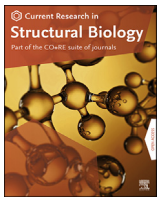
Current Research in Structural Biology 5 (2023) 100099

Contents lists available at ScienceDirect

Current Research in Structural Biology

journal homepage: www.journals.elsevier.com/current-research-in-structural-biology



Research Article

A cytosine-patch sequence motif identified in the conserved region of lincRNA-p21 interacts with the KH3 domain of hnRNPK

Aditi Maulik [1,**], Devleena Bandopadhyay [1], Mahavir Singh [*]

Molecular Biophysics Unit, Indian Institute of Science, Bengaluru, 560012, India

ARTICLE INFO

Handling Editor: Dr A Wlodawer

Keywords:
lincRNA-p21
Sequence conservation
Bioinformatics
hnRNPK
NMR CSP
ITC

ABSTRACT

Long Intergenic Non-coding RNAs (lincRNAs) are the largest class of long non-coding RNAs in eukaryotes, originating from the genome's intergenic regions. A ~4 kb long lincRNA-p21 is derived from a transcription unit next to the p21/Cdkn1a gene locus. LincRNA-p21 plays regulatory roles in p53-dependent transcriptional and translational repression through its physical association with proteins such as hnRNPK and HuR. It is also involved in the aberrant gene expression in different cancers. In this study, we have carried out a bioinformatics-based gene analysis and annotation of lincRNA-p21 to show that it is highly conserved in primates and identified two conserved domains in its sequence at the 5′ and 3′ terminal regions. hnRNPK has previously been shown to interact specifically with the 5′ conserved region of lincRNA-p21. hnRNPK is known to bind preferentially to the pyrimidine-rich (poly C) nucleotide sequences in RNAs. Interestingly, we observed a single occurrence of a cytosine-rich patch (C-patch) consisting of a CUCCCGC sequence in the 5′ conserved region of human lincRNA-p21, making it a putative hnRNPK binding motif. Using NMR and ITC experiments, we showed that the single-stranded C-patch containing RNA sequence motif interacts specifically with the KH3 domain of hnRNPK.

1. Introduction

Long Intergenic Non-coding RNAs (lincRNAs) are one of the highly abundant and functionally important classes of long non-coding RNAs (lncRNAs) in the eukaryotes, which originate from the intergenic regions in the genome (Ma et al., 2013). Although more than 3000 human lincRNAs have been discovered using transcriptomic data and bioinformatics analysis (Cabili et al., 2011; Derrien et al., 2012; Hrdlickova et al., 2014), only a subset (less than 1%) have been functionally characterized, as determining the functions of individual lincRNAs remains a challenge (Khalil et al., 2009). Nuclear lincRNAs function by targeting chromatin-modifying complexes and regulating the transcription (Guttman et al., 2011; Khalil et al., 2009; Yoon et al., 2012), whereas cytoplasmic lincRNAs regulate translational control of gene expression and transcript stability by binding with specific targets or acting as competing endogenous RNAs (Cesana et al., 2011; Kretz et al., 2013; Poliseno et al., 2010; Sumazin et al., 2011; Yoon et al., 2012). A few well-characterized examples are HOTAIR, XIST, and MALAT-1 lncRNAs that play roles in chromatin maintenance, X-chromosome inactivation, transcription regulation, cell motility, etc., showing that lincRNAs are a key regulator of diverse cellular processes (Gupta et al., 2010; He et al., 2014; Huarte and Rinn, 2010; Ji et al., 2003; Penny et al., 1996). Several lincRNAs (e.g., MALAT1, H19) are also involved in carcinogenic processes by interacting with cancer-associated genes or other non-coding RNAs (Huarte, 2015; Matouk et al., 2007; Raveh et al., 2015; Zhang et al., 2017a, 2017b).

LincRNA-p21 is a ~4 kb long lncRNA derived from a transcription unit next to the p21/Cdkn1A gene locus (hence named lincRNA-p21) (Guttman et al., 2009, 2011). The 'guardian of the genome' tumour suppressor p53 plays a key role in maintaining genomic integrity (Vazquez et al., 2008). Upon DNA damage, p53 triggers a transcriptional response resulting in either cell cycle arrest or apoptosis (Riley et al., 2008). While p53 is known to transcriptionally activate numerous genes directly, the mechanism by which p53 causes gene repression involves its interaction with other factors. For example, several lincRNAs that are physically linked with repressive chromatin-modifying complexes have been shown to act as repressors in p53 transcriptional regulatory networks (Khalil et al., 2009). LincRNA-p21 has been identified as one among such p53-activated lincRNAs (Huarte et al., 2010). p53 binds to the conserved canonical p53-binding motifs in the promoter regions of

* Corresponding author.
** Corresponding author.
E-mail addresses: aditi.maulik@gmail.com (A. Maulik), singh@iisc.ac.in (M. Singh).
[1] Equal contribution.

https://doi.org/10.1016/j.crstbi.2023.100099
Received 27 May 2022; Received in revised form 7 December 2022; Accepted 28 February 2023

lincRNA-p21, driving its expression (Dimitrova et al., 2014; el-Deiry et al., 1992; Funk et al., 1992; Winkler et al., 2022). LincRNA-p21, in turn, functions as a downstream transcriptional repressor of several genes. Several reports have also shown misregulation and differential expression of lincRNA-p21 in several cancers that include prostate, colorectal, chronic lymphatic leukaemia, and atherosclerosis (Blume et al., 2015; Isin et al., 2015; Tang et al., 2015; Wang et al., 2014; Zhai et al., 2013).

LincRNA-p21 mediates transcriptional repression through its association with heterogeneous nuclear ribonucleoprotein K (hnRNPK). hnRNPK was previously identified as a component of the repressor complex that acts in the p53 pathway (Kim et al., 2008; Moumen et al., 2005). It can bind both ssDNA and ssRNA via its three KH (hnRNPK homology) domains (Backe et al., 2005). Furthermore, lincRNA-p21 was shown to bind hnRNPK through a conserved 780 nucleotides long 5′ region. This interaction was required for the localization of lincRNA-p21 and subsequent induction of apoptosis via transcriptional repression of several p53-regulated genes (Huarte et al., 2010) to mediate the p53-dependent transcriptional repression of several genes. In a recent study, full-length transcription, splicing, and accumulation of lincRNA-p21 were dispensable for the chromatin organization of the locus and cis-regulation. Instead, 5' conserved elements in the exon 1 in nascent lincRNA-p21 were found to be sufficient to contribute to cis-activation, underscoring the functional importance of this region (Winkler et al., 2022).

In general, lncRNAs show poor sequence conservation across species showing scattered conserved regions surrounded by large seemingly unconstrained sequences (Zampetaki et al., 2018). Therefore, sequence and structure-based analysis of lncRNAs is important and can reveal the conserved and functionally important regions in RNA. A functional domain in lncRNA is likely to be conserved in animals and adopt a possibly shared structure by the orthologue sequences. This conservation pattern allows computational genome analysis that can be used to identify lncRNAs from different organisms and define their functional domains. This approach has been adopted to understand the origin and evolution of lincRNAs such as XIST and HOTAIR (Duret et al., 2006; Elisaphenko et al., 2008; He et al., 2011). Here, we have investigated lincRNA-p21 that functions in both cis and trans to understand its sequence and structural conservation using a similar approach (Dimitrova et al., 2014; Huarte et al., 2010). We addressed the questions, whether lincRNA-p21 exists and shows evolutionary conservation in mammals or vertebrates. To address this, we looked into 13 vertebrate genomes using Infernal (INFERence of RNA ALignment), a structure-based RNA homology search program (Nawrocki and Eddy, 2013). Our results showed that orthologues sequences of lincRNA-p21 exist only in mammals. The sequence analysis further revealed a single occurrence of a C-patch sequence motif (CUCCCGC) in the 5' conserved region of primate lincRNA-p21 that we hypothesize could be a putative hnRNPK binding site. The interaction between hnRNPK and C-patch sequence motif was probed using NMR and ITC methods. The results showed that the KH3 domain of hnRNPK primarily interacts with the C-patch sequence motif in a specific manner.

2. Materials and methods

2.1. Data

The sequence of human lincRNA-p21 long isoform (accession number KU881768.1) and mouse lincRNA-p21 (accession number NC_000083.6) were acquired from the National Center for Biotechnology Information (NCBI) database. The unmasked genome data of a human (GRCh38, Dec. 2013), chimpanzee (Pan_tro_3.0, May. 2016), rhesus monkey (Mmul_8.0.1, Nov. 2015), gorilla (gorGor4, Dec. 2014), cow (ARS-UCD1.2, Apt. 2018), horse (EquCab3.0, Jan. 2018), dolphin (turTru1, Jul. 2008), cat (Felis_catus_9.0, Nov. 2017), mouse (GRCm38, Jan. 2012), rat (Rnor_6.0, Jul. 2014), platypus (OANA5, Dec 2005), chicken (GRCg6a, Mar 2018), and zebrafish (GRCz11, May. 2017) were downloaded from Ensembl (release 97).

2.2. Acquiring sequences orthologues to human lincRNA-p21 through genome search

Human and mouse lincRNA-p21 sequences were aligned in Emboss (Madeira et al., 2019). The human lincRNA-p21 sequence ranging from 1 to 198 nucleotides and 199 to 3898 nucleotides were found to align with mouse lincRNA-p21 exon 1 and exon 2 sequences, respectively (Fig. 1). Accordingly, for the purpose of our study, the human lincRNA-p21 sequence was divided into two segments: 1–198 and 199–3898 nucleotides sequence spans. These two consecutive sequence ranges in human lincRNA-p21 were considered as mouse equivalent of human exon 1 and exon 2, respectively, and termed as segment 1 and segment 2 for human lincRNA-p21 (Fig. 1). Each of these two segments was separately used as a query to search the genomes of the human, chimpanzee, rhesus monkey, gorilla, cow, horse, dolphin, cat, mouse, rat, platypus, chicken, and zebrafish as several representative genomes in Ensembl using BLASTN (Chen et al., 2015). For each query, the best hit obtained in a particular genome was considered as the sequence ortholog of human lincRNA-p21 for the corresponding species. Likewise, the obtained sequence orthologues in humans, rhesus monkey, and cat were aligned using LocARNA (Will et al., 2007). Based on these alignment results, two queries (query 1 and query 2) were built using cmbuild and cmcalibrate functions of Infernal [v1.1] (Nawrocki and Eddy, 2013). Cmbuild is a program that builds a covariance model from an input multiple alignment and cmcalibrate calibrates E-value parameters for the covariance model. The calibrated models were used to search the whole genomes of 13 vertebrates using the cmsearch function of Infernal, a program that searches for a covariance model against any sequence database.

2.3. Protein expression and purification

463 residues long full length (FL) human hnRNPK (UniProtKB ID P61978) was cloned in the pET-16b vector that imparts an N-terminal hexahistidine purification tag to the protein. The tandem KH1+KH2 domain (residues 39–216) and KH3 domain (residues 384−463) of human hnRNPK were separately cloned in the pET-28a(+) vector with C-terminal hexahistidine purification tag. The proteins were expressed in *Escherichia coli BL21 (DE3)-Rosetta* cells using 1 mM isopropyl β-D-1-thiogalactopyranoside for induction at 37 °C for 3 h for hnRNPK FL and KH3 domain and at 18 °C for 16 h for KH1+KH2 domain. In the case of KH1+KH2, cells were pelleted and treated with Triton X-100 (0.1% v/v) and lysozyme (1 mg/ml). Cell lysates were sonicated, centrifuged and the supernatant was used for further protein purification. The first purification step involved His-tag Ni^{2+} affinity chromatography followed by size-exclusion (SEC) chromatography. In the case of hnRNPK FL and KH1+KH2, the final SEC chromatography was performed under a buffer consisting of 50 mM NaH_2PO_4, 300 mM NaCl, pH 8, 0.025 mM EDTA. In the case of KH3, the final SEC chromatography was performed under a buffer consisting of 50 mM NaH_2PO_4, 500 mM NaCl, pH 6, 0.025 mM EDTA. The uniformly ^{15}N-labeled KH3 was expressed in modified M9 medium using $^{15}NH_4Cl$ as the sole source of nitrogen. Protein purification protocol as detailed above was followed for the purification of ^{15}N-labeled KH3 domain.

2.4. Preparation of RNA for ITC and NMR titrations

The C-patch containing 29 nucleotides RNA fragment (FR1) was in vitro transcribed and purified using 15% urea-PAGE. RNA was electroeluted and further purified using an anion-exchange column (Mono Q column from GE). The heptamer (CUCCCGC) and mutant heptamer (CUAAAGC) RNA used in this study were commercially synthesized (Sigma). The RNA samples were dissolved in protein buffer

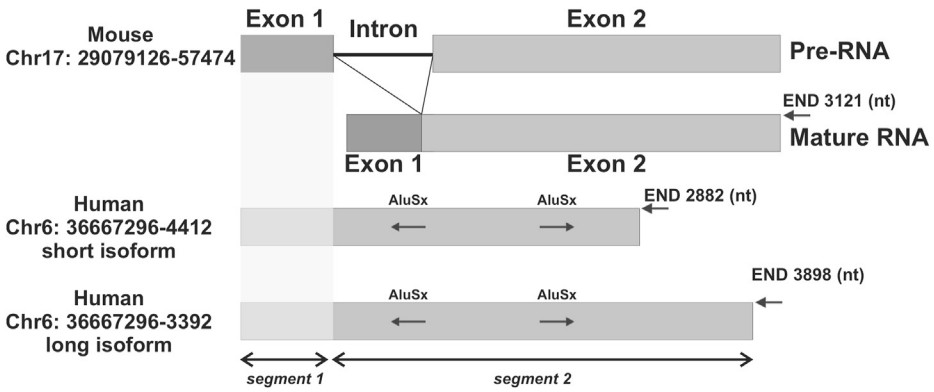


Fig. 1. Schematic of precursor (Pre-RNA) and mature mouse lincRNA-p21 (2 exons lncRNA), short and long isoform of human lincRNA-p21 (single exon lncRNA). The shaded box represents a region of sequence homology between mouse and human lincRNA-p21 (both SIsoE1 and LIsoE2). The Alu repeats found in the human lincRNA-p21 isoforms are mentioned. The two segments that span 1–198 and 199–3898 nucleotides in human lincRNA-p21 sequence are marked as segment 1 and segment 2 respectively.

(50 mM NaH_2PO_4, 300 mM NaCl, pH 8, 0.025 mM EDTA), heated at 95 °C for 3 min, and snap-cooled at 4 °C on ice.

2.5. Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) experiments were performed using a VP-ITC instrument (MicroCal). For the titration experiments, the sample cell was filled with 10 μM RNA and titrated with 200 μM protein. Twenty-nine injections of the titrant were performed with an interval of 180s between each injection. The data were integrated either automatically or manually using ORIGIN software provided with the VP-ITC (MicroCal). The data were fit for a one-site binding model. All the parameters were kept floating during data fitting.

2.6. NMR titration experiments

We recorded 2D ^{1}H–^{15}N HSQC spectra of KH3 in the free form and in complex with RNAs (CUCCCGC and mutant CUAAAGC) at protein-to-RNA ratios of 1:0, 1:0.5, 1:1 and 1:2 at 25 °C on 700 MHz Bruker NMR spectrometer equipped with a room temperature probe. NMR data were processed using the Bruker program Topspin 3.1, and spectra were

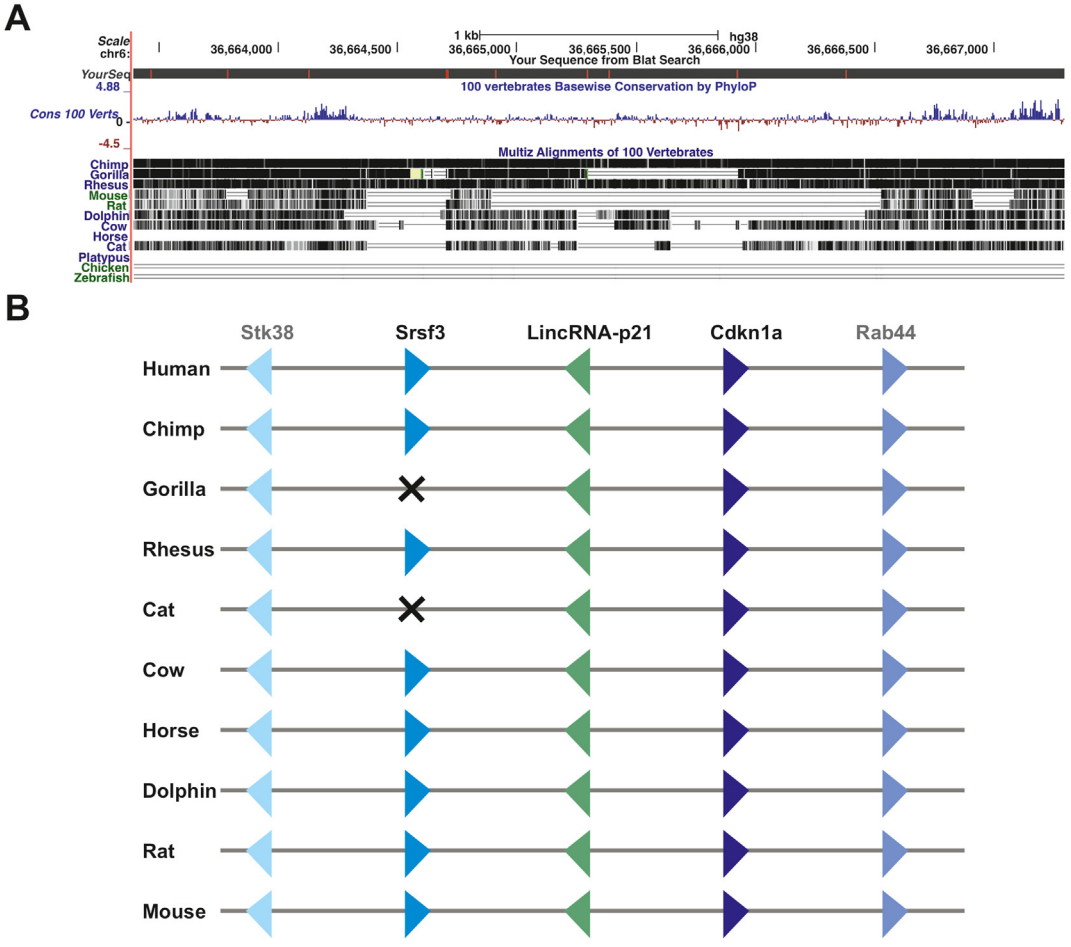


Fig. 2. Sequence conservation of lincRNA-p21 orthologs in mammals. (A) The sequences of lincRNA-p21 orthologs are conserved in primates but less conserved in other animals (from UCSC Genome browser). (B) The order and orientation of lincRNA-p21 genes and its neighboring protein coding genes in mammals (X represents the absence of Srsf3 gene).

analyzed using NMRFAM-SPARKY (Lee et al., 2015). The assignment of the KH3 domain was taken from a previously published study (Backe et al., 2005), where the assignment of the KH3 was deposited as BMRB entry 4405.

3. Results

3.1. The sequences of lincRNA-p21 orthologs show poor conservation among vertebrates

LincRNA-p21 is located between Srsf3 and Cdkn1a protein-coding genes on chromosome 6 in humans. In mice, the mature lincRNA-p21 is found on chromosome 17 and consists of two exons: exon 1 and exon 2 while, human lincRNA-p21 contains only a single exon that aligns with combined exon 1 and exon 2 of mouse lincRNA-p21 (Fig. 1). We searched several vertebrate genomes in the UCSC genome browser (http://genome.ucsc.edu) (Kent et al., 2002) for matches to lincRNA-p21. The whole sequence of human lincRNA-p21 showed apparent conservation among mammalian orthologs (Fig. 2A and B). Individually, when mouse exon 1 and exon 2 equivalent human lincRNA-p21 sequences (defined as segment 1 and segment 2, respectively) (Fig. 1) were searched using BLAT (Kent, 2002) and BLAST (Altschul et al., 1990) tools across different representative vertebrate genomes in Ensembl databases, they returned several hits with high to low scores. We assumed that if a hit is found in between Srsf3 and Cdkn1a genes with an E-value less than 1e-05, it can be considered as orthologs to the human lincRNA-p21 sequence. Notably, for both queries, the hits produced with the highest score (least E-value) were always located between the two protein-coding genes for all the different representative mammalian genomes studied here. This result suggests that lincRNA-p21 has orthologs in mammals. The top-scoring hits from each genome are listed in Table 1. The hits found in primates were high scoring, but hits from all other non-primate mammals had comparatively poor scores. In other words, close matches were observed only in primates and not in other mammals. Eventually, for other vertebrates, like platypus, chicken, and zebrafish, no hits were found between the two protein-coding genes. This finding implies that if lincRNA-p21 has orthologs in non-mammalian vertebrates, they may have moderate to low sequence conservation to be identified or revealed solely by the sequence search method.

3.2. Conserved orthologues regions of lincRNA-p21 exists in mammal and show high conservation among primates

LncRNAs are characterized not only by varying sequences but also by conserved structures. Therefore, we extended our investigation through structural search in different vertebrate genomes to confirm the existence of lincRNA-p21 only in mammals. Infernal was used to search different genomes for matches to lincRNA-p21. Infernal is a local RNA alignment and search program that uses the combination of sequence consensus and secondary structure conservation in RNA to generate a covariance model (Nawrocki and Eddy, 2013). To construct the query for building covariance model of lincRNA-p21 (necessarily a representative structure), we used human lincRNA-p21 segment 1 and segment 2 corresponding sequences and their identified orthologues sequences in rhesus monkey (BLASTN score: 336, E-value: 3e-90 for segment 1 and BLASTN score: 5174, E-value: 0 for segment 2), a primate that is distantly related to human and cat (BLASTN score: 60, E-value: 4e-07 for segment 1 and BLASTN score: 174, E-value: 3e-40 for segment 2) which is a non-primate mammal. Two queries (query 1 and query 2) were built with these sequences using the cmbuild and cmcalibrate functions of the Infernal tool (Nawrocki and Eddy, 2013). These two queries were used to search the complete genome of ten representative placental mammals (human, chimpanzee, rhesus monkey, gorilla, cow, horse, cat, dolphin, mouse, and rat), the extant mammal platypus, and two other vertebrates (non--mammals) chicken and zebrafish (Fig. 3) through the cmsearch function of Infernal. Orthologues sequences with secondary structural conservation of the lincRNA-p21 segments (hits located between Srsf3 and Cdkn1a with high scores) were obtained through Infernal search only in all the placental mammals but not in platypus or other vertebrates. Notably, each query produced just one high-scoring hit in the mammalian genomes (Table S1, Fig. 3). All other random hits were large in number and had low Infernal (cmsearch) scores. For example, in the gorilla genome, the score from the top hit to the next hit had drastically reduced from 246.2 (E-value 3.9e-66) to 23.7 (E-value 0.53) for query 1 and from 202.4 (E-value 6.5e-43) to 57.9 (E-value 9.4e-08) for query 2. The infernal score for the top hits for query 1 ranged from 246.2 (E-value 3.9e-66) in gorilla to 84.8 (E-value 2.9e-18) in the mouse genome. Similarly, for query 2, values for the top hits ranged between 203.6 (E-value 3.5e-43) in rhesus monkey to 71.4 (E-value 3.9e-11) in horse. These hits were located between Stk38/Srsf3 and Cdkn1a/Rab44 (note: Srsf3 is absent in some mammals, Fig. 2B). Overall, query 1 did not produce any high-scoring hits in rodents. Similarly, query 2 did not result in any high-scoring hits in ungulates like horse or dolphin. Moreover, both the queries produced good matches in primates but poor matches in the remaining mammals (Fig. 3 and Table S1).

The top hits from non-mammalian vertebrates (platypus, chicken, zebrafish) had less scores with high E-values (Table S1). For example, in chicken, the highest-scoring hit had an Infernal score as less as 25.2 (E-value 0.0666) and 25.9 (E-value 2.0) for query 1 and query 2, respectively. Insignificant scores implied the absence of any sequence or structural conservation of lincRNA-p21 in non-mammalian vertebrates. Moreover, the low-scoring hits were not found between Srsf3 and Cdkn1a genes. Thus, a combination of sequence and structural search confirmed that there are orthologues regions of lincRNA-p21 that exist only in mammals, and they have evolved further to become highly conserved in primates.

Table 1
BLASTN result from Ensemble genome browser.

Ensemble Hits			Segment 1		Segment 2	
			Score	E-value	Score	E-value
Mammals	Primates	Human (chr6)	392	5e-106	7198	0.0
		Chimp (chr6)	392	3e-107	6784	0.0
		Gorilla (chr6)	392	3e-107	6597	0.0
		Rhesus monkey (chr4)	336	3e-90	5174	0.0
	Ungulates	Cow (chr23)	73.8	3e-11	246	7e-62
		Horse (chr20)	71.9	1e-10	174	2e-40
		Dolphin	67.9	2e-09	281	1e-72
	Carnivorous	Cat (chrB2)	60	4e-07	174	3e-40
	Rodents	Rat (chr20)	61.9	1e-07	176	6e-41
		Mouse (chr17)	65.7	8e-09	152	1e-33
	Extant mammal	Platypus	No hit found with Srsf3 and/Cdkn1a			
Other vertebrates	Bird	Chicken	No hit found with Srsf3 and/Cdkn1a			
	Fish	Zebrafish	No hit found		No hit found with Srsf3 and/Cdkn1a	

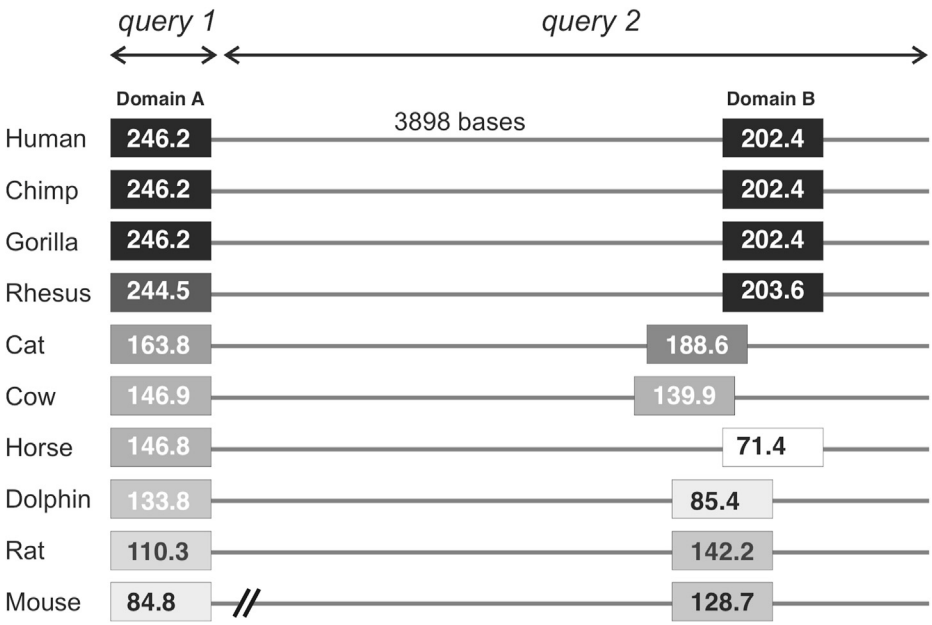


Fig. 3. The orthologs of lincRNA-p21 exist in mammals. Depiction of Infernal search result where each grey box represents hit with highest score (scores are written in the boxes) found for the respective genomes. Better conservation (better scores) is indicated by the darkness of the boxes. In each of the searched genomes two conserved domains are found for lincRNA-p21, domain A at 5′ terminal region that spans for the entire query 1 and domain B at 3′ terminal region, much shorter fragment than the query 2. Black lines indicate the unmatched part in the Infernal search. The double slashes in the schematic of the mouse gene indicate long intron region.

3.3. Two conserved domains identified in the 5′ and 3' terminal regions were found unique to lincRNA-p21

Besides one high-scoring hit located between Srsf3 and Cdkn1a, several scattered low-scoring hits of queries (query 1 and 2) were widely obtained in mammalian as well as other vertebrate genomes. In different genomes, several non-specific hits varied across human, chimpanzee, gorilla, rhesus monkey, cow, horse, dolphin, rat, mouse, platypus, chicken, and zebrafish for both query 1 and query 2. The 3707 nucleotides-long query 2 produced more numbers of low-scoring hits compared to 198 nucleotides-long query 1 that received far fewer numbers of hits. However, whether these hits have any functional roles is not clear. These low-scoring hits are expected to be insignificant and random. Considering that the best hit was less conserved in non-primate mammals and much shorter in all other genomes (Fig. 3), we inferred that for query 2, functionally conserved domain(s) in mammals must be much shorter than 3707 nucleotides. However, as the best hits for query 1 spanned the entire length for all the mammals (Fig. 3), we assume that the entire region of 198 nucleotides long segment 1 may have a structurally conserved function. The data in Fig. 3 show that the best structural hits were found as shorter fragments, one at the 5′ terminal region (domain A) and another towards the 3′ terminal region (domain B) of full-length lincRNA-p21. The result of a single high scoring hit, found in each case, also suggests that these two domains in segment 1 and segment 2 (henceforth called domain A and domain B, respectively) could form the unique, minimum functional, structural unit of lincRNA-p21, which is not shared by other lncRNAs. Thus, we conclude that the functional domain(s) of lincRNA-p21 conserved in mammals should be much shorter than the ~4 Kb length of this RNA. The orthologues of domains A and B were conserved in primates with Infernal score ranging from 246.2 (E−value 3.9e-66) to 244.5 (E−value 1.3e-65) and from 203.6 (E−value 6.5e-43) to 202.4 (E−value 3.5e-43), respectively and were less conserved in other mammals with 163.8 (E-value 4.3e-42) to 84.8 (E-value 2.9e-18) and 188.6 (E-value 1.2e-39) to 71.4 (E-value 3.9e-11) Infernal scores. Furthermore, while the orthologues of domain A were found to be much less conserved in rodents with an Infernal score of 110.3 (E-value 2.3e-26) and 84.8 (E-value 2.9e-18) for rat and mouse (Fig. 3, Table S1), respectively, the orthologues of domains B were less conserved in ungulates like dolphin and horse with 85.4 (E-value 1.6e-14) and 71.4 (E-value 3.9e-11) Infernal score, respectively. This observation suggests that the two domains have experienced different evolutionary constraints.

3.4. Single occurrence of a C-patch sequence containing, potential KH binding motif in 5' terminal conserved domain of human lincRNA-p21

A cytosine patch (C-patch) is defined as three cytosine residues or more in a row present in a RNA sequence (Nakamoto et al., 2020). We found a single occurrence of C-patch containing sequence motif (CUCCCGC) within the 5′ conserved domain (domain A) of human lincRNA-p21. hnRNPK has three RNA binding KH domains (KH1-3). KH domains, in general, have been shown to have binding specificity towards a core motif of four, predominantly C-rich pyrimidine nucleotide sequences containing motifs in single-stranded DNA and RNA (Braddock et al., 2002). To inspect the conservation of the identified C-patch in human lincRNA-p21, we generated multiple sequence alignments for human, and other primates along with mouse. Sequence analysis showed a C-patch motif at the extreme 5′ end of lincRNA-p21 sequence in the primates, including human, chimpanzee, gorilla, and monkey (Supplemental Fig. S1A). Sequence alignment of 5′ conserved domain of mouse lincRNA-p21 with the primates shows an all pyrimidine UCUC motif at the equivalent position (Supplemental Fig. S1A). Often, the hnRNPK binding motifs are conserved for the pyrimidine sequences and contain U nucleotides along with C. Interestingly, however, mouse lincRNA-p21 has an extended 5' extension that contains multiple C-rich patches (Supplemental Fig. S1A), besides the UCUC sequence, which may also act as hnRNPK binding sites. Therefore, we speculate that these C-patch sequences in primates and mouse can be the hnRNPK binding sites in the conserved domain A of lincRNA-p21.

In the Mfold predicted minimum free energy structure at 5′ terminal domain A of human lincRNA-p21, the C-patch containing sequence motif (CUCCCGC) was found in the single-stranded loop region as shown in Fig. 4A. No C-patch containing sequence motif was found in any single-stranded region within the Mfold predicted minimum free energy structure of 3' domain B of human lincRNA-p21 (Fig. 4B).

We also used R-scape (Rivas et al., 2017) to generate a covariance structure of domain A by taking a multiple-sequence alignment as an input file. A multiple sequence alignment of human, chimpanzee, gorilla, monkey, and mouse lincRNA-p21 domain A was performed using LocARNA webserver (Will et al., 2007) and used as an input to generate

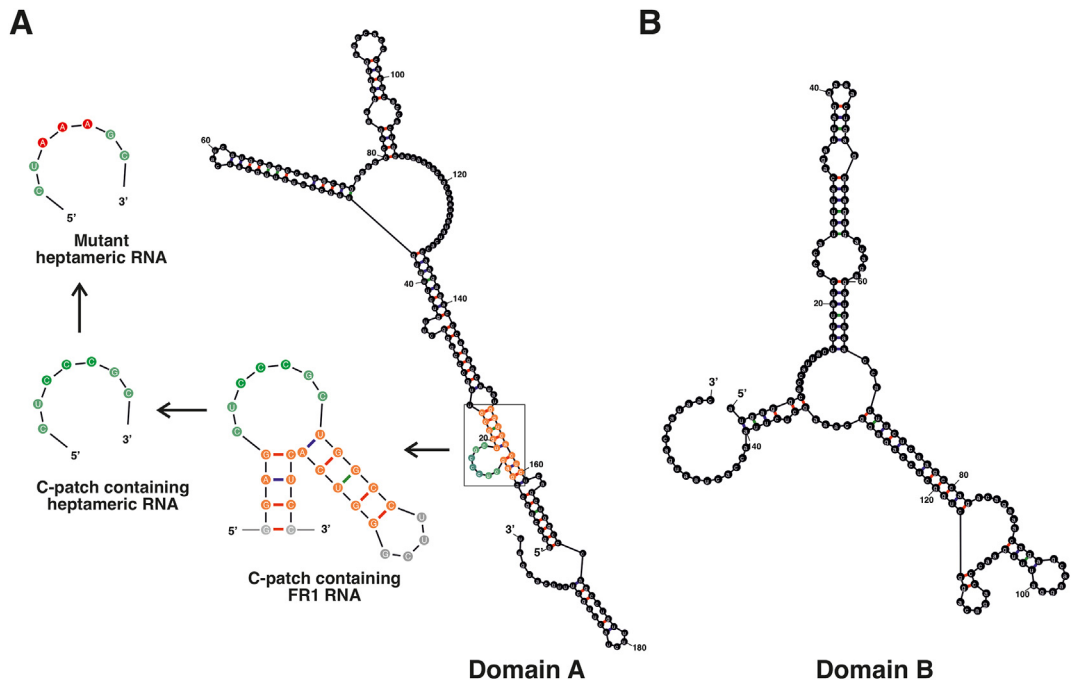


Fig. 4. Mfold predicted structure of conserved domains of lincRNA-p21. (A) Mfold predicted minimum free energy structure of domain A present in the 5′ terminal region of lincRNA-p21. The region marked with a rectangular box depicts the 29 nt long FR1 fragment (in color) that contains a single stranded loop region consisting of a C-patch sequence motif (CUCCCGC) (in green). Mfold predicted structure of FR1 is shown. This fragment contains no-native UUCG tetraloop and G.U base pair at the start (extra 5′ G nucleotide was added to increase the in vitro transcription yield by T7 RNA polymerase). Single stranded, C-patch containing and mutant heptameric sequences used in this study are also depicted. (B) Mfold predicted minimum free energy structure of domain B in the 3′ terminal region of lincRNA-p21.

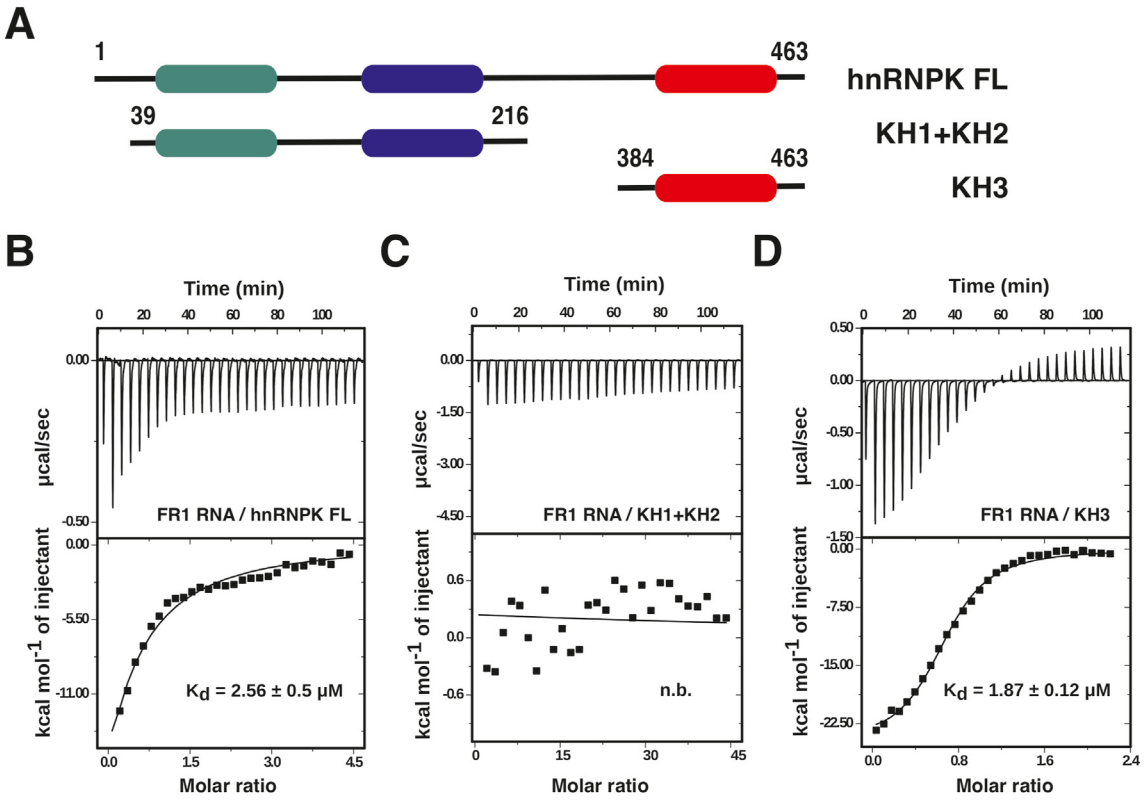


Fig. 5. Interaction of FR1 RNA with the hnRNPK domains were monitored using ITC experiments at 25 °C. (A) Domain architecture of human hnRNPK. Different protein constructs used in this study are depicted. (B–D) Interaction of FR1 RNA with (B) hnRNPK FL. (C) KH1+KH2 tandem domain, and (D) KH3 domain of hnRNPK. The top portion of each panel shows the raw titration data, and the bottom portion of each panel shows the integrated heats fit to single-binding site isotherms. The calculated K_d values are shown in each panel.

covariance RNA structure using R-scape. The R-scape generated model also confirmed the presence of C-patch in a single-stranded loop region in the primates and mouse (Supplemental Figs. S1B and C). Thus, we hypothesize that the C-patch containing fragment (FR1) in domain A of human lincRNA-p21 (Fig. 4A) may constitute the potential binding motif for hnRNPK (Huarte et al., 2010).

3.5. Isothermal titration calorimetry experiments revealed that KH3 of hnRNPK is the main domain that interacts with the C-patch containing fragment of lincRNA-p21

We overexpressed and purified the full-length (FL) hnRNPK, KH1+KH2, and KH3 domains from *E. coli* (Fig. 5A and Supplemental Figs. S2A, B, C). The 29 nt long C-patch containing FR1 RNA was in vitro transcribed and purified (Supplemental Fig. S2D). FR1 adopts a stem-loop structure as predicted by Mfold consisting of the C-patch containing CUCCCGC heptameric loop (Fig. 4A).

ITC experiments were carried out to study the interaction of hnRNPK FL, KH1+KH2 domain, and KH3 domain (Fig. 5A) with the FR1 fragment of lincRNA-p21. At 25 °C, an enthalpy-driven ($\Delta H = -11.05$ kcal/mol) binding with an apparent K_d of 2.56 μM was observed for the FR1 RNA and hnRNPK FL interaction (Table 2 and Fig. 5B). The titration of FR1 RNA with tandem KH1+KH2 domains resulted in heat change which did not saturate even at higher ratio of RNA to protein (Supplemental Fig. S3A). Most of the observed heat change was due to the dilution of the KH1+KH2 into the buffer, as observed in the control KH1+KH2 into buffer titration (Supplemental Fig. S3B). Upon subtracting this heat of dilution from the integrated heat change of FR1 RNA and KH1+KH2 titration, the net heat change for the protein – RNA titration was small in magnitude (Fig. 5C). Therefore, we conclude that KH1+KH2 does not bind to FR1 RNA with a significant affinity. Next, we probed the interaction of the KH3 domain with the C-patch containing FR1 RNA using ITC. ITC result showed that FR1 RNA binds the KH3 domain of hnRNPK in an enthalpy-driven ($\Delta H = -29.88$ kcal/mol) binding with an apparent K_d of 1.87 μM (Table 2 and Fig. 5D). Therefore, taken together, we conclude that the KH3 domain constitutes the main domain of hnRNPK that binds to FR1 from lincRNA-p21, which harbours a C-patch containing single-stranded sequence motif.

3.6. Isothermal titration calorimetry experiments revealed the specific nature of the interaction between KH3 and C-patch-containing sequences

In the next set of ITC experiments, we used a single-stranded, C-patch containing heptameric sequence (CUCCCGC) to probe its interaction with the KH3 domain. The ITC experiments were carried out in a buffer containing 0 and 100 mM NaCl with 50 mM NaH_2PO_4, pH 6 at 25 °C.

Table 2
Thermodynamic parameters of interactions of hnRNPK FL, KH1+KH2, and KH3 with different RNAs sequences at 25 °C were obtained using ITC experiments.

	K_d (μM)	ΔG (kcal/mol)	ΔH (kcal/mol)	TΔS (kcal/mol)
hnRNPK-FL/FR1 RNA	2.56 ± 0.5	−7.62 ± 1.7	−11.05 ± 2.5	−3.43 ± 4.3
KH3/FR1 RNA	1.87 ± 0.12	−7.80 ± 0.5	−29.88 ± 0.5	−22.1 ± 1
KH1+KH2/FR1 RNA	n.b.	n.b.	n.b.	n.b.
KH3/heptamer RNA(in 0 mM NaCl)	4.31 ± 0.68	−7.31 ± 1.2	−14.06 ± 1.1	−6.75 ± 2.2
KH3/heptamer RNA(in 100 mM NaCl)	4.74 ± 1	−7.26 ± 2.2	−7.48 ± 1.1	−0.23 ± 3.2
KH3/mutant heptamer RNA (in 0 mM NaCl)	n.b.	n.b.	n.b.	n.b.
KH3/mutant heptamer RNA (in 100 mM NaCl)	n.b.	n.b.	n.b.	n.b.

n.b. no binding observed

Table 2 shows the results obtained from ITC titration experiments. The KH3 of hnRNPK interacts with the heptamer RNA with K_d of 4.31 μM and 4.74 μM, at 0 and 100 mM NaCl concentration, respectively (Table 2 and Fig. 6A and B). The K_d values for this interaction are in the range similar to the KH3 and FR1 interaction (1.87 μM), suggesting that heptameric single-stranded, C-patch containing loop of FR1 likely constitutes the main KH3 binding region in FR1 RNA.

In the next experiment, we titrated the KH3 to a mutant heptamer RNA sequence, where three consecutive cytosines were mutated to adenines (CUCCCGC to CUAAAGC). We observed very small heat change in the mutant heptamer and KH3 titration experiments at both 0 and 100 mM NaCl (Fig. 6C and D), showing that the mutant heptamer sequence does not interact with the KH3 domain. This result revealed that the KH3 – C-patch interaction is specific for the C-rich sequence.

3.7. NMR chemical shift perturbation experiments showed that hnRNPK KH3 domain interacts specifically with the C-patch containing RNA

A 0.2 mM uniformly ^{15}N labeled KH3 sample was titrated with FR1 RNA and 2D ^{1}H–^{15}N HSQC of KH3 was recorded at each titration step. Resonance cross peaks resulting from the backbone NHs from each residue, except proline, can be followed using these spectra. We started with a 1:0.1 protein: RNA molar ratio and increased the RNA concentration at subsequent steps (Fig. 7A). At a 1:0.2 ratio, we found distinct chemical shift perturbation in G404, K405, G406, G407 residues, which are part of the GXXG motif. Peaks corresponding to G404 and G407 were perturbed, whereas K405 and G406 had disappeared and could not be traced. We observed severe peak broadening beyond the 1:0.2 (protein to RNA) titration resulting in the disappearances of the majority of the peaks. This is likely due to the intermediate exchange of residues involved in FR1 binding. These observations prove that FR1 binds to the KH3 domain of hnRNPK and the conserved GXXG sequence motif residues are involved in the RNA binding (Fig. 7A).

We further performed NMR CSP experiments to understand the interaction of single-stranded C-patch containing heptameric RNA sequence and KH3 domain of hnRNPK. A 0.2 mM uniformly ^{15}N labeled KH3 sample was titrated with heptamer RNA at 1:0.5, 1:1 and 1:2 protein:RNA molar ratios, and the 2D ^{1}H–^{15}N HSQC NMR spectrum was recorded at each step of the titration (Fig. 7B). We observed distinct CSPs in selected residues in KH3 upon RNA binding starting from a 1:0.5 to 1:2 protein:RNA ratio in an intermediate to slow exchange time scale.

In the next experiment, the uniformly ^{15}N labeled KH3 was titrated with mutant heptamer RNA and a 2D ^{1}H–^{15}N HSQC NMR spectrum was recorded at each titration step. We did not observe any significant CSPs in the spectra of KH3, even at 1:2 protein-to-RNA molar ratios (Fig. 7C). This corroborated the ITC results and confirmed the sequence-specific nature of KH3 and single-stranded C-patch RNA interaction.

Nucleic acid binding in KH domains has been shown to occur through a narrow groove formed by two unstructured loops (Fig. 7D). Loop 1, which is present between α-helix 1 and 2, contains the G-X-X-G motif (G = glycine, X = any amino acid) which is crucial for oligonucleotide binding. In the case of the hnRNPK KH3 domain, the sequence of the G-X-X-G motif is G404-K405-G406-G407 (Fig. 7D). Loop 2, between β strands 2 and 3, is of variable length and sequence in different KH domains and flanks the nucleic acid binding groove. The core DNA/RNA C-rich motif and the water molecules create a strong network of molecular interactions comprising hydrogen bonds, hydrophobic interactions, and base stacking with KH domain binding (Backe et al., 2005). In the invariant G-X-X-G motif-containing loop in KH3 of hnRNPK, the G406 is involved in hydrogen bonding with the phosphate group of a pyrimidine residue (mostly cytosine). The G404 and G407 are involved in van der Waals contact with the core recognition motif of an oligonucleotide bound to loop 1 of the KH3 domain. The K405 which is part of the G-X-X-G motif is involved in electrostatic interactions with the phosphate backbone. We observed large chemical shift perturbation in all the residues G404, K405, G406 and G407 from the G-X-X-G motif in loop 1 of

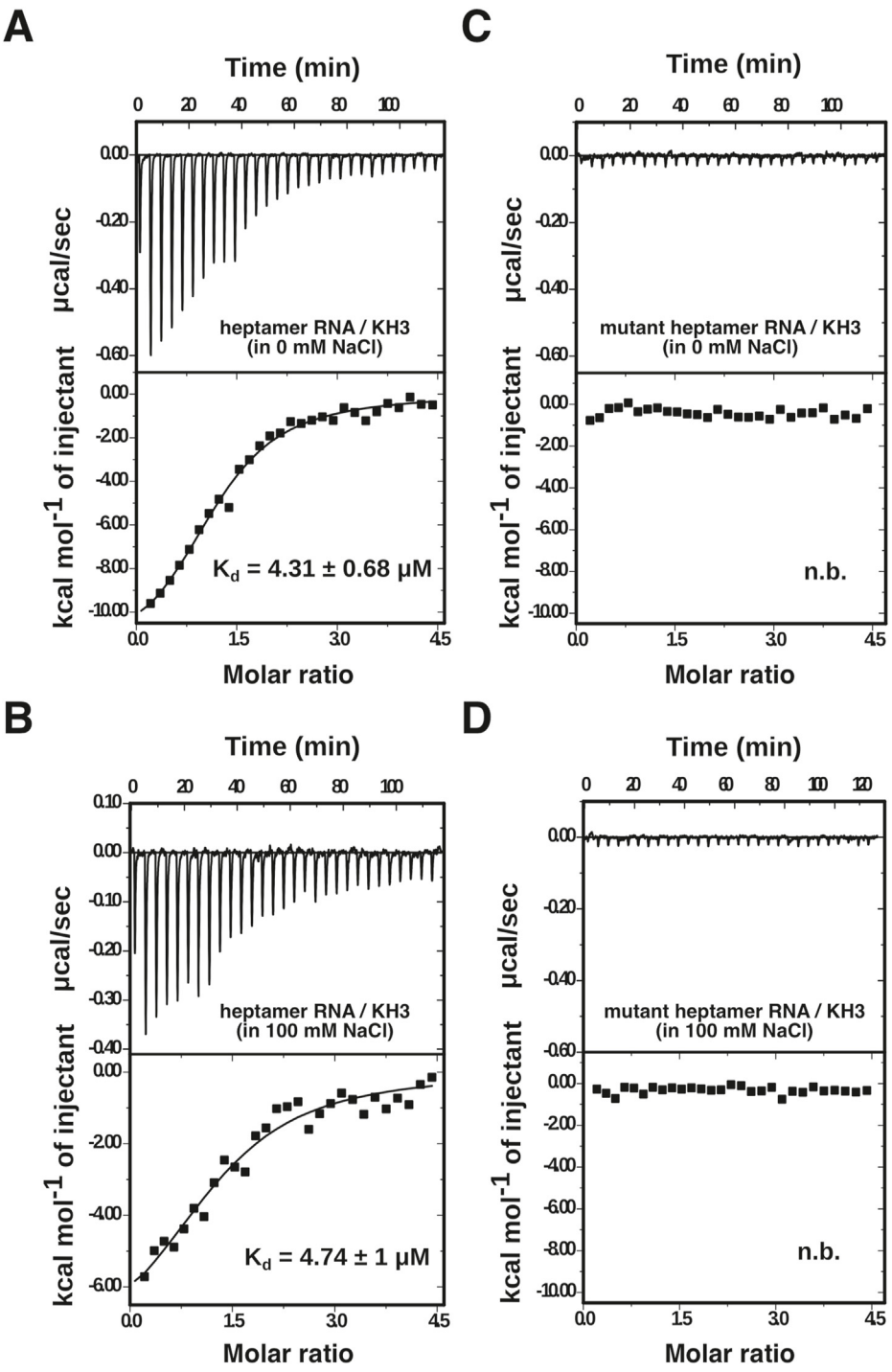


Fig. 6. Heptameric RNA/KH3 domain interaction monitored using ITC experiments at 25 °C. (A, B) Interaction of KH3 domain with heptameric RNA (CUCCCGC) in presence of (A) 0 mM NaCl and (B) 100 mM NaCl salt concentration. (C, D) Interaction of KH3 domain with mutant heptamer RNA sequence (CUAAAGC) in presence of (C) 0 mM NaCl and (D) 100 mM NaCl salt concentration. The top portion of each panel shows the raw titration data, and the bottom portion of each panel shows the integrated heats fit to single-binding site isotherms. The calculated K_d values are shown in each panel.

KH3 when titrated with the FR1 or heptamer RNA from lincRNA-p21 (Fig. 7A and B). The cross-peak of G406 is not visible in the free KH3 spectra, however it appears at a distinct chemical shift position in the heptamer bound KH3 spectra, similar to an observation that has been reported previously (Backe et al., 2005). The assignment of these residues is taken from a previously published study (Backe et al., 2005). The C-patch binding cleft on KH3 is narrow, which confers the specificity for binding pyrimidine bases (C and U nucleotides). On the other hand, larger purine nucleobase would not fit in the narrow cleft; hence the C-rich sequence binds the KH3 domain, specifically (Sidiqi et al., 2005).

4. Discussion

Our bioinformatic-based gene analysis and annotation showed that lincRNA-p21 is highly conserved in primates and harbours two conserved domains in its sequence at the 5′ and 3’ terminal regions. The orthologues sequences of lincRNA-p21 identified in different organisms, using the RNA homology search software Infernal, contain sequence mismatch with gaps; hence we hypothesize that lincRNA-p21 may have conserved structural domains despite divergent sequences in different mammals. This is a general feature of lncRNAs (Ulitsky and Bartel, 2013). The

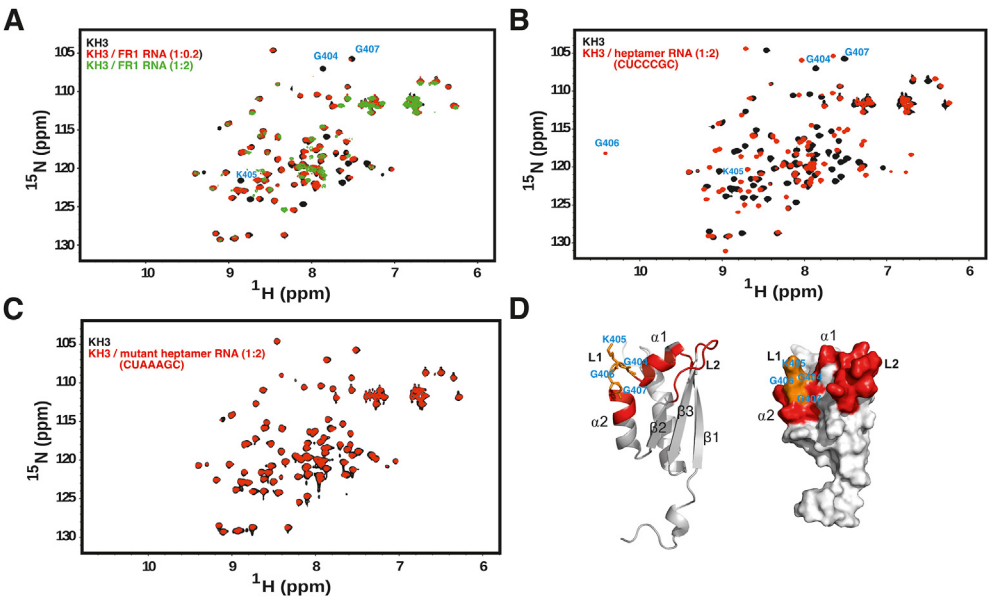


Fig. 7. Chemical shift perturbations observed in the KH3 domain upon titration with C-patch containing FR1 and heptameric RNAs. (A) Overlay of 2D $^{1}H–^{15}N$ HSQC spectra of the free KH3 domain (black) and KH3/FR1 RNA at 1:0.2 (red) and 1:2 (green) protein to RNA molar ratios. Residues G404, K405, and G407 from G-X-X-G motif in loop 1 that show chemical shift perturbations and peak broadening are marked. (B) Overlay of 2D $^{1}H–^{15}N$ HSQC spectra of the free KH3 domain (black) and KH3/heptamer RNA (red) at 1:2 protein to RNA molar ratio. Residues G404, K405, G406 and G407 from G-X-X-G motif in loop 1 that show large chemical shift perturbations are marked. (C). Overlay of 2D $^{1}H–^{15}N$ HSQC spectra of the free KH3 domain (black) and KH3/mutant heptamer RNA (red) at 1:2 protein to RNA molar ratio. (D) Ribbon and surface model structures of KH3 domain showing secondary structural elements. The C-patch binding region is shown in red, and within that, the G-X-X-G motif is shown in orange and residues are marked. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

compensatory mutations are commonly found in ncRNAs that change the primary sequence but maintain the secondary structures, a feature postulated to be one of the reasons behind poor primary sequence conservation of ncRNAs (Pang et al., 2006; Torarinsson et al., 2006). For example, XIST and HOTAIR lncRNAs contain fast-evolving sequences and highly conserved structures (He et al., 2011; Nesterova et al., 2001).

An Infernal search found just one high-scoring hit of lincRNA-p21 in each mammal located between Cdkn1a and Srsf3 genes. Thus, it can be concluded that conserved orthologues regions of lincRNA-p21 exist only in mammals. This also suggests that these domains could harbour functional motifs in lincRNA-p21. However, additionally, several low-scoring hits were produced in many other places in mammalian and other vertebrate genomes, which suggests that there may exist other lncRNAs that share functional domains similar to lincRNA-p21.

Recent studies have revealed that hnRNPK interacts with several long non-coding RNAs, including lincRNA-p21 to control their functions and nuclear localization. hnRNPK has been shown to have a binding preference for multi-cytosine-patch (C-patch) present in lncRNAs (Nakamoto et al., 2020). We noticed an occurrence of a single-stranded C-patch containing sequence in the 5′ terminal conserved region (domain A) of human lincRNA-p21. hnRNPK had been shown to recognize C-patch containing four nucleotide core motifs specifically. Indeed recently, Nakamoto et al. analyzed the publicly available enhanced crosslinking immunoprecipitation (eCLIP) data for hnRNPK in two different cell lines, HepG2 and K562 and identified enriched hnRNPK binding motifs (Nakamoto et al., 2020). All the identified motifs from both cell types contained multiple cytosine patches (as defined by three or more cytosines in a row) interspersed with one or two other nucleotides (Nakamoto et al., 2020). Therefore, we hypothesize that observed C-patch containing sequence motif in the 5′ terminal domain A of lincRNA-p21 could be a putative hnRNPK binding motif. ITC and NMR titration experiments showed that hnRNPK interacts with the C-patch containing sequence motif at the 5’ terminal of lincRNA-p21 mainly using its KH3 domain. The mutation of three cytosines to adenines in the C-patch resulted in a complete loss of binding to the KH3 domain, showing that the interaction between KH3 and the C-patch is specific.

CRediT authorship contribution statement

Aditi Maulik: conceived the research, designed the research, performed the research, analyzed the results and wrote the manuscript. **Devleena Bandopadhyay:** performed the research. **Mahavir Singh:** conceived the research, designed the research, analyzed the results and wrote the manuscript, All the authors reviewed the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Any data or material reported in this article will be made available by the corresponding author upon a reasonable request.

Acknowledgement

This work is supported by an Early Career Fellowship from Wellcome Trust/DBT India Alliance to A.M. (IA/E/15/1/502321). Research in M.S. lab is supported by grants from the Department of Biotechnology (DBT), India (BT/PR15829/BRB/10/1469/2015) and the IISc-DBT partnership program. The authors also acknowledge funding for infrastructural support from the following programs of the Government of India: DST-FIST, UGC-CAS, and the DBT-IISc partnership program. M.S. is a recipient of Ramalingaswami Fellowship from DBT, India and STAR award from SERB, DST, India (award number STR/2021/000015). We thank Dr. Stephen Cusack for the hnRNPK DNA-containing plasmid.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crstbi.2023.100099.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Backe, P.H., Messias, A.C., Ravelli, R.B., Sattler, M., Cusack, S., 2005. X-ray crystallographic and NMR studies of the third KH domain of hnRNP K in complex with single-stranded nucleic acids. Structure 13, 1055–1067.

Blume, C.J., Hotz-Wagenblatt, A., Hullein, J., Sellner, L., Jethwa, A., Stolz, T., Slabicki, M., Lee, K., Sharathchandra, A., Benner, A., et al., 2015. p53-dependent Non-coding RNA Networks in Chronic Lymphocytic Leukemia (Leukemia).

Braddock, D.T., Baber, J.L., Levens, D., Clore, G.M., 2002. Molecular basis of sequence-specific single-stranded DNA recognition by KH domains: solution structure of a complex between hnRNP K KH3 and single-stranded DNA. EMBO J 21, 3476–3485.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., Rinn, J.L., 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Gene Dev. 25, 1915–1927.

Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A., Bozzoni, I., 2011. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell 147, 358–369.

Chen, Y., Ye, W., Zhang, Y., Xu, Y., 2015. High speed BLASTN: an accelerated MegaBLAST search tool. Nucleic Acids Res. 43, 7762–7768.

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 Res. 22, 1775–1789.

Dimitrova, N., Zamudio, J.R., Jong, R.M., Soukup, D., Resnick, R., Sarma, K., Ward, A.J., Raj, A., Lee, J.T., Sharp, P.A., et al., 2014. LincRNA-p21 activates p21 in cis to promote Polycomb target gene expression and to enforce the G1/S checkpoint. Mol. Cell 54, 777–790.

Duret, L., Chureau, C., Samain, S., Weissenbach, J., Avner, P., 2006. The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. Science 312, 1653–1655.

el, Deiry, W.S., Kern, S.E., Pietenpol, J.A., Kinzler, K.W., Vogelstein, B., 1992. Definition of a consensus binding site for p53. Nat. Genet. 1, 45–49.

Elisaphenko, E.A., Kolesnikov, N.N., Shevchenko, A.I., Rogozin, I.B., Nesterova, T.B., Brockdorff, N., Zakian, S.M., 2008. A dual origin of the Xist gene from a protein-coding gene and a set of transposable elements. PLoS One 3, e2521.

Funk, W.D., Pak, D.T., Karas, R.H., Wright, W.E., Shay, J.W., 1992. A transcriptionally active DNA-binding site for human p53 protein complexes. Mol. Cell Biol. 12, 2866–2871.

Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.C., Hung, T., Argani, P., Rinn, J.L., et al., 2010. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464, 1071–1076.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., et al., 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458, 223–227.

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., et al., 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature 477, 295–300.

He, J.H., Han, Z.P., Li, Y.G., 2014. Association between long non-coding RNA and human rare diseases (Review). Biomed. Rep. 2, 19–23.

He, S., Liu, S., Zhu, H., 2011. The sequence, structure and evolutionary features of HOTAIR in mammals. BMC Evol. Biol. 11, 102.

Hrdlickova, B., Kumar, V., Kanduri, K., Zhernakova, D.V., Tripathi, S., Karjalainen, J., Lund, R.J., Li, Y., Ullah, U., Modderman, R., et al., 2014. Expression profiles of long non-coding RNAs located in autoimmune disease-associated regions reveal immune cell-type specificity. Genome Med. 6, 88.

Huarte, M., 2015. The emerging role of lncRNAs in cancer. Nat. Med. 21, 1253–1261.

Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M., et al., 2010. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell 142, 409–419.

Huarte, M., Rinn, J.L., 2010. Large non-coding RNAs: missing links in cancer? Hum. Mol. Genet. 19, R152–R161.

Isin, M., Uysaler, E., Ozgur, E., Koseoglu, H., Sanli, O., Yucel, O.B., Gezer, U., Dalay, N., 2015. Exosomal lncRNA-p21 levels may help to distinguish prostate cancer from benign disease. Front. Genet. 6, 168.

Ji, P., Diederichs, S., Wang, W., Boing, S., Metzger, R., Schneider, P.M., Tidow, N., Brandt, B., Buerger, H., Bulk, E., et al., 2003. MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene 22, 8031–8041.

Kent, W.J., 2002. BLAT–the BLAST-like alignment tool. Genome Res. 12, 656–664.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler, D., 2002. The human genome browser at UCSC. Genome Res. 12, 996–1006.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., et al., 2009. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. In: Proceedings of the National Academy of Sciences of the United States of America, vol. 106, pp. 11667–11672.

Kim, K., Choi, J., Heo, K., Kim, H., Levens, D., Kohno, K., Johnson, E.M., Brock, H.W., An, W., 2008. Isolation and characterization of a novel H1.2 complex that acts as a repressor of p53-mediated transcription. J. Biol. Chem. 283, 9113–9126.

Kretz, M., Siprashvili, Z., Chu, C., Webster, D.E., Zehnder, A., Qu, K., Lee, C.S., Flockhart, R.J., Groff, A.F., Chow, J., et al., 2013. Control of somatic tissue differentiation by the long non-coding RNA TINCR. Nature 493, 231–235.

Lee, W., Tonelli, M., Markley, J.L., 2015. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. Bioinformatics 31, 1325–1327.

Ma, L., Bajic, V.B., Zhang, Z., 2013. On the classification of long non-coding RNAs. RNA Biol. 10, 925–933.

Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., et al., 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 47, W636–W641.

Matouk, I.J., DeGroot, N., Mezan, S., Ayesh, S., Abu-lail, R., Hochberg, A., Galun, E., 2007. The H19 non-coding RNA is essential for human tumor growth. PLoS One 2, e845.

Moumen, A., Masterson, P., O'Connor, M.J., Jackson, S.P., 2005. hnRNP K: an HDM2 target and transcriptional coactivator of p53 in response to DNA damage. Cell 123, 1065–1078.

Nakamoto, M.Y., Lammer, N.C., Batey, R.T., Wuttke, D.S., 2020. hnRNPK recognition of the B motif of Xist and other biological RNAs. Nucleic Acids Res. 48, 9320–9335.

Nawrocki, E.P., Eddy, S.R., 2013. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29, 2933–2935.

Nesterova, T.B., Slobodyanyuk, S.Y., Elisaphenko, E.A., Shevchenko, A.I., Johnston, C., Pavlova, M.E., Rogozin, I.B., Kolesnikov, N.N., Brockdorff, N., Zakian, S.M., 2001. Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. Genome Res. 11, 833–849.

Pang, K.C., Frith, M.C., Mattick, J.S., 2006. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. Trends Genet. 22, 1–5.

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., Brockdorff, N., 1996. Requirement for Xist in X chromosome inactivation. Nature 379, 131–137.

Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W.J., Pandolfi, P.P., 2010. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 465, 1033–1038.

Raveh, E., Matouk, I.J., Gilon, M., Hochberg, A., 2015. The H19 Long non-coding RNA in cancer initiation, progression and metastasis - a proposed unifying theory. Mol. Cancer 14, 184.

Riley, T., Sontag, E., Chen, P., Levine, A., 2008. Transcriptional control of human p53-regulated genes. Nat. Rev. Mol. Cell Biol. 9, 402–412.

Rivas, E., Clements, J., Eddy, S.R., 2017. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. Nat. Methods 14, 45–48.

Sidiqi, M., Wilce, J.A., Vivian, J.P., Porter, C.J., Barker, A., Leedman, P.J., Wilce, M.C., 2005. Structure and RNA binding of the third KH domain of poly(C)-binding protein 1. Nucleic Acids Res. 33, 1213–1221.

Sumazin, P., Yang, X., Chiu, H.S., Chung, W.J., Iyer, A., Llobet-Navas, D., Rajbhandari, P., Bansal, M., Guarnieri, P., Silva, J., et al., 2011. An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. Cell 147, 370–381.

Tang, S.S., Zheng, B.Y., Xiong, X.D., 2015. LincRNA-p21: implications in human diseases. Int. J. Mol. Sci. 16, 18732–18740.

Torarinsson, E., Sawera, M., Havgaard, J.H., Fredholm, M., Gorodkin, J., 2006. Thousands of corresponding human and mouse genomic regions unalignable in primary sequence contain common RNA structure. Genome Res. 16, 885–889.

Ulitsky, I., Bartel, D.P., 2013. lincRNAs: genomics, evolution, and mechanisms. Cell 154, 26–46.

Vazquez, A., Bond, E.E., Levine, A.J., Bond, G.L., 2008. The genetics of the p53 pathway, apoptosis and cancer therapy. Nat. Rev. Drug Discov. 7, 979–987.

Wang, G., Li, Z., Zhao, Q., Zhu, Y., Zhao, C., Li, X., Ma, Z., Li, X., Zhang, Y., 2014. LincRNA-p21 enhances the sensitivity of radiotherapy for human colorectal cancer by targeting the Wnt/beta-catenin signaling pathway. Oncol. Rep. 31, 1839–1845.

Will, S., Reiche, K., Hofacker, I.L., Stadler, P.F., Backofen, R., 2007. Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. PLoS Comput. Biol. 3, e65.

Winkler, L., Jimenez, M., Zimmer, J.T., Williams, A., Simon, M.D., Dimitrova, N., 2022. Functional elements of the cis-regulatory lincRNA-p21. Cell Rep. 39, 110687.

Yoon, J.H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J.L., De, S., Huarte, M., Zhan, M., Becker, K.G., Gorospe, M., 2012. LincRNA-p21 suppresses target mRNA translation. Mol. Cell 47, 648–655.

Zampetaki, A., Albrecht, A., Steinhofel, K., 2018. Long non-coding RNA structure and function: is there a link? Front. Physiol. 9, 1201.

Zhai, H., Fesler, A., Schee, K., Fodstad, O., Flatmark, K., Ju, J., 2013. Clinical significance of long intergenic noncoding RNA-p21 in colorectal cancer. Clin. Colorectal Cancer 12, 261–266.

Zhang, L., Zhou, Y., Huang, T., Cheng, A.S., Yu, J., Kang, W., To, K.F., 2017a. The interplay of LncRNA-H19 and its binding partners in physiological process and gastric carcinogenesis. Int. J. Mol. Sci. 18.

Zhang, X., Hamblin, M.H., Yin, K.J., 2017b. The long noncoding RNA Malat1: its physiological and pathophysiological functions. RNA Biol. 14, 1705–1714.