Conserved Structure Modules within the lncRNA SChLAP1 Mediate Protein Recognition Implicated in Aggressive Prostate Cancer

Emily J. McFadden[†1], James P. Falese[†1], and Amanda E. Hargrove[*1,2]

Affiliations
1. Department of Biochemistry, Duke University School of Medicine, Durham, North Carolina, 27710, USA
2. Department of Chemistry, Duke University, Durham, North Carolina, 27708, USA

†: Joint Authors
* To whom correspondence should be addressed: amanda.hargrove@duke.edu

ABSTRACT

The lncRNA Second Chromosome Locus Associated with Prostate 1 (SChLAP1) was previously identified as a predictive biomarker and potential driver of aggressive prostate cancer. Recent work suggests that SChLAP1 may bind the SWI/SNF chromatin remodeling complex to promote prostate cancer metastasis, though the exact role of SWI/SNF recognition is debated. To date, there are no detailed biochemical studies of *apo* SChLAP1 or SChLAP1:protein complexes. Herein, we report the first secondary structure model of SChLAP1 using SHAPE-MaP and DMS-MaPseq both *in vitro* and *in cellulo*. Comparison of the *in vitro* and *in cellulo* data via ΔSHAPE identified putative protein binding sites within SChLAP1, specifically to evolutionarily conserved exons of the transcript. We also demonstrate that global SChLAP1 secondary structure is sensitive to both purification method and magnesium concentration. Furthermore, we identified a 3′–fragment of SChLAP1 ($\text{SChLAP1}_{\text{Frag}}$) that harbors multiple potential protein binding sites and presents a robustly folded secondary structure, supporting a functional role for this region. This work lays the foundation for future efforts to selectively target and disrupt the SChLAP1:protein interface and to develop new therapeutic avenues in prostate cancer treatment.

INTRODUCTION

Prostate cancer is one of the most commonly occurring cancers (one in four new cancer cases) and is the second leading cause of cancer death in American men over the age of 60 (1). One in nine American men will be diagnosed with prostate cancer in their lifetime, and Black men are 60% more likely to develop prostate cancer than White men (2). The main treatments for prostate cancer include: surveillance, surgery such as prostatectomy, and non-specific treatments such as radiation, chemotherapy, cryotherapy, and hormone deprivation therapies. None of these treatments provides a cure for prostate cancer, however, and aggressive prostate cancer rapidly becomes treatment resistant. While more than 99% of patients with primary prostate cancer enter remission, the remission rate drops to less than 30% after metastasis (1). There is thus an urgent unmet need for specific therapeutic strategies that target molecular drivers of aggressive prostate cancer.

While the majority of the human genome is transcribed, less than 2% of the human genome encodes for proteins (3). The remaining RNA, the non-coding transcriptome, is thus an underexplored realm for potential therapeutics. Long non-coding RNAs (lncRNAs), generally defined as non-translated transcripts ≥200 nucleotides (n.t.) in length, are often differentially expressed throughout developmental stages, tissue types, and disease states (4-8). Following the ENCODE project (3) and the identification of thousands of new lncRNAs, the FANTOM consortium (9) began annotating these transcripts and found biochemical indices of function historically ascribed solely to proteins. Although there are over 170,000 identified lncRNA, only ~20,000 of these are functionally annotated transcripts, and an even smaller portion of these have been biochemically characterized.

The lncRNA Second Chromosome Locus Associated with Prostate-1 (SChLAP1) is a prime example of the incongruity between the identification of cancer-associated lncRNAs and characterization of their biochemical function. In 2011, Prensner et al. identified the transcript PCAT-114 (later renamed SChLAP1) as over-expressed in tumor prostate tissue compared to benign prostate tissue (10). In other work, Gerashchenko et al. determined that tumors with high SChLAP1 transcript levels correlated with a Gleason score of 9, where 10 is the highest possible score and most at-risk group (11). SChLAP1 levels also correlated with high levels of epithelial-to-mesenchymal transition (EMT) markers such as vimentin (VIM), fibronectin (FN1), and matrix metalloproteinase 2 (MMP2), further supporting the correlation between SChLAP1 over-expression and prostate cancer metastasis found in other clinical studies (11-14). Mehra et al. determined that approximately 16% of clinically localized prostate cancers in American men exhibit high SChLAP1 transcript levels, suggesting use of SChLAP1 as an early detection biomarker (15).

SChLAP1 overexpression was also observed in multiple prostate cancer cell lines. In model systems, SChLAP1 depletion reduced cellular invasion *in vitro* and formation of distal metastases in severe combined immunodeficiency (SCID) mice (16). Using Gene Set Enrichment Analysis (GSEA) in two different prostate cancer cell lines, Prensner et al. found that SChLAP1 inversely regulated genes that were regulated by the SWItch/Sucrose Non-Fermentable (SWI/SNF) chromatin remodeling complex, also known as the BRG1/BRM-associated factor (BAF) complex. The 11-15–mer SWI/SNF complex specifically catalyzes ATP-dependent chromatin remodeling through sliding or ejection of nucleosomes from DNA, and it is dysregulated and/or houses mutations in its respective subunit genes in nearly 25% of all cancers (17-20). Using the normal prostate cell line RWPE-1 over-expressing SChLAP1 via lentiviral transfection as well as 22Rv1 and LNCaP prostate cancer cells lines with nascent high levels of SChLAP1 expression, Prensner et al. observed direct binding between the SWI/SNF complex and SChLAP1 via RNA immunoprecipitation (RIP), specifically between the SChLAP1 transcript and the SMARCB1 subunit (also known as SNF5 or INI1) (16). This interaction resulted in genome-wide depletion of SWI/SNF from the chromatin and inhibition of normal SWI/SNF function. However, follow up work from Raab et al. did not recapitulate this global loss in SMARCB1 genomic occupancy (21). In addition, Raab et al. observed widespread RNA interaction by the SWI/SNF complex, which has also been documented in more recent work (21-23). While SChLAP1 still bound to SMARCB1 and altered chromatin accessibility at a subset of genomic sites with known SWI/SNF interactors or pathways, Raab et al. proposed that SChLAP1 may function in a SWI/SNF-independent manner. Indeed, several other proteins and/or protein complexes have been proposed to interact with SChLAP1, including: Polycomb repressive complex 2 (PRC2), a histone methyltransferase (24); DNA (cytosine-5)-methyltransferase 3A (DNMT3A), a DNA methyltransferase (24); Heterogeneous nuclear ribonucleoprotein L (HNRNPL), which regulates RNA processing (25); and Heterogeneous nuclear ribonucleoprotein D0 (HNRNPD), which regulates RNA stability (26). While additional studies of lncRNA:SWI/SNF recognition observed crucial interactions in both healthy and disease states (22,23,27-33), the role for SWI/SNF recognition by SChLAP1 remains disputed.

Regardless of known or disputed protein binding partners, elucidating the sequence- and/or structure-function relationships within SChLAP1 relative to aggressive prostate cancer is critical to both our understanding of prostate cancer metastasis and of the potential for SChLAP1 as a therapeutic target. Herein, we report first insights into the role of a sequence-structure relationship to SChLAP1 function. First, we found three exons that are highly conserved across all human SChLAP1 isoforms and non-human primates. We then completed the first *in vitro* secondary structural analysis of SChLAP1 wild-type (WT) isoform 1 and a synthetic deletion mutant (D5) via selective 2′-Hydroxyl Acylation Analyzed by Primer Extension and Mutational Profiling (SHAPE-MaP) (34,35) and Dimethyl Sulfate Mutational Profiling and sequencing (DMS-MaPseq) (36), which allowed us to identify multiple conserved structures across SChLAP1. Using *in cellulo* probing and analysis via ΔSHAPE, we mapped putative protein binding regions to the SChLAP1 WT transcript and found them localized to the evolutionarily conserved exons. Finally, we identified an evolutionarily conserved putative independent folding domain, SChLAP1$_{\text{Frag}}$, which shows high indications for protein binding in our *in cellulo* data. We show that while *global* SChLAP1 secondary structure may change in response to purification method and magnesium concentration, SChLAP1$_{\text{Frag}}$ secondary structure is unperturbed in these varying conditions, suggesting it is a key functional motif. As the first detailed biochemical and structural analysis of SChLAP1, this work proposes an important functional role of SChLAP1$_{\text{Frag}}$ in SChLAP1:protein recognition and the metastatic phenotype.

MATERIAL AND METHODS

Cell culture. LNCaP cells were obtained from the Duke University Cell Culture Facility and were strain and mycoplasma tested. Cells were grown in RPMI 1640 media (Gibco) with 10% fetal bovine serum (FBS, Gibco) at 37 °C in a humidified atmosphere with 5% CO_2. All experiments were performed on cells less than 20 passages before retrieving a fresh vial from cryopreservation.

DNA Template and Primers. All oligos were purchased from Integrated DNA Technologies (IDT). The full-length sequence for the SChLAP1 isoform 1 wild-type (WT) and Deletion #5 (D5) sequences were inserted downstream of a bacteriophage T7 RNA polymerase promoter and upstream of the BAMHI restriction site. We scaled up plasmid growth and purification via transformation into NEB5α competent cells (New England Biolabs, USA) following manufacturer's instructions and selection LB agar plates with ampicillin (100 μg/mL final) for overnight growth at 37 °C. A single colony was propagated in LB broth with ampicillin selection and plasmids were isolated using the Qiagen Plasmid Kit. The plasmid was linearized using BAMHI-HF (New England Biolabs, USA) following manufacturer protocol. Linearized plasmid was purified using Qiagen DNA Mini Kit. For SChLAP1$_{Frag}$, primers were designed for to create a transcription-suitable template from the full-length SChLAP1 template, where the forward primer contained a T7 RNA polymerase promoter overhang. See SI Table 1 for sequence information. PCR reactions were performed using Q5 High-Fidelity DNA polymerase (New England Biolabs, USA). The following thermocycler settings were used: 1 cycle, 98 °C for 45 seconds; 30 cycles of: 98 °C for 10 seconds, T_{ann} °C for 30 seconds, 72 °C for 15 seconds; 1 cycle of 72 °C for 5 minutes; hold at 4-10 °C (see SI Table 1 for T_{ann} temperature for individual sequences). DNA was purified using Zymo DNA Clean and Concentrator (5).

In Vitro Transcription. *In vitro* transcription (IVT) for SChLAP1 Isoform 1 WT, SChLAP1 D5, and SChLAP1$_{Frag}$ were completed following the procedure from Adams et al. with some modifications.(37) T7 RNA polymerase was a generous gift from Blanton Tolbert's lab (Case Western). No RNase inhibitor was used in any of the steps. IVT of these constructs was performed by mixing: 200 μL 10X Transcription buffer (400 mM Tris-HCl pH 8.0, 100 mM NaCl, 120 mM $MgCl_2$, 20 mM spermidine, 0.1% Triton X-100), 200 μL rNTPs (25 mM equimolar mix), 25 μL T7 RNA polymerase (custom preparation), 25 μL Yeast Inorganic Pyrophosphatase (YIPP, 2 kU/mL, New England Biolabs, USA), 50 μg PCR-amplified DNA template; 100 μL molecular biology grade DMSO (5% final for SChLAP1 WT and SChLAP1 D5 only), and nuclease-free water up to 2 mL. This mixture was aliquoted into 1.5 mL Eppendorf tubes at 500 μL each and incubated at 37 °C for 2-4 hours. DNase I, Proteinase K treatments, and RNA concentration were followed as outlined in Adams et al. Following concentration of the reaction with 100 kDa MWCO Amicon filter to a final volume of 1 mL, size exclusion chromatography was performed at room temperature using Bio-Rad NGC FPLC. A Cytiva (formerly GE Healthcare) HiPrep Sephacryl 16/60 S-500 column was used for SChLAP1 WT and SChLAP1 D5, and Bio-Rad ENrich™ SEC 650 24 mL column was used for SChLAP1$_{Frag}$. An isocratic method was employed using 1X filtration buffer (FB; 50 mM K-HEPES, pH 7.5, 150 mM KCl, 100 µM EDTA pH 8.0). Prior to use, columns were washed with 3 column volumes (CV) of 1:1 RNase ZAP (Ambion) followed by 3 CV nuclease-free water (DEPC-treated MilliQ water), and finally equilibrated with 3 CV 1X FB. Flow rates were between 0.5–0.75 mL/min and 0.5 mL fractions were collected. RNA peaks were monitored using UV$_{255}$ absorbance. For WT and D5, the largest absorbance of the product peak and two surrounding fractions were used for downstream experiments. For SChLAP1$_{Frag}$, one fraction gave sufficient RNA quantities for probing. Nanodrop and/or Qubit confirmed RNA concentration and the RNA sample, and purity was verified using agarose gel electrophoresis before proceeding.

SHAPE Chemical Probing. SChLAP1 WT, D5, and SChLAP1$_{Frag}$ were probed in one biological replicate for each condition (i.e. magnesium concentration or purification method).

SChLAP1 WT and SChLAP1 D5 were typically purified by SEC at approximately ~90 ng/µL for all probing reactions and were diluted with 1X FB if necessary to achieve this concentration. For the semi-native purification (“Native”), RNA from the FPLC was maintained at room temperature prior to addition of $MgCl_2$ to a final concentration of 0, 5, or 20 mM. For denaturing/re-annealing (“Annealed”), RNA from the FPLC was incubated at 95 °C for 3 minutes and placed on ice for 10 minutes before addition of magnesium. Following either prep, 2.5 µL of 10X magnesium concentration was added to 20 µL of ~90 ng/µL RNA for a total of 22.5 µL. These reactions were then incubated at 37 ˚C for 30 minutes. 5NIA (Millipore-Sigma) was dissolved into anhydrous DMSO and were prepared immediately prior to use to limit oxidation. Following this incubation, 5NIA was added to each reaction, flicked to mix, and incubated at 37 °C for 10 minutes. At 10 minutes, each reaction was quenched by adding 33% final volume BME (Sigma-Aldrich) and placed on an ice block pre-chilled to -20 °C. Prior to ethanol precipitation, Sephadex G-50 columns (GE Healthcare) were used to remove the hydrolyzed 5NAI reagent following the manufacturer instructions.

DMS Chemical Probing. *In vitro* DMS probing was completed following protocols from Rouskin lab.(36) PCR, IVT, and FPLC purification of SChLAP1 WT and SChLAP1$_{Frag}$ were completed as described above. Following FPLC purification, RNA samples were adjusted to 300 mM HEPES alongside the addition of magnesium before incubation at 37 °C for 30 minutes prior to DMS probing. DMS (EMD Millipore) was diluted into 100% molecular biology grade ethanol (MBG

EtOH) for a final concentration of 2% upon addition to the RNA sample and was prepared immediately prior to use to limit oxidation. DMS was incubated with the RNA at 37 °C for 5 minutes while shaking at 500 rpm. The reaction was quenched after 5 minutes by adding 33% final volume BME (Sigma-Aldrich) and placed on an ice block pre-chilled to -20 °C.

RNA Reverse Transcription. For SChLAP1$_{Frag}$, only one cDNA and resultant dsDNA amplicon was needed for sequencing. To sequence SChLAP1 WT and D5, we performed reverse transcription using four unique primer sets to generate four overlapping amplicons of maximum 600 bp, as Illumina sequencing has a limit of 600 bp. Each amplicon overlapped with the adjacent amplicons via a 100 base-pair window. Sequences for all primers and amplicons can be found in SI Table 2. For reverse transcription of SChLAP1 WT or SChLAP1 D5, 1 µg of RNA was diluted into 16 µL of nuclease-free water, such that 250 ng RNA went into each RT reaction. For SChLAP1$_{Frag}$, 500 ng of RNA was used for RT. To each reaction, 1 µL 1 µM respective primer was added and incubated at 65 °C for 5 minutes before the reaction was placed on ice. Then, 8 µL of 2.5X MaP Buffer (125 mM Tris-HCl pH 8.0, 187.5 mM KCl, 25 mM DTT, 1.25 mM dNTPs, and 15 mM $MnCl_2$) was added and incubated at 42 °C for 2 minutes before addition of 1 µL SuperScript II Reverse Transcriptase (ThermoFisher) and incubation at 42 °C for 3 hours. Samples were heat inactivated at 70 °C for 15 minutes.

For DMS-treated samples 0.5 µg RNA was mixed with 2 µL 10X FSB (500 mM Tris pH 8.0, 750 mM KCl, 100 mM DTT), 1 µL dNTPs (10 mM equimolar mix), 1 µL TGIRT-III Reverse Transcriptase (200 U/µL, InGex), 1 µL 10 µM respective reverse primer, 1 µL 1M DTT, and brought to 20 uL final with nuclease-free water. For SChLAP1 WT, the reactions were incubated at 65 °C for 90 minutes and inactivated at 85 °C for 5 minutes. For SChLAP1$_{Frag}$, reverse transcription occurred for 3 hours at 55 °C. The reactions were then treated with RNAse H (New England Biolabs, USA) and incubated at 37 °C for 20 minutes (60 minutes for SChLAP1$_{Frag}$).

Library Preparations

Self-made Libraries. Following reverse transcription, SHAPE reactions were prepared for sequencing using a two-step PCR reaction to add on Illumina primers following the amplicon workflow.(38) Samples were individually quantified and diluted to 2 nM each using the High Sensitivity (HS) DNA kit for the Qubit (ThermoFisher) before submitting to the Duke University School of Medicine Sequencing and Genomic Technologies Shared Resource. Samples were pooled on a MiSeq Reagent Kit v2 Nano (2 x 300 bp), and loaded on a MiSeq Illumina sequence per manufacturer's instructions.

Core-made Libraries. Following reverse transcription (outlined above), DMS-probed samples were amplified into dsDNA using blunt-end primers. These samples were submitted to the

Whitehead Institute for Biomedical Research Genome Technology Core for tagmentation prior to sequencing as described above.

In Cellulo Chemical Probing. Probing was performed as previously published(39). Specifically, LNCaP cells were plated at 5 x 10^5 cells per well in a 6-well plate and grown for approximately 2 days. The day of the experiment, cells were washed with 1 mL PBS (Geneclone), and 900 uL complete media was added to each well. To control wells, 100 uL anhydrous DMSO (Invitrogen) was added to the well. For treatment wells, 100 uL of freshly-prepared 250 mM 5NIA was added. Gentle swirling was used to evenly distribute the SHAPE reagent. The reactions were incubated in an incubator at 37 ˚C for 15 minutes. Media was removed, and cells were washed with 1 mL warm PBS. Total RNA from each reaction was extracted with TRIzol reagent and resuspended in 88 µL nuclease-free water. The solutions were treated with TURBO DNAse for 30 minutes at 37C and thereafter purified using RNA Clean/Concentrator 25 columns (Zymo). Reverse transcription reactions contained 0.75 µg of cellular RNA, 2M betaine, 0.5 mM dNTPs, 6 mM $MnCl_2$, and 5 uL SuperScript II in 100 uL reactions. cDNA were purified with Agencourt RNAClean XP beads (Beckman Coulter). dsDNA was generated via PCR using Q5 high-fidelity polymerase using with 12-18 amplification cycles. PCR products were gel-purified using 1% or 2% agarose E-gels (Invitrogen) and Zymoclean Gel DNA Recovery Kit (Zymo). The same overlapping amplicons employed for *in vitro* SChLAP1 were utilized for *in cellulo* probing, and sequencing libraries were assembled using the aforementioned two-step PCR approach. Eluted product was quantified with Qubit dsDNS HS assay. Samples were individually diluted and quantified to 2 nM each using the High Sensitivity (HS) DNA kit for the Qubit (ThermoFisher) before submitting to the Duke University School of Medicine Sequencing and Genomic Technologies Shared Resource. Samples were pooled using the MiSeq Reagent Kit v3 (2 x 300 bp) and loaded on a MiSeq Illumina sequencer per manufacturer's instructions. In cell probing was performed in one biological replicate.

Bioinformatics Pipeline. SHAPE reactivity profiles, error estimates, mutation counts, and sequencing depths were obtained using the ShapeMapper pipeline (v 2.1.5) developed by the Weeks lab (UNC-Chapel Hill) (40). All default parameters were used. Samples were filtered at ≥1,000 nucleotide read depth. Amplicons for a single transcript, i.e. SChLAP1 WT or D5, were manually concatenated, where overlapping chemical reactivity regions were averaged within a single treatment condition to produce a final .map file. This final .map file was then submitted to SuperFold for secondary structure prediction, base-pairing probability, and Shannon entropies. These final .map files were also used in ΔSHAPE calculations using default parameters to produce ΔSHAPE plots.

For DMS probing, mutation counts were calculated in the DMS-MaPSeq program from the Rouskin Group (36). DMSO-only mutation counts were subtracted from DMS-probed mutation counts. The mutation counts for adenosine and cytosine nucleotides were normalized to the same scale as SHAPE data following previously established methods.(41) Specifically, the mutation counts 1.5 times above the interquartile range were removed, and the remaining top 10% of values were averaged to obtain a normalization factor. Thereafter, the mutation counts of all adenosine and cytosine nucleotides were divided by this normalization factor to obtain a normalized reactivity. Normalized reactivities were manually overlaid onto the SHAPE-informed minimum free energy structures.

Correlation Analyses. Spearman correlation coefficients were calculated in Graphpad Prism (Version 9). Before comparison, nucleotides within primer binding sites or undefined reactivities were removed. If a nucleotide had an undefined reactivity in one SHAPE profile, it was removed from all reactivity profiles before comparison.

Phylogeny Analysis. Identified or predicted sequences of SChLAP1 in all primates were aligned using Clustal Omega. All sequences are aligned to hsSChLAP1 isoform 4 as this contains all available exons. Accession numbers for these sequences can be found in SI Table 4. Phylogenetic tree was made through Clustal Omega with the following parameters: guidetreeout: true; dismatout: false; dealign: true; mbed: true; mbediteration: true; iterations: 0 gtiterations: -1; hmmiterations: -1; outfmt: clustal_num; order: aligned; stype: rna

RESULTS

Sequence alignments reveal potentially functional regions of SChLAP1

Well-conserved and highly structured RNA motifs within viruses, bacteria, and human endogenous transcripts are typically indicative of critical function (42-45). We thus performed an nBLAST alignment on the entire SChLAP1 WT (isoform 1) sequence to look for conserved SChLAP1 sequences across other species. This alignment returned 14 non-human primate sequences with >85% sequence conservation (Fig. 1A-B and Supplementary Table 3), all of which are predicted to be non-coding RNA but have no documented functional annotation to date; this search did not return any lower eukaryotic or prokaryotic species. In gorilla and chimpanzee, our two closest primate relatives with a putative SChLAP1 homologue, these sequences are also located on chromosome 2. Alignments of the 14 identified sequences with the other seven human SChLAP1 isoforms indicated higher conservation for some exons over others. In particular, exons 1, 2, and 7 show the strongest conservation across all species and isoforms (Fig. 1B), supporting their functional importance. Covariation analyses performed using R-scape(46) showed no significant differences among all 14 sequences (data not shown), in line with the high level of sequence conservation. Although the temporal proximity in evolution limits our ability to perform covariation analyses, we hypothesize that these conserved exons sequences and/or structures may be crucial to SChLAP1 function in normal and diseased states.

The nBLAST search further identified an interesting sequence alignment between SChLAP1 exon 1 and ERV-9 LTR U5 in orangutan: the 3′–end of ERV-9 LTR U5 (~320 n.t.) overlaps with SChLAP1 exon 1 with >90% sequence identity (Fig. 1C), suggesting this exon was acquired by retroviral inclusion. Liu and Eiden found that this LTR was inserted into the primate genome during evolution between gibbons and orangutans (~15-18 million years ago) and persisted across evolution through to humans (47). ERV-9 LTRs are currently only found in the primate genome and could explain why SChLAP1 homologs are not found in lower, non-primate species in our search. This hypothesis is further supported by the general observation that a greater number of lncRNAs are observed in higher-order species (48).

Identification of *in vitro* structure and putatively functional regions within SChLAP1 by SHAPE-MaP

We next set out to determine the first *in vitro* secondary structure model of SChLAP1 and to identify important structures in the transcript. We performed SHAPE-MaP on *in vitro* transcribed SChLAP1 WT isoform 1 (SChLAP1 WT) to produce the first experimentally informed secondary structure model of SChLAP1. We chose the SHAPE reagent 5-nitroisatoic anhydride (5NIA) for consistency with downstream *in cellulo* work, as it has enhanced cell permeability compared to other SHAPE reagents and precedence for use in prostate cancer cells (39). We used a semi-native purification protocol for the transcribed RNA, where heat denaturing or harsh buffer exchanges are avoided, as this approach was previously found to maintain a homogenously-folded RNA compared to denaturing protocols (Fig. S1) (37).

The resulting structure models were generated using the SHAPEMapper (v.2.1.5) (40) and SuperFold (38) pipelines (Fig. 2). A varied combination of secondary structures and single stranded regions are observed across the length of the transcript (Fig. 3). For example, at the 5′–end of SChLAP1 WT (1-155 n.t.) we observed a four-way junction (4WJ), an RNA structural motif that commonly functions as a protein scaffold (49). This 4WJ is contained within exon 1, one of three well-conserved exons across the above primate sequences, and also contained within the aforementioned retroviral alignment, which indicates a highly conserved structural and likely functional relevance. We also observed a large central loop structure (828-871 n.t.) and a highly extended stem loop structure towards the 3′–end (975-1130 n.t.), both of which being present in the conserved exon 7. SHAPEMapper is also capable of identifying putative pseudoknots based on sequence context and chemical reactivity, and our analysis did not identify any putative pseudoknots. We also employed the program G4Hunter (50), which predicted several G-Quadruplexes (GQs) within SChLAP1, including in the D5 region discussed below, which may play a functional role in protein recognition. Although initially identified in DNA, GQs are increasingly recognized as important functional tertiary structures within RNA transcripts, including long non-coding RNA,(51) and have the potential for small molecule targeting (52).

We observed a lack of chemical reactivity in a poly(A) stretch of SChLAP1 WT (1088-1104 n.t.), which is consistent with work from Kladwang et al. who discovered that chemical modifications in poly(A) regions are bypassed by reverse transcriptases, resulting in incongruous chemical modification frequency and mutational profiling results (53). To avoid biasing our resultant model, we manually set this poly(A) stretch to “undefined” in the input .map file for all SuperFold and ΔSHAPE calculations of SChLAP1 WT and SChLAP1$_{\mathrm{Frag}}$ (reported below); setting these nucleotides to undefined removes all chemical reactivity data so that the programs use only RNA folding thermodynamics to predict base-pairing.

We then used our *in vitro* SHAPE data to identify putative functional structures within the SChLAP1 transcript. RNA structural elements that are: 1) highly structured (low SHAPE reactivity) and 2) well-determined (low Shannon entropy, i.e. low shape/shannon (lowSS) regions), were initially characterized in HIV-1 genomic RNA for significant enrichment in previously unknown functional roles (35). This metric has since been used to identify novel functional regions, including in the Dengue virus RNA genome (54) and XIST lncRNA (55). In our SChLAP1 model (Fig. 2B), we identified two major lowSS regions: the 5′–end of the transcript (1-474 n.t.) and between 989-

1054 n.t at the 3′–end (Fig. 2B). The 5′–end region corresponds to exons 1 and 2, and the 3′–end region corresponds to exon 7 within the D5 region (see below). Thus, the most robustly conserved exons within SChLAP1 all contain lowSS regions, further supporting their functional significance.

To further validate our model, we performed independent probing experiments with dimethyl sulfate (DMS), which preferentially modifies the Watson-Crick-Franklin face of adenines and cytidines over guanines and uridines at neutral pH (56). The DMS probing results supported the SHAPE probing results and agree with the models derived from SuperFold for both this experiment and the other conditions reported below (Fig. 3 and S6).

Identification of protein-binding sites within SChLAP1 using SHAPE-MaP

We next performed *in cellulo* SHAPE probing to identify putative protein binding sites within SChLAP1 (Fig. 4). Comparing the *in cellulo* reactivity data to *in vitro* reactivity data has previously identified protein binding sites in other lncRNAs and viral RNAs (55,57-59). We performed SHAPE probing in LNCaP prostate cancer cells because they are derived from supraclavicular lymph nodes (with lymph nodes being a frequent site of metastatic prostate cancer localization) (60), have high SChLAP1 expression (16), and have a known metastatic phenotype. Multiple motifs within the minimum free energy (MFE) structure are consistent between *in cellulo* and *in vitro* models, namely the 4WJ, large loop structure and extended stem loop structure (Fig. S2). These results supported the literature-based premise that differences between SChLAP1 *in cellulo* and *in vitro* reactivities are due to *in cellulo* protein binding and thus identify candidate protein binding sites (61). Specifically, protein binding sites are identified by reduced nucleotide reactivity with the chemical probing adduct *in cellulo*, i.e. nucleotides that are less flexible or are masked due to protein binding have fewer instances of chemical probing adduct formation *in cellulo* as compared to *in vitro*.

In our ΔSHAPE analysis, we observed that the majority of in-cell protections identified for SChLAP1 were within exons 1, 2, and 7 (111/114 instances; Fig. 4C), further supporting the functional significance inferred from phylogenetic analysis (Fig. 1) and the presence of lowSS regions (Fig. 2). Within exon 1, many in-cell protections are localized to the aforementioned 4WJ immediately adjacent to the 5′–end of the transcript (Fig. 4C-D) and within the retroviral alignment (Fig. 1C), supporting the functional relevance of this structure. *In cellulo* protections in SChLAP1 exon 2 are consistent with recent work from Ji et al. where binding between SChLAP1 exon 2 and HNRNPL facilitated activation of the NF-κB pathway in glioblastoma (25).

Significant protein binding with exon 7 is in line with previous work from the Chinnaiyan group. Preliminary work from Sahu et al. revealed that a deletion of 250 nucleotides (n.t.), from position 1001-1250 of SChLAP1 isoform 1 (1,436 n.t.), inhibited SChLAP1-driven invasion and binding to the SWI/SNF complex (62). These results suggested that inhibition of SChLAP1:SWI/SNF complexation by removing a sequence and/or structure specific to this region (termed D5) reduced the invasive cancer phenotype. Within Exon 7, the Deletion #5 (D5) region shows particularly abundant in-cell protections (Fig. 4C-D) and contains a lowSS region (Fig. 2), further supporting its role in protein binding and potentially the metastatic phenotype as proposed by Sahu et al. (62).

The secondary structure of SChLAP1 is sensitive to preparation and magnesium concentration

We proposed that the *in vitro* structure model presented above is a physiologically relevant structure through the use of native RNA folding and physiological magnesium concentrations. To evaluate this assumption, we performed parallel analyses of *in vitro* SChLAP1 under different purification methods and magnesium concentrations. Harsh purification (denaturing RNA followed by either slow- or fast-cooling) is a method commonly used prior to RNA chemical probing to remove RNA aggregates and produce a more homogenous structure population. While we avoided denaturing methods in the model generated above, we were curious how this process would affect the secondary structure of SChLAP1, particularly the well-conserved and highly structured regions we identified by probing completed after semi-native purification. To this end, we denatured and snap-cooled purified SChLAP1 WT (3 minutes at 95 ˚C, 10 minutes on ice, and 30 minutes at 37 ˚C) prior to probing.

In addition, we explored the role of magnesium in both purification approaches, as it is known to play crucial roles in RNA folding (63) and stabilization, and is the most populous divalent metal ion found in RNA structures (64). As RNA co-transcriptionally folds as it exits the RNA polymerase, magnesium interacts with the structure to facilitate the formation of the most energetically favorable conformation(s) (63). Free magnesium can diffusely interact with RNA or can be coordinated, either directly or mediated by water molecules, by the RNA backbone or specific nucleobase moieties, alleviating charge repulsion in a folded RNA state (63). In these experiments, magnesium was titrated into the RNA sample at 0, 5, or 20 mM magnesium final concentrations before probing.

Comparison of the SChLAP1 WT Native versus the snap-cooled (Annealed) samples across the three different magnesium concentrations revealed varying structural landscapes (Fig. 5). Correlation coefficients were calculated between the chemical reactivities of all conditions, i.e. Native and Annealed SChLAP1 WT at 0, 5, and 20 mM magnesium (Fig. 5C). For Native SChLAP1, altering the magnesium concentration did not globally impact the structure, as the Spearman correlation coefficients between each natively-folded condition are high (over 0.7). For comparison, a study by Frank et al. considered a correlation coefficient of approximately 0.70 to be robustly similar when comparing several probing experiments of the lncRNA GAS5 (59). However, reannealing resulted in significant perturbation of the structure, as demonstrated by significant reduction in correlation coefficients between Native and Annealed SChLAP1 (Fig. 5C). This result is also reflected through significant changes in the arc diagrams in the Annealed samples (Fig. 5A), as well as alterations in the MFE structures from these experiments (Fig. S3, S4, and S5). Interestingly, incubation in 20 mM magnesium after reannealing shows higher correlation coefficients to natively folded SChLAP1, though alterations in the arc diagrams are again observed. In addition, we performed independent DMS-MaPseq experiments for these snap-cooled conditions (Fig. S6) and again found that DMS data supported the SHAPE-informed structure models for all experiments (Fig. S3, S4, and S5).

Surprisingly, four of the six WT structures did not meet the chemical probing reactivity threshold (8%). This threshold is documented in the SHAPEMapper workflow as empirically established, and failure to meet this threshold is presumed to mean that the RNA is highly structured (Fig. 5A) (40). Both of the two acceptable structures were from natively-purified SChLAP1, suggesting that the Annealed preparation forms an RNA that is less sensitive to chemical probing as compared to the Native preparation. This result is consistent with observations from the Weeks group, where reannealing of Dengue virus genomic RNA resulted in a higher proportion of nucleotides with low SHAPE reactivity (65). Indeed, ΔSHAPE analysis of Native versus Annealed for all three magnesium concentrations revealed localized differences in chemical reactivities and supports that the reannealing process introduced structures not present in the semi-natively purified samples (Fig. 5B).

The magnesium titration approach for semi-natively folded SChLAP1 WT facilitated the identification of local magnesium-dependent regions in the transcript. ΔSHAPE identified significant changes in SHAPE reactivity between 0 and 5 mM magnesium, supporting the presence of magnesium-dependent structures in physiologically-relevant concentrations (Fig. 5E). Of note, a region at the 3′–end of the transcript (1,235-1,285 n.t), which is within the D5 region, shows reduced chemical reactivity in the presence of 5 mM magnesium, suggesting a highly folded and thus functionally relevant structure. While these changes in reactivity do not distinguish between direct magnesium binding or through-space base pairing induced by magnesium, these findings support that magnesium can play crucial roles in particular SChLAP1 regions.

Altogether, our comparison of the different preparations and magnesium concentrations support that the structure derived from SChLAP1 WT Native preparation is the most relevant for our studies, and that magnesium plays an important role in local secondary structures, particularly in the D5 region. Moving forward in our analyses, we thus decided to focus on comparison of WT and the aforementioned D5 construct using natively-folded RNA.

Removal of the D5 Region does not globally perturb SChLAP1 folding

With evidence for functionally important structures in the D5 region, we investigated if removal of this region, i.e. Deletion 5 construct (D5), results in global changes in chemical probing reactivity. A global change in D5 chemical reactivities as compared to WT would suggest that this region impacts the global folding of SChLAP1, whereas similar reactivities would suggest that the observed change in phenotype can be localized to this specific region of the transcript. We thus performed SHAPE-MaP on the SChLAP1 D5 transcript. The transcript was semi-natively purified in an identical manner to SChLAP1 WT and also incubated in several magnesium concentrations before probing to evaluate magnesium responsiveness. Notably, the 4WJ and large loop structure identified in the WT construct persist in the D5 construct (Fig. S7). The sequence that forms the 3′–extended stem loop structure is deleted in the D5 mutant. We completed ΔSHAPE analysis of the first two amplicons (1-903 n.t.) for Native versions of WT and D5 based on the hypothesis that we would see a large change in the first two amplicons if removal of the D5 region caused significant structural perturbation across the transcript (Fig. S7). We tested all three magnesium concentrations as different magnesium concentrations would be expected to amplify any differences. For 0 and 5 mM magnesium conditions, we observed minor alterations in reactivity for respective WT versus D5 mutants, and most alterations were located at open loops or did not significantly impact the predicted structure, e.g. a reactivity change on one side of a stem not reciprocated on the other side of a stem structure. Significant differences, however, were observed at 20 mM magnesium. We calculated correlation coefficients for the first two amplicons of WT and D5: In 0 mM and 5 mM magnesium, the correlation coefficients were 0.69 and 0.65, supporting the similarities in these structures (Fig. S7). In support of our ΔSHAPE analysis, correlations dropped to 0.43 in the presence of 20 mM magnesium (Fig. S7).

Collectively, these data indicate that structural rearrangements at the 3′–end of the transcript due to the 250 n.t. deletion do not induce long-range secondary structure rearrangement, and, more broadly, that the phenotypic change seen with the D5 mutant as compared to the WT can be localized to the absence of specific RNA structures at the 3′–end. While this work does not rule out the possibility of long-range tertiary structures or structural rearrangement within SChLAP1, it supports the role of D5–localized secondary structures in protein recognition and metastasis.

SChLAP1$_{\text{Frag}}$ forms a stable structure insensitive to preparation or magnesium concentration

We investigated the exon 7/D5 region for specific structural features that might be responsible for protein recognition based on the data above, namely that: 1) exon 7 is highly conserved across human isoforms and non-human primates (Fig. 1); 2) ΔSHAPE data supports protein binding within exon 7 (Fig. 4); and 3) the previous deletion study supported the role of exon 7 in SWI/SNF recognition. In the arc diagrams for full-length, natively purified SChLAP1 (Fig. 2A), we observed that the D5 region (1000-1250 n.t.) appears to be participating in a longer-range structure encompassing approximately the last 500 nucleotides of SChLAP1. Using RNAStructure(66), we observed a robustly predicted structure within exon 7 ranging from 949-1428 n.t. This region, termed SChLAP1$_{\text{Frag}}$, was independently transcribed and subjected to chemical probing in the same manner as *in vitro* SChLAP1 WT (Fig. S1) to see whether it is indeed a structurally robust motif that folds independently of the surrounding SChLAP1 sequence.

The structural architecture for SChLAP1$_{\text{Frag}}$ (natively purified in 5 mM magnesium) was consistent with the structure of this region within the context of the full-length transcript (Fig. 6A, also natively purified in 5 mM magnesium). We note that the predicted probabilities for several base-pair arcs are higher in SChLAP1$_{\text{Frag}}$ than in full-length SChLAP1 WT, and we hypothesize that this might be due to the disruption of several poorly predicted, long-range base pairing interactions, thereby allowing localized base pairs to more stably form in the context of the fragment.

We next used Spearman correlation coefficients to compare the reactivities between SChLAP1$_{\text{Frag}}$ and the respective nucleotides in SChLAP1 WT as a way to evaluate structural similarity. Across all preparations and magnesium concentrations, the Spearman correlation between the fragment and Native WT transcript was consistently ~0.7, suggesting a strong correlation between these chemical reactivities (Fig. 6B). This finding supports that the resultant structure model of SChLAP1$_{\text{Frag}}$ is highly similar to the same sequence found within natively purified, full-length SChLAP1. These correlations drop precipitously when SChLAP1$_{\text{Frag}}$ is compared to reactivities in Annealed WT, suggesting that thermal denaturation and reannealing of the full-length transcript disrupts the co-transcriptional folding of this region.

We then evaluated whether varying magnesium and annealing preparation conditions impacted the structure of SChLAP1$_{\text{Frag}}$ itself, similar to the aforementioned experiments with SChLAP1 WT, as a means to interpret the stability of the structure and thus its functional importance. After semi-native purification, SChLAP1$_{\text{Frag}}$ was either maintained at room temperature or snap-cooled and was incubated at either 0, 5, or 20 mM magnesium, yielding 6 conditions as was the case for SChLAP1 WT. We again used Spearman correlation coefficients to make pairwise comparisons between conditions. Interestingly, we observed high correlation coefficients (Spearman's $\rho>0.9$) across all SChLAP1$_{\text{Frag}}$ vs SChLAP1$_{\text{Frag}}$ conditions (Fig. 6B), indicating that the secondary structure of SChLAP1$_{\text{Frag}}$ is not significantly impacted by magnesium concentration or RNA preparation. Similarly, high correlation coefficients were observed for semi-natively purified SChLAP1$_{\text{Frag}}$ with independent DMS probing experiments. These correlation coefficients remained high when compared to snap-cooled SChLAP1$_{\text{Frag}}$ refolded in 5 mM magnesium (Fig. S6D), further supporting that the folding of SChLAP1$_{\text{Frag}}$ secondary structure is stable regardless of preparation.

With this data in hand, we propose that SChLAP1$_{\text{Frag}}$ is a critical, functional RNA structure module within SChLAP1. In addition, given that this domain has high in-cell protections from our *in* *cellulo* data, we propose that this region within SChLAP1 may be a hub for protein recognition, of both SWI/SNF and other proteins, which facilitates prostate cancer progression.

DISCUSSION

The study of lncRNA structure-function relationships is of great interest to the scientific community from both a basic science perspective as well as therapeutic one. While three-dimensional analysis of lncRNA structure by methods such as x-ray diffraction (XRD), nuclear magnetic resonance (NMR), and cryo-electron microscopy (cryo-EM) are limited, in large part due to lncRNA size and conformational dynamics,(67) chemical probing methods such as SHAPE-MaP combined with phylogenetic analyses have provided insight into these large biomolecules (68-70). Importantly, these chemical probing methods have often informed and/or reproduced three-dimensional structure findings and even expanded upon them as seen in with rRNA (71) and XIST (72). Our phylogenetic analyses and chemical probing studies have produced the first secondary structure model of the lncRNA SChLAP1, which serves as a foundation for further biochemical and biophysical analyses.

A potential mechanism for SChLAP1 function

Comparison of SChLAP1 human and non-human primates sequences revealed a high degree of sequence conservation, particularly among exons 1, 2, and 7. These regions exhibited well-conserved secondary structures in SChLAP1 WT, indicating their importance in SChLAP1 function, although the normal function of SChLAP1 remains unknown. The recognition sites identified here are in agreement with previous studies that have characterized SChLAP1-protein interactions. In particular, our work aligns with the Sahu et al. deletion studies (62) as well as other work that has uncovered functional roles for RNA:SWI/SNF recognition (22,27-33) in spite of noted promiscuous binding for SWI/SNF.(21-23) However, more mechanistic characterization of SChLAP1:SWI/SNF interaction is needed.

Analysis of SChLAP1 exon 1 indicates this sequence was recently incorporated into the human genome via endogenous retrovirus 9 (ERV-9) insertion, specifically to lncRNA genes, prior to humans splitting from orangutans (73). This insertion was recently shown to be important for lncRNA activity as Alfeghaly et al. reported the ERV-dependent *trans* activity of ANRIL (74). Specifically, retrovirally incorporated Exon 8 within ANRIL facilitated its ability to bind chromatin at specific genes and altering their expression. Comparative analysis of SChLAP1 via chemical probing indicated reduced reactivity in exon 1 in prostate cancer cells versus *in vitro*. While it is generally accepted that lower chemical reactivity *in cellulo* as compared to *in vitro* indicates an RNA:protein interface, this could also indicate an RNA:DNA interface as lncRNA:DNA hybridization is a known lncRNA function.

Given the in-cell protections within exons 1 and 7, and the alignment observed between exon 1 and a ERV-9 insertion with known roles in chromatin recognition, we propose that SChLAP1 may sometimes function through chromatin recognition mediated by exon 1 and scaffold proteins at these genes mediated by RNA structures within Exon 7 / SChLAP1$_{Frag}$ (Fig. 7). SChLAP1-mediated regulation of transcription via chromatin binding is also supported by a recent study by Huang and Tang, which found that SChLAP1 promoted the binding of EZH2 and DNMT3A to the promoters of several tumor-suppressing miRNAs, reducing their expression (24). While there are still remaining questions regarding the intermolecular interactions that facilitate SChLAP1's role in oncogenic progression, our sequence alignments and structural approach support a model where SChLAP1 interacts with various proteins, including the aforementioned SWI/SNF, PRC2, DNMT3A, AUF1, and PRC2, via highly conserved and structured exons, supporting SChLAP1's function and favoring aggressive cancer phenotypes. While it is likely that SChLAP1$_{Frag}$ interacts with multiple proteins, we hypothesize that this domain within SChLAP1 also contains a recognition element for the SWI/SNF complex. We propose that chromatin binding within exon 1 and protein binding within exon 7 may be a more general mechanism for SChLAP1 function on the basis of the data presented herein.

Insights into SChLAP1 and SChLAP1$_{Frag}$ garnered by varying probing conditions

While we considered the structure of natively folded SChLAP1 in 5 mM magnesium as the most physiologically relevant structure based on literature precedence (37), we explored denaturing preparations and other magnesium concentrations to observe how methods commonly used in the field could impact SChLAP1 structure. Significant changes in SHAPE reactivity were observed between 0 and 5 mM magnesium, supporting the formation of magnesium-dependent structures at physiologically relevant magnesium concentrations. Additional alterations in chemical reactivity are observed at 20 mM magnesium as compared to 5 mM magnesium. While 20 mM is significantly higher than physiological concentrations, these results suggest that magnesium concentration plays important roles in chemical probing experiments and that a magnesium screen should be completed for all *in vitro* chemical probing experiments, as has been previously suggested by the Pyle lab (37).

In previous work, native purification was found to be necessary for homogenous isolation of the lncRNA HOTAIR (69). In our work, we found that reannealing full-length SChLAP1 also results in structural rearrangement of the RNA (Fig. 5). In contrast, the secondary structure of SChLAP1$_{Frag}$ appears unaffected/unperturbed by denaturing preparation. The consistency of chemical reactivity for SChLAP1$_{Frag}$ across various preparations is remarkable and poises SChLAP1$_{Frag}$ for use in various downstream applications, including X-ray crystallography or cryo-electron microscopy. In addition, SChLAP1$_{Frag}$'s robust secondary structure supports its relevance in SChLAP1 function in metastatic prostate cancer, as highly conserved and well-folded regions are known to be important in a wide range of endogenous and pathogenic RNA (75,76). Specifically, we observed that this fragment is 1) conserved across non-human primates, 2) highly structured, and 3) has strong protein binding indications. Given that protein interactions by SChLAP1 have been previously linked to SChLAP1's function in aggressive prostate cancer, we propose that there is a likely role for $\text{SChLAP1}_{\text{Frag}}$ in mediating oncogenic progression via protein recognition. We also propose that this fragment is a putative independent folding domain, though further work is needed to assess the limits of this domain on the adjacent 5′–end.

In conclusion, we observe that SChLAP1 has a complex secondary structure around which multiple lines of evidence support relationships between structured regions and functional roles. Our analysis of evolutionary conservation and chemical probing identified a set of structures putatively crucial to SChLAP1's role in oncogenic progression. With fragment folding analysis alongside in-cell probing and previous deletion studies, our work identifies $\text{SChLAP1}_{\text{Frag}}$ as a tractable proxy for structural analysis and screening inhibitors of RNA:protein complexation. We believe that the insights developed here will facilitate both fundamental understanding of prostate cancer progression and the development of specific therapeutic strategies against SChLAP1.

SUPPLEMENTARY DATA

Supplementary Data are available online.

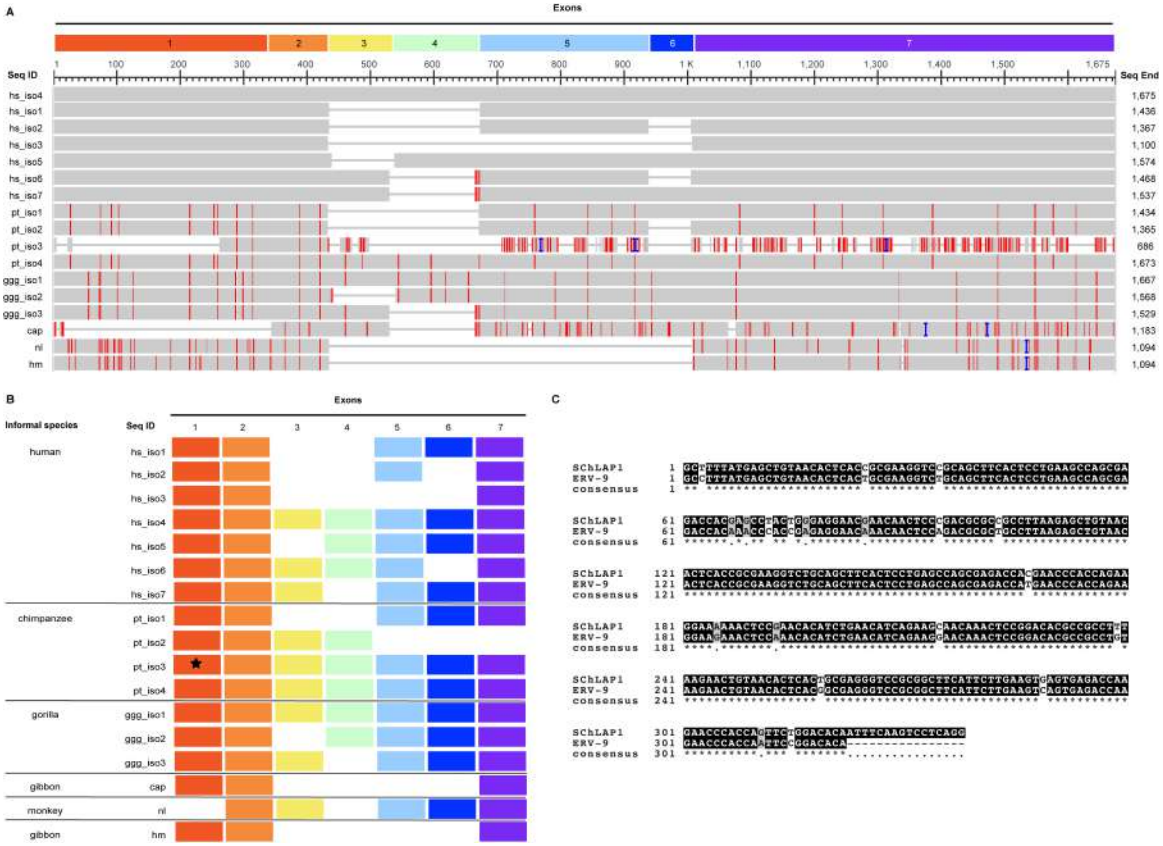
ACKNOWLEDGEMENT

The authors would like to acknowledge all members of the Hargrove lab, past and present, for invaluable feedback and support. The authors also acknowledge Bill Day, Michael Peterson, and İrem Altan for assistance with computation. The authors thank Whitehead Institute for Biomedical Research Genome Technology Core for DMS sample preparation and sequencing, as well as Silvi Rouskin and Fengrui Zhang for help with DMS data processing. We thank the Duke University School of Medicine for the use of the Sequencing and Genomic Technologies Shared Resource, which provided MiSeq service, as well as the Duke Compute Cluster for use of their cluster.

FUNDING

The authors acknowledge financial support from the Prostate Cancer Foundation Young Investigator Award and the Office of the Assistant Secretary of Defense for Health Affairs through the Prostate Cancer Research Program [grant number W81XWH2010188]. E.J.M. was supported in part by Duke University Center for Biomolecular and Tissue Engineering Ruth L. Kirschstein [grant number T32GM008555]. J.P.F. was supported in part by Duke University Department of

Biochemistry fellowship and Duke University School of Medicine scholarship funding. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the Department of Defense. In the conduct of research utilizing recombinant DNA, the investigator adhered to NIH Guidelines for research involving recombinant DNA molecules.



```
SChLAP1     1 GCTTTTATGAGCTGTAACACTCACCGCGAAGGTCCGCAGCTTCACTCCTGAAGCCAGCGA
ERV-9       1 GCCTTTATGAGCTGTAACACTCACTGCGAAGGTCTGCAGCTTCACTCCTGAAGCCAGCGA
consensus   1 ** ********************* ********* *************************

SChLAP1    61 GACCACGAGCCTACTGGGAGGAACGAACAACTCCCGACGCGCCGCCTTAAGAGCTGTAAC
ERV-9      61 GACCACAAACCCACCGAGAGGAACAAACAACTCCAGACGCGCTGCCTTAAGAGCTGTAAC
consensus  61 ******.*.** ** *.*******.********* ******* *****************

SChLAP1   121 ACTCACCGCGAAGGTCTGCAGCTTCACTCCTGAGCCAGCGAGACCACGAACCCACCAGAA
ERV-9     121 ACTCACCGCGAAGGTCTGCAGCTTCACTCCTGAGCCAGCGAGACCATGAACCCACCAGAA
consensus 121 ********************************************** *************

SChLAP1   181 GGAAAAAACTCCGAACACATCTGAACATCAGAAGCAACAAACTCCGGACACGCCGCCTTT
ERV-9     181 GGAAGAAACTCCAAACACATCTGAACATCAGAAGGAACAAACTCCGGACACGCCGCCTGT
consensus 181 ****.*******.********************* *********************** *

SChLAP1   241 AAGAACTGTAACACTCACTGCGAGGGTCCGCGGCTTCATTCTTGAAGTGAGTGAGACCAA
ERV-9     241 AAGAACTGTAACACTCACGGCGAGGGTCCGCGGCTTCATTCTTGAAGTCAGTGAGACCAA
consensus 241 ****************** ***************************** ***********

SChLAP1   301 GAACCCACCAGTTCTGGACACAATTTCAAGTCCTCAGG
ERV-9     301 GAACCCACCAATTCCGGACACA----------------
consensus 301 **********.*** *******................
```

Figure 1. Sequence alignment (Clustal Omega) of all primate genes aligned to human SChLAP1. A) Sequences were aligned to human SChLAP1 (hsSChLAP1) isoform 4 (iso 4) as this isoform contains all possible exons (top). Gray is overlapping sequence and red is mutations/deletions. B) Exons in rainbow order correspond to exons in (A). All accession codes can be found in SI Table 1. Star = contains 86/338 nucleotides for exon 1 C) Alignment between SChLAP1 exon 1 and ERV-9 LTR sequence.

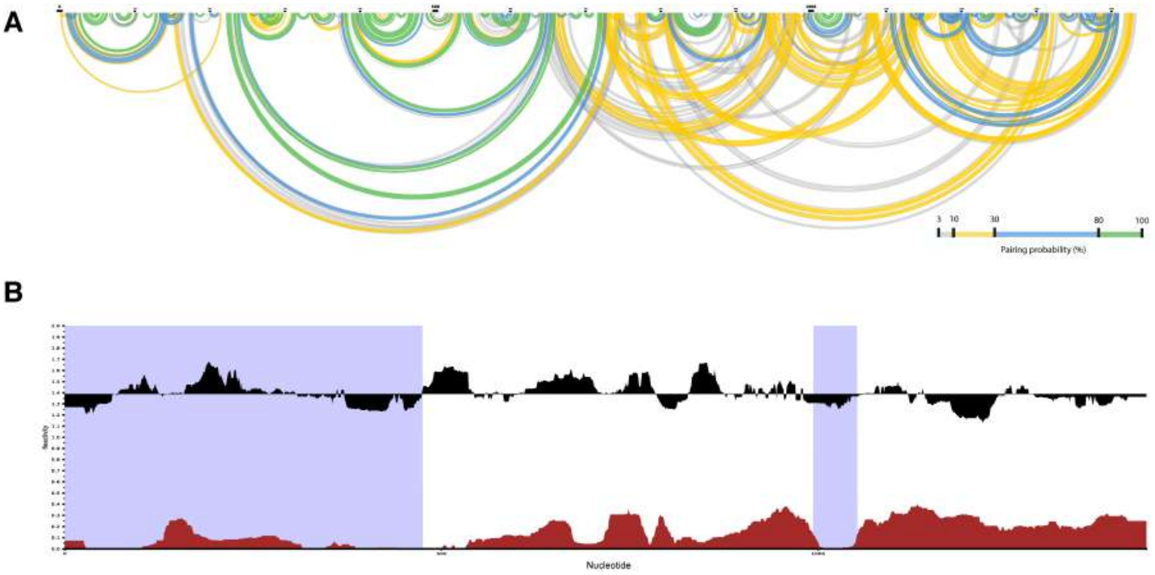


Figure 2. SHAPEMapper analysis of *in vitro* SChLAP1 WT, semi-natively purified in 5 mM magnesium. A) Arc diagram for SChLAP1 WT, where predicted base pairs are shown as arcs along the length of the transcript. B) Shannon Entropy plot. Black plots are chemical probing reactivities relative to the median reactivity, and red plots are Shannon entropy. Purple bars indicate regions of low SHAPE reactivity and low Shannon entropy (lowSS).

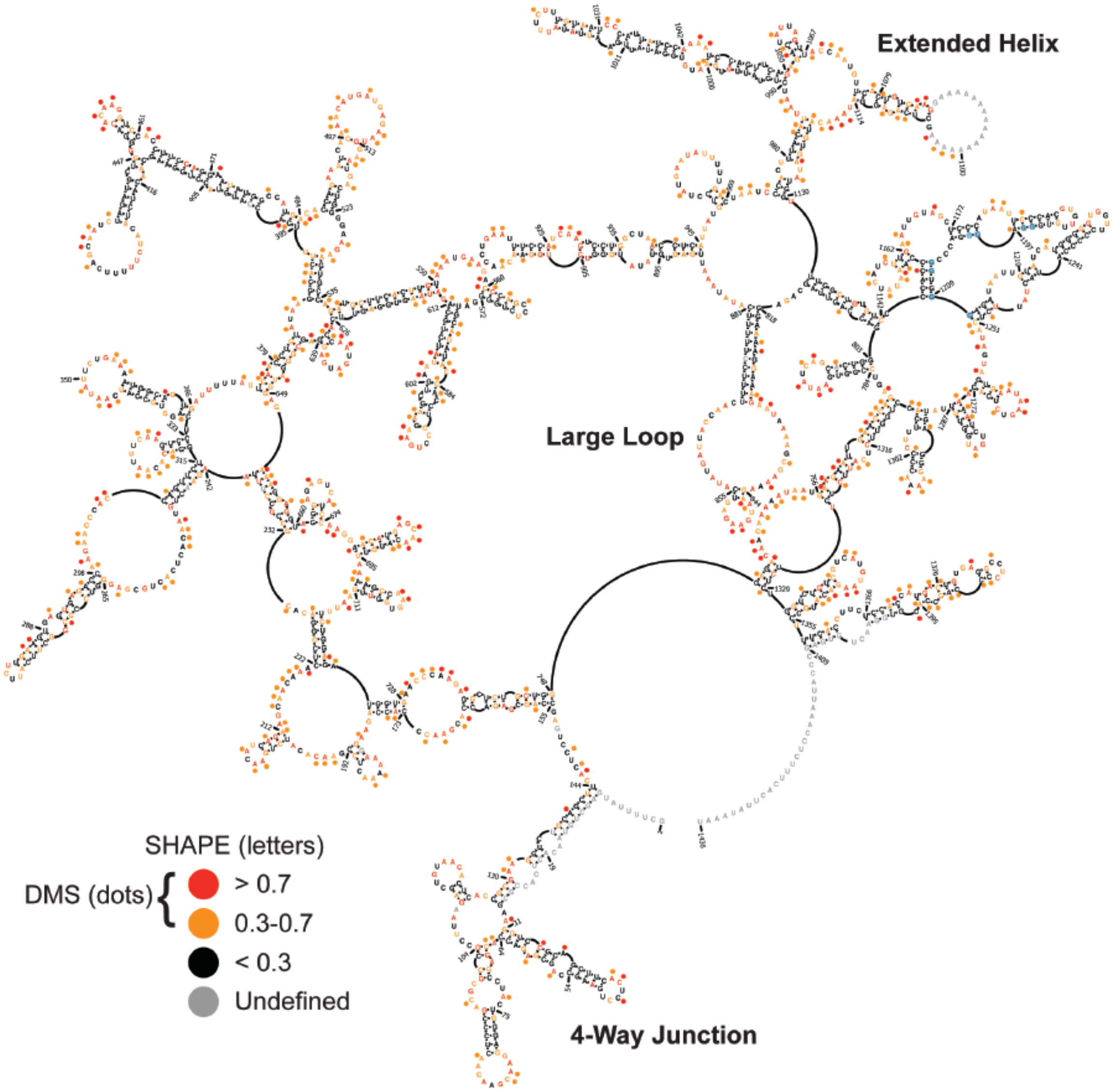


Figure 3. Minimum free energy (MFE) structure of SChLAP1 WT with SHAPE and DMS reactivities annotated. Blue dots (nts 1195-1210) indicate G-quadruplex prediction by G4Hunter (50).

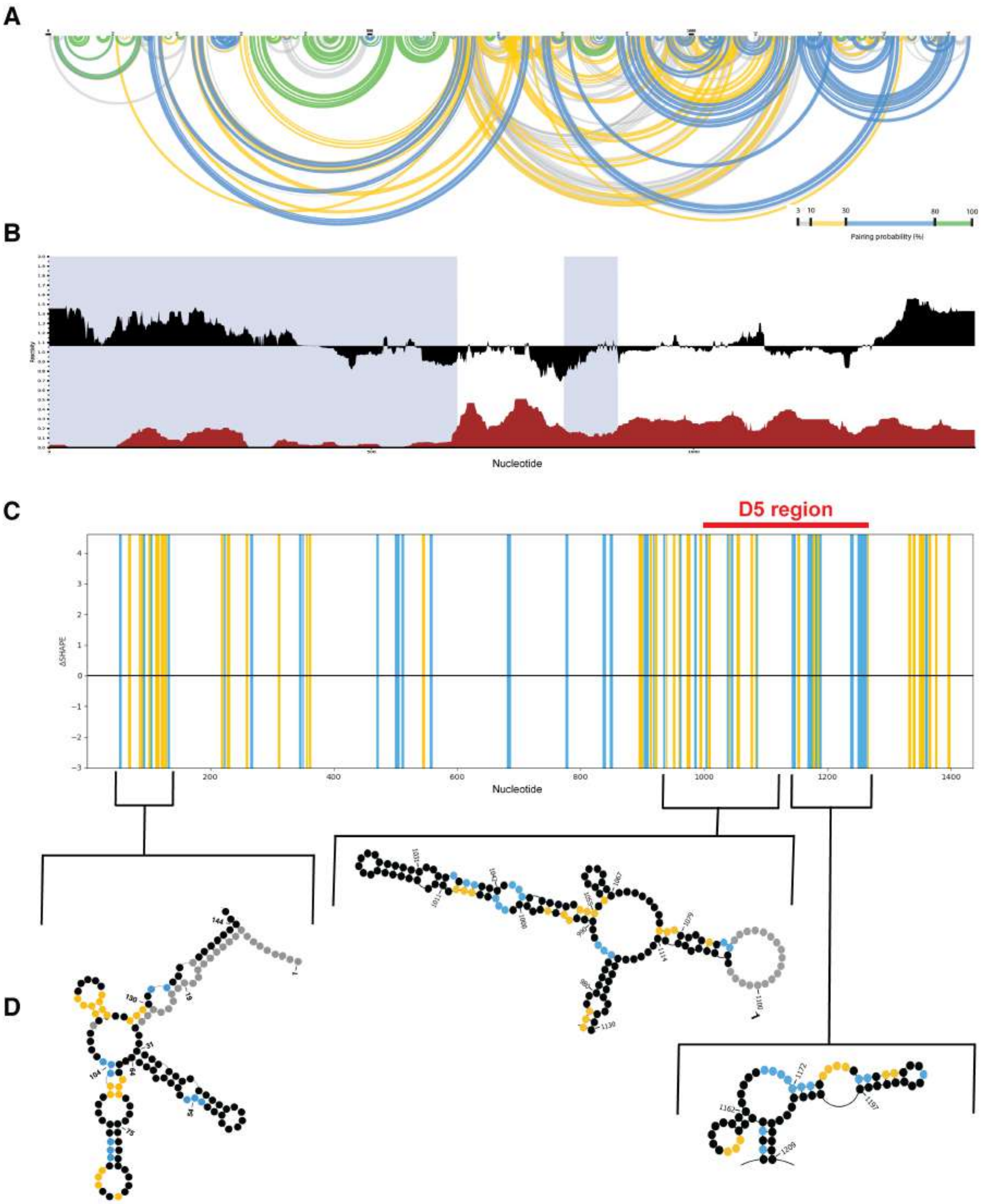


Figure 4. *In cellulo* SHAPE probing of endogenous SChLAP1 in LNCaP cells. A) Arc diagram for *in cellulo* SChLAP1. B) Shannon entropy plot. C) ΔSHAPE of SChLAP1 isoform 1 *in vitro* and SChLAP1 *in cellulo*. D5 region highlighted in red above. Yellow represents regions with lower reactivity in cells; blue represents regions with higher reactivity in cells. D) Selected structures of lower reactivity/more protected in cells, notably the 4-way junction at the 5'–end of SChLAP1 as well as structures within the D5 region and towards the 3'–end of SChLAP1. Circles indicate changes in reactivity: black are no change; blue are more reactive in cells; yellow are less

reactive in cells; gray are no data available.

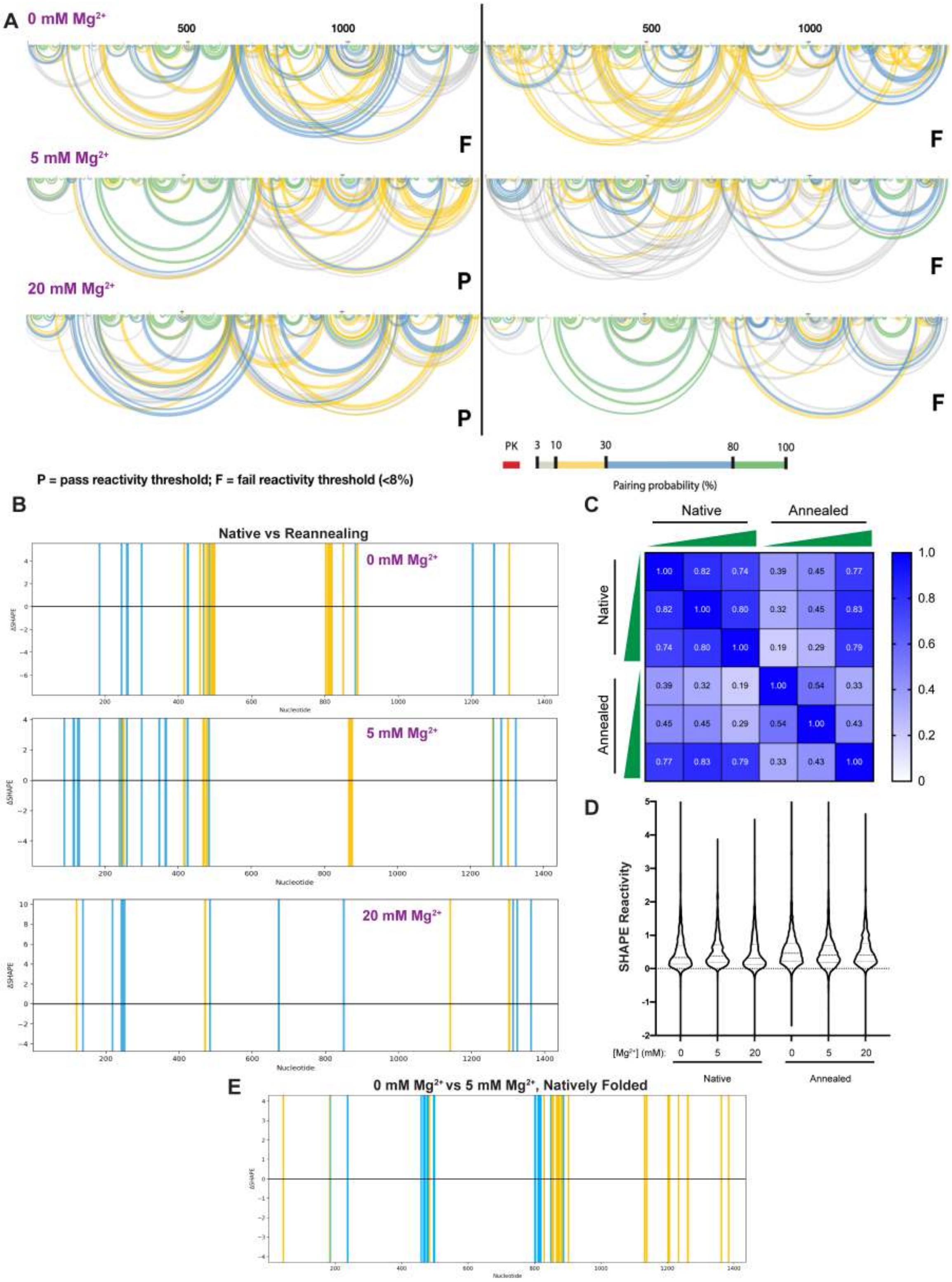
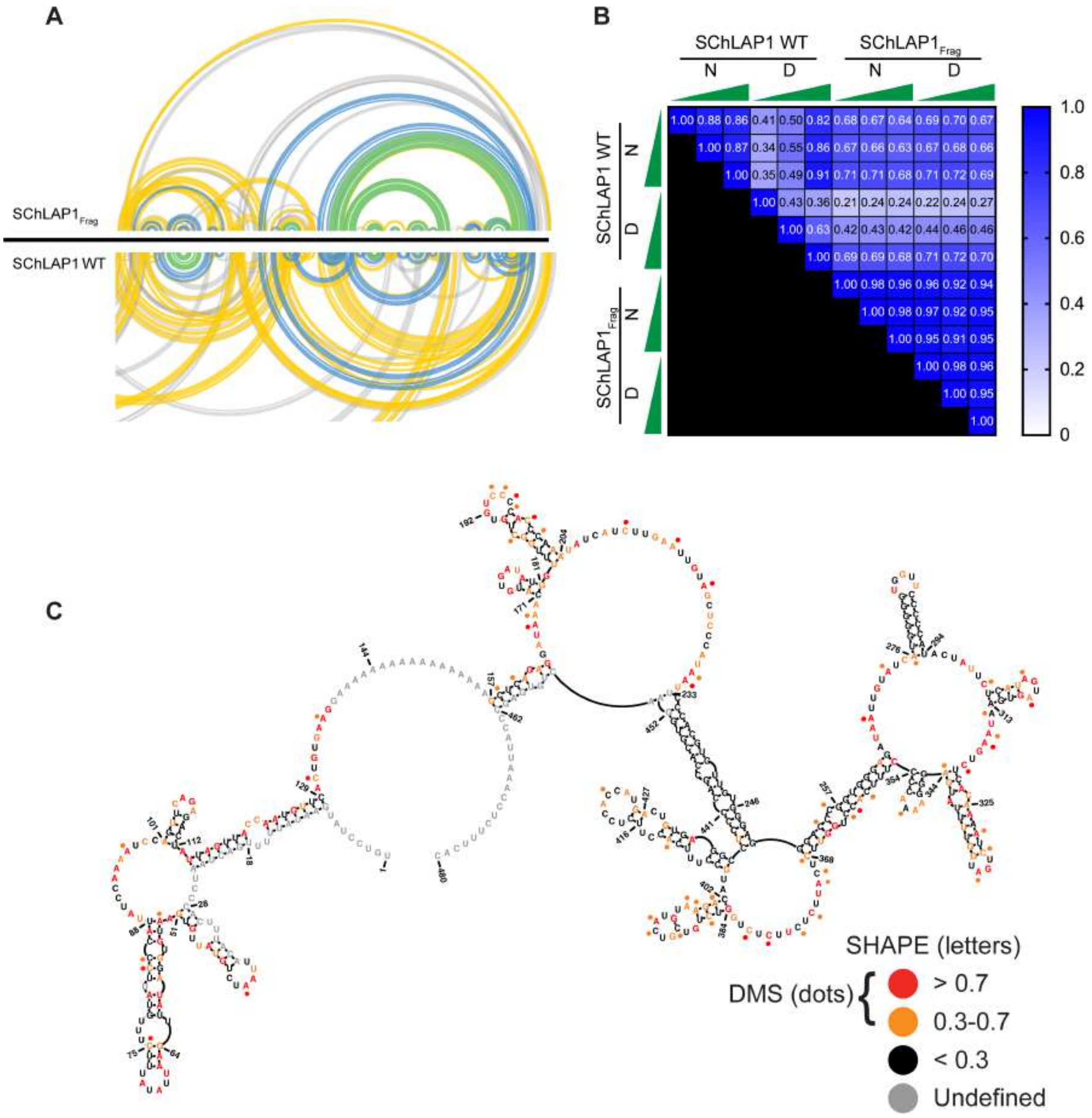


Figure 5. Influence of preparation (Native versus Annealed) and magnesium concentration on SChLAP1 WT *in vitro* secondary structure. A) Arc diagrams for SChLAP1 WT across two different preparations, Native (left) and Annealed (right) as well as at three different magnesium concentrations (0, 5, and 20 mM). P = pass and F = fail refer to an empirically derived threshold for passing reactivities in SHAPEMapper (8%). B) ΔSHAPE comparison of Native versus Annealed for all three magnesium concentrations, where blue bars indicate higher reactivity in Native versus Annealed and yellow bars indicate higher reactivity in Annealed versus Native. C) Spearman correlation coefficients of SHAPE reactivities for native and annealed SChLAP1 WT. Green triangles denote increasing magnesium concentration. D) Distribution of SHAPE reactivities for each condition for SChLAP1 WT. Values below -2 and above 5 are cropped out. E) ΔSHAPE analysis of natively folded SChLAP1 WT in 0 and 5 mM magnesium. Yellow = more protected with 5 mM Mg, and blue = more reactive in 5 mM Mg.



		SChLAP1 WT						SChLAP1 Frag					
		N			D			N			D		
SChLAP1 WT	N	1.00	0.88	0.86	0.41	0.50	0.82	0.68	0.67	0.64	0.69	0.70	0.67
			1.00	0.87	0.34	0.55	0.86	0.67	0.66	0.63	0.67	0.68	0.66
				1.00	0.35	0.49	0.91	0.71	0.71	0.68	0.71	0.72	0.69
	D				1.00	0.43	0.36	0.21	0.24	0.24	0.22	0.24	0.27
						1.00	0.63	0.42	0.43	0.42	0.44	0.46	0.46
							1.00	0.69	0.69	0.68	0.71	0.72	0.70
SChLAP1 Frag	N							1.00	0.98	0.96	0.96	0.92	0.94
									1.00	0.98	0.97	0.92	0.95
										1.00	0.95	0.91	0.95
	D										1.00	0.98	0.96
												1.00	0.95
													1.00

Figure 6. SChLAP1$_{\text{Frag}}$ forms a stable structure that mimics natively folded *in vitro* SChLAP1. A) Arc diagram of natively folded SChLAP1$_{\text{Frag}}$ in 5 mM magnesium (top) compared to arc diagram of natively folded SChLAP1 WT in 5 mM magnesium (bottom). B) Spearman correlation coefficients for SChLAP1$_{\text{Frag}}$ and respective nucleotides in SChLAP1 WT. Green triangles denote magnesium concentration (0, 5, or 20 mM). N = native. D = denatured and reannealed. C) MFE structure for SChLAP1$_{\text{Frag}}$.

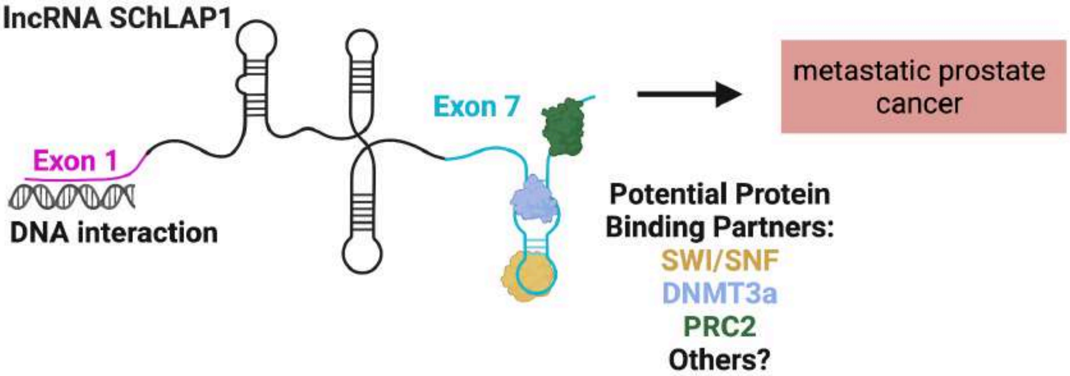


Figure 7. Proposed mechanism for SChLAP1. We propose that SChLAP1 interacts with both chromatin and proteins for some genes. Chromatin recognition is hypothesized to be facilitated by retrovirally incorporated exon 1, while other exons, particularly exon 7, is a hub for protein recognition. These proposed recognition events by conserved RNA structures within SChLAP1 facilitate metastatic cancer progression and ultimately poor patient outcomes. This figure was made with BioRender.

References

1. Siegel, R.L., Miller, K.D., Fuchs, H.E. and Jemal, A. (2021) Cancer Statistics, 2021. *CA Cancer J Clin*, **71**, 7-33.
2. Siegel, R.L., Miller, K.D. and Jemal, A. (2019) Cancer statistics, 2019. *CA Cancer J Clin*, **69**, 7-34.
3. Thomas, D.J., Rosenbloom, K.R., Clawson, H., Hinrichs, A.S., Trumbower, H., Raney, B.J., Karolchik, D., Barber, G.P., Harte, R.A., Hillman-Jackson, J. *et al.* (2007) The ENCODE Project at UC Santa Cruz. *Nucleic acids research*, **35**, D663-667.
4. Willingham, A.T., Orth, A.P., Batalov, S., Peters, E.C., Wen, B.G., Aza-Blanc, P., Hogenesch, J.B. and Schultz, P.G. (2005) A Strategy for Probing the Function of Noncoding RNAs Finds a Repressor of NFAT. *Science*, **309**, 1570-1573.
5. Ghildiyal, M. and Zamore, P.D. (2009) Small silencing RNAs: an expanding universe. *Nature Reviews Genetics*, **10**, 94-108.
6. Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E. and Chang, H.Y. (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science*, **329**, 689-693.
7. Lee, J.T. (2012) Epigenetic regulation by long noncoding RNAs. *Science*, **338**, 1435-1439.
8. Mercer, T.R. and Mattick, J.S. (2013) Structure and function of long noncoding RNAs in epigenetic regulation. *Nature structural & molecular biology*, **20**, 300-307.
9. de Hoon, M., Shin, J.W. and Carninci, P. (2015) Paradigm shifts in genomics through the FANTOM projects. *Mamm Genome*, **26**, 391-402.
10. Prensner, J.R., Iyer, M.K., Balbin, O.A., Dhanasekaran, S.M., Cao, Q., Brenner, J.C., Laxman, B., Asangani, I.A., Grasso, C.S., Kominsky, H.D. *et al.* (2011) Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an

unannotated lincRNA implicated in disease progression. *Nat Biotechnol*, **29**, 742-749.

11. Gerashchenko, G.V., Mevs, L.V., Chashchina, L.I., Pikul, M.V., Gryzodub, O.P., Stakhovsky, E.O. and Kashuba, V.I. (2018) Expression of steroid and peptide hormone receptors, metabolic enzymes and EMT-related genes in prostate tumors in relation to the presence of the TMPRSS2/ERG fusion. *Experimental oncology*, **40**, 101-108.
12. Chua, M.L.K., Lo, W., Pintilie, M., Murgic, J., Lalonde, E., Bhandari, V., Mahamud, O., Gopalan, A., Kweldam, C.F., van Leenders, G. *et al.* (2017) A Prostate Cancer "Nimbosus": Genomic Instability and SChLAP1 Dysregulation Underpin Aggression of Intraductal and Cribriform Subpathologies. *European urology*, **72**, 665-674.
13. Prensner, J.R., Zhao, S., Erho, N., Schipper, M., Iyer, M.K., Dhanasekaran, S.M., Magi-Galluzzi, C., Mehra, R., Sahu, A., Siddiqui, J. *et al.* (2014) RNA biomarkers associated with metastatic progression in prostate cancer: a multi-institutional high-throughput analysis of SChLAP1. *Lancet Oncol*, **15**, 1469-1480.
14. Kidd, S.G., Carm, K.T., Bogaard, M., Olsen, L.G., Bakken, A.C., Lovf, M., Lothe, R.A., Axcrona, K., Axcrona, U. and Skotheim, R.I. (2021) High expression of SCHLAP1 in primary prostate cancer is an independent predictor of biochemical recurrence, despite substantial heterogeneity. *Neoplasia*, **23**, 634-641.
15. Mehra, R., Udager, A.M., Ahearn, T.U., Cao, X., Feng, F.Y., Loda, M., Petimar, J.S., Kantoff, P., Mucci, L.A. and Chinnaiyan, A.M. (2016) Overexpression of the Long Non-coding RNA SChLAP1 Independently Predicts Lethal Prostate Cancer. *European urology*, **70**, 549-552.
16. Prensner, J.R., Iyer, M.K., Sahu, A., Asangani, I.A., Cao, Q., Patel, L., Vergara, I.A., Davicioni, E., Erho, N., Ghadessi, M. *et al.* (2013) The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nature genetics*, **45**, 1392-1398.
17. Mittal, P. and Roberts, C.W.M. (2020) The SWI/SNF complex in cancer - biology, biomarkers and therapy. *Nat Rev Clin Oncol*, **17**, 435-448.
18. Kadoch, C., Copeland, R.A. and Keilhack, H. (2016) PRC2 and SWI/SNF Chromatin Remodeling Complexes in Health and Disease. *Biochemistry*, **55**, 1600-1614.
19. Shain, A.H. and Pollack, J.R. (2013) The spectrum of SWI/SNF mutations, ubiquitous in human cancers. *PLoS One*, **8**, e55119.
20. Roberts, C.W.M.O., S.H. (2004) The SWI/SNF complex — chromatin and cancer. *Nat Rev Cancer*, **4**, 133–142.
21. Raab, J.R., Smith, K.N., Spear, C.C., Manner, C.J., Calabrese, J.M. and Magnuson, T. (2019) SWI/SNF remains localized to chromatin in the presence of SCHLAP1. *Nature genetics*, **51**, 26-29.
22. Elena Grossi, I.R., Enrique Goñi, Jovanna González, Francesco P Marchese, Vicente Chapaprieta, José I Martín-Subero, Shuling Guo, Maite Huarte. (2020) A lncRNA-SWI/SNF complex crosstalk controls transcriptional activation at specific promoter regions. *Nat. Commun.*, **11**.
23. Skalska, L.B., V.; Beltran, M.; Lukauskas, S.; Khandelwal, G.; Faull, P.; Bhamra, A.; Tavares, M.; Wellman, R.; Tvardovskiy, A.; Foster, B.M.; Ruiz de Los Mozos, I.; Herrero, J.; Surinova, S.; Snijders, A.P.; Bartke, T.; Jenner, R.G. (2021) Nascent RNA antagonizes the interaction of a set of regulatory proteins with chromatin. *Mol. Cell*, **81**, 2944-2959.e2910.
24. Huang, K. and Tang, Y. (2021) SChLAP1 promotes prostate cancer development through interacting with EZH2 to mediate promoter methylation modification of

multiple miRNAs of chromosome 5 with a DNMT3a-feedback loop. *Cell Death Dis*, **12**, 188.

25. Ji, J., Xu, R., Ding, K., Bao, G., Zhang, X., Huang, B., Wang, X., Martinez, A., Wang, X., Li, G. *et al.* (2019) Long Noncoding RNA SChLAP1 Forms a Growth-Promoting Complex with HNRNPL in Human Glioblastoma through Stabilization of ACTN4 and Activation of NF-κB Signaling. *Clinical cancer research : an official journal of the American Association for Cancer Research*, **25**, 6868-6881.
26. Du, Z., Niu, S., Wang, J., Wu, J., Li, S. and Yi, X. (2021) SChLAP1 contributes to non-small cell lung cancer cell progression and immune evasion through regulating the AUF1/PD-L1 axis. *Autoimmunity*, **54**, 225-233.
27. Jegu, T., Blum, R., Cochrane, J.C., Yang, L., Wang, C.Y., Gilles, M.E., Colognori, D., Szanto, A., Marr, S.K., Kingston, R.E. *et al.* (2019) Xist RNA antagonizes the SWI/SNF chromatin remodeler BRG1 on the inactive X chromosome. *Nature structural & molecular biology*, **26**, 96-109.
28. Lino Cardenas, C.L.K., C. W.; Cheng, Y.; MacDonald, C.; MacGillivray, T.; Ghoshhajra, B.; Huleihel, L.; Nuri, S.; Yeri, A. S.; Jaffer, F. A.; Kaminski, N.; Ellinor, P.; Weintraub, N. L.; Malhotra, R.; Isselbacher, E. M.; Lindsay, M. E. (2018) An HDAC9-MALAT1-BRG1 complex mediates smooth muscle dysfunction in thoracic aortic aneurysm. *Nat Commun*, **9**, 1009.
29. Wang, Y., He, L., Du, Y., Zhu, P., Huang, G., Luo, J., Yan, X., Ye, B., Li, C., Xia, P. *et al.* (2015) The long noncoding RNA lncTCF7 promotes self-renewal of human liver cancer stem cells through activation of Wnt signaling. *Cell Stem Cell*, **16**, 413-425.
30. Kawaguchi, T., Tanigawa, A., Naganuma, T., Ohkawa, Y., Souquere, S., Pierron, G. and Hirose, T. (2015) SWI/SNF chromatin-remodeling complexes function in noncoding RNA-dependent assembly of nuclear bodies. *Proc Natl Acad Sci U S A*, **112**, 4304-4309.
31. Cajigas, I.L., D.E.; Cochrane, J.; Luo, H.; Swyter, K.R.; Chen, S.; Clark, B.S.; Thompson, J.; Yates, J.R. 3rd; Kingston, R.E.; Kohtz, J.D. (2015) Evf2 lncRNA/BRG1/DLX1 interactions reveal RNA-dependent inhibition of chromatin remodeling. *Development*, **142**, 2641–2652.
32. Hu, G.G., A.Y.; Wang, Y.; Ma, S.; Chen, X.; Chen, J.; Su, C.J.; Shibata, A.; Strauss-Soukup, J.K.; Drescher, K.M.; Chen, X.M.;. (2016) LincRNA-Cox2 Promotes Late Inflammatory Gene Transcription in Macrophages through Modulating SWI/SNF-mediated Chromatin Remodeling. *J Immunol.*, **196**, 2799–2808.
33. Huang, M., Wang, H., Hu, X. and Cao, X. (2019) lncRNA MALAT1 binds chromatin remodeling subunit BRG1 to epigenetically promote inflammation-related hepatocellular carcinoma progression. *Oncoimmunology*, **8**, e1518628.
34. Smola, M.J., Calabrese, J.M. and Weeks, K.M. (2015) Detection of RNA-Protein Interactions in Living Cells with SHAPE. *Biochemistry*, **54**, 6867-6875.
35. Siegfried, N.A., Busan, S., Rice, G.M., Nelson, J.A. and Weeks, K.M. (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods*, **11**, 959-965.
36. Zubradt, M., Gupta, P., Persad, S., Lambowitz, A.M., Weissman, J.S. and Rouskin, S. (2017) DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat Methods*, **14**, 75-82.
37. Adams, R.L., Huston, N.C., Tavares, R.C.A. and Pyle, A.M. (2019) Sensitive detection of structural features and rearrangements in long, structured RNA molecules. *Methods Enzymol*.

38. Smola, M.J., Rice, G.M., Busan, S., Siegfried, N.A. and Weeks, K.M. (2015) Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat Protoc*, **10**, 1643-1669.
39. Busan, S., Weidmann, C.A., Sengupta, A. and Weeks, K.M. (2019) Guidelines for SHAPE Reagent Choice and Detection Strategy for RNA Structure Probing Studies. *Biochemistry*, **58**, 2655-2664.
40. Busan, S. and Weeks, K.M. (2018) Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA*, **24**, 143-148.
41. Low, J.T. and Weeks, K.M. (2010) SHAPE-directed RNA secondary structure prediction. *Methods*, **52**, 150-158.
42. Breaker, R.R. (2012) Riboswitches and the RNA world. *Cold Spring Harb Perspect Biol*, **4**.
43. Rangan, R., Zheludev, I.N., Hagey, R.J., Pham, E.A., Wayment-Steele, H.K., Glenn, J.S. and Das, R. (2020) RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses: a first look. *RNA*, **26**, 937-959.
44. Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W., Jr., Swanstrom, R., Burch, C.L. and Weeks, K.M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, **460**, 711-716.
45. Wan, Y., Qu, K., Zhang, Q.C., Flynn, R.A., Manor, O., Ouyang, Z., Zhang, J., Spitale, R.C., Snyder, M.P., Segal, E. *et al.* (2014) Landscape and variation of RNA secondary structure across the human transcriptome. *Nature*, **505**, 706-709.
46. Rivas, E., Clements, J. and Eddy, S.R. (2017) A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat Methods*, **14**, 45-48.
47. Liu, M. and Eiden, M.V. (2011) Role of human endogenous retroviral long terminal repeats (LTRs) in maintaining the integrity of the human germ line. *Viruses*, **3**, 901-905.
48. Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G. *et al.* (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research*, **22**, 1775-1789.
49. Laing, C. and Schlick, T. (2009) Analysis of four-way junctions in RNA structures. *J Mol Biol*, **390**, 547-559.
50. Brazda, V., Kolomaznik, J., Lysek, J., Bartas, M., Fojta, M., Stastny, J. and Mergny, J.L. (2019) G4Hunter web application: a web server for G-quadruplex prediction. *Bioinformatics*, **35**, 3493-3495.
51. Matsumura, K., Kawasaki, Y., Miyamoto, M., Kamoshida, Y., Nakamura, J., Negishi, L., Suda, S. and Akiyama, T. (2017) The novel G-quadruplex-containing long non-coding RNA GSEC antagonizes DHX36 and modulates colon cancer cell migration. *Oncogene*, **36**, 1191-1199.
52. Zafferani, M. and Hargrove, A.E. (2021) Small molecule targeting of biologically relevant RNA tertiary and quaternary structures. *Cell chemical biology*, **28**, 594-609.
53. Kladwang, W., Topkar, V.V., Liu, B., Rangan, R., Hodges, T.L., Keane, S.C., Al-Hashimi, H. and Das, R. (2020) Anomalous Reverse Transcription through Chemical Modifications in Polyadenosine Stretches. *Biochemistry*, **59**, 2154-2170.
54. Dethoff, E.A., Boerneke, M.A., Gokhale, N.S., Muhire, B.M., Martina, D.P., Sacco, M.T., McFadden, M.J., Weinstein, J.B., Messer, W.B., Horner, S.M. *et al.* (2018)

Pervasive tertiary structure in the dengue virus RNA genome. *Proc Natl Acad Sci U S A*.

55. Smola, M.J., Christy, T.W., Inoue, K., Nicholson, C.O., Friedersdorf, M., Keene, J.D., Lee, D.M., Calabrese, J.M. and Weeks, K.M. (2016) SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells.
56. Peattie, D.A. and Gilbert, W. (1980) Chemical probes for higher-order structure in RNA. *Proc Natl Acad Sci U S A*, **77**, 4679-4682.
57. Jones, A.N., Pisignano, G., Pavelitz, T., White, J., Kinisu, M., Forino, N., Albin, D. and Varani, G. (2020) An evolutionarily conserved RNA structure in the functional core of the lincRNA Cyrano. *RNA*, **26**, 1234-1246.
58. Schmidt, K., Weidmann, C.A., Hilimire, T.A., Yee, E., Hatfield, B.M., Schneekloth, J.S., Jr., Weeks, K.M. and Novina, C.D. (2020) Targeting the Oncogenic Long Non-coding RNA SLNCR1 by Blocking Its Sequence-Specific Binding to the Androgen Receptor. *Cell Rep*, **30**, 541-554.e545.
59. Frank, F., Kavousi, N., Bountali, A., Dammer, E.B., Mourtada-Maarabouni, M. and Ortlund, E.A. (2020) The lncRNA Growth Arrest Specific 5 Regulates Cell Survival via Distinct Structural Modules with Independent Functions. *Cell Rep*, **32**, 107933.
60. Horoszewicz, J.S., Leong, S.S., Kawinski, E., Karr, J.P., Rosenthal, H., Chu, T.M., Mirand, E.A. and Murphy, G.P. (1983) LNCaP model of human prostatic carcinoma. *Cancer Res*, **43**, 1809-1818.
61. Smola, M.J. and Weeks, K.M. (2018) In-cell RNA structure probing with SHAPE-MaP. *Nature protocols*, **13**.
62. Sahu, A. (2015), University of Michigan.
63. Misra, V.K. and Draper, D.E. (1998) On the role of magnesium ions in RNA stability. *Biopolymers*, **48**, 113-135.
64. Zheng, H., Shabalin, I.G., Handing, K.B., Bujnicki, J.M. and Minor, W. (2015) Magnesium-binding architectures in RNA crystal structures: validation, binding preferences, classification and motif detection. *Nucleic acids research*, **43**, 3789-3801.
65. Dethoff, E.A. and Weeks, K.M. (2019) Effects of Refolding on Large-Scale RNA Structure. *Biochemistry*, **58**, 3069-3077.
66. Bellaousov, S., Reuter, J.S., Seetin, M.G. and Mathews, D.H. (2013) RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic acids research*, **41**, W471-474.
67. Pyle, A.M. (2014) Looking at LncRNAs with the ribozyme toolkit. *Molecular cell*, **56**, 13-17.
68. Zhang, B., Mao, Y.S., Diermeier, S.D., Novikova, I.V., Nawrocki, E.P., Jones, T.A., Lazar, Z., Tung, C.S., Luo, W., Eddy, S.R. *et al.* (2017) Identification and Characterization of a Class of MALAT1-like Genomic Loci. *Cell Rep*, **19**, 1723-1738.
69. Somarowthu, S., Legiewicz, M., Chillon, I., Marcia, M., Liu, F. and Pyle, A.M. (2015) HOTAIR forms an intricate and modular secondary structure. *Molecular cell*, **58**, 353-361.
70. Tavares, R.C.A., Pyle, A.M. and Somarowthu, S. (2019) Phylogenetic Analysis with Improved Parameters Reveals Conservation in lncRNA Structures. *J Mol Biol*, **431**, 1592-1603.
71. Mustoe, A.M., Lama, N.N., Irving, P.S., Olson, S.W. and Weeks, K.M. (2019) RNA base-pairing complexity in living cells visualized by correlated chemical probing. *Proc Natl Acad Sci U S A*, **116**, 24574-24582.

72. Fang, R., Moss, W.N., Rutenberg-Schoenberg, M. and Simon, M.D. (2015) Probing Xist RNA Structure in Cells Using Targeted Structure-Seq. *PLoS Genet*, **11**, e1005668.
73. Hu, T., Pi, W., Zhu, X., Yu, M., Ha, H., Shi, H., Choi, J.H. and Tuan, D. (2017) Long non-coding RNAs transcribed by ERV-9 LTR retrotransposon act in cis to modulate long-range LTR enhancer function. *Nucleic acids research*, **45**, 4479-4492.
74. Alfeghaly, C., Sanchez, A., Rouget, R., Thuillier, Q., Igel-Bourguignon, V., Marchand, V., Branlant, C., Motorin, Y., Behm-Ansmant, I. and Maenner, S. (2021) Implication of repeat insertion domains in the trans-activity of the long non-coding RNA ANRIL. *Nucleic acids research*, **49**, 4954-4970.
75. Bernat, V. and Disney, M.D. (2015) RNA Structures as Mediators of Neurological Diseases and as Drug Targets. *Neuron*, **87**, 28-46.
76. Boerneke, M.A., Ehrhardt, J.E. and Weeks, K.M. (2019) Physical and Functional Analysis of Viral RNA Genomes by SHAPE. *Annu Rev Virol*, **6**, 93-117.