNF chromatin-modifying complex. These results suggest that SChLAP1 contributes to the development of lethal cancer at least in part by antagonizing the tumor-suppressive functions of the SWI/SNF complex.

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

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

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

Of these two lncRNAs, PCAT-114 was expressed at higher levels in prostate cell lines, and, in the PCAT-114 region, we defined a 1.4-kb polyadenylated gene composed of up to seven exons and spanning nearly 200 kb on chromosome 2q31.3 (Fig. 1b and Supplementary Fig. 2a). We named this gene second chromosome locus associated with prostate-1 (SChLAP1) after its genomic location. Published prostate cancer chromatin immunoprecipitation and sequencing (ChIP-seq) data12 confirmed that the transcriptional start site (TSS) of SChLAP1 was marked by tri-methylation of histone H3 at lysine 36 (H3K36me3) (Fig. 2a) and that its gene body harbored trimethylation of histone H3 at lysine 36 (H3K36me3) (Fig. 1b), an epigenetic signature consistent with lncRNAs13. We observed numerous SChLAP1 splicing isoforms, of which three (termed isoforms 1, 2 and 3) constituted the vast majority (>90%) of transcripts in the cell (Supplementary Fig. 2b,c).

Using quantitative PCR (qPCR), we confirmed that SChLAP1 was highly expressed in ~25% of prostate cancers (Fig. 1c). SChLAP1 was found to be expressed more frequently in metastatic compared to localized prostate cancers, and its expression was associated with ETS gene fusions in this cohort but not with other molecular events (Supplementary Fig. 2d,e). A computational analysis of the SChLAP1 sequence suggested no coding potential, which was confirmed

1Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, Michigan, USA. 2Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA. 3Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA. 4GenomeDx Biosciences, Inc., Vancouver, British Columbia, Canada. 5Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA. 6Department of Radiation Oncology, University of Michigan, Ann Arbor, Michigan, USA. 7Department of Pathology, University of Michigan, Ann Arbor, Michigan, USA. 8Comprehensive Cancer Center, University of Michigan, Ann Arbor, Michigan, USA. 9Howard Hughes Medical Institute, University of Michigan, Ann Arbor, Michigan, USA. 10These authors contributed equally to this work. Correspondence should be addressed to A.M.C. (arul@med.umich.edu).

Received 18 February; accepted 30 August; published online 29 September 2013; doi:10.1038/ng.2771
Figure 1  Identification of SChLAP1 as a prostate cancer–associated lncRNA. (a) COPA for intergenic lncRNAs (lncRNAs defined in ref. 11). (b) Representation of the SChLAP1 gene and its annotations in current databases. An aggregated representation of current gene annotations for Ensembl, the Encyclopedia of DNA Elements (ENCODE), the UCSC Genome Browser, RefSeq 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 demonstrate its expression. (c) 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 are normalized to the average of (GAPDH + HMBS) and are represented as standardized expression values. (d) Fractionation of prostate cell lysates demonstrates nuclear expression of SChLAP1. U1 RNA serves as a positive control for nuclear gene expression. Error bars, s.e.m. (e) In situ hybridization of SChLAP1 in human prostate cancer. Histological scores for localized cancer samples are indicated, with the first number representing the score and the second number representing the minor Gleason score. SChLAP1 staining is shown for both localized and metastatic tissues.

experimentally by in vitro translation assays of the three SChLAP1 isoforms (Supplementary Fig. 3). Additionally, we found that SChLAP1 transcripts were located in the nucleus (Fig. 1d). We confirmed the nuclear localization of SChLAP1 transcripts in human samples (Fig. 1e) using an in situ hybridization assay in formalin-fixed, paraffin-embedded prostate cancer samples (Supplementary Fig. 4a,b and Supplementary Note).

An analysis of SChLAP1 expression in localized tumors demonstrated a strong correlation with higher Gleason scores, a histopathological measure of aggressiveness (Supplementary Fig. 4c,d and Supplementary Table 2). Next, we performed a network analysis of prostate cancer microarray data in the Oncomine database using signatures of SChLAP1-correlated or SChLAP1-anticorrelated genes, as SChLAP1 itself is not measured by expression microarrays (Online Methods and Supplementary Table 3a). We found a striking association with enriched concepts related to prostate cancer progression (Fig. 2a and Supplementary Table 3b). For comparison, we next incorporated disease signatures using prostate RNA-seq data and additional known prostate cancer genes, including EZH2 (a metastasis gene) PCA3 (a lncRNA biomarker) and AMACR (a tissue biomarker), as well as ACTB (encoding β-actin) as a control (Supplementary Fig. 5, Supplementary Table 3c–i and Supplementary Note). A heatmap visualization of significant comparisons confirmed a strong association of SChLAP1-correlated genes but not of
PCA3- and AMACR-correlated genes with high-grade and metastatic cancers (Fig. 2b). Kaplan-Meier analysis similarly showed significant associations between the SchLAP1 signature and biochemical recurrence\(^6\) and overall survival\(^11\) (Supplementary Fig. 6a,b).

To directly evaluate the relationship between SchLAP1 levels and clinical outcome, we next used SchLAP1 expression to stratify 235 samples from individuals with localized prostate cancer who underwent radical prostatectomy at the Mayo Clinic\(^19\) (Online Methods and Supplementary Fig. 6c). We evaluated samples for three clinical endpoints: biochemical recurrence, clinical progression to systemic disease and prostate cancer–specific mortality (Supplementary Table 4). At the time of this analysis, subjects had a median follow-up time of 8.1 years.

SchLAP1 was a powerful single-gene predictor of aggressive prostate cancer (Fig. 2c-e). SchLAP1 expression was highly significant when distinguishing disease with clinical progression and prostate cancer–specific mortality (\(P = 0.00005\) and 0.002, respectively; Fig. 2d,e). For the biochemical recurrence endpoint, high SchLAP1 expression was associated with a shorter median time to progression (1.9 versus 5.5 years for individuals with high and low expression of SchLAP1, respectively; Fig. 2c). We further confirmed this association with rapid biochemical recurrence using an independent cohort (Supplementary Fig. 6d). Multivariate and univariate regression analyses of the Mayo Clinic data demonstrated that SchLAP1 expression is an independent predictor of prostate cancer aggressiveness, with highly significant hazard ratios for predicting biochemical recurrence, clinical progression and prostate cancer–specific mortality (hazard ratios of 3.045, 3.563 and 4.339, respectively; \(P < 0.01\)), which are comparable to those for other clinical factors such as advanced clinical stage and Gleason histopathological score (Supplementary Fig. 7 and Supplementary Note).

To explore the functional role of SchLAP1, we performed small interfering RNA (siRNA)-mediated knockdowns to compare the impact of SchLAP1 depletion to that of EZH2, which is essential for cancer cell aggressiveness\(^15\). Notably, knockdown of SchLAP1 dramatically impaired cell invasion and proliferation in vitro to
Figure 3  SChLAP1 coordinates cancer cell invasion in vitro and metastatic seeding in vivo. (a) siRNA knockdown of SChLAP1 in vitro in three prostate cell lines (LNCaP, 22Rv1, Du145) impairs cellular invasion through Matrigel in a Boyden chamber assay. EZH2 siRNA serves as a positive control for impaired invasion. Images to the right show representative Boyden chamber membranes stained with crystal violet following invasion. All images were captured at the same magnification, as indicated by the scale bars. (b) Overexpression of SChLAP1 in RWPE cells results in increased cellular invasion through Matrigel. Images to the right show Boyden chambers as in a in indicated RWPE cell lines; scale bars represent equal magnification of images. (c) Intracardiac injection of severe combined immunodeficient (SCID) mice with 22Rv1 cells with stable SChLAP1 knockdown. Representative Luciferase bioluminescence images (BLI) are shown for mice 5 weeks after intracardiac injection with 22Rv1 cells expressing non-targeting shRNA 1 (n = 14) or SChLAP1 shRNA 2 (n = 14). Mouse IDs are given above each image. (d) Relative intensity of whole-mouse luciferase signal is plotted for mice with intracardiac injection of 22Rv1 cells expressing control, non-targeting shRNA (n = 9), SChLAP1 shRNA 1 (n = 14) or SChLAP1 shRNA 2 (n = 14). (e) Number of gross metastatic sites observed by luciferase signal in mice injected with 22Rv1 cells expressing SChLAP1 shRNA or control, non-targeting shRNA. Independent foci with luciferase signal were averaged for mice injected with 22Rv1 cells expressing control shRNA (n = 9), SChLAP1 shRNA 1 (n = 14) or SChLAP1 shRNA 2 (n = 14). (f) Invasion of 22Rv1 cells expressing non-targeting or SChLAP1 shRNA in chick CAM assays. 22Rv1 cells are labeled with GFP. Images are counterstained for chicken collagen IV to visualize vasculature (RFP) and for luciferase signal were averaged for mice injected with 22Rv1 cells expressing control shRNA 1 (n = 14) or SChLAP1 shRNA 2 (n = 14). (g) Quantification of intravasation of RWPE cells expressing LacZ or SChLAP1 in the CAM assay. All data in bar plots are presented as mean ± s.e.m. Statistical significance was determined by two-tailed Student’s t test: *P < 0.05. Box plots in e,i display box-and-whisker plots with the midpoint line indicating the median, box boundaries showing 25th and 75th quartile ranges and whiskers displaying the minimum and maximum values.

To test SChLAP1 in vivo, we performed intracardiac injection of CB-17 SCID mice with 22Rv1 cells stably knocking down SChLAP1 (Supplementary Fig. 9a) and observed that SChLAP1 depletion impaired metastatic seeding and growth, as measured by luciferase signaling at both proximal (lungs) and distal sites (Fig. 3c,d). Indeed, compared to mice injected with 22Rv1 cells expressing a non-targeting control, mice injected with 22Rv1 cells stably expressing short hairpin RNA (shRNA) against SChLAP1 had both fewer gross metastatic sites overall as well as smaller metastatic tumors when they did form (Fig. 3d,e). Histopathological analysis of the metastatic 22Rv1 tumors, regardless of SChLAP1 knockdown, showed uniformly high-grade epithelial cancer (Supplementary Fig. 9b). Interestingly, subcutaneous xenografts with stable knockdown of SChLAP1 showed slower tumor progression; however, this was due to delayed tumor engraftment rather than to decreased tumor growth kinetics, with no change in Ki67 staining observed between cells expressing SChLAP1 shRNA and control cells expressing non-targeting shRNA (Supplementary Fig. 9c–i).

Next, using the chick chorioallantoic membrane (CAM) assay19, we found that 22Rv1 cells expressing SChLAP1 shRNA 2, which have depleted expression of both isoforms 1 and 2, had greatly reduced ability to invade, intravasate and metastasize to distant organs (Fig. 3f–h). Additionally, cells with knockdown of SChLAP1 also
resulted in decreased tumor growth (Fig. 3i). Notably, RWPE cells with overexpression of SChLAP1 isoform 1 partially supported these results, showing a markedly increased ability to intravasate (Fig. 3j). RWPE cells overexpressing SChLAP1 did not generate distant metastases or cause altered tumor growth in this model (data not shown). Together, the mouse metastasis and CAM data strongly implicate SChLAP1 in tumor invasion and metastasis through activity in cancer cell intravasation, extravasation and subsequent tumor cell seeding.

To elucidate the mechanisms of SChLAP1 function, we profiled 22Rv1 and LNCaP cells with SChLAP1 knockdown, identifying 165 upregulated and 264 downregulated genes (q value < 0.001) (Supplementary Fig. 10a and Supplementary Table 5a). After ranking genes according to differential expression q, we employed Gene Set Enrichment Analysis (GSEA) to search for enrichment across the Molecular Signatures Database (MSigDB). Among the highest ranked concepts, we noticed genes positively or negatively correlated with the SWI/SNF complex, and this association was independently confirmed using gene signatures generated from our RNA-seq data (Supplementary Fig. 10b–e and Supplementary Table 5b,c).

Notably, SChLAP1-regulated genes were inversely correlated with these data sets, suggesting that SChLAP1 and the SWI/SNF complex function in opposing manners. The SWI/SNF complex regulates gene transcription as a multiprotein system that physically moves nucleosomes at gene promoters. SWI/SNF complex mutations do occur in prostate cancer, albeit not commonly. SWI/SNF complex mutations do occur in prostate cancer, albeit not commonly, and downregulation of SWI/SNF complex members
characterizes subsets of prostate cancer. Thus, antagonism of SWI/SNF complex 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-mediated knockdown of SMARCB1 (which encodes the SNF5 protein) (Supplementary Fig. 10f), an essential subunit that facilitates SWI/SNF complex binding to histone proteins, and confirmed predicted expression changes for several SChLAP1- or SNF5-regulated genes (Supplementary Fig. 10g,h). A comparison of genes whose expression was altered by knockdown of SMARCB1 to those regulated by SChLAP1 demonstrated an antagonistic relationship in which SChLAP1 knockdown affected the same genes as SMARCB1 knockdown but with opposing directions of effect (Fig. 4a and Supplementary Table 5d–h). We used GSEA to quantify and verify the significance of these findings (false discovery rate (FDR) < 0.05) (Supplementary Fig. 10i–k). Furthermore, a shared SMARCB1-SChLAP1 signature of coregulated genes was highly enriched for prostate cancer clinical signatures for disease aggressiveness (Supplementary Fig. 11 and Supplementary Table S1).

Mechanistically, although SChLAP1 and SMARCB1 mRNA levels were comparable (Supplementary Fig. 12a), SChLAP1 knockdown or overexpression did not alter SNF5 protein abundance (Supplementary Fig. 12b), suggesting that SChLAP1 regulates SWI/SNF activity post-translationally. To explore this possibility, we performed RNA immunoprecipitation assays (RIPs) for SNF5. We found that endogenous SChLAP1 but not other cytoplasmic or nuclear lncRNAs robustly coimmunoprecipitated with SNF5 under native conditions (Fig. 4b) and with use of UV cross-linking (Supplementary Fig. 12c), and coimmunoprecipitation was also observed with a second antibody to SNF5 (Supplementary Fig. 12d). In contrast, SChLAP1 did not coimmunoprecipitate with androgen receptor (Fig. 4b). Furthermore, both SChLAP1 isoform 1 and isoform 2 coimmunoprecipitated with SNF5 in RWPE overexpression models (Fig. 4c and Supplementary Fig. 12e). SNRNP70 binding to U1 RNA was used as a technical control in all cell lines (Supplementary Fig. 12f,g). Finally, pulldown of SChLAP1 RNA in RWPE cells overexpressing SChLAP1 isoform 1 robustly recovered SNF5 protein, confirming this interaction (Fig. 4d and Supplementary Fig. 12h).

To address whether SChLAP1 modulates SWI/SNF genomic binding, we performed ChIP-seq for SNF5 in RWPE cells expressing LacZ or SChLAP1 and called significantly enriched peaks with respect to an IgG control (Online Methods and Supplementary Table 6a). Protein blot validation confirmed SNF5 pulldown by ChIP (Supplementary Fig. 13a). After aggregating called peaks from all samples, we found 6,235 genome-wide binding sites for SNF5 (FDR < 0.05; Supplementary Table 6b), which were highly enriched for sites near gene promoters (Supplementary Fig. 13b), supporting results from previous studies of SWI/SNF binding.

A comparison of SNF5 binding across these 6,235 genomic sites characterizes subsets of prostate cancer. Thus, antagonism of SWI/SNF complex 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-mediated knockdown of SMARCB1 (which encodes the SNF5 protein) (Supplementary Fig. 10f), an essential subunit that facilitates SWI/SNF complex binding to histone proteins, and confirmed predicted expression changes for several SChLAP1- or SNF5-regulated genes (Supplementary Fig. 10g,h). A comparison of genes whose expression was altered by knockdown of SMARCB1 to those regulated by SChLAP1 demonstrated an antagonistic relationship in which SChLAP1 knockdown affected the same genes as SMARCB1 knockdown but with opposing directions of effect (Fig. 4a and Supplementary Table 5d–h). We used GSEA to quantify and verify the significance of these findings (false discovery rate (FDR) < 0.05) (Supplementary Fig. 10i–k). Furthermore, a shared SMARCB1-SChLAP1 signature of coregulated genes was highly enriched for prostate cancer clinical signatures for disease aggressiveness (Supplementary Fig. 11 and Supplementary Table S1).

Mechanistically, although SChLAP1 and SMARCB1 mRNA levels were comparable (Supplementary Fig. 12a), SChLAP1 knockdown or overexpression did not alter SNF5 protein abundance (Supplementary Fig. 12b), suggesting that SChLAP1 regulates SWI/SNF activity post-translationally. To explore this possibility, we performed RNA immunoprecipitation assays (RIPs) for SNF5. We found that endogenous SChLAP1 but not other cytoplasmic or nuclear lncRNAs robustly coimmunoprecipitated with SNF5 under native conditions (Fig. 4b) and with use of UV cross-linking (Supplementary Fig. 12c), and coimmunoprecipitation was also observed with a second antibody to SNF5 (Supplementary Fig. 12d). In contrast, SChLAP1 did not coimmunoprecipitate with androgen receptor (Fig. 4b). Furthermore, both SChLAP1 isoform 1 and isoform 2 coimmunoprecipitated with SNF5 in RWPE overexpression models (Fig. 4c and Supplementary Fig. 12e). SNRNP70 binding to U1 RNA was used as a technical control in all cell lines (Supplementary Fig. 12f,g). Finally, pulldown of SChLAP1 RNA in RWPE cells overexpressing SChLAP1 isoform 1 robustly recovered SNF5 protein, confirming this interaction (Fig. 4d and Supplementary Fig. 12h).

To address whether SChLAP1 modulates SWI/SNF genomic binding, we performed ChIP-seq for SNF5 in RWPE cells expressing LacZ or SChLAP1 and called significantly enriched peaks with respect to an IgG control (Online Methods and Supplementary Table 6a). Protein blot validation confirmed SNF5 pulldown by ChIP (Supplementary Fig. 13a). After aggregating called peaks from all samples, we found 6,235 genome-wide binding sites for SNF5 (FDR < 0.05; Supplementary Table 6b), which were highly enriched for sites near gene promoters (Supplementary Fig. 13b), supporting results from previous studies of SWI/SNF binding.

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. 4e,f and Supplementary Fig. 13c). Of the 1,299 SNF5 peaks occurring within 1 kb of a gene TSS, 390 showed relative SNF5 binding that was decreased by ≥2-fold with SChLAP1 overexpression (Supplementary Fig. 13d and Supplementary Table 6c). To verify these findings independently, we performed ChIP for SNF5 in 22Rv1 cells expressing shRNA to SChLAP1, 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 with knockdown of SChLAP1 (Supplementary Fig. 14a), confirming our predictions.

Finally, we used expression profiling of RWPE cells expressing LacZ or SChLAP1 to characterize the relationship between SNF5 binding and SChLAP1-mediated changes in gene expression. After identifying a gene signature with highly significant changes in expression (Supplementary Table 6d), we intersected this signature with the ChIP-seq data. We observed that a substantial subset of genes with ≥2-fold relative decrease in SNF5 genomic binding were dysregulated when SChLAP1 was overexpressed (Supplementary Fig. 14b). Decreased SNF5 binding was primarily associated with the downregulation of target gene expression (Supplementary Table 6e), although the SWI/SNF complex is known to regulate expression in either direction. Integrative GSEA of the microarray and SNF5 ChIP-seq data demonstrated significant enrichment for genes that were repressed when SChLAP1 was overexpressed (q value = 0.003; Fig. 4g). 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 properly regulate gene expression.

Here we have discovered SChLAP1, a highly prognostic lncRNA that is abundantly expressed in ~25% of prostate cancers and that aids in the discrimination of aggressive tumors from indolent forms of the disease. Mechanistically, we find that 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 regulation of gene expression and genomic binding (Supplementary Fig. 14c). Thus, whereas 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, SChLAP1 is the first lncRNA, to our knowledge, that impairs a major epigenetic complex with well-documented tumor suppressor function. 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.

URLs. Stellaris probe designer, http://www.singlemoleculefish.com; HT-Seq, http://www-huber.embl.de/users/anders/HTSeq; BioVenn, http://www.cmbi.ru.nl/cdd/biovenn; Galaxy, http://usegalaxy.org/.

METHODS

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

Accession codes. Sequences for SChLAP1 isoforms 1–7 have been deposited in GenBank under accessions JX117418, JX117419, JX117420, JX117421, JX117422, JX117423 and JX117424. Microarray data have been deposited in the Gene Expression Omnibus (GEO) under accession GSE40386.

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

ACKNOWLEDGMENTS

We thank O.A. Balbin, S.A. Tomlins, C. Brenner, S. Deroo and S. Roychowdhury for helpful discussions. This work was supported in part by US National Institutes of Health (NIH) Prostate Specialized Program of Research Excellence grant P50CA69568, Early Detection Research Network grant U01 CA111275, US NIH grant R01CA132874-01A1 and US Department of Defense grant PC100171 (A.M.C.). A.M.C. is supported by a Doris Duke Charitable Foundation Clinical Scientist Award, by the Prostate Cancer Foundation and by the Howard Hughes Medical Institute. A.M.C. is an American Cancer Society Research Professor. A.M.C. is a Taubman Scholar of the University of Michigan. F.Y.F. was supported by the Prostate Cancer Foundation and by US Department of Defense grant PC094231, Q.C. was supported by US Department of Defense Postdoctoral
LETTERS

Fellowship PC094725. J.R.P. was supported by US Department of Defense Predoctoral Fellowship PC094290. M.K.I. was supported by US Department of Defense Predoctoral Fellowship BC100238. A.S. was supported by NIH Predoctoral Fellowship 1F36CA180376-01, J.R.P., M.K.I. and A.S. are Fellows of the University of Michigan Medical Scientist Training Program.

AUTHOR CONTRIBUTIONS
J.R.P., M.K.I., A.S. and A.M.C. designed the project and directed experimental studies. J.R.P., Q.C., W.C., S.M.D., B.C., S.H., R.M., L.P., T.J.T. and E.D. performed CAM assays. J.R.P., M.K.I., A.S. and A.M.C. interpreted expression in the Mayo Clinic cohort. E.D., N.E., M.G. and I.A.V. performed statistical analyses of data and wrote the manuscript. 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.

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

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

1. Etzioni, R., Cha, R., Feuer, E.J. & Davidov, O. Asymptomatic incidence and duration of prostate cancer. Am. J. Epidemiol. 148, 775–785 (1998).
2. Cooperberg, M.R., Moul, J.W. & Carroll, P.R. The changing face of prostate cancer. Cancer 105, 2642–2649 (2005).
3. Grasso, C.S. et al. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. J. Natl. Cancer Inst. 100, 815–825 (2008).
4. Nakagawa, T. et al. A tissue biomarker panel predicting systemic progression after PSA recurrence post-definitive prostate cancer therapy. PLoS ONE 3, e2318 (2008).

© 2013 Nature America, Inc. All rights reserved.
ONLINE METHODS

Cell lines. All cell lines were obtained from the American Type Culture Collection. Cell lines were maintained using standard media and conditions. Specifically, VCaP and DU145 cells were maintained in DMEM (Invitrogen) supplemented with 10% FBS and 1% penicillin-streptomycin. LNCaP and 22Rv1 cells were maintained in RPMI 1640 (Invitrogen) supplemented with 10% FBS and 1% penicillin-streptomycin. RWPE cells were maintained in KSF medium (Invitrogen) supplemented with 10 ng/ml epidermal growth factor (EGF; Sigma) and bovine pituitary extract (BPE) and with 1% penicillin-streptomycin. All cell lines were grown at 37 °C in a 5% CO₂ cell culture incubator. All cell lines were genotyped for identity at the University of Michigan Sequencing Core and were tested routinely for Mycoplasma contamination.

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

Tissue samples. Prostate tissues were obtained from the radical prostatectomy series and Rapid Autopsy Program at the University of Michigan tissue core. These programs are part of the University of Michigan Prostate Cancer Specialized Program of Research Excellence (SPRE). All tissue samples were collected with informed consent under an institutional review board (IRB)-approved protocol at the University of Michigan (SPRE in Prostate Cancer (Tissue/Serum/Urine) Bank Institutional Review Board 1994-0481).

RNA isolation and cDNA synthesis. Total RNA was isolated using TRIzol (Invitrogen) and an RNaseasy kit (Qiagen) with DNase I digestion according to the manufacturers' instructions. RNA integrity was verified on an Agilent Bioanalyzer 2100 (Agilent Technologies). cDNA was synthesized from total RNA using Superscript III (Invitrogen) and random primers (Invitrogen).

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

RT-PCR. RT-PCR was performed for primer pairs using Platinum Taq High-Fidelity polymerase (Invitrogen). PCR products were resolved on a 1.0% agarose gel. PCR products were then either sequenced directly (if only a single product was observed) or appropriate gel products were extracted using a Gel Extraction kit (Qiagen) and cloned into pcR4-TOPO vector (Invitrogen). PCR products were bidirectionally sequenced at the University of Michigan Sequencing Core using either gene-specific primers or M13 forward and reverse primers for cloned PCR products. All oligonucleotide primers were obtained from IDT, and primer sequences are listed in Supplementary Table 7a.

RACE. 5' and 3' RACE were performed using the GeneRacer RLM-RACE kit (Invitrogen) according to the manufacturer’s instructions. RACE PCR products were obtained using Platinum Taq High-Fidelity polymerase, the supplied GeneRacer primers and the appropriate gene-specific primers indicated in Supplementary Table 7a. RACE PCR products were separated on a 1.5% agarose gel. Gel products were extracted with a Gel Extraction kit, cloned into pcR4-TOPO vectors and sequenced bidirectionally using M13 forward and reverse primers at the University of Michigan Sequencing Core. At least three colonies were sequenced for every RACE PCR product that was gel purified.

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

Overexpression. Full-length SChLAP1 transcript was amplified from LNCaP cells and cloned into the pLent6 vector along with LacZ control sequence. Insert sequences were confirmed by Sanger sequencing at the University of Michigan Sequencing Core. Lentiviruses were generated at the University of Michigan Vector Core. The benign immortalized prostate cell line RWPE was infected with lentiviruses expressing SChLAP1 or LacZ, and stable pools and clones were generated by selection with blasticidin. Similarly, the immortalized cancer cell line 22Rv1 was infected with lentiviruses expressing SChLAP1 or LacZ, and stable pools were generated by selection with blasticidin.

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

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

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

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

Mouse intracardiac and subcutaneous in vivo models. All experimental procedures were approved by the University of Michigan Committee for the Use and Care of Animals (UCUCA).

For the intracardiac injection model, 5 × 10⁵ cells from 1 of 3 experimental cell lines (22Rv1-shSChLAP1-1 or 22Rv1-shSChLAP1-2) (two cell lines expressing SChLAP1 shRNA) or 22Rv1-shNT (expressing control vector), all with luciferase constructs incorporated) were introduced into C57BL/6 SCID mice at 6 weeks of age. Female mice were used to minimize endogenous androgen production that might stimulate xenografted prostate cells. We used 15 mice per cell line to ensure adequate statistical power to distinguish phenotypes between groups. Mice used in these studies were randomized by double-blind injection of cell line samples into mice and were monitored for tumor growth by researchers blinded to the study design. Beginning 1 week after injection, bioluminescent imaging of mice was performed weekly using a CCD IVIS...
system with a 50-mm lens (Xenogen), and the results were analyzed using LivingImage software (Xenogen). When the a mouse reached the determined end point, defined as whole-body region of interest (ROI) of $1 \times 10^{15}$ photons, or became fatally ill, it was euthanized, and the lung and liver were resected. Half of the resected specimen was placed in an immunohistochemistry cassette, incubated in 10% buffered formalin phosphate (Fisher Scientific) for 24 h and transferred to 70% ethanol until further analysis. The other half of each specimen was snap frozen in liquid nitrogen and stored at −80 °C. A specimen was disregarded if the tumor was localized only in the heart. After accounting for these considerations, there were 9 mice analyzed for 22Rv1-shNT cells and 14 mice each analyzed for 22Rv1-shSChLAP1-1 and 22Rv1-shSChLAP1-2 cells.

For the subcutaneous injection model, $1 \times 10^6$ cells from 1 of the 3 previously described experimental cell lines were introduced into mice (CB-17 SCID), aged 5–7 weeks, with a Matrigel scaffold (BD Matrigel Matrix, BD Biosciences) in the posterior dorsal flank region (n = 10 per cell line). Tumors were measured weekly using a digital caliper, and the end point was defined by tumor volume of 1,000 mm$^3$. When a mouse reached the end point or became fatally ill, it was euthanized, and the primary tumor was resected. The resected specimen was divided in half: one half was placed in 10% buffer formalin, and the other half was snap frozen. For histological analyses, formalin-fixed, paraffin-embedded mouse livers and livers were sectioned on a microtome into 5-µm sections on glass slides. Slides were stained with hematoxylin and eosin using standard methods and were analyzed by a board-certified pathologist (L.P.K.).

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

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

RIP assays. RIP assays were performed using a Millipore EZ-Magna RIP RNA-Binding Protein Immunoprecipitation kit (Millipore, 70-701) according to the manufacturer’s instructions. RIP PCR was performed as qPCR, as described above, using total RNA as input controls. We used 1/150 volume of the RIP RNA product per PCR reaction. Antibodies used for RIP included rabbit polyclonal IgG (Millipore, PP64) and antibodies to SNRNP70 (Millipore, CS203216), SNF5 (Millipore, ABD22, rabbit), SNF5 (1:1,000 dilution; Abcam, ab58209, mouse), and androgen receptor (Millipore, 06-680, rabbit).

ChIP assays. ChIP assays were performed as described previously$^{11,12}$ using antibody for SNF5 (Millipore, ABD22, rabbit) and rabbit IgG (Millipore, PP64B). Briefly, approximately 1 million cells were cross-linked per antibody for 10–15 min with 1% formaldehyde, and crosslinking was inactivated by incubation with 0.125 M glycine for 5 min at room temperature. Cells were rinsed with cold PBS three times, and cell pellets were resuspended in lysis buffer supplemented with protease inhibitors. Chromatin was sonicated to an average length of 500 bp and centrifuged to remove debris, and supernatants containing chromatin fragments were incubated with protein A or protein G beads to reduce non-specific binding. Beads were then removed, and supernatants were incubated with 6 µg of antibody overnight at 4 °C. Fresh beads were added and incubated with protein-chromatin-antibody complexes for 2 h at 4 °C, washed twice with 1× dialysis buffer and four times with IP wash buffer, and eluted in 150 µl of IP elution buffer$^{11,12}$. One-tenth of the ChIP reaction was taken for protein evaluation for validation of pulldown. Cross-linking was reversed by incubating eluted products with 0.3 M NaCl at 65 °C overnight. ChIP products were cleaned with the USB PrepEase kit. ChIP experiments were validated for specificity of the antibody by protein blotting.

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

CAM assays. CAM assays were performed as previously described$^{39}$. Briefly, fertilized chicken eggs were incubated in a rotary humidified incubator at 38 °C for 10 d. CAM was released by applying a mild amount of pressure to the hole over the air sac and cutting a 1-cm$^2$ window encompassing a second hole near the allantoic vein. Approximately 2 million cells in 50 µl of medium were implanted in each egg, windows were sealed, and eggs were returned to a stationary incubator.

For local invasion and intravasation experiments, the upper and lower CAMs were isolated after 72 h. Upper CAMs were processed and stained for chicken collagen IV (immunofluorescence) or human cytokeratin (immunohistochemistry) as previously described$^{39}$.

For metastasis assays, embryonic livers were isolated on day 18 of embryonic growth and analyzed for the presence of tumor cells by quantitative human Alu-specific PCR. Genomic DNA isolates from lower CAMs and livers were prepared using the Puregene DNA purification system (QIagen), and quantification with human Alu-specific PCR was performed as described$^{39}$. Fluoresgenic TaqMan qPCR probes were generated as described above and used to determine DNA copy number.

For xenograft growth assays with RWPE cells, embryos were sacrificed on day 18, and extraembryonic xenografts were excised and weighed.

In situ hybridization. In situ hybridization assays were performed as a commercial service from Advanced Cell Diagnostics, Inc. Briefly, cells in the clinical specimens were fixed and permeabilized using xylene, ethanol and protease to allow for probe access. Slides were boiled in pretreatment buffer for 15 min and rinsed in water. Next, two independent target probes were hybridized to SChLAP1 RNA at 40 °C for 2 h, with this pair of probes creating a binding site for a preamplifier. After this incubation, the preamplifier was hybridized to the target probes at 30 °C and amplified with six cycles of hybridization followed by two washes. Cells were counterstained to visualize signal. Finally, slides were stained with hematoxylin and eosin, dehydrated with 100% ethanol and mounted in a xylene-based mounting medium.

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

ChIRP assays. ChIRP assays were performed as previously described$^{48}$. Briefly, antisense DNA probes targeting the full-length SChLAP1 sequence.
were designed using the online designer at Stellaris (see URLs). Fifteen probes spanning the entire transcript and unique to the SChLAP1 sequence were chosen. Additionally, ten probes were designed against TERC RNA as a positive control, and 24 probes were designed against Lacz RNA as a negative control. All probes were synthesized with 3′ biotinylated (IDT). Sequences of all probes are listed in Supplementary Table 8. RWPE cells overexpressing SChLAP1 isoform 1 were grown to 80% confluency in 100-mm cell culture dishes. Two dishes were used for each probe set. Before being collected, cells were rinsed with 1× PBS and cross-linked with 1% glutaraldehyde (Sigma) for 10 min at room temperature. Cross-linking was quenched by incubation with 0.125 M glycine for 5 min at room temperature. Cells were rinsed twice with 1× PBS, collected and pelleted at 1,500g for 5 min. Nuclei were isolated using the Pierce NE-PER Nuclear Protein Extraction kit. Nuclear pellets were resuspended in 100 mg/ml cell lysis buffer (50 mM Tris, pH 7.0, 10 mM EDTA, 1% SDS, and, added before use, 1 mM dithiothreitol (DTT), phenylmethylsulphonyl fluoride (PMSF), protease inhibitor and Superase-In (Invitrogen)). Lysates were placed on ice for 10 min and sonicated using a Bioruptor (Diagenode) at the highest setting with 30-s on and 45-s off cycles until lysates were completely solubilized. Cell lysates were diluted in twice the volume of hybridization buffer (500 mM NaCl, 1% SDS, 100 mM Tris, pH 7.0, 10 mM EDTA, 15% formamide, and, added before use, DTT, PMSF, protease inhibitor and Superase-In). 100 nM probes were added to the diluted lysates. Hybridization was carried out by end-over-end rotation at 37°C for 4 h. Magnetic streptavidin C1 beads were prepared by washing three times in cell lysis buffer and were then added to each hybridization reaction at a concentration of 100 µl per 100 pmol of probe. Reactions were incubated at 37°C for 30 min with end-over-end rotation. Read-probe RNA complexes were captured with magnetic racks (Millipore) and washed five times with 1 ml wash buffer (2× SSC, 0.5% SDS, fresh PMSF added). After the final wash, 20% of the sample was used for RNA isolation, and 80% of the sample was used for protein isolation. For RNA elution, beads were resuspended in 200 µl of RNA proteinase K buffer (100 mM NaCl, 10 mM Tris, pH 7.0, 1 mM EDTA, 0.5% SDS) and 1 mg/ml proteinase K (Ambion). Samples were incubated at 50°C for 45 min and then boiled for 10 min. RNA was isolated using 500 µl of TRIzol reagent and the miRNeasy kit (Qiagen) with on-column DNAse digestion (Qiagen). RNA was eluted with 10 µl of water and then analyzed by quantitative RT-PCR for the detection of enriched transcripts. For protein elution, beads were resuspended in three times the original volume of DNase buffer (100 mM NaCl, 0.1% NP-40), and protein was eluted with a cocktail of 100 µg/ml RNase A (Sigma-Aldrich), 0.1 U/ml RNase H (Epicenter) and 100 U/ml DNase I (Invitrogen) at 37°C for 30 min. Eluted protein samples were supplemented with NuPAGE LDS Sample Buffer (Novex) and NuPAGE Sample Reducing Agent (Novex) to a final concentration of 1× each and then boiled for 10 min before SDS-PAGE protein blot analysis using an antibody to SNF5 (Millipore).

RNA-seq library preparation. Total RNA was extracted from healthy and cancer cell lines and subject tissues, and RNA quality was assessed via Agilent Bioanalyzer. Transcriptome libraries from the mRNA fractions were generated following the RNA-seq protocol (Illumina). Each sample was sequenced in a single lane with the Illumina Genome Analyzer II (with a 40- to 80-nt read length) or with the Illumina HiSeq 2000 (with a 100-nt read length) according to published protocols. For strand-specific library construction, we employed the dUTP method of second-strand marking as described previously. Statistical analyses for experimental studies. All data are presented as means ± s.e.m. All experimental assays were performed in duplicate or triplicate. Statistical analyses shown in figures represent Fisher’s exact tests or two-tailed t tests, as indicated. For details regarding the statistical methods employed during microarray, RNA-seq and ChiP-seq data analysis, see below.

Nomination of SChLAP1 as an outlier using RNA-seq data. We nominated SChLAP1 as a prostate cancer outlier as described. Briefly, a modified COPA analysis was performed on the 81 tissue samples in the cohort. Reads per kilobase per million mapped reads (RPKM) expression values were used and shifted by 1.0 to avoid division by zero. COPA analysis included the following steps: (i) gene expression values were median centered, using the median expression value for the gene across all samples in the cohort, which sets the gene’s median to zero; (ii) the median absolute deviation (MAD) was calculated for each gene, and each gene expression value was then scaled by its MAD; (iii) the 80th, 85th, 90th and 98th percentiles of the transformed expression values were calculated for each gene, the average of those four values was taken, and genes were then ranked according to this ‘average percentile’, which generated a list of outlier genes arranged by importance; and (iv) finally, genes showing an outlier profile in the benign samples were discarded.

LNCaP ChiP-seq data. Sequencing data from GSE14097 were downloaded from GEO. Reads from the LNCAP H3K4me3 and H3K36me3 ChiP-seq samples were mapped to human genome version hg19 using BWA 0.5.9 (ref. 51). Peak calling was performed using MACS2 according to published protocols. Data were visualized using the UCSC Genome Browser.

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

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

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

SChLAP1 siRNA knockdown microarrays. For the 22Rv1 and LNCaP SChLAP1 knockdown experiments, we ran the GseaPreRanked tool to discover enriched gene sets in MSigDB version 3.0. Lists of positively and negatively enriched concepts were interpreted manually.
SMARCB1 siRNA knockdown microarrays. For each SMARCB1 knockdown experiment, we nominated genes that were altered by an average of at least twofold. 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 whose expression changed by an average of twofold or greater across SMARCB1 knockdown experiments and quantified the enrichment for SChLAP1 target genes using GSEA.

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

RNA-seq data. We assembled an RNA-seq cohort from prostate cancer tissues sequenced at multiple institutions. We included data from 12 primary tumors and 5 benign tissues published in GEO (GSE21260)27, from 16 primary tumors and 3 benign tissues released in the database of Genotypes and Phenotypes (dbGaP) (phs000310.v1.p1)28 and from 17 benign, 57 primary and 14 metastatic tumors sequenced by our own institution and released in dbGaP (phs000443.v1.p1). Sample information is shown in Supplementary Table 1a, and sequencing library information is shown in Supplementary Table 1b.

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

RNA-seq differential expression analysis. Differential expression analysis was performed using R package DESeq2 version 1.6.1. Read counts were normalized using the `estimateSizeFactors` function, and variance was modeled by the ‘estimateDispersion’ function. Statistics on differential expression were computed by the ‘nbomtest’ function. We called differentially expressed genes by imposing adjusted P-value cutoffs for cancer versus benign samples ($P_{adj} < 0.05$), metastasis versus primary samples ($P_{adj} < 0.05$) and Gleason score of 8+ versus 6 ($P_{adj} < 0.10$). Heatmap visualizations of these analyses are presented as Supplementary Figure 5.

RNA-seq correlation analysis. Read count data were normalized using functions from the R package DESeq version 1.6.1. Adjustments for library size were made using the ‘estimateSizeFactors’ function, and variance was modeled using the ‘estimateDispersion’ function using the parameters ‘method=blind’ and ‘sharingModel=fit-only’. Next, raw read count data were converted to pseudocounts using the ‘getVarianceStabilizedData’ function. Gene expression levels were then mean centered and standardized using the ‘scale’ function in R. Pearson’s correlation coefficients were computed between each gene of interest and all other genes. Statistical significance of Pearson’s correlations was determined by comparison to correlation coefficients achieved with 1,000 random permutations of the expression data. We controlled for multiple-hypothesis testing using the ‘qvalue’ package in R. We performed a large meta-analysis of these correlation signatures across Oncomine data sets corresponding to disease outcome (Glinsky Prostate and Setlur Prostate), metastatic disease (Holzbeierlein Prostate, Lapointe Prostate, LaTulippe Prostate, Taylor Prostate 3, Vanaja Prostate, Varambally Prostate and Yu Prostate), advanced Gleason score (Bittner Prostate, Glinsky Prostate, Lapointe Prostate, LaTulippe Prostate, Setlur Prostate, Taylor Prostate 3 and Yu Prostate) and localized cancer (Arredouani Prostate, Holzbeierlein Prostate, Lapointe Prostate, LaTulippe Prostate, Taylor Prostate 3, Varambally Prostate and Yu Prostate). We also incorporated our own concept signatures for metastasis, advanced Gleason score and localized cancer determined from our RNA-seq data. For each concept, we downloaded the gene signatures corresponding to the top 5% of genes up- and downregulated. Pairwise signature comparisons were performed using a one-sided Fisher’s exact test. We controlled for multiple-hypothesis testing using the ‘qvalue’ package in R. We considered concept pairs with $q < 0.01$ and odds ratio > 2.0 as significant. In cases where a gene signature associated with both the over- and underexpression gene sets from a single concept, only the most significant result (as determined by odds ratio) is shown.

Analysis of SChLAP1 and SMARCB1 expression signatures. Gene signatures obtained with knockdown of SChLAP1 and SMARCB1 were generated from Agilent gene expression microarray data sets. For each cell line, we obtained a single vector of per-gene fold changes by averaging technical replicates and then taking the median across biological replicates. We merged the results from individual cell line using the median of the changes in 22RV1 and LNCaP cells. Venn diagram plots were produced using the BioVenn website56. We then compared the top 10% of upregulated and downregulated genes with knockdown of SChLAP1 and SMARCB1 to gene signatures downloaded from the Taylor Prostate 3 data set in the Oncomine database. We performed signature comparison using one-sided Fisher’s exact tests and controlled for multiple testing using the R package ‘qvalue’. Signature comparisons with $q < 0.05$ were considered significantly enriched. We plotted the odds ratios from significant comparisons using the ‘heatmap.2’ function in the ‘ggplot’ R package.

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

Additionally, we imported the 253-gene SChLAP1 signature into Oncomine to download the expression data for 167 of the 253 genes profiled by the Glinsky prostate data set26. We assigned SChLAP1 expression scores in a similar fashion and designated the top 20% of patients as having ‘high’ SChLAP1 scores. We performed survival analysis using the time to biochemical prostate-specific antigen (PSA) recurrence and computed statistical significance as described above.

PhyloCSF analysis. We obtained 46-way multi-alignment FASTA files for SChLAP1, HOTAIR, GAPDH and ACTB using the ‘Stitch Gene blocks’ tool

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

Oncomine concepts analysis of the SChLAP1 signature. We separated the 253 genes with expression levels significantly correlated with SChLAP1 into positively and negatively correlated gene lists. We imported these gene lists into Oncomine as custom concepts. We then nominated significantly associated prostate cancer concepts with odds ratio $> 3.0$ and $P < 1 \times 10^{-6}$. We exported these results as the nodes and edges of a concept association network and visualized the network using Cytoscape version 2.8.2. Node positions were computed using the Force-Directed Layout algorithm in Cytoscape using the odds ratio as the edge weight. Node positions were subtly altered manually to enable better visualization of node labels.
within the Galaxy bioinformatics framework. We evaluated each gene for the likelihood that it represented a protein-coding region using PhyloCSF software (version released 28 October 2012). Each gene was evaluated using the phylogeny from 29 mammals (available by default within PhyloCSF) in any of the 3 reading frames. Scores are measured in decibans and represent the likelihood ratio that a sequence is protein-coding rather than noncoding.

**Mayo Clinic cohort analyses.** Subjects were selected from a cohort of individuals from the Mayo Clinic with high-risk prostate cancer who had undergone radical prostatectomy. The cohort was defined as 1,010 men with high-risk prostate cancer who underwent radical prostatectomy between 2000 and 2006, of whom 73 developed clinical progression (defined as individuals with systemic disease as evidenced by positive bone or computed tomography (CT) scan)\(^5\). High risk of recurrence was defined by preoperative PSA levels of >20 ng/ml, pathological Gleason score of 8–10, seminal vesicle invasion (SVI) or Gleason, PSA, seminal vesicle and margin (GPSM) score of ≥10 (ref. 62). The subcohort incorporated all 73 subjects with clinical progression to systemic disease and a random sampling of 20% of the entire cohort (202 men, including 19 with clinical progression). The total case-cohort included 256 subjects, and tissue specimens were available from 235 subjects. The subcohort was previously used to validate a genomic classifier for predicting clinical progression\(^6\).

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

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

**Microarray expression analysis.** Normalization and summarization of the microarray samples was performed with the frozen Robust Multiarray Average (RMA) algorithm using custom frozen vectors. These custom vectors were created using the vector creation methods described previously\(^6\). Quantile normalization and robust weighted average methods were used for normalization and summarization, respectively, as implemented in RMA.

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

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

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

46. Rubin, M.A. et al. Rapid (“warm”) autopsy study for procurement of metastatic prostate cancer. Clin. Cancer Res. 6, 1038–1045 (2000).
47. Tomlins, S.A. et al. Role of the TMRPSS2-ERG gene fusion in prostate cancer. Neoplasia 10, 177–188 (2008).
48. Chu, C., Qu, K., Zhong, F.L., Artandi, S.E. & Chang, H.Y. Genomic maps of long non-coding RNA occupants reveal principal RNA-chromatin interactions. Mol. Cell 44, 667–678 (2011).
49. Maher, C.A. et al. Chimeric transcript discovery by paired-end transcriptome sequencing. Proc. Natl. Acad. Sci. USA 106, 12353–12358 (2009).
50. Levin, J.Z. et al. Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat. Methods 7, 709–715 (2010).
51. Li, H. & Durbin, R. Fast and accurate short read alignment with Bumrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).
52. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137 (2008).
53. Feng, J., Liu, T. & Zhang, Y. Using MACS to identify peaks from ChIP-Seq data. Curr. Protoc. Bioinformatics Chapter 2 Unit 2.14 (2011).
54. Kent, W.J. et al. The human genome browser at UCSC. Genome Res. 12, 996–1006 (2002).
55. Shin, H., Liu, T., Mannai, A.K. & Liu, X.S. CES: cis-regulatory element annotation system. Bioinformatics 25, 2605–2606 (2009).
56. Anders, S. & Huber, W. Differential expression analysis for sequence count data. Genome Biol. 11, R106 (2010).
57. Kannan, K. et al. Recurrent chimeric RNAs enriched in human prostate cancer identified by deep sequencing. Proc. Natl. Acad. Sci. USA 108, 9172–9177 (2011).
58. 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).
59. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111 (2009).
60. 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).
61. Karnes, R.J. et al. Validation of a genomic classifier that predicts metastasis following radical prostatectomy in an at-risk patient population. J. Urol. doi:10.1016/j.juro.2013.06.017 (11 June 2013).
62. 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).
63. Vergara, I.A. et al. Genomic “dark matter” in prostate cancer: exploring the clinical utility of ncRNA as biomarkers. Front. Genet. 3, 23 (2012).