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

Emily J. McFadden[†1,2], James P. Falese[†,2], and Amanda E. Hargrove[*2,3]

Affiliations
1. Yale School of Medicine, Department of Molecular Biophysics and Biochemistry, New Haven, Connecticut, USA
2. Duke University School of Medicine, Department of Biochemistry, Durham, North Carolina, USA
3. Duke University, Department of Chemistry, Durham, North Carolina, USA

†: These authors contributed equally to this work.
*Corresponding author: amanda.hargrove@duke.edu

Abstract

The lncRNA Second Chromosome Locus Associated with Prostate 1 (SChLAP1) was previously identified as a predictive biomarker and 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 the SChLAP1:SWI/SNF complex. Herein, we report the first secondary structure model of SChLAP1 utilizing SHAPE-MaP 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. Further, we identified a 3′–fragment of SChLAP1 ($SChLAP1_{Frag}$) 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 in selective targeting and disruption of the SChLAP1:protein interface and the development of new therapeutic avenues in prostate cancer treatment.

Introduction

Prostate cancer is one of the most commonly occurring cancers (one in five new cancer cases) and is the second leading cause of cancer death in American men. 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. 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.[2] The remaining RNA, the non-coding transcriptome, is thus an underexplored realm for potential therapeutics. Long noncoding 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.[3-7] Following the ENCODE project[2] and the identification of thousands of new lncRNAs, the FANTOM consortium[8] (RIKEN, Japan) began annotating these transcripts and found biochemical indices of function historically ascribed solely to proteins. Although there are over 170,000 identified lncRNA, approximately 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 their biochemical characterization. In 2011, Prensner et al. identified PCAT-114 (later renamed SChLAP1) as a transcript over-expressed in tumor prostate tissue as compared to benign prostate tissue.[9] In other work, Gerashchenko et al. determined that tumors with a high SChLAP1 levels correlated with a Gleason score of 9 and also correlated with high levels epithelial-to-mesenchymal transition (EMT) markers such as vimentin (VIM), fibronectin (FN1), and matrix metalloproteinase 2 (MMP2), supporting the correlation between SChLAP1 overexpression and prostate cancer metastasis found in other clinical studies.[10-13] Mehra et al. determined that approximately 16% of clinically localized prostate cancers in American men exhibit high levels of SChLAP1 expression, suggesting use of SChLAP1 as an early detection biomarker.[14]

In model systems, knockdown of SChLAP1 reduced migration in multiple prostate cancer cell lines and reduced the formation of distal metastases in SCID mice.[15] 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 Switching/Sucrose non-fermenting (SWI/SNF) chromatin remodeling complex. The 11-15-mer SWI/SNF complex is dysregulated and/or houses mutations in its respective subunit genes in 20% of all cancers.[16,17] SWI/SNF specifically catalyzes ATP-dependent chromatin remodeling through sliding or ejection of nucleosomes from the DNA.[18] Prensner et al. observed direct binding between SWI/SNF and SChLAP1 via RNA immunoprecipitation (RIP), specifically with the SMARCB1 subunit (also known as SNF5 and INI1), resulting in genome-wide depletion of SWI/SNF from the chromatin. While follow up work from Raab et al. did not recapitulate this global loss in SMARCB1 genomic occupancy, SChLAP1 still bound to SMARCB1 and altered chromatin accessibility at a subset of genomic sites, with motifs at these sites being enriched for known SWI/SNF interactors or pathways.[19-21] Raab et al. proposed that SChLAP1 may function in a SWI/SNF-independent manner, yet additional studies in lncRNA:SWI/SNF recognition observed crucial interactions in both healthy and disease states.[22-30]

Elucidating the sequence- and/or structure-function relationships in SChLAP1 lncRNA relative to aggressive prostate cancer is critical to both our understanding of

prostate cancer metastasis and of the potential of SChLAP1 as a therapeutic target. Toward this goal, preliminary unpublished 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 the formation of a SWI/SNF complex.[31] 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.

We herein report first insights into the role of sequence and structure in SChLAP1. First, we found that specific exons within SChLAP1 are highly conserved across non-human primates. We then completed *in vitro* secondary structural analysis of SChLAP1 isoform 1 WT and D5 via selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP)[32,33] to identify an evolutionarily conserved putative independent folding domain, $SChLAP1_{Frag}$, which exhibits changes in chemical reactivities via ΔSHAPE consistent with protein association. We show that while global SChLAP1 secondary structure may change in response to purification method and magnesium concentration, $SCHLAP1_{Frag}$ secondary structure is unperturbed in these varying conditions, supporting the presence of a functionally critical structure. As the first detailed biochemical and structural analysis of SCHLAP1, this work proposes an important functional role of $SChLAP1_{Frag}$ in SChLAP1:protein recognition and the metastatic phenotype.

Results

Sequence alignments reveal potentially functional regions of SChLAP1

Highly structured RNA motifs within viruses, bacteria, and human endogenous transcripts are indicative of critical functional motifs and are well conserved across evolution.[34-37] We ran nBLAST on the entire SChLAP1 WT isoform 1 sequence to look for critical sequences of SChLAP1 across other species. This returned 14 non-human primate sequences with >85% sequence conservation (Fig. 1), all of which are predicted to be non-coding RNA but have no documented functional annotation to date. In gorilla and chimpanzee, our two closest primate relatives with a putative SChLAP1 homologue, these sequences are also located on chromosome 2. Alignments of these sequences with all 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, supporting their functional importance. The D5 region found critical to metastasis by Sahu et al. is housed within exon 7.[31] Covariation analyses performed using R-scape showed no significant differences among all 14 sequences, in line with the high level of sequence conservation (Fig. S1). Although the temporal proximity in evolution limits our ability to perform covariation analyses, these data support the claim that certain exons may be crucial to SChLAP1's function in healthy cells and in disease.

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, 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.[38] 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 fact that a greater number of lncRNAs are observed in higher-order species.[39]

We then mined existing data sets for evidence that the lncRNA SChLAP1 regulates gene expression via scaffolding chromatin modifying protein complexes. Evidence for differential SNF5 chromatin binding between RWPE cells overexpressing SChLAP1 or LacZ is supported by chromatin immunoprecipitation (ChIP) experiments reported by the Prensner et al.[15] We filtered through the corresponding SChLAP1 depletion microarray dataset for genes that overlap with the ERV-9-derived exon 1 on chromosome 2 – the SChLAP1 gene locus location – to identify potential *cis*-regulatory roles and then focused on the top 100 genes with >95% sequence alignment. This analysis returned eight protein-coding genes on Chromosome 2 (Fig. S2). The microarray data deposited by Prensner revealed that lower levels of SChLAP1 caused dysregulation of both PTPRN (downregulated) and NPHP1 (upregulated) proteins.[15] Likewise, Raab et al.'s ChIP-seq data found SWI/SNF subunits SMARCB1, SMARCA4, and SMARCA2 cross-linked to regions adjacent to PTPRN and NPHP1 genes.[19] NPHP1, or nephrocystin 1, plays a regulatory role in cell polarity and maintains epithelial cell-cell junctions.[40] PTPRN, or protein tyrosine phosphatase receptor type N (also known as IA-2), is a member of the PTP family of proteins involved in cell growth, differentiation, and oncogenic transformation.[41] According to The Protein Atlas (www.proteinatlas.org), PTPRN is dysregulated in glioma, a cancer where SChLAP1 over-expression was previously reported.[42] The normal functions of these proteins, and their potential dysregulation by SChLAP1, would be supportive of the metastatic phenotype seen with SChLAP1 over-expression in a clinical setting.

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.[43] Interestingly, Huang and Tang found that miR-145 had reduced chromatin accessibility upon SChLAP1 overexpression in the RWPE-1 cell line, although SWI/SNF binding was still observed at this site.[19, 43] Overall, these analyses support chromatin recognition as a potential mechanism for SChLAP1 function in both a normal and disease state.

Identification of *in vitro* structure and protein-binding sites within SChLAP1 by SHAPE-MaP

We next set out to elucidate the *in vitro* structure of SChLAP1 and to identify important structures in the transcript. We performed *in vitro* SHAPE-MaP of SCHLAP1 isoform 1 (SChLAP1 WT; Fig. 2A) to produce the first experimentally informed secondary structure model of SChLAP1. We chose the SHAPE reagent 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.[44]

We used a semi-native purification *in vitro* protocol, 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.[45]

The resulting structure models were generated using the SHAPEMapper (v.2.1.5)[46] and SuperFold[47] pipelines. A varied combination of secondary structures and single stranded regions are observed across the length of the transcript. For example, at the 5'-end of SCHLAP1 WT (1-155 n.t.) we see a four-way junction (4WJ), a structural motif known to function as a protein scaffold in other RNA classes.[48] We also observe a large central loop structure (828-871 n.t.) and a highly extended stem loop structure towards the 3′–end (935-1137 n.t.). SHAPEMapper analysis did not identify any pseudoknots. The program G4Hunter predicted G-Quadruplexes (GQs) within SCHLAP1, including the D5 region (Fig. S3), which may play a functional role in protein recognition.[49] We also observed a lack of chemical reactivity in a poly(A) stretch of SChLAP1 WT (1088-1104 n.t.), consistent with work from Kladwang et al., where chemical modifications in poly(A) regions are bypassed by reverse transcriptases.[50] We accounted for this in our structural models for full-length and $SChLAP1_{Frag}$ (reported below) by manually setting these nucleotides as undefined in the input .map file for Superfold and ΔSHAPE.

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.[33] This metric has since been used to identify functional regions, including in the Dengue virus RNA genome[51] and XIST lncRNA.[52] In our SChLAP1 model, we identified two major lowSS regions: the 5′–end of the transcript (approximately 1-480 n.t.) and between approximately 1000-1050 n.t at the 3′–end (Fig. 2A). The 5′–end region corresponds to exons 1 and 2, and the 3′–end region corresponds to exon 7 within the D5 region. Thus, the most robustly conserved exons within SChLAP1 all contain lowSS regions, further supporting their functional significance.

We next performed *in cellulo* probing to identify putative protein binding sites within SChLAP1 by comparing the reactivity data to the structure model generated above, a method that has yielded insights into protein binding for multiple other lncRNAs to date (Fig. 2B).[52-57] This was performed in the LNCaP prostate cancer cell line, which is derived from supraclavicular lymph nodes,[58] has high SChLAP1 expression,[15] and a known metastatic phenotype. Multiple aspects of the minimum free energy (MFE) structure are consistent *in cellulo* as *in vitro*, namely the 4WJ, large loop structure, and extended stem loop structure. These results supported the assumption that changes in *in cellulo* SChLAP1 structure due to protein binding would be reflected by localized changes in nucleotide reactivity as opposed to global changes in RNA structure.[44] *In cellulo* reactivity profiles were quantitatively compared to *in vitro* reactivity profiles using ΔSHAPE, and nucleotides that showed significant reductions in *in cellulo* reactivity as compared to *in vitro* were considered candidate protein binding sites.

The majority of in-cell protections (111/114 instances) identified by ΔSHAPE were within exons 1, 2, and 7 (Fig. 3A), further supporting the functional significance

inferred from phylogenetic analysis (Fig. 1) and the presence of lowSS regions. Within exon 1, many in-cell protections are localized to the aforementioned 4WJ immediately adjacent to the 5′–end of the transcript (Fig. 3B). This finding is consistent with the potential role for exon 1 in chromatin recognition, which may induce similar protections as protein binding. In-cell protections in SChLAP1 exon 2 are consistent with recent work from Ji et al., wherein they observed binding between SChLAP1 exon 2 and HNRNPL, which facilitated activation of the NF-κB pathway in glioblastoma.[42] Within exon 7, the D5 region shows particularly abundant in-cell protections (Fig. 3B), further supporting its role in protein binding and potentially the metastatic phenotype as proposed by Sahu et al.[31]

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.[45] Comparison of the SChLAP1 WT Native versus heat denatured and reannealed (Annealed) samples across the three different magnesium concentrations revealed varying structural landscapes (Fig. 4A and Fig S4). Correlation coefficients were calculated between all conditions, i.e. Native and Annealed SChLAP1 WT at 0, 5, and 20 mM magnesium (Fig. S4A). For Native SChLAP1, altering the magnesium concentration did not globally impact the structure, as the correlation coefficients between each natively-folded condition are high (over 0.7). However, reannealing results in significant perturbation of the structure, as demonstrated by significant reduction in correlation coefficients between Native and Annealed SChLAP1. This result is also reflected through significant changes in the arc diagrams in the Annealed samples (Fig. 4A), as well as alterations in the MFE structures from these experiments. 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.

Surprisingly, four of these 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.[59] Both of the two passing structures were from natively-purified SChLAP1, suggesting that the annealing process forms an RNA that is less prone 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.[51] 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 native-purified samples (Fig. 4B)..

We then explored the role of magnesium concentration on transcript structure using ΔSHAPE. We identified regions of SChLAP1 that had significant changes in

SHAPE reactivity between 0 and 5 mM magnesium, suggesting that magnesium-dependent structures occur in physiologically-relevant concentrations (Fig. S5). Of note, a region at the 3′–end of the transcript (1,235-1,285 n.t), shows significant protection in the presence of 5 mM magnesium. 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 plays crucial roles in particular SChLAP1 regions.

Altogether, our findings from ΔSHAPE comparison of the different preparations and magnesium concentrations support that the SChLAP1 WT Native preparation is the most relevant for our studies, and that magnesium plays an important role in local secondary structures. Moving forward in our analyses, we thus decided to focus on comparison of WT and D5, as well as eventual *in cellulo* analyses, with only natively-folded RNA.

Removal of the D5 Region does not globally perturb SChLAP1 folding

With evidence for structure 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 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 a D5 region structure. We thus performed SHAPE-MaP on the SChLAP1 D5 transcript. The transcript was 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. The sequence of 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 at all three magnesium concentrations, with 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. 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. S6). 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.[57] In support of our ΔSHAPE analysis, correlations dropped to 0.43 in the presence of 20 mM magnesium (Fig. S6).

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 requirement of D5–localized secondary structures for protein recognition and metastasis.

SChLAP1$_{Frag}$ 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 and non-human primates (Fig. 1); 2) ΔSHAPE data support protein binding within exon 7 (Fig. 3); 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 the last approximately 500 nucleotides of SChLAP1. Using RNAStructure,[60] we observed a robustly predicted structure within exon 7 ranging from 949-1428 n.t. This region, termed SChLAP1$_{Frag}$, was independently transcribed and subjected to chemical probing in the same manner as *in vitro* SChLAP1 WT.

The overall architecture for natively-prepared 5 mM magnesium SChLAP1$_{Frag}$ is consistent with the structure of this region within the full-length transcript prepared in the same conditions (Fig. 5A). We note that the predicted probabilities of the base-pair arcs are higher in SChLAP1$_{Frag}$ than in these nucleotides for full-length SChLAP1 WT and 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 compared the raw reactivities between SChLAP1$_{Frag}$ and these nucleotide coordinates in SChLAP1 WT as a way to evaluate nucleotide environment 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. 5B). This finding suggests that the structure of SChLAP1$_{Frag}$ is highly similar to the same sequence found within the context of the natively purified, full-length SChLAP1. These correlations drop precipitously when SChLAP1$_{Frag}$ is compared to reactivities in Annealed WT, suggesting that thermal denaturation of the full-length transcript disrupts the folding of this region.

We next evaluated whether varying magnesium and preparation conditions impacted the structure of SChLAP1$_{Frag}$ as a means to interpret the stability of the structure and thus its functional importance. We generated a correlation matrix to make all possible pairwise comparisons between conditions. We observed high correlation coefficients (Spearman's $\rho>0.9$) across all conditions (Fig. 5B), indicating that the secondary structure of SChLAP1$_{Frag}$ is not significantly impacted by magnesium concentration or RNA preparation. Overall, SChLAP1$_{Frag}$ is proposed to be a critical structure module within SChLAP1 that mediates protein binding previously linked to aggressive cancer.

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,[61] chemical probing methods such as SHAPE-MaP combined with phylogenetic analyses have provided insight into these large biomolecules.[62-64] 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[65] and XIST.[66] 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.

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. 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 off from orangutans.[67] This insertion was recently shown to be important for lncRNA activity as Alfeghaly et al. reported the ERV-dependent *trans* activity of ANRIL.[68] 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. While follow up experiments are needed, we hypothesize that ERV-derived exon 1 may serve as a chromatin recognition element for SChLAP1 to localize to specific chromosome 2 genes, perhaps through RNA:DNA sequence-specific hybridization.

Comparative analysis of SChLAP1 in prostate cancer cells also supported protein binding in evolutionarily conserved exon 7. 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[31] as well as other work that has uncovered functional roles for RNA:SWI/SNF recognition[22-30] in spite of noted promiscuous binding for SWI/SNF.[22, 28, 30] Additionally, we found that a structured fragment within exon 7 ($SChLAP1_{Frag}$), which contains multiple putative protein bindings sites, forms a remarkably stable and context-independent structure. Furthermore, structures within $SChLAP1_{Frag}$ were similar to predicted structures identified by Wu et al. to be important in recognition of SNF5 by another lncRNA, Stem-Cell Derived Angiogenic lncRNA (SCDAL).[69] The structure-dependent recognition yet modestly promiscuous binding of RNA by SWI/SNF warrants further investigation and would be consistent with other chromatin remodeling complexes such as PRC2.[70]

While we considered the structure of natively folded SChLAP1 in 5 mM magnesium as the most physiologically relevant structure based on literature

precedence,[45] 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 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.[45]

In previous work, native purification was found to be necessary for homogenous isolation of the lncRNA HOTAIR.[63] In our work, we found that reannealing full-length SChLAP1 also results in structural rearrangement of the RNA (Fig. 4). In contrast, the structure of SChLAP1$_{Frag}$ appears unaffected/unperturbed by denaturing preparation. This method allowed us to identify a structurally robust element of SChLAP1 that is likely important for protein recognition and will be more amenable to downstream biophysical analyses, including X-ray crystallography or cryo-electron microscopy.

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, including evolutionary conservation, deletion studies, chemical probing, and fragment folding analysis. Additionally, our work identifies SChLAP1$_{Frag}$ as a tractable proxy for 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.

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 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. Successful insertion of the remaining adenosine was confirmed via Sanger Sequencing (Eton Biosciences) using the T7 promoter for primer binding. 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 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, Tann ˚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 Tann temperature for individual sequences). DNA was purified using Zymo DNA Clean and Concentrator (5).

In Vitro Transcription. *In vitro* transcription for SChLAP1 Isoform 1 WT, SChLAP1 D5, and SChLAP1$_{Frag}$ were completed following the procedure from Adams et al. with some modifications.[45] 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, 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), 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 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. For the semi-native purification ("Native"), RNA from the FPLC was maintained at room temperature prior to magnesium addition. 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. 5NIA (Astatech) was dissolved into anhydrous DMSO and were prepared immediately prior to use to limit oxidation.

SChLAP1 WT and SChLAP1 D5 were typically purified by SEC at approximately at ~90 ng/µL for all probing reactions and were diluted with 1X FB if necessary to achieve this concentration. 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. Following this incubation, 5NIA in DMSO 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.

RNA Reverse Transcription. To facilitate sequencing of full-length SChLAP1, reverse transcription was performed using four unique primer sets, which generated four overlapping amplicons. Specifically, each amplicon was no more than 600 base-pairs in length and had a 100 base-pair overlapping window with the neighboring fragments. Sequences for all primers 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 µg RNA went into each RT reaction. 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 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.

Library Preparations

Self-made Libraries. These separate reactions were then processed using the Smola et al two-step PCR reaction to add on Illumina primers following the amplicon workflow (see Appendix A for primer information). 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 150 bp), and loaded on a MiSeq Illumina sequence per manufacturer's instructions.

In Cellulo Chemical Probing. Probing was performed as previously published. 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 transcriptions contained at least 0.75 µg of cellular RNA and were performed in the presence of 2M betaine. cDNA were purified with Agencourt RNAClean XP beads (Beckman Coulter). Some samples were further purified by ethanol precipitation. 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). 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.

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). 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.

Correlation Analyses. 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 for 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

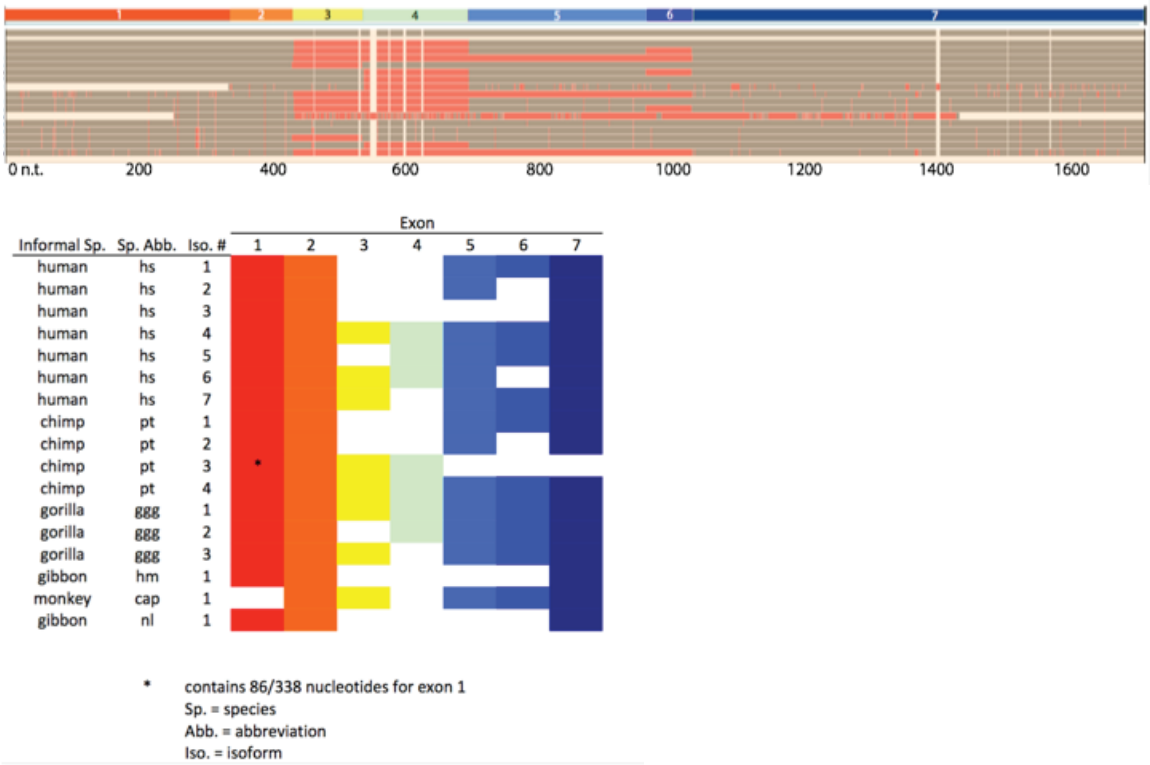

Figure 1. Sequence alignment (Clustal Omega) of all primate genes aligned to human SChLAP1 isoform 4. Sequences were aligned to human SChLAP1 (hsSChLAP1) isoform 4 (iso 4) as this isoform contains all exons (top). Brown is overlapping sequence and red is mutations/deletions. Exons in rainbow order on top correspond to exons in bottom table. All accession codes can be found in SI Table 4. Bottom shows simplified exon arrangement.

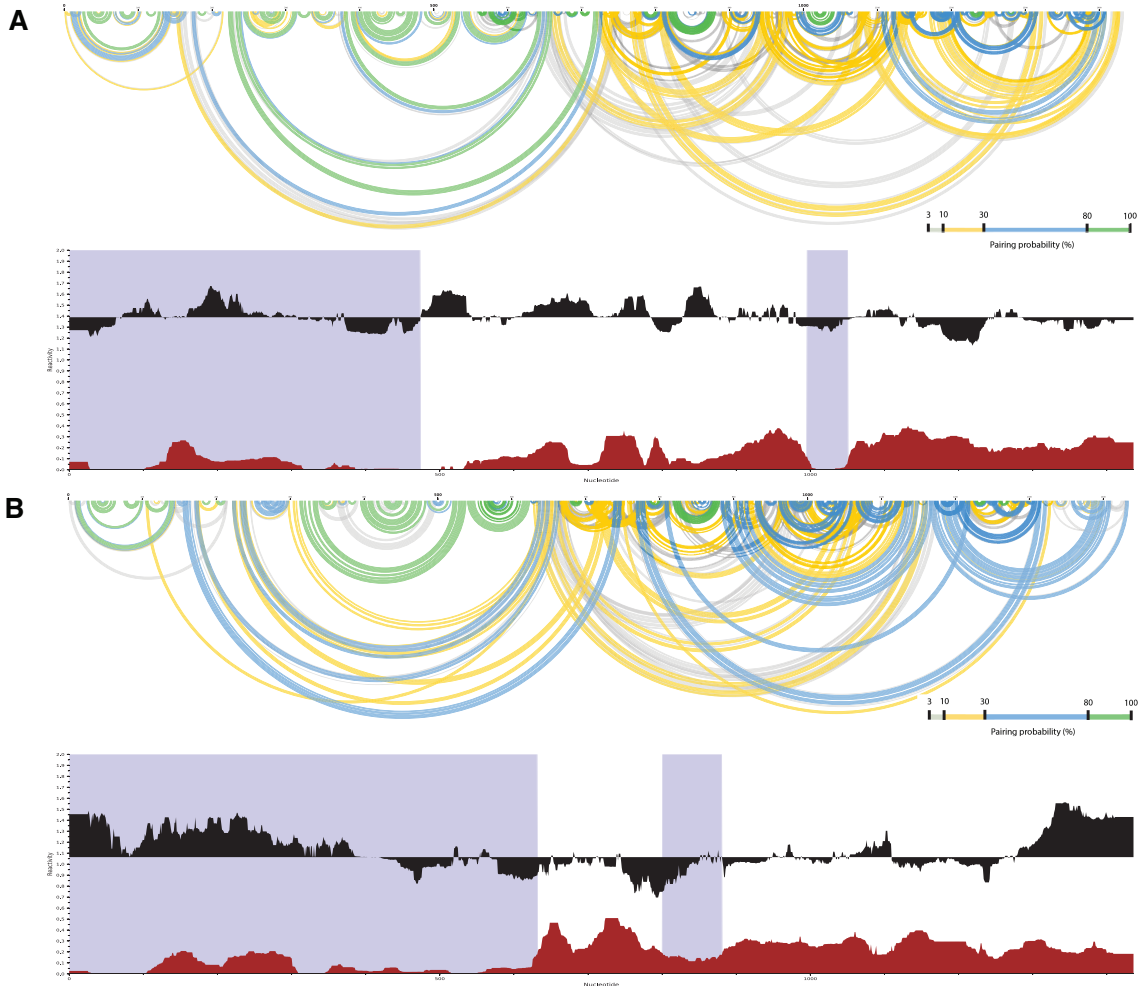

Figure 2. Secondary structure analysis and comparison of SChLAP1 *in vitro* and *in cellulo*. Secondary structure models of (A) SChLAP1 isoform 1 *in vitro* and (B) SChLAP1 *in cellulo*. Arc map representing linear secondary structure model with pairing probabilities. Black plots are chemical probing reactivities and red plots are Shannon entropy. Purple bars indicate regions of low Shannon entropy.

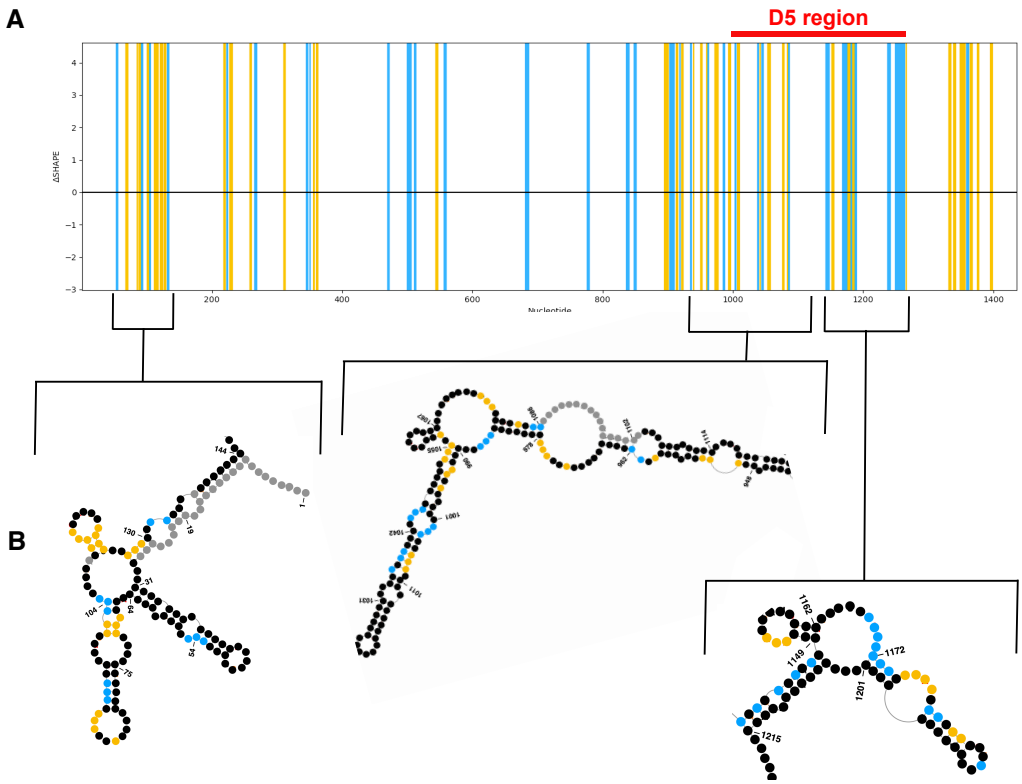

Figure 3. A) ΔSHAPE of SChLAP1 isoform 1 *in vitro* and SChLAP1 *in cellulo*. D5 region highlighted in red above. Yellow represents regions with lower reactivity in cells; Blue represents regions with higher reactivity in cell. B) Selected structures of lower reactivity/more protected in cells, notably 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.

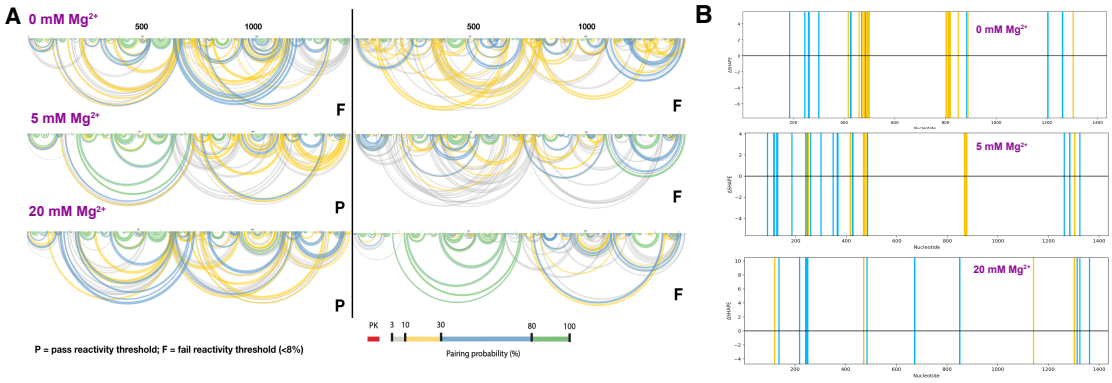

Figure 4. Influence of preparation (Native versus Annealed) and magnesium concentration on SChLAP1 isoform 1 WT *in vitro* secondary structure. A) Shows linear secondary structure probabilities for SChLAP1 isoform 1 WT across two different preparations (Native and Annealed) as well as at three different magnesium concentrations (0, 5, and 20 mM). P and F refer to 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.

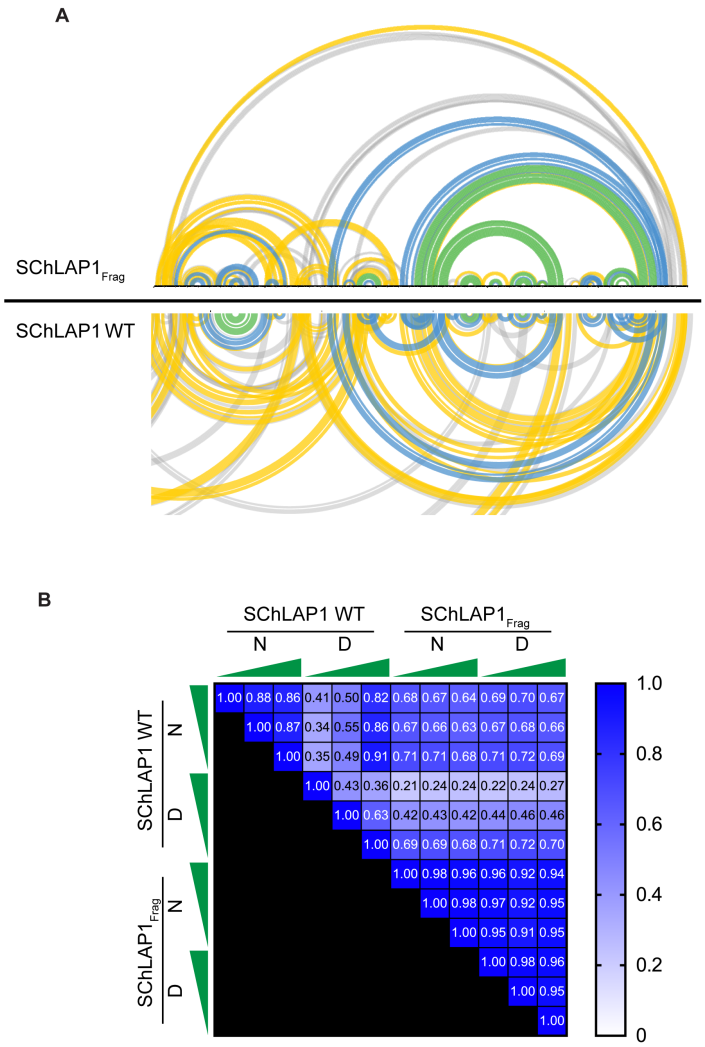

Figure 5. SChLAP1$_{Frag}$ forms a stable structure that mimics natively folded *in vitro* SChLAP1. A) Arc diagram of natively folded SChLAP1$_{Frag}$ in 5 mM magnesium (top) compared to arc diagram of natively folded SChLAP1 WT in 5 mM magnesium. B) Spearman correlation coefficients for SChLAP1$_{Frag}$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

Acknowledgements

The authors would like to acknowledge all members of the Hargrove lab, past and present, for invaluable feedback and support. The authors would also like to acknowledge Bill Day, Michael Peterson, and İrem Altan for assistance with computation. 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. 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 under Award No. W81XWH2010188. E.J.M. was supported by Duke University Center for Biomolecular and Tissue Engineering Ruth L. Kirschstein T32GM008555. J.P.F. was supported 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.

References

[1] Siegel, R. L., *et al.* **2019**. Cancer statistics, 2019. *CA Cancer J Clin.*

[2] Thomas, D. J., *et al.* **2007**. The ENCODE Project at UC Santa Cruz. *Nucleic acids research.*

[3] Willingham, A. T., *et al.* **2005**. A Strategy for Probing the Function of Noncoding RNAs Finds a Repressor of NFAT. *Science.*

[4] Ghildiyal, M., and Zamore, P. D. **2009**. Small silencing RNAs: an expanding universe. *Nature Reviews Genetics.*

[5] Tsai, M. C., *et al.* **2010**. Long noncoding RNA as modular scaffold of histone modification complexes. *Science.*

[6] Lee, J. T. **2012**. Epigenetic regulation by long noncoding RNAs. *Science.*

[7] Mercer, T. R., and Mattick, J. S. **2013**. Structure and function of long noncoding RNAs in epigenetic regulation. *Nature structural & molecular biology.*

[8] de Hoon, M., *et al.* **2015**. Paradigm shifts in genomics through the FANTOM projects. *Mamm Genome.*

[9] Prensner, J. R., *et al.* **2011**. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol.*

[10] Gerashchenko, G. V., *et al.* **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.*

[11] Chua, M. L. K., *et al.* **2017**. A Prostate Cancer "Nimbosus": Genomic Instability and SChLAP1 Dysregulation Underpin Aggression of Intraductal and Cribriform Subpathologies. *European urology.*

[12] Prensner, J. R., *et al.* **2014**. RNA biomarkers associated with metastatic progression in prostate cancer: a multi-institutional high-throughput analysis of SChLAP1. *The Lancet. Oncology.*

[13] Kidd, S. G., *et al.* **2021**. High expression of SCHLAP1 in primary prostate cancer is an independent predictor of biochemical recurrence, despite substantial heterogeneity. *Neoplasia (New York, N.Y.).*

[14] Mehra, R., *et al.* **2016**. Overexpression of the Long Non-coding RNA SChLAP1 Independently Predicts Lethal Prostate Cancer. *European urology.*

[15] Prensner, J. R., *et al.* **2013**. The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nature genetics.*

[16] Kadoch, C., *et al.* **2016**. PRC2 and SWI/SNF Chromatin Remodeling Complexes in Health and Disease. *Biochemistry.*

[17] Shain, A. H., and Pollack, J. R. **2013**. The spectrum of SWI/SNF mutations, ubiquitous in human cancers. *PloS one.*

[18] Roberts, C.W.M. and Orkin, S.H. **2004**. The SWI/SNF complex — chromatin and cancer. *Nat Rev Cancer.*

[19] Raab, J. R., *et al.* **2019**. SWI/SNF remains localized to chromatin in the presence of SCHLAP1. *Nature genetics.*

[20] Vierbuchen, T. *et al*. **2017** AP-1 transcription factors and the SWI/SNF complex mediate signal-dependent enhancer selection, pp 1067–1082.e1012, Mol. Cell.

[21] Chang, J., *et al.* **2018**. MALAT1 silencing suppresses prostate cancer progression by upregulating miR-1 and downregulating KRAS. *OncoTargets and therapy.*

[22] Grossi, E., *et al.* **2020**. A lncRNA-SWI/SNF complex crosstalk controls transcriptional activation at specific promoter regions. *Nat Commun.*

[23] Jegu, T., *et al.* **2019**. Xist RNA antagonizes the SWI/SNF chromatin remodeler BRG1 on the inactive X chromosome. *Nature structural & molecular biology.*
[24] Lino Cardenas, C. L., *et al.* **2018**. An HDAC9-MALAT1-BRG1 complex mediates smooth muscle dysfunction in thoracic aortic aneurysm. *Nat Commun.*
[25] Huang, M., *et al.* **2019**. lncRNA MALAT1 binds chromatin remodeling subunit BRG1 to epigenetically promote inflammation-related hepatocellular carcinoma progression. *Oncoimmunology.*
[26] Wang, Y., *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.*
[27] Kawaguchi, T., *et al.* **2015**. SWI/SNF chromatin-remodeling complexes function in noncoding RNA-dependent assembly of nuclear bodies. *Proc Natl Acad Sci U S A.*
[28] Cajigas, I. *et al*. **2015**. Evf2 lncRNA/BRG1/DLX1 interactions reveal RNA-dependent inhibition of chromatin remodeling. *Development.*
[29] Hu, G. *et al*. **2016**. LincRNA-Cox2 Promotes Late Inflammatory Gene Transcription in Macrophages through Modulating SWI/SNF-mediated Chromatin Remodeling. *J Immunol.*
[30] Skalska, L. *et al*. **2021**. Nascent RNA antagonizes the interaction of a set of regulatory proteins with chromatin. *Mol. Cell.*
[31] Sahu, A. (2015) The role of long noncoding RNA *SChLAP1* in prostate cancer, In *Molecular and Cellular Pathology*, University of Michigan.
[32] Smola, M. J., *et al.* **2015**. Detection of RNA-Protein Interactions in Living Cells with SHAPE. *Biochemistry.*
[33] Siegfried, N. A., *et al.* **2014**. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods.*
[34] Breaker, R. R. **2012**. Riboswitches and the RNA world. *Cold Spring Harb Perspect Biol.*
[35] Rangan, R., *et al.* **2020**. RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses: a first look. *RNA.*
[36] Watts, J. M., *et al.* **2009**. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature.*
[37] Wan, Y., *et al.* **2014**. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature.*
[38] 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.*
[39] Derrien, T., *et al.* **2012**. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research.*
[40] Srivastava, S., *et al.* **2017**. Many Genes-One Disease? Genetics of Nephronophthisis (NPHP) and NPHP-Associated Disorders. *Front Pediatr.*
[41] Reiterer, V., *et al.* **2020**. The dead phosphatases society: a review of the emerging roles of pseudophosphatases. *FEBS J.*
[42] Ji, J., *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.*
[43] 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.*
[44] Busan, S., *et al.* **2019**. Guidelines for SHAPE Reagent Choice and Detection Strategy for RNA Structure Probing Studies. *Biochemistry.*

[45] Adams, R. L., *et al.* **2019**. Sensitive detection of structural features and rearrangements in long, structured RNA molecules. *Methods Enzymol.*
[46] Weeks, S. B. a. K. M. **2018**. Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA.*
[47] Smola, M. J., *et al.* **2015**. Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat Protoc.*
[48] Laing, C., and Schlick, T. **2009**. Analysis of four-way junctions in RNA structures. *J Mol Biol.*
[49] Brazda, V., *et al.* **2019**. G4Hunter web application: a web server for G-quadruplex prediction. *Bioinformatics.*
[50] Kladwang, W., *et al.* **2020**. Anomalous Reverse Transcription through Chemical Modifications in Polyadenosine Stretches. *Biochemistry.*
[51] Dethoff, E. A., *et al.* **2018**. Pervasive tertiary structure in the dengue virus RNA genome. *Proc Natl Acad Sci U S A.*
[52] Smola, M. J., *et al.* **2016**. SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells.
[53] Jones, A. N., *et al.* **2020**. An evolutionarily conserved RNA structure in the functional core of the lincRNA Cyrano. *RNA.*
[54] Uroda, T., *et al.* **2019**. Conserved Pseudoknots in lncRNA MEG3 Are Essential for Stimulation of the p53 Pathway. *Molecular cell.*
[55] Sherpa, C., *et al.* **2018**. Structural characterization of maternally expressed gene 3 RNA reveals conserved motifs and potential sites of interaction with polycomb repressive complex 2. *Nucleic acids research.*
[56] Schmidt, K., *et al.* **2020**. Targeting the Oncogenic Long Non-coding RNA SLNCR1 by Blocking Its Sequence-Specific Binding to the Androgen Receptor. *Cell Rep.*
[57] Frank, F., *et al.* **2020**. The lncRNA Growth Arrest Specific 5 Regulates Cell Survival via Distinct Structural Modules with Independent Functions. *Cell Rep.*
[58] Horoszewicz, J. S., *et al.* **1983**. LNCaP model of human prostatic carcinoma. *Cancer Res.*
[59] Busan, S., and Weeks, K. M. **2018**. Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA.*
[60] Bellaousov, S., *et al.* **2013**. RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res.*
[61] Pyle, A. M. **2014**. Looking at LncRNAs with the ribozyme toolkit. *Molecular cell.*
[62] Zhang, B., *et al.* **2017**. Identification and Characterization of a Class of MALAT1-like Genomic Loci. *Cell Rep.*
[63] Somarowthu, S., *et al.* **2015**. HOTAIR forms an intricate and modular secondary structure. *Mol Cell.*
[64] Tavares, R. C. A., *et al.* **2019**. Phylogenetic Analysis with Improved Parameters Reveals Conservation in lncRNA Structures. *J Mol Biol.*
[65] Mustoe, A. M., *et al.* **2019**. RNA base-pairing complexity in living cells visualized by correlated chemical probing. *Proc Natl Acad Sci U S A.*
[66] Fang, R., *et al.* **2015**. Probing Xist RNA Structure in Cells Using Targeted Structure-Seq. *PLoS Genet.*
[67] Hu, T., *et al.* **2017**. Long non-coding RNAs transcribed by ERV-9 LTR retrotransposon act in cis to modulate long-range LTR enhancer function. *Nucleic acids research.*
[68] Alfeghaly, C., *et al.* **2021**. Implication of repeat insertion domains in the trans-activity of the long non-coding RNA ANRIL. *Nucleic Acids Research.*

[69] Wu, R., *et al.* **2021**. A Novel Human Long Noncoding RNA SCDAL Promotes Angiogenesis through SNF5-Mediated GDF6 Expression. *Adv Sci (Weinh).*
[70] Wang, X., *et al.* **2017**. Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines. *Mol. Cell.*