

RESEARCH ARTICLE

Mitochondrial sequencing identifies long noncoding RNA features that promote binding to PNPase

 Andrew D. Taylor,^{1,2}  Quincy A. Hathaway,^{1,3,4}  Amina Kunovac,^{1,2} Mark V. Pinti,^{2,5} Mackenzie S. Newman,⁶ Chris C. Cook,⁷ Evan R. Cramer,⁸ Sarah A. Starcovic,⁸ Michael T. Winters,⁹ Emily S. Westemeier-Rice,⁹ Garrett K. Fink,¹  Andrya J. Durr,^{1,2} Saira Rizwan,^{1,2} Danielle L. Shepherd,^{1,2} Aaron R. Robart,⁸ Ivan Martinez,⁹ and  John M. Hollander^{1,2}

¹Division of Exercise Physiology, West Virginia University School of Medicine, Morgantown, West Virginia, United States;

²Mitochondria, Metabolism, and Bioenergetics Working Group, West Virginia University School of Medicine, Morgantown, West Virginia, United States; ³Heart and Vascular Institute, West Virginia University, Morgantown, West Virginia, United States;

⁴Department of Medical Education, West Virginia University School of Medicine, Morgantown, West Virginia, United States; ⁵West Virginia University School of Pharmacy, Morgantown, West Virginia, United States; ⁶Department of Physiology

and Pharmacology, West Virginia University School of Medicine, Morgantown, West Virginia, United States; ⁷Cardiovascular

and Thoracic Surgery, West Virginia University School of Medicine, Morgantown, West Virginia, United States; ⁸Department of

Biochemistry, West Virginia University School of Medicine, Morgantown, West Virginia, United States; and ⁹Department of

Microbiology, Immunology, and Cell Biology, West Virginia University Cancer Institute, School of Medicine, Morgantown, West Virginia, United States

Abstract

Extranuclear localization of long noncoding RNAs (lncRNAs) is poorly understood. Based on machine learning evaluations, we propose a lncRNA-mitochondrial interaction pathway where polynucleotide phosphorylase (PNPase), through domains that provide specificity for primary sequence and secondary structure, binds nuclear-encoded lncRNAs to facilitate mitochondrial import. Using FVB/NJ mouse and human cardiac tissues, RNA from isolated subcellular compartments (cytoplasmic and mitochondrial) and cross-linked immunoprecipitate (CLIP) with PNPase within the mitochondrion were sequenced on the Illumina HiSeq and MiSeq, respectively. lncRNA sequence and structure were evaluated through supervised [classification and regression trees (CART) and support vector machines (SVM)] machine learning algorithms. In HL-1 cells, quantitative PCR of PNPase CLIP knock-out mutants (KH and S1) was performed. In vitro fluorescence assays assessed PNPase RNA binding capacity and verified with PNPase CLIP. One hundred twelve (mouse) and 1,548 (human) lncRNAs were identified in the mitochondrion with Malat1 being the most abundant. Most noncoding RNAs binding PNPase were lncRNAs, including Malat1. lncRNA fragments bound to PNPase compared against randomly generated sequences of similar length showed stratification with SVM and CART algorithms. The lncRNAs bound to PNPase were used to create a criterion for binding, with experimental validation revealing increased binding affinity of RNA designed to bind PNPase compared to control RNA. The binding of lncRNAs to PNPase was decreased through the knockout of RNA binding domains KH and S1. In conclusion, sequence and secondary structural features identified by machine learning enhance the likelihood of nuclear-encoded lncRNAs binding to PNPase and undergoing import into the mitochondrion.

NEW & NOTEWORTHY Long noncoding RNAs (lncRNAs) are relatively novel RNAs with increasingly prominent roles in regulating genetic expression, mainly in the nucleus but more recently in regions such as the mitochondrion. This study explores how lncRNAs interact with polynucleotide phosphorylase (PNPase), a protein that regulates RNA import into the mitochondrion. Machine learning identified several RNA structural features that improved lncRNA binding to PNPase, which may be useful in targeting RNA therapeutics to the mitochondrion.

bioinformatics; heart; lncRNA; machine learning; mitochondria

INTRODUCTION

Noncoding RNAs (ncRNAs) are a diverse group of transcribed genes that do not generate proteins but instead possess

unique regulatory roles. Long noncoding RNAs (lncRNAs) are a class of noncoding RNAs that assume multiple responsibilities within the cell, including regulating chromatin structure, histone modification, modifying mRNA transcription, microRNA

(miRNA) sponging, and regulation of protein translation and modification (1, 2). In general, lncRNAs have reduced sequential conservation across species, but as the field has grown, more and more functional orthologues are being discovered that lack alignment but possess positional, splicing, and/or structural homology (3). Current explorations of lncRNA activity are primarily focused on cytoplasmic and nuclear localization, while other subcellular regions have remained largely ignored. Moreover, procedures examining localization of lncRNAs often lump nuclear and mitochondrial subcellular fractions together (4). The mitochondrion represents an isolated, spatially dense region containing a distinct genome that can serve as a stoichiometrically favorable compartment for interactions between lncRNAs and other RNA species (5). Recent investigations from our laboratory as well as others have assessed the involvement of lncRNAs in altering mitochondrial function and gene expression (6–8), but the direct interactions of lncRNA within the mitochondrion remain predominantly uncertain. The mechanism by which nuclear-encoded lncRNAs pass through the mitochondrial outer and inner membranes is even more unclear.

Polynucleotide phosphorylase (PNPase) is a ribonuclease that is conserved across bacterial, archaeal, and eukaryotic domains of life (9, 10). In bacteria, PNPase is a primary regulator of mRNA degradation (11, 12), modifying the cellular response to temperature change (13) and biofilm formation (14), as well as functioning as a processing enzyme for small RNA (15). Progressing through the evolutionary timeline, eukaryotic PNPase is now localized specifically to the intermembrane space of the mitochondrion, seemingly lacking its bacterial functional roles (16–18). In 2010, the first established RNA import function for PNPase in mammalian systems (19) revealed that tRNA and ribozyme machinery are transported by PNPase from the intermembrane space of the mitochondrion into the matrix, and it was hypothesized that two external domains of PNPase (annotated as KH and S1) recognize stem-loop structures for RNA import. Further evaluation of these domains has shown that their deletion does not alter PNPase's mitochondrial localization but will drastically reduce RNA binding (9, 16, 20–23). Mammalian PNPase has been shown to form a trimer structure that embeds in the inner mitochondrial membrane with a positively charged central pore that has been hypothesized to serve as a channel for RNAs (see Supplemental Fig. S11C) (21). Additionally, changes in PNPase expression have resulted in alterations in cellular function and miRNA import without directly altering mitochondrial-encoded RNA content (24–26). PNPase, through its KH and S1 RNA binding domains, may provide a functional context for noncoding RNA (ncRNA) import and regulation in the mitochondrion.

In the current study, we investigated ncRNA presence in cardiac mitochondria from mice and humans. Through comparison of both cytoplasmic and isolated mitochondrial ncRNA populations, we identified hundreds of nuclear-encoded lncRNAs within the mitochondrion, including the lncRNA metastasis-associated lung adenocarcinoma transcript 1 (Malat1). We further examined the mechanisms controlling the localization of lncRNAs into mitochondria. PNPase binding to lncRNAs is greater when specific primary and secondary structural conformations are present. We performed experimental validation of these hypotheses through

gel shift assays, revealing increased binding of a predicted binding construct to PNPase. Additionally, initiation of binding by the lncRNA Malat1 to PNPase is significantly impaired when the KH and S1 RNA-binding domains are perturbed. With the present data, we propose that nuclear-encoded lncRNAs with specific primary sequence and secondary structural features identified by machine learning have a greater propensity to bind to PNPase and undergo import into the mitochondrion than lncRNAs that lack those features.

EXPERIMENTAL PROCEDURES

Study Approval and Patient Population

The West Virginia University Institutional Review Board and Institutional Biosafety Committee approved the studies and data generated from this work (IRB No. 1812394926), including all tissue and patient information acquired (27). Informed written consent was obtained from all patients by the Heart and Vascular Institute, J.W. Ruby Memorial Hospital at West Virginia University School of Medicine. Right atrial appendages were removed during open-heart and/or valvular surgeries, and all tissue and data were stored in a double-deidentified process. There was no incentive provided for patients. Patients' tissue was used irrespective of sex, race, or ethnicity. Patients enrolled in the study were not part of a clinical trial. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Murine Model

The West Virginia University Animal Care and Use Committee approved all animal studies, including animal housing, sedation, euthanasia, and experimentation. These studies conformed to the most current National Institutes of Health (NIH) *Guidelines for the Care and Use of Laboratory Animals*. FVB/NJ (wild-type) male and female mice were housed in the West Virginia University Health Sciences Center Animal Facility and given access to a rodent diet and water ad libitum. Animals were aged to 25 wk and were euthanized. Both sexes were evaluated indiscriminately.

Mitochondrial Isolations

Mitochondrial subpopulations were isolated for analyses as previously described (28), with modifications by our laboratory (29–31). Briefly, mitochondria were isolated from both human right atrial tissue and mouse whole heart. *Buffer 1* (100 KCl mmol/L, 50 MOPS mmol/L, 5 MgSO₄·7H₂O mmol/L, 1 EGTA mmol/L, and 1 ATP mmol/L at pH 7.4) was used at a 1:10 (wt:vol) ratio to homogenize samples at 4°C. Samples were then centrifuged at 700 g for 10 min. Following centrifugation, the supernatant was extracted and recentrifuged at 10,000 g. The precipitated pellet was washed in *buffer 1* and centrifuged at 10,000 g twice. The precipitant from the 700 g spin was further processed by resuspending in a solution of KCl-MOPS-EGTA buffer (100 KCl mmol/L, 50 MOPS mmol/L, and 0.5 EGTA mmol/L at pH 7.4) and digested with 5 mg/g of trypsin for 10 min. Following the 10-min incubation, a

protease inhibitor cocktail (Biovision, Mountain View, CA) was added to the mixture to stop the digestion and centrifuged at 700 g for 10 min. The supernatant was collected and centrifuged at 10,000 g for 10 min. Both sets of isolated mitochondria (from the initial supernatant and the digested pellet) were combined in a sucrose buffer containing 220 sucrose mmol/L, 70 mannitol mmol/L, 10 Tris-HCl mmol/L, and 1 EDTA mmol/L (pH 7.4).

Mitochondria were isolated from both human right atrial tissue and mouse whole heart. Differential centrifugation allowed for the compartmentalization of nuclear, cytoplasmic, and mitochondrial fractions. Mitochondrial subpopulations, subsarcolemmal and interfibrillar, were combined to form a total mitochondrial population. Mitochondria were further purified through the use of a sucrose gradient (23%, 15%, 10%, and 3% Percoll solution) by centrifugation in a Beckman Optima MAX-XP Ultracentrifuge (Beckman Coulter, Fullerton, CA) at 32,000 g for 8 min, as previously described (32). Final mitochondrial isolates were stored in a KME buffer (100 mM KCL, 50 mM MOPS, and 0.5 mM EGTA (pH 7.4).

PNPase Overexpression and Knockdown Constructs

Express Cloning Vectors were synthesized through GenScript (Piscataway, NJ) using the pcDNA3.1 + N-eGFP backbone. Sequences for the full-length PNPase protein (FL), KH domain knockout of exon 23 of PNPase (KH), and the S1 domain knockout of the COOH terminus of PNPase (S1) were cloned into vectors and the complete sequence information is provided (Supplemental Fig. S1). Vectors were then transfected into a bacterial cell line (DH5 α), grown through antibiotic selection, and plasmid DNA was isolated for transfection. Immortalized HL-1 murine cardiomyocytes were cultured at 37°C, 5% CO₂, as previously described (24, 32, 33). Cells were supplemented with Claycomb media (Sigma-Aldrich, St. Louis, MO) containing 10% FBS, glutamine (2 mM), penicillin/streptomycin, and norepinephrine (0.1 mM). Cells were seeded and transfection occurred at 70–80% confluence. Lipofectamine 3000 (Thermo Fisher, Waltham, MA) was used to transfect cells, per the manufacturer's instructions.

Briefly, cells were divided into five groups; no plasmid control (NP), pcDNA3.1 + N-eGFP only (GFP), FL, KH, and S1. In the GFP, FL, KH, and S1 sets, 10 μ g of each respective plasmid were transfected. Media, with the Lipofectamine 3000, remained on the cells for 48 h followed by cell imaging using the EVO FL Auto Imaging System (Thermo Fisher) on the GFP fluorescent and phase contrast channels. Cells were washed with PBS, dissociated with 0.05% trypsin, and preserved at –80°C. For knockdown experiments of PNPase, HL-1 cells were seeded at 70–80% confluence. Lipofectamine 3000 was mixed with Opti-MEM Medium and 10 μ M stock solution of Silencer siRNA for Pnpt1 (cat. no. AM16708, assay ID no. 177160, Thermo Fisher) targeting exon 3. The siRNA-lipid complex was added to cells with a final concentration of 50 nM siRNA and incubated for 2 days. Cells were washed with PBS, dissociated with 0.05% trypsin, and preserved at –80°C.

Malat1 Fluorescent In Situ Hybridization

HL-1 cells were seeded on 18-mm round no. 1 coverslips in 6-well dishes and were allowed to reach 70–80% confluency,

with growth conditions provided above. For fluorescent in situ hybridization (FISH), a modified version of the Stellaris RNA FISH Protocol for Adherent Cells was used. Stellaris FISH Probes against Malat1 and Gapdh (Biosearch Technologies, Inc., Petaluma, CA) were utilized in the hybridization, and NucBlue Fixed Cell Stain ReadyProbes reagent (Thermo Fisher) was used to stain nuclei. Following MitoTracker Deep Red FM (Thermo Fisher) incubation, coverglass was washed with 1 mL 1 $\times$ phosphate-buffered saline (PBS), and 1 mL Stellaris RNA FISH Fixation Buffer [3.7% (vol/vol) formaldehyde in 1 $\times$ PBS] was then added to the six-well dishes and incubated at room temperature for 10 min. The cells were washed twice with 1 mL 1 $\times$ PBS and permeabilized with 1 mL 70% ethanol for 16 h at 4°C. The ethanol was eluted and the cells were incubated with Stellaris RNA FISH Wash Buffer A [Biosearch Technologies; 10% (vol/vol) formamide in Wash Buffer A] for 5 min at room temperature.

The cells were transferred, on the coverglass, to a humidified chamber and incubated with 100 μ L of Stellaris RNA FISH Hybridization Buffer [Biosearch Technologies; 10% (vol/vol) formamide in Hybridization Buffer] containing either a Malat1 (no. SMF-3008-1) or Gapdh (no. SMF-3002-1) probe (Stellaris FISH Probes with Quasar 570 Dye) for 16 h at 37°C in darkness. Each Stellaris FISH Probe has ~80 hybridization sites on the target RNA. The coverglass was then incubated with 1 mL Stellaris RNA FISH Wash Buffer A for 30 min at 37°C in darkness, the buffer was aspirated, and one drop of NucBlue Fixed Cell Stain ReadyProbes reagent (Thermo Fisher; DAPI stain) was added. The cells were then incubated for 30 min at 37°C in darkness. The NucBlue stain was removed, 1 mL Stellaris RNA FISH Wash Buffer B (Biosearch Technologies) was added, and the cells were allowed to incubate at room temperature for 5 min. A drop of Vectashield Mounting Medium (Vector Laboratories, San Diego, CA) was added to a microscope slide, and the coverglass was mounted. The mounted cells were sealed with clear nail polish and stored at 4°C.

Cells were incubated with 250 nM of MitoTracker Deep Red FM (Thermo Fisher) in Opti-MEM (Thermo Fisher) growth media for 30 min and then washed with Claycomb media. The cells were incubated with the Stellaris RNA FISH Hybridization Buffer containing either a Malat1 or Gapdh probe for 16 h at 37°C in darkness. Cells were imaged using the Nikon AIR/SIM (Nikon, Minato City, Tokyo, Japan) confocal microscope. Images were processed using the NIS Elements AR (Nikon) software.

Cross-Linking Immunoprecipitation

Cell culture.

Cells in the NP, GFP, FL, KH, and S1 groups were resuspended in PBS and transferred to a 24-well plate on ice. Additionally, cells without plasmid transfection (i.e., NP) were assigned to an immunoprecipitation (IP) control (i.e., no cross linking) and antibody control (i.e., no antibody). Cross linking was performed using the CX-2000 Crosslinker (Analytik Jena, Upland, CA) at (400 mJ/cm²) five times, as previously described (32), in all groups with the exception of the IP control. After cross linking, cells were centrifuged (1,200 $\times$ g) for 7 min and subsequently frozen (–80°C) for further use. Then, 75 μ L per sample of Dynabeads Protein G (Thermo Fisher) were prepared through three washes with

NP-40 buffer (20 mM Tris at pH 8.0, 137 mM NaCl, 10% glycerol, 1% Triton X100, 2 mM EDTA, and 0.1 mM PMSF). After the third wash, the supernatant on the beads was discarded and 5 μ g of anti-GFP was added, along with NP-40 buffer, to reach a total volume of 100 μ L. Beads were incubated overnight at 4°C. The next day, each sample was resuspended in NP-40 buffer to a total volume of 1 mL, and protein concentrations were determined through the Bradford Method (34). Then, 100 μ g of protein from each sample was added to a new tube followed by RNase I treatment (1:500 dilution, Ambion RNase I, Thermo Fisher) for 3 min at 37°C, which allowed for partial digestion of RNA bound to PNPase. The reaction was stopped by incubating the samples on ice. The supernatant on the beads solution was then removed, and the protein mixture was added and incubated at 4°C for 4 h. Following the incubation, beads were washed three times in NP-40 buffer, the final supernatant was removed, and a mixture of water/NuPAGE LDS Sample Buffer (4 $\times$; Thermo Fisher) was added to the beads. The beads solution was heated at 70°C for 15 min and assessed through immunoblotting (see *Western Blotting*) and sequenced (see *High-Throughput Sequencing Cross-Linking Immunoprecipitation* below).

Isolated mitochondria.

Cross-linking immunoprecipitation (CLIP) was also performed on isolated mitochondria (see *Mitochondrial Isolations*) from human and mouse cardiac tissue. To procure sufficient protein, ($n = 3$) biological replicates were pooled together four separate times (i.e., $n = 4$ pooled groups of mitochondria, each containing $n = 3$ biological replicates). From pooled mitochondrial samples in human ($n = 4$) and mouse ($n = 4$), 5% of each sample volume was saved for an IP control (i.e., no cross linking) and 5% volume for an antibody control (i.e., no antibody). The CLIP protocol was performed as described above, including cross linking, centrifugation, bead preparation, assessment of protein concentration, RNase treatment, and incubation and denaturation steps. With the tissue samples, in place of incubation with anti-GFP (as this was used for GFP-tagged proteins in the plasmids), samples were incubated with either 5 μ g anti-PNPase (human), 5 μ g anti-PNPase (mouse), or 5 μ g anti-Ago2 (mouse). Differential expression was not performed as both human and mouse samples have no comparison groups in the current study; rather the CLIP analysis was performed for descriptive purposes. Immunoblotting (see *Western Blotting*) and sequencing (see *High-Throughput Sequencing Cross-Linking Immunoprecipitation*) provide further information.

In vitro binding assay.

CLIP with FAM-labeled RNA probes was performed on isolated mitochondria from mouse cardiac tissue as described above, with an additional 4 h incubation of RNA and lysate before cross linking. Following incubation with beads, the supernatant was kept and marked as unbound RNA. After three washes, beads as well as their bound content were added to a black 96-well plate in duplicate along with an unbound RNA sample, RNA was added to NP-40 buffer, and NP-40 with beads was added for normalization. The plate was read at wavelengths 485–528, with NP-40 buffer used as a blank.

Western Blotting

Using 4–12% gradient gels, immunoblotting was performed through MOPS SDS-PAGE, as previously described (35–37). Using the Bradford Method, protein concentrations were normalized. Primary antibodies (1:1,000 dilution) implemented in the study included the following: anti-GFP (B-2: sc-9996, SCBT, Dallas, TX), anti-PNPase (C-5: sc-271973, SCBT), anti-PNPase (D-1: sc-271479, SCBT), anti-GAPDH (ab8245, Abcam, Cambridge, MA), anti-VDAC1 (voltage-dependent anion channel; SAB5201374, Sigma), and anti-cytochrome *c* oxidase subunit 4 (anti-COX IV; ab16056, Abcam). Rabbit anti-mouse IgG (H&L) horseradish peroxidase (HRP) conjugate 1:10,000 (ab6728, Abcam) and goat anti-rabbit (H&L) HRP conjugate 1:10,000 (ab6721, Abcam) were used as secondary antibodies when appropriate. Normalization of protein content was through GAPDH (cytoplasmic) and VDAC (mitochondrial) expression. Chemiluminescence was quantified with Radiance Chemiluminescent Substrate (Azure Biosystems, Dublin, CA), per manufacturer's instructions, and imaged using the G:Box Bioimaging system (Syngene, Frederick, MD). Images were taken with GeneSnap/GeneTools software (Syngene). Densitometry was analyzed using ImageJ and Fiji Software (NIH, Bethesda, MD). For CLIP samples, the area of nitrocellulose membrane containing the GFP (cells) or PNPase (human/mouse) fluorescent region was excised and saved for RNA analysis.

RNA Isolation/Quantitative PCR

Using the miRNeasy Mini Kit (product no. 217004, Qiagen, Hilden, Germany), per manufacturer's instructions, RNA was isolated from 20 mg of human right atrial tissue and mouse whole heart, as well as from samples of nitrocellulose membrane containing RNA derived from CLIP samples. Each sample was homogenized in QIAzol lysis reagent before proceeding. Total RNA was isolated for each sample and RNA to be analyzed through quantitative (q)PCR was converted to cDNA through the High-Capacity RNA-to-cDNA Kit (Thermo Fisher), per manufacturer's instructions. Malat1 levels were also measured in HL-1 cells following PNPase CLIP (as described above). Experiments were performed on the Applied Biosystems 7500 Fast Real-Time PCR system (Applied Biosystems, Foster City, CA), using 2 $\times$ SYBR Green Master Mix. Quantification was achieved using the $2^{-\Delta\Delta Ct}$ method (38), standardized to U6 expression where appropriate. Primer sequences are provided (Supplemental Table S1) and were designed through NCBI Primer BLAST and Primer3 (39).

Long Noncoding RNA Sequencing

RNA, isolated from human atrial appendages and mouse whole heart (as described above), was sequenced through the West Virginia University Genomics Core Facility. The complete differential expression profiles of mitochondrial and cytoplasmic groups are included (Additional File 1 in the Supplemental Material). Briefly, RNA was selected for lncRNA analysis using a Ribo-Zero rRNA Removal Kit (Illumina, San Diego, CA). Samples were run on the HiSeq 2500 (Illumina) in 51 bp paired-end reads achieving ~ 30 million clustered reads per sample. Generated Fastq files were processed through HISAT2 (40, 41) under standard parameters, with the exception of quality scoring control specified under “-phred33.”

The reference genome used in generating BAM files included fasta formatted DNA from Ensemble release 95 for human (GRCh38) and mouse (GRCm38), including novel transcripts that at the time of print are not annotated. Comparative alignment was performed in the R environment using the *AdaLiftOver* package (42). Differential gene expression was performed in the R (v3.5.3) environment, through *DESeq2* (43). Calculation of fragments per kilobase per million fragments mapped (FPKM) and transcripts per kilobase million (TPM) reads was performed in the R environment using the *countToFPKM* package (44) and *samtools* (45). Visualization was performed through the packages *ggplot2* (46), *limma* (47), *vidger* (48), and *EnhancedVolcano* (49). Additional programs, such as *Seqmonk* (v1.45.4) and *Cytoscape* (v3.7.2) (50) were implemented to display data.

High-Throughput Sequencing Cross-Linking Immunoprecipitation

CLIP RNA (as described above) from human atrial appendages (anti-PNPase) and mouse whole heart (anti-PNPase, anti-Ago2) was sequenced through the West Virginia University Genomics Core Facility. The complete transcript abundance profiles, along with the raw count values for each sample are included (Additional File 1 in the Supplemental Material). RNA was selected for using the NEXTFLEX Small RNA-Seq Kit v3 (Bioo Scientific, Austin, TX). Samples were run on the MiSeq in 36-bp single-end reads achieving ~1 million clustered reads per sample. Adapter trimming through *cutadapt* (51) (5'-TGAATTCTCGGGTGCCAAGG -3') allowed for isolation of genomic reads. Generated Fastq files were processed through *Bowtie* (52) using the parameters, “-a -best -chunkmbs 250,” which reported all valid alignments ranked from best to worst. This permitted identification of cross-linking-induced biases while permitting accurate gene alignment. The reference genome used in generating BAM files included fasta formatted DNA from Ensemble release 103 for human (GRCh38) and mouse (GRCm38). Calculation of FPKM and TPM reads was performed as described above. Additional programs, such as *matplotlib's* *plotly* package (53), *wiggleplotR* (54), and *rtracklayer* (55), were implemented to display CLIP data.

Machine Learning CLIP Sequencing lncRNA Input

To determine if lncRNAs that were found in the mitochondrion and associated with PNPase had sequence or secondary structure homology, *RNAfold* through the *ViennaRNA* Package was employed (56). *RNAfold* was used in conjunction with *LncFinder* (57) and *seqinR* (58) to evaluate folding parameters of both the full-length lncRNA, as well as isolated regions (15–22 nucleotides) shown to bind to PNPase in CLIP analyses. Isolated regions were identified for 60 human and 60 mouse lncRNA labeled “positive” (found in CLIP sequencing) and 120 randomly generated sequences of equal length labeled “negative” (59). The isolated regions from genes identified within the PNPase CLIP were flanked by surrounding nucleotides within the genome to achieve 60 nucleotide sequences implemented in the subsequent *RNAfold* and machine learning applications. The length of 60 nucleotides for each target was selected to include enough information about the surrounding region to generate an appropriate secondary structure but also

succinct enough to be used in computational models for designing RNA secondary structures (60). The randomly generated sequences also consisted of 60 nucleotides. The sequences directly identified in the CLIP sequencing, as well as flanking regions, are provided for each positive control gene (Additional File 1 in the Supplemental Material). For comparisons within the positive group, lncRNAs identified in the PNPase CLIP were categorized by total read count into increased interactions (>1,500 human or >750 mouse) and decreased interactions (<1,500 human or <750 mouse).

Machine Learning for RNA Sequences

Using 10-fold cross validation, classification of “positive” and “negative” 60 nucleotide isolated regions was assessed through classification and regression trees (CART) in *randomForest* (61) and *rfUtilities* (62). Machine learning algorithms were performed on 135 features, which included primary sequence and secondary structure information (Additional File 1 in the Supplemental Material). Briefly, *rf.crossValidation* was used with a *randomForest* object to provide 10-fold cross validation, with a 70% training and 30% holdout/testing set. As a confirmatory measure, support vector machines (SVM), using the *svm_cv* function of *LncFinder*, were implemented to determine 10-fold cross validation of full length and isolated regions. Distance referred to in this context is defined as linear length calculated for two data points on the model's most optimal decision surface, one data point being a sequence and the other being a feature defined by *LncFinder*. Power spectrum signal refers to a vector of numerical values representing the single electron-ion interaction potential (EIIP) of each nucleotide, which is most frequently used for appropriate DNA-to-signal mapping (88).

Expression and Purification of PNPase Protein

PNPase was transformed and expressed in pET11a plasmids and expressed in Rosetta (DE3) cells. Expression was achieved in 3-liter lysogeny broth autoinduction media (63) containing 100 ug/mL carbenicillin and 20 ug/mL chloramphenicol grown at 25°C for 24 h. Cells were harvested and pelleted at 5,000 g, resuspended in a total volume of 22 mL lysis buffer (50 mM Tris-HCl pH 7.5, 1 M NaCl, and 10 mM imidazole), and treated with PMSF protease inhibitor at 1/1,000 dilution. Cells were lysed via sonication with cleared lysate being applied over a Ni-NTA column, washed with additional lysis buffer, and eluted in 20 mM HEPES (pH 7.5), 300 mM KCl, and 300 mM Imidazole. Purified protein was then centrifuged in a 100-kDa cutoff filter to remove purification tags, PNPase monomers, and imidazole before being stored in 20 mM HEPES (pH 7.5) with 300 mM KCl until use. The Native Shift Assay was run using Predicted RNA oligos generated from the ML and control RNA oligos that were 5' labeled with fluorescein (IDT) and combined with pure PNPase samples at specified molar ratios with a constant concentration (250 nM) of RNA probe. The reaction was carried out in 20 mM Tris-HCl (pH 7.4), 50 mM KCl, 0.1 mM EDTA, 1 mM DTT, 10% glycerol, and 2 mM MgCl₂ and incubated at 37°C for 30 min before separating on a 5% polyacrylamide gel in native conditions at 4°C. The resulting gels were then imaged using a GE Typhoon for Cy2 fluorescence (21).

Protein Modeling and Prediction

Predicted models of full-length, KH-deleted, and S1-deleted PNPase were generated using Colabfold through Cosmic² (64) and superpositioned using ProteinDataBank's Mol*3D Viewer. Electrostatic potential surface modeling was performed using human PNPase (Protein Data Bank ID: 3U1K) imported into Pymol (65) and evaluated using APBS Tools plugin.

Statistics

A two-sided Student's *t* test or one-way analysis of variance (ANOVA) was used to determine statistical differences, where appropriate. Multiple groups were assessed through Tukey's multiple comparisons test following the ANOVA. Differences between groups were considered statistically significant if $P \leq 0.05$. Data are presented as the means $\pm$ SE, when appropriate. For all sequencing data, raw counts >1 were considered in statistical analyses. The false-discovery rate was set to 0.05, and all significance was determined through a *P* adjusted value <0.05 using Wald testing.

RESULTS

Mitochondrial Isolation

To evaluate the subcellular localization of ncRNAs, purity of cellular fractionation was assessed. Tissue homogenization,

differential centrifugation, and sucrose gradient separation allowed for the enrichment of pure populations of cardiac mitochondria from both mice and humans (Fig. 1A). Mitochondrial contamination was determined through the presence of cytoplasmic ribosomal RNA (rRNA) 18S and 28S within mitochondrial RNA fractions (Fig. 1B). Low RNA integrity numbers (RIN) for cytoplasmic rRNAs in mitochondrial fractions, while typically used in the context of RNA degradation, verify the purity of mitochondrial RNA isolation through the absence of cytoplasmic contaminants (32, 66) (Fig. 1C). Further, the presence of 12S and 16S ribosomal RNA peaks (Supplemental Fig. S2, A and B), without the inclusion of 18S or 28S (Supplemental Fig. S2, C and D), helped confirm the identity of the DNA as mitochondrial on gel electrophoresis. qPCR in mouse and human mitochondria (Supplemental Fig. S3, A and B) also revealed that 12S and 16S rRNA transcript levels were significantly higher than nuclear rRNA contaminants, indicated by U6 snRNA. Immunoblotting further confirmed purification of mitochondrial fractions, indicated by an absence of cytoplasmic proteins (GAPDH), with inclusion of mitochondrial proteins (COX IV) (Fig. 1D). The presence of COX IV in cytoplasmic fractions is likely due to nuclear-encoded mitochondrial proteins, such as COX IV, being transported to the mitochondrion following translation in the cytoplasm (67).

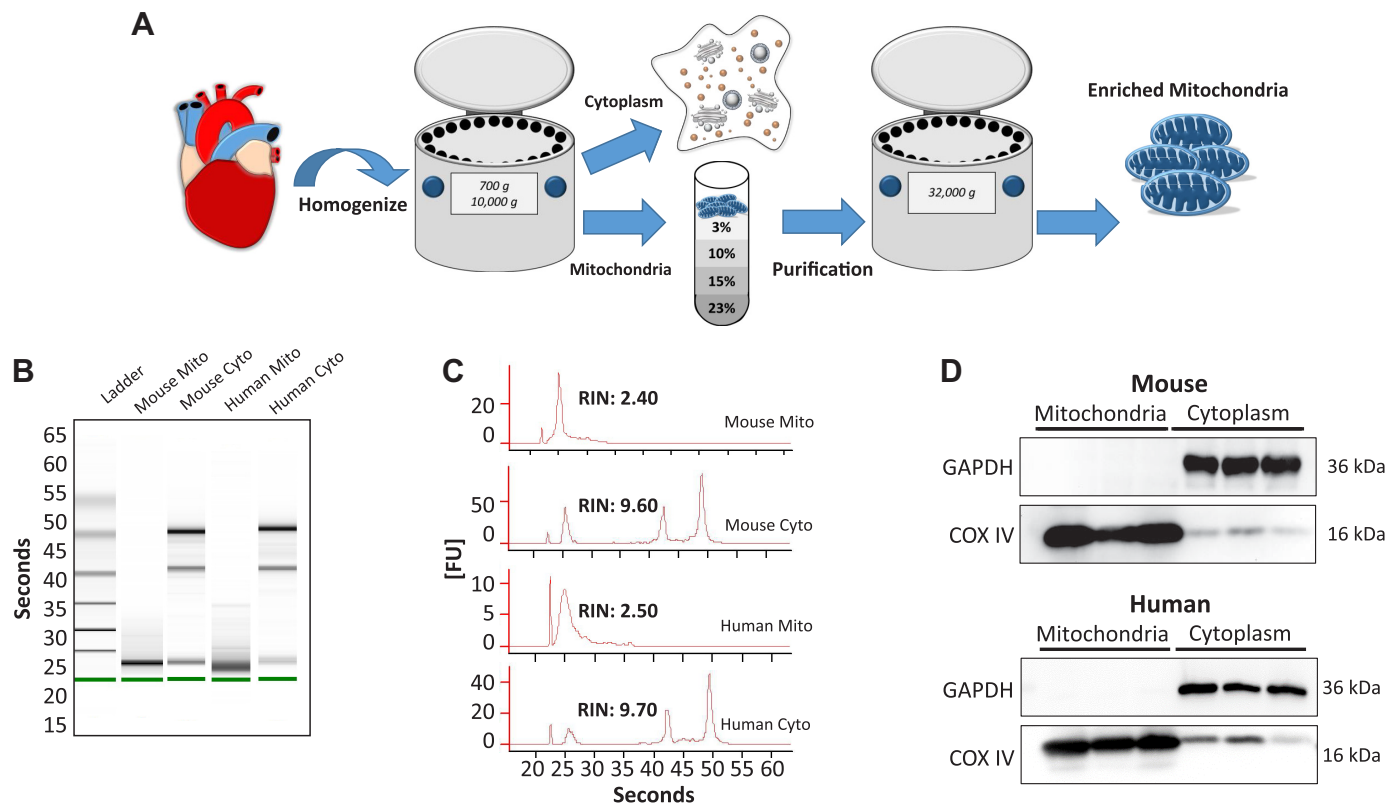


Figure 1. Enrichment of mitochondria from cardiac tissue. **A:** from human atrial appendages and whole mouse hearts, samples are homogenized and differentially centrifuged to produce subcellular fractionations; mitochondria are further sucrose-gradient purified to remove cytoplasmic contamination. **B:** electrophoresis gel showing purity of cytoplasmic (Cyto) and mitochondrial (Mito) RNA assessed through the Bioanalyzer (Agilent Technologies). **C:** electropherogram trace for calculating RNA integrity number (RIN) indicating limited or no presence of cytoplasmic ribosomal RNA. **D:** immunoblotting of isolated mitochondria for cytoplasmic [glyceraldehyde phosphate dehydrogenase (GAPDH)] and mitochondrial [cytochrome c oxidase subunit (COX IV)] proteins in mouse (top) and human (bottom). FU, fluorescence units.

Nuclear-Encoded lncRNA in Mitochondria

Using a lncRNA sequencing approach on sucrose-purified mitochondria, we identified a variety of lncRNAs that were compartmentalized in mitochondria from mouse (Fig. 2A, top) and human (Fig. 2A, bottom) hearts relative to the cytoplasm. To confirm minimal nuclear contamination, which is crucial for evaluating lncRNAs, GAPDH mRNA in mouse and human lncRNA sequencing of the cytoplasmic fraction was identified to be immensely more highly expressed relative to small nuclear RNAs such as U6 and 7SK (Supplemental Fig. S3C). In total, 112 (mouse) and 1,548 (human) lncRNAs with 0.1 or more mean TPM counts were identified in the mitochondrion, with over 30% of mouse lncRNAs identified to have annotated human orthologues. Of the lncRNAs identified in the mitochondrion, Malat1 was expressed in the greatest abundance for both mouse (Fig. 2B, top) and human (Fig. 2B, bottom) mitochondrial sequencing results, presented as the top 6 most abundant genes. Additionally, other highly expressed nuclear-localized lncRNAs such as Haglr, Hotair, Hottip, Fendrr, Xact, Disc2, and Anril were not identified in mitochondrial sequencing, confirming the specificity of these results (68, 69). FISH revealed that Malat1, while predominately present in the nucleus, can be identified in the cytoplasm (Fig. 2C; Supplemental Fig. S4A). Further, coupling FISH with MitoTracker staining suggested that a colocalization of Malat1 with mitochondria also occurred (Supplemental Fig. S4B), similar to recent findings in hepatocellular carcinoma cells (70). This indicates that Malat1, although classically considered to be restricted to the nucleus, can be detected interacting with the mitochondrion. Additional evaluations of Malat1 localization support its potential to translocate to the mitochondrion, although these evaluations have never been performed in cardiac tissue or with a sequencing approach optimized for the mitochondrion (71, 72).

PNPase CLIP Sequencing

The complete pathway facilitating ncRNA transport into the mitochondrial matrix is not fully elucidated, though work from our laboratory (24) as well as others (19, 25) has implicated PNPase as a mediator of miRNA and tRNA import (Fig. 3A). Using purified mitochondria from mouse and human cardiac tissue, we performed CLIP analysis on PNPase. CLIP revealed that a significant portion of ncRNA PNPase binding targets were lncRNAs in both mouse (~60%) and human (~69%) mitochondria (Fig. 3B). While additional ncRNAs such as miRNAs and miscRNAs also appear as PNPase binding targets, lncRNAs are by far the most prominent target ncRNA type. Protein-coding RNAs are also present in CLIP sequencing results with the greatest quantity of gene species among all the RNAs (Supplemental Fig. S5A). However, at least in human results, lncRNAs are more prominent by a margin of 6% when scaled by percentage to total contribution to TPM, (Supplemental Fig. S5B). This does not hold true for mouse PNPase CLIP sequencing, which may reflect the lack of characterization of mouse ncRNA. Wiggle plot analysis of CLIP sequencing shows that lncRNAs such as Malat1, Neat1, and H19 have defined peak densities of reads, indicating that PNPase targets specific regions (Supplemental Fig. S6). The presence of additional smaller peaks may indicate the involvement of tertiary structures.

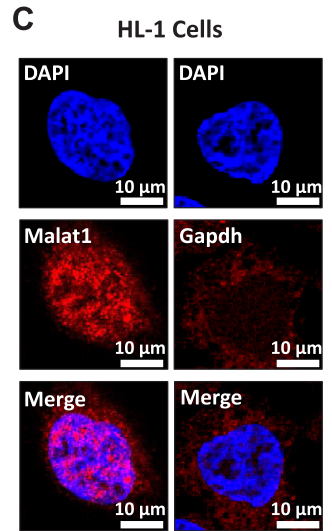
To determine that the noncoding RNA binding profile of PNPase was not fortuitous, CLIP analysis was also performed on noncoding RNA-binding protein Ago2, which has been reported to localize to mitochondria (32, 73) and also has the capability to bind to lncRNAs (Supplemental Fig. S7A) (74, 75). We hypothesized that if the binding of lncRNAs to PNPase was merely fortuitous, the profile of lncRNAs binding to Ago2 would be similar. The comparison between the two proteins in mouse cardiac mitochondria shows that a low proportion of PNPase-binding lncRNA targets have the capacity to bind to Ago2 as well, indicating that there may be characteristics of these lncRNAs that provide specificity for PNPase that Ago2 lack (Supplemental Fig. S7B). These data also indicate that not all lncRNAs imported into mitochondria engage in Ago2-related RNA-induced silencing complex (RISC) activity.

Stratification of lncRNA Fragments by Primary and Secondary Features

CLIP is a technique that takes advantage of UV cross linking to examine short RNA regions bound to a protein of interest, leading to both the identification of the bound RNA and the specific region(s) involved in binding (76) (Fig. 4A). Examining the most prominent peaks of gene coverage in the CLIP PNPase sequencing data, we wanted to determine if a mitochondrial RNA targeting sequence or secondary structure exists that would predict the binding of lncRNA to PNPase and trigger import activity. For secondary structure predictions relying on base pairing energies to be accurate, the more of the endogenous sequence included the better. However, sequences like Malat1, which are multiple kilobases in length, are difficult to manage and appropriately analyze.

To define a better understanding of the primary sequence and secondary structure of the RNA bound to PNPase, we needed to increase the number of nucleotides in the evaluated RNA fragments. Concomitantly, we wanted to limit the amount of bias introduced into the data by including additional sequence information from surrounding regions of the lncRNA found to be bound to PNPase. We used the receiver operating characteristics (ROC) area under the curve (AUC) values derived from the LncFinder pipeline to choose how many additional nucleotides would be optimal to maximize folding potential and limit unnecessary extension of the reads. The largest increase in the slope of AUC values is between 50 and 60 nucleotides (Supplemental Fig. S9A). After 60 nucleotides, the margin of improvement of the model significantly decreases. lncRNA reads (ranging from 15 to 24 bases) from CLIP sequencing were identified and endogenous flanking nucleotides were included at the 5'- and 3'-ends to achieve a total sequence length of 60 nucleotides (Fig. 4A). These sequences were analyzed using RNAfold from the ViennaRNA Package (56).

Machine learning is a powerful tool for identifying commonalities between large datasets and forming predictions from those commonalities. Established algorithms for evaluating RNA-protein interactions, while externally validated, are often limited to short motifs that cannot provide clarity on secondary structure and additionally lack ncRNA-specific feature descriptions (57, 77,78). As previous literature has demonstrated the advantages of using machine learning in



With each k-fold validation, accuracy in the subset validation data (Fig. 4B, left) and across the entire model (Fig. 4B, right) increased. Out-of-bag (OoB) error rates remained consistent in the subset validation data (Fig. 4C, left) and across the entire model (Fig. 4C, right). Stratification of lncRNA fragments from random sequences was robustly shown using CART (AUC: 0.89; Fig. 4D) and SVM (accuracy: 82%; Supplemental Table S2). The top features influencing classification of lncRNAs from random nucleotide sequences are

Quantification of lncRNA Properties Associated with Interactions with PNPase

After identifying the unique properties of lncRNAs bound to PNPase from random sequences, we wanted to assess if

After identifying the unique properties of lncRNAs bound to PNPase from random sequences, we wanted to assess if

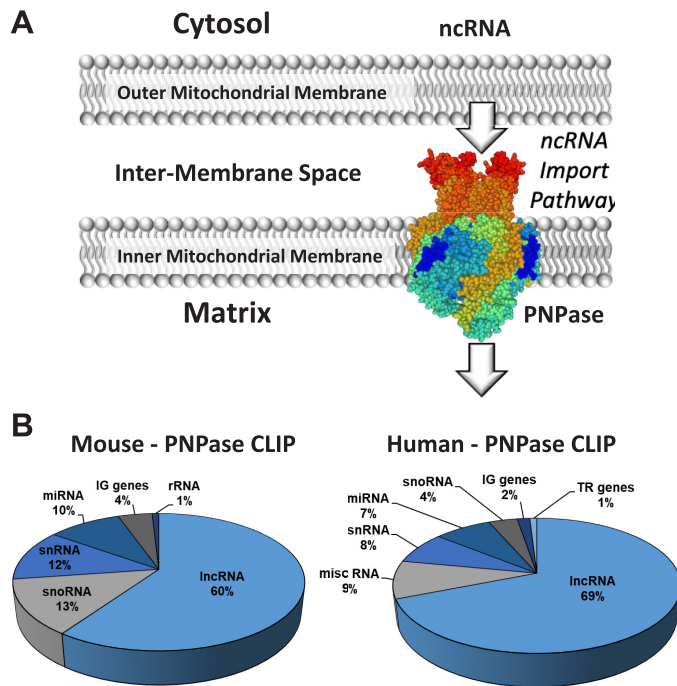


Figure 3. Expression profiles of noncoding RNA (ncRNA) bound to polynucleotide phosphorylase (PNPase). Mouse whole heart ($n = 4$) and human atrial tissue ($n = 4$) cross-linking immunoprecipitation (CLIP) sequencing of PNPase. **A:** illustration of noncoding RNA (ncRNA) passage through the mitochondrial membrane and interaction with PNPase. **B:** pie charts comparing percentages of ncRNA species identified in PNPase CLIP for mouse (*left*) and human (*right*). IG gene, immunoglobulin gene; IncRNA, long noncoding RNA; miRNA, microRNA; miscRNA, miscellaneous non-coding RNA; rRNA, ribosomal RNA; snoRNA, small nucleolar RNA; snRNA, small nuclear RNA; TR gene, T-cell receptor gene.

the identified lncRNAs could be separated into biologically significant groupings. The top 60 lncRNAs identified to bind to PNPase were arbitrarily selected based on total read counts for both human and mouse to eliminate noise from transient interactions with PNPase. Comparison between the features of these 2 groups of 60 shows no significant difference, indicating that the predominant features are likely to be conserved. Additionally, the top 60 human lncRNAs were evaluated using AdaLiftOver to verify that while only 25% of them had previously annotated mouse orthologues, the remainder possessed orthologous regulatory elements (Supplemental Fig. S8). The resulting 120 lncRNAs were split into the top and bottom 50%, which were designated as high versus low binding potential for PNPase and input into machine learning to identify differences that might increase affinity relative to random sequences. This lncRNA input was characterized by mitochondrial presence in lncRNA sequencing to identify if they were imported or just binding to PNPase. Approximately one-third of the machine learning input lncRNAs were designated as high interaction potential and imported and represented ~50% of all the machine learning input lncRNAs that were imported. Another third of the input lncRNAs had low interaction potential but were still imported, and the remaining third had predominantly low interaction potential and were not imported (Fig. 5A; Additional File 1 in the Supplemental Material). This indicates that lncRNAs with increased interaction potential were

more likely to appear in lncRNA mitochondrial sequencing as being imported and that this selection was an optimal input for machine learning to differentiate between high and low interaction potential in relation to import. Features influencing classification of increased interaction lncRNAs from decreased interaction lncRNAs identified by machine learning are provided (Table 2). A representative lncRNA from each group is provided (Fig. 5B), highlighting the major differences in primary sequence and secondary structure, including GC content and various k-mers, as well as distribution of unpaired nucleotides. Specifically, the lncFinder algorithm utilized was designed to identify nucleotide pairing distributions either unique to lncRNAs or as a ratio between lncRNAs and protein-coding transcripts (57). On a superficial level, analysis of these groups using RNAfold indicates that stem-loops with at least seven unpaired nucleotides in the loop are significantly more prevalent in increased interaction lncRNAs. Additionally, minimum free energy assessed to not be significantly different between high or low interaction potential lncRNAs and random sequences indicates that this loop is unstable and unlikely to resist linearization, which would be optimal for import as it has been suggested that structured RNAs are linearized to be threaded through the central channel of PNPase (22, 80). Based on these results, lncRNAs possessing primary sequence features, including reduced GC content and increased prevalence of k-mers like CA, result in a secondary structure with an increased relative unbound nucleotide distribution represented by a ≥ 7 -mer loop that may enhance binding to PNPase.

Additional performance metrics are included for the total lncRNA, lncRNA with increased interaction, and lncRNA with decreased interaction (Supplemental Table S2), including AUCs for increased interaction lncRNAs against random sequences (0.89; Supplemental Fig. S9B), decreased interaction lncRNAs against random sequences (0.91; Supplemental Fig. S9C), and increased interaction lncRNAs against decreased interaction lncRNAs (0.60; Supplemental Fig. S9D). Due to our restriction of the folded RNA fragments to 60 nucleotides, we also showcase the preservation of secondary structure in Supplemental Fig. S10. As an example, full-length MALAT1 was folded using RNAfold, and the structure of the region containing the PNPase CLIP sequence was extracted. This extracted region (Supplemental Fig. S10) retains the stem-loop motif(s) that were found in our 60 nucleotide regions used throughout our analysis and matches the criteria for improved binding to PNPase. The lncRNAs bound to PNPase have different properties involving their primary sequence and secondary structure, which can be utilized to develop a criterion necessary to engineer PNPase binding sequences.

Importance of PNPase RNA Binding Domains

At the interface of the mitochondrion, alterations to the structural integrity of PNPase may influence lncRNA binding and eventual import. PNPase may be involved in shuttling Malat1, and other lncRNAs, through its KH and S1 RNA binding domains (Fig. 6A). Using an HL-1 immortalized mouse cardiomyocyte cell line, we transfected cells with vectors overexpressing full-length PNPase, PNPase with the KH domain knocked out, or PNPase with the S1 domain knocked out (Fig. 6B). Overexpression of constructs was demonstrated through

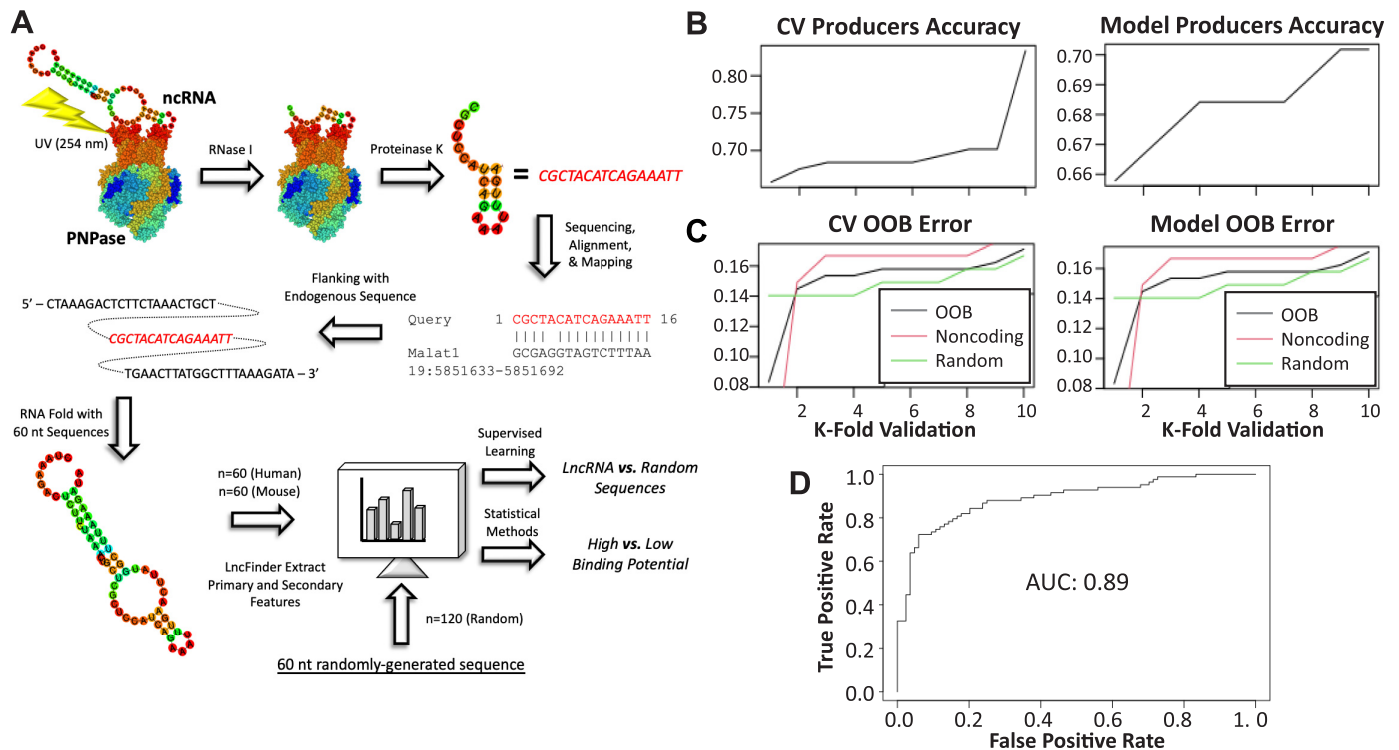


Figure 4. Polynucleotide phosphorylase (PNPase) cross-linking immunoprecipitation (CLIP) analysis pipeline and machine learning implementation. **A:** illustration highlighting critical steps in the CLIP procedure and analysis of reads bound to PNPase. **B:** producer's accuracy for validation subsets (left) and model (right) in the prediction of lncRNAs against random nucleotide sequences. **C:** out of bag (OOB) errors for validation subsets (left) and model (right) in the prediction of long noncoding RNAs (lncRNAs) against random nucleotide sequences. **D:** receiver operating characteristics showing area under curve (AUC) for prediction of lncRNAs against random nucleotide sequences. CV, cross validation; ncRNA, noncoding RNA.

immunoblotting, with a shift in size between the full length, KH domain knockout, and S1 domain knockout (Fig. 6C). Using AlphaFold predictions, KH and S1 domain knockouts were confirmed to not grossly misfold and create confounding variables (Supplemental Fig. S11A), which is supported by other assessments of KH and S1 deletion (9, 16, 20–23). Mitochondrial localization of Malat1 and Neat1 was confirmed through qPCR in HL-1 cells with the GFP, FL, KH, and S1 overexpression vectors (Supplemental Fig. S12A). Additionally, multiple isoforms that align with diverse areas of Malat1 were shown to have altered presence in the mitochondrion (Supplemental Fig. S12B). Knockdown of PNPase expression was shown to reduce Malat1 import into the mitochondrion, with Neat1 showing similar but less significant trends (Supplemental Fig. S12C).

Using CLIP for PNPase in each of the cell variants, RNA was purified from each of the overexpression groups through SDS-PAGE gel electrophoresis, with a representative gel illustrated (Fig. 6D). Examination of the RNA bound to PNPase revealed that Malat1 levels were significantly downregulated when the S1 domain was removed from PNPase (Fig. 6E), suggesting its role in RNA-protein interactions. While not significant in ANOVA analyses, Malat1 levels were also visibly diminished with KH domain removal ($P \leq 0.09$, Student's t test), which may indicate a role in regulating lncRNA import through the KH domain as well. These results suggest that PNPase has a binding affinity for Malat1, most likely through the S1 and/or KH RNA binding domains. They also verify the importance of secondary structure recognition for lncRNA import.

Table 1. The top 10 most significant features between lncRNA and random nucleotide sequences

Feature	P Value	Long Noncoding RNA	Random Nucleotide Sequence
CG	5.17E-24	0.0005	0.0015
Log distance acgu-ACGU lncRNA	8.43E-13	−6.3753	−6.1339
CGA	1.30E-11	0.0004	0.0020
Log distance acgu-ACGU protein-coding transcripts	6.22E-11	−6.3537	−6.1407
ACG	9.88E-10	0.0006	0.0019
Distance ratio acgu-ACGU	6.94E-08	1.0034	0.9989
CCG	3.22E-07	0.0007	0.0022
Distance ratio acguD	2.24E-06	0.9996	0.9965
CGC	2.80E-06	0.0007	0.0020
CA	3.04E-06	0.0022	0.0017

Data are reported as the mean. The values for k-mers are listed as frequencies and not counts. A two-sided Student's t test was implemented for determining statistical significance, with a Bonferroni-adjusted alpha < 0.000373 . acgu-ACGU, proportion of uppercase letters annotating paired nucleotides; acguD, proportion of dots annotating unpaired nucleotides; lncRNA, long noncoding RNA.

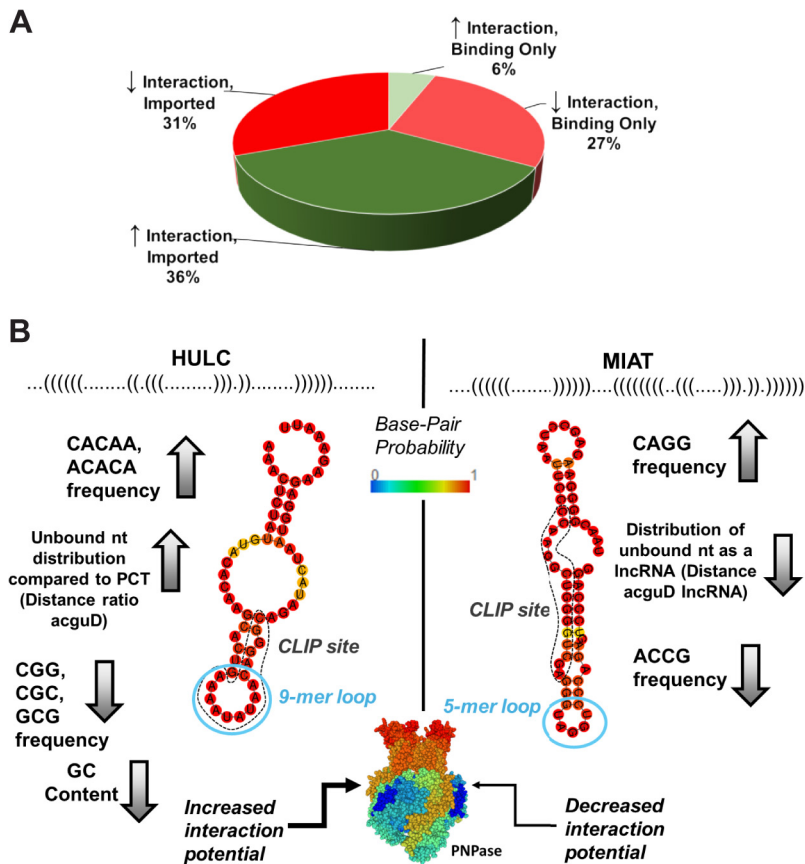


Figure 5. Increased and decreased interacting long noncoding RNA (lncRNA) characteristics. **A:** pie chart illustrating division of increased and decreased interaction potential lncRNA input into machine learning as well as presence (imported) or absence (binding only) in lncRNA mitochondrial sequencing. **B:** representative 60 nucleotide human lncRNA fragments from the increased interacting [hepatocellular carcinoma up-regulated long noncoding RNAH (HULC); left] and decreased interacting [myocardial infarction associate transcript (MIAT); right] groups. The color of each nucleotide signifies the probability of binding, or remaining unpaired, in the diagram. acguD, proportion of dots annotating unpaired nucleotides; CLIP, cross-linking immunoprecipitation; PCT, protein coding transcript; PNPase, polynucleotide phosphorylase.

Design and Validation of PNPase Binding Potential for RNA

With information from the mitochondrial lncRNA and PNPase CLIP sequencing and machine learning applications,

we designed a 60-nucleotide RNA fragment that should have the propensity to bind PNPase (Fig. 7A). This predicted sequence contained a unique sequence structure (e.g., GC content and CA frequency) as well as characteristics regarding its secondary structure (e.g., a high distribution of

Table 2. Significant features for increased and decreased interaction lncRNA fragments subsets compared to random lncRNA fragments

Feature	↑ Interaction P Value	↑ Interaction Mean	↓ Interaction P Value	↓ Interaction Mean	Random Mean
CG	4.99756E-21	0.000347224	3.36924E-14	0.000620281	0.001530936
Distance ratio acgu-ACGU	1.6297E-08	1.00490202	0.000288139	1.002302427	0.998883974
ACG	8.07509E-08	0.000336071	1.51827E-05	0.000730169	0.001860001
Log distance acgu-ACGU lncRNA	9.43243E-08	-6.351869094	3.19177E-11	-6.391983821	-6.133939417
CGA	1.68631E-07	0.000239047	7.71784E-07	0.000522188	0.002002316
CCG	2.6595E-06	0.00048908	0.000339854	0.000893047	0.002179283
Log distance acgu-ACGU protein-coding transcripts	4.47231E-06	-6.32078765	3.18596E-10	-6.37713348	-6.140681014
CA	1.96269E-09	0.002491291	NS	0.001960414	0.00165374
Distance ratio acguD	1.17543E-05	1.000437813	NS	0.999026841	0.996540097
CGG	1.91646E-05	0.000615433	NS	0.001089223	0.002040002
GCG	2.51624E-05	0.000493545	NS	0.000872332	0.002020788
CACA	5.5346E-05	0.007606291	NS	0.004237178	0.002149228
GCG	5.93167E-05	0.000570784	NS	0.000966878	0.002025162
CACAA	8.92206E-05	0.017195084	NS	0.003215303	0.002338989
ACACA	0.000267668	0.010455789	NS	0.00828333	0.001248782
ACA	0.000323641	0.003894203	NS	0.002617873	0.002260456
GC content	0.00036204	0.597227355	NS	0.669249925	0.650958698
CAGG	NS	0.003589873	6.74433E-05	0.006717755	0.002310656
Log distance acguD lncRNA	NS	-8.882641233	0.000296033	-8.959120344	-8.686657734
ACCG	NS	0.000600641	0.000358731	0.000272817	0.003254588

Data are reported as the means. The values for k-mers are listed as frequencies and not counts. A two-sided Student's *t*-test was implemented for determining statistical significance between either the high or low interaction potential subset against random lncRNAs, with a Bonferroni-adjusted alpha < 0.000373. NS, not significant, i.e. a *P* value above the determined Bonferroni-adjusted alpha. acgu-ACGU, proportion of upper-case letters annotating paired nucleotides; acguD, proportion of dots annotating unpaired nucleotides; lncRNA, long noncoding RNA.

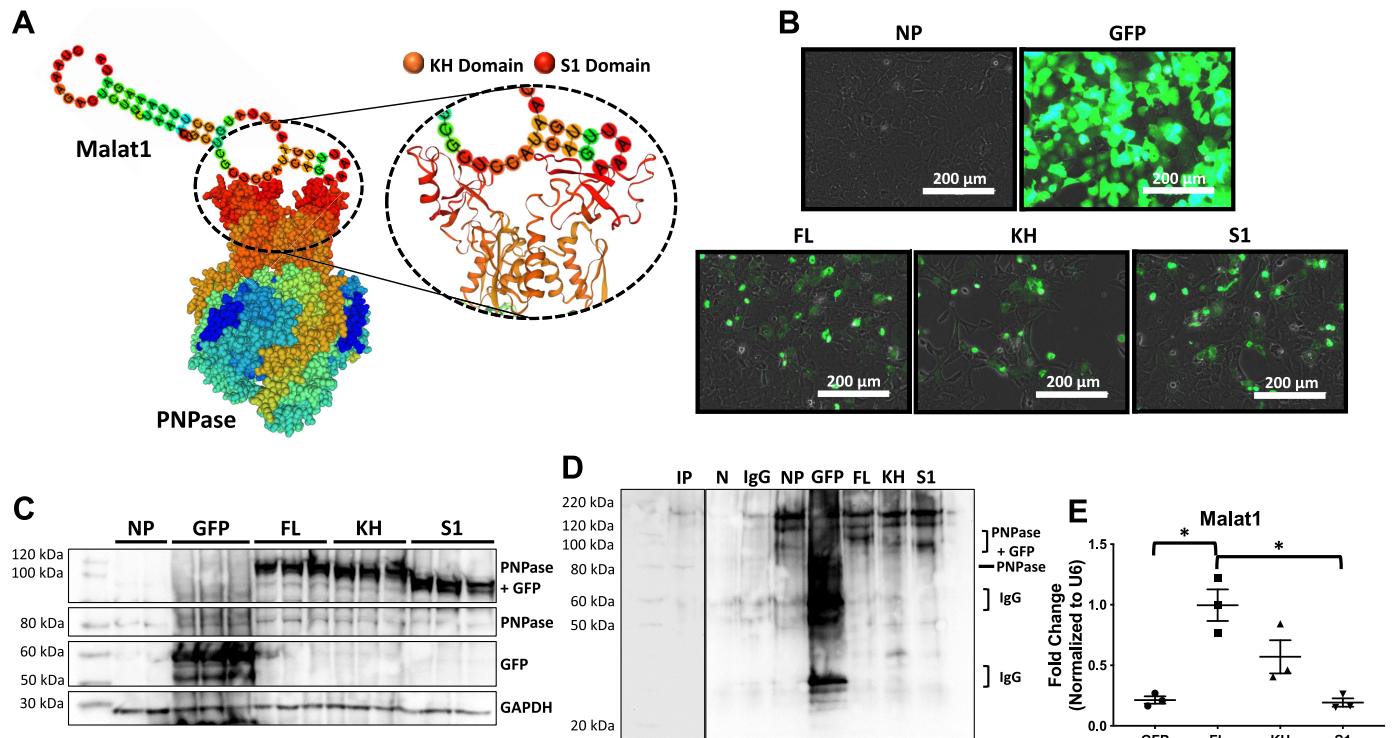


Figure 6. Polynucleotide phosphorylase (PNPase) RNA binding domain interaction with metastasis-associated lung adenocarcinoma transcript 1 (Malat1). **A:** Illustration of plausible KH and S1 interaction with long noncoding RNAs (lncRNAs), not experimentally validated. **B:** representative images for HL-1 cells in each group ($n = 10$ per group) showing fluorescence induced by pcDNA3.1 + N-eGFP overexpression. **C:** immunoblotting depicting the overexpression in HL-1 cells, as well as shift in size, of the constructs. **D:** representative image of SDS-PAGE following cross-linking immunoprecipitation (CLIP) of HL-1 cells to retrieve RNA specifically bound to PNPase. **E:** real-time PCR on isolated RNA to assess binding affinity of lncRNA Malat1 ($n = 3$ each group) to PNPase variants. Differences were considered statistically different at $*P \leq 0.05$. A one-way ANOVA with Tukey's multiple comparisons was implemented for determining significance of quantitative PCR. All data are presented as the means $\pm$ SE. GFP, pcDNA3.1 + N-eGFP backbone only; FL, full-length PNPase open reading frame in pcDNA3.1 + N-eGFP; IP, immunoprecipitation; KH, exon 23 removed from full-length PNPase in pcDNA3.1 + N-eGFP; N, no antibody; NP, no plasmid control; S1, COOH terminus removed from full-length PNPase in pcDNA3.1 + N-eGFP.

unbound nucleotides). We then tested our hypothesis through a fluorescence native gel shift assay to determine the percentage of binding of PNPase with our RNA constructs (Fig. 7B). A Poly-U RNA sequence of equal length was used as a negative control to represent the binding potential of a completely unstructured RNA, which revealed minimal binding to PNPase with increasing concentrations (11.3% RNA shifted; Fig. 7C). The minimal binding observed may have been due to the presence of bacterial monomeric ribonuclease PNPase endogenously present in the native gel shift assay, which is why it was crucial to utilize an unstructured RNA to set an accurate baseline as opposed to a scrambled RNA. Additionally, the 60-nucleotide sequence of Malat1 identified in the PNPase CLIP sequencing was utilized as an approximation of a biologically relevant positive control for binding to PNPase, revealing increased binding affinity to PNPase (51.4% RNA shifted), compared to control. When comparing the Predicted (74.3% RNA shifted) and Poly-U RNA sequences, a strong preference for the Predicted construct was observed (558% increased binding affinity compared to Poly-U; Fig. 7C). A PNPase CLIP was run with the same Predicted and Poly-U RNAs incubated with mouse mitochondrial lysate to confirm that the interaction potentials identified with purified PNPase were consistent with PNPase in its native environment. Cross-linked precipitate compared to unbound RNA remaining in solution indicated that Predicted RNA had a high affinity for ex vivo

PNPase while Poly-U RNA was entirely absent and showed no indication of binding to PNPase (Fig. 7D).

DISCUSSION

In the current article, we observed the presence of numerous nuclear-encoded lncRNAs localized within mitochondria, a large percentage of which were previously considered to be restricted to the nucleus (81, 82). Our goal was to understand how these lncRNAs can enter the mitochondrion and if a defined sequence or structural motif is necessary to facilitate binding. Only recently has the concept of nuclear-encoded lncRNA localization into the mitochondrion been explored (70, 83), challenging our understanding of the import mechanisms required to facilitate this process and the extent of lncRNAs being sequestered in extranuclear compartments. However, this study is the first to directly compare nuclear-encoded lncRNA sequence identity with mitochondrial binding through PNPase.

When not directly accessible to the cytoplasm or mitochondrial matrix, studies have reported the ability of PNPase to interact with and transport RNAs into the mitochondrion (19, 84). The KH and S1 RNA binding domains of PNPase are suggested as the mechanism for channeling ncRNA through the mitochondrial intermembrane space into the matrix (19).

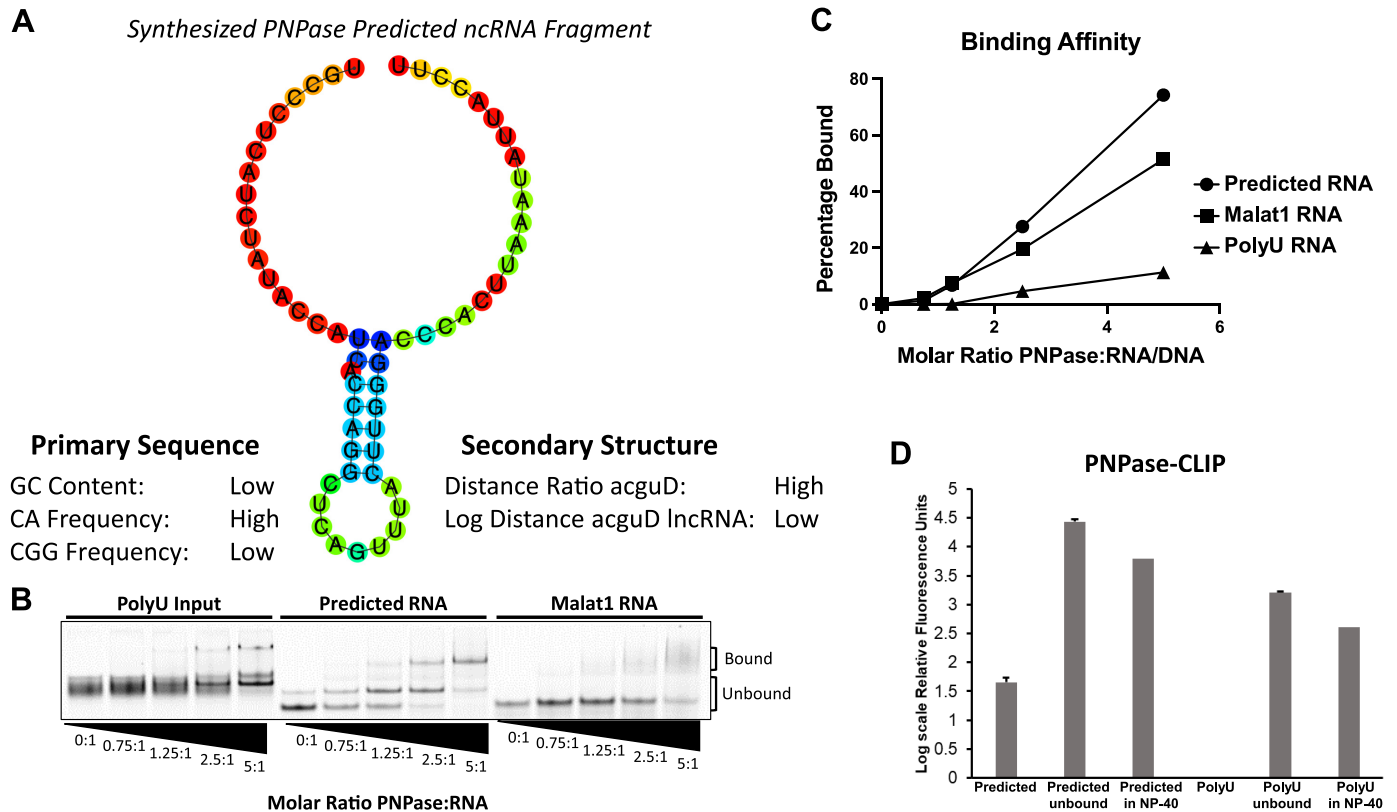


Figure 7. Designed long noncoding RNA fragment with specific criteria for increased binding affinity. **A:** using RNAfold, a 60-nucleotide predicted RNA fragment was engineered based on the criteria of low GC content (43%) and electron-ion interaction pseudopotential (EIIP; 0.068), increased minimum free energy (more positive, -4.8), and k-mer specific base combinations (CA and AGG). The resultant structure is provided and serves as an illustration of motifs that can be designed to aid in the incorporation of noncoding RNA sequences into the mitochondrion. **B:** fluorescence native gel shift assay identifying differences in polynucleotide phosphorylase (PNPase) protein binding between a 60-nucleotide sequence of Poly-U RNA, the Predicted 60-nucleotide sequence construct, and a 60-nucleotide portion of metastasis-associated lung adenocarcinoma transcript 1 (Malat1) RNA selected based on the PNPase cross-linking immunoprecipitation (CLIP) added at increasing ratios of protein relative to probe. **C:** band intensity of bound bands compared to unbound and expressed as a percentage of binding affinity plotted against molar ratio of protein to probe. **D:** PNPase CLIP fluorescence assay verifying ex vivo binding between PNPase and Predicted RNA probe construct (Predicted) relative to Poly-U RNA probe as well as probes in CLIP buffer ($n = 3$ for all groups except for positive probe controls). All data are presented as the means $\pm$ SE. acguD, proportion of dots annotating unpaired nucleotides.

Importantly, it has been suggested that a stem-loop sequence on the RNA was necessary for mitochondrial import (19). These findings support a mechanism whereby PNPase recognition requires RNA secondary structure as opposed to sequence specificity but does not provide further detail regarding the structure, size, or number of the stem loop(s).

We separated lncRNAs bound to PNPase into clusters representing increased and decreased potential for interacting with PNPase. By utilizing features generated for each cluster like GC content and nucleotide pairing distribution, pseudorandom sequences can be produced that visually retain secondary structures similar to what are observed in endogenous lncRNAs identified in the PNPase CLIP sequences. Our data suggest that lncRNAs with lower GC content and increased unpaired nucleotide distribution as well as certain frequencies of various k-mers are most likely to have an increased interaction potential for PNPase. Low GC content (85) and increased distance ratio acguD are both characteristic qualities of lncRNAs, as higher GC content reflects protein-coding regions and the presence of trinucleotides required for amino acid assembly. This suggests that the more “lncRNA-like” a sequence is, the higher the potential binding to PNPase. Lower frequencies of k-mer sequences

such as CGG but higher frequencies of ACACA and CACAA indicate the potential for sequence-driven motifs that complement secondary structures. Overall, these data support the hypothesis that PNPase has specificity for lncRNAs in a manner more complicated than generic stem-loops or simple primary sequences.

Additionally, we validated that a construct designed from these predicted features can interact with PNPase. Future research could explore the degree of affinity between PNPase and the Predicted RNA probe and if point mutations can disrupt that affinity. Future applications of this research could also evaluate if a construct containing a sequence with high PNPase binding potential can be tracked passing into the mitochondrion. These future experiments can include in vivo chemical modifications, such as dimethyl sulfate mutational profiling with sequencing (DMS-MaPseq), that can allow for tracking even low abundance lncRNAs (86). Additionally, the machine learning models trained in our application only capture a single read from the PNPase CLIP for each lncRNA represented. While we use the most abundant and correctly mapped read as our surrogate, it could be that additional fragments from the same lncRNA further

increase suitable binding sequences to PNPase. While this and other research indicate that PNPase possesses RNA structural specificity, the role of charge-based binding should not be ignored, and future evaluations should explore PNPase's electrostatic features (87). Finally, although novel lncRNAs that were annotated but have not been functionally characterized (i.e., Gm- or -Rik) were included in sequencing analyses, multiple other datasets exist that incorporate additional novel lncRNAs, which may afford a broader understanding of lncRNA interaction with PNPase.

Conclusions

The framework from LncFinder and the machine learning algorithms begins to provide an archetype for designing lncRNA sequences to target to eukaryotic mitochondrial PNPase. Specifically, through constructing a lncRNA with low GC content and increased unpaired nucleotide distribution as well as k-mer specific base combinations, an engineered ncRNA optimized for binding to PNPase can be constructed (Fig. 7). Beyond the therapeutic possibilities, this pipeline of lncRNA sequencing combined with CLIP-Seq, machine learning, and probe design assessments shows promise in optimizing lncRNA-binding protein analyses.

DATA AVAILABILITY

The datasets and computer code produced in this study are available in the following databases: HITS-CLIP data: Sequence Read Archive: PRJNA553501; lncRNA-Seq data: Sequence Read Archive: PRJNA553501; Differential Gene Expression: Github: qahathaway/Mitochondria_PNPase_lncRNA; and Source Code: Github: qahathaway/Mitochondria_PNPase_lncRNA.

SUPPLEMENTAL MATERIAL

Supplemental Figs. S1–S12 and Tables S1 and S2: <https://doi.org/10.6084/m9.figshare.25605657>.

ACKNOWLEDGMENTS

We acknowledge the West Virginia University (WVU) Genomics Core Facility, Morgantown, WV, as well as the Marshall University Genomics Core Facility, Huntington, WV, for the support provided to help make this publication possible. We thank the West Virginia University Ruby Memorial Hospital Heart and Vascular Institute for collaborating on this work.

GRANTS

This work was supported by National Heart, Lung, and Blood Institute Grants R01 HL-128485 and R01 HL-168290 (to J.M.H.), National Institute of Environmental Health Sciences Grant R01 ES-034628 (to J.M.H.), American Heart Association Grant 829079 (to A.D.T.), American Heart Association Grant AHA-17PRE33660333 (to Q.A.H.), American Heart Association Grant AHA-20PRE35080170 (to A.K.), National Science Foundation Grant 2125872 (to E.R.C.), National Institute of General Medical Sciences Grant R01GM133857 (to A.R.R.), CoBRE Tumor Microenvironment supported by National Institute of General Medical Sciences Grant P20GM121322 (to I.M.), West Virginia IDeA Network of Biomedical Research WV-INBRE supported by National Institute of General Medical Sciences Grant P20GM103434, WVU Genomics Core Facility supported by Clinical and Translational Science Institute Grant U54GM104942, WVU

Microscopy Imaging Facility for use of the Nikon A1R/N SIM-E and Dell Workstation (U54GM104942, P20GM103434, and P30GM103488), and the Community Foundation for the Ohio Valley Whipkey Trust (to J.M.H.).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

A.D.T. and Q.A.H. conceived and designed research; A.D.T., Q.A.H., A.K., M.V.P., M.S.N., C.C.C., E.R.C., S.A.S., M.T.W., E.S.W.-R., G.K.F., A.J.D., and D.L.S. performed experiments; A.D.T., Q.A.H., A.K., E.R.C., S.A.S., M.T.W., E.S.W.-R., and S.R. analyzed data; A.D.T., Q.A.H., A.K., M.V.P., M.S.N., E.R.C., S.A.S., A.R.R., I.M., and J.M.H. interpreted results of experiments; A.D.T., Q.A.H., E.R.C., and S.A.S. prepared figures; A.D.T. and Q.A.H. drafted manuscript; A.D.T., Q.A.H., A.K., M.V.P., E.R.C., A.J.D., S.R., A.R.R., I.M., and J.M.H. edited and revised manuscript; A.D.T., Q.A.H., and J.M.H. approved final version of manuscript.

REFERENCES

- Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* 23: 1494–1504, 2009. doi:10.1101/gad.1800909.
- Yao RW, Wang Y, Chen LL. Cellular functions of long noncoding RNAs. *Nat Cell Biol* 21: 542–551, 2019. doi:10.1038/s41556-019-0311-8.
- Szczeniak MW, Kubiak MR, Wanowska E, Makałowska I. Comparative genomics in the search for conserved long noncoding RNAs. *Essays Biochem* 65: 741–749, 2021. doi:10.1042/EBC20200069.
- Zaghlool A, Niaz A, Bjorklund AK, Westholm JO, Ameer A, Feuk L. Characterization of the nuclear and cytosolic transcriptomes in human brain tissue reveals new insights into the subcellular distribution of RNA transcripts. *Sci Rep* 11: 4076, 2021. doi:10.1038/s41598-021-83541-1.
- Thomson DW, Dinger ME. Endogenous microRNA sponges: evidence and controversy. *Nat Rev Genet* 17: 272–283, 2016. doi:10.1038/nrg.2016.20.
- Hathaway QA, Pinti MV, Durr AJ, Waris S, Shepherd DL, Hollander JM. Regulating microRNA expression: at the heart of diabetes mellitus and the mitochondrion. *Am J Physiol Heart Circ Physiol* 314: H293–H310, 2018. doi:10.1152/ajpheart.00520.2017.
- Sun W, Lu Y, Zhang H, Zhang J, Fang X, Wang J, Li M. Mitochondrial non-coding RNAs are potential mediators of mitochondrial homeostasis. *Biomolecules* 12: 1863, 2022. doi:10.3390/biom12121863.
- Durr AJ, Hathaway QA, Kunovac A, Taylor AD, Pinti MV, Rizwan S, Shepherd DL, Cook CC, Fink GK, Hollander JM. Manipulation of the miR-378a/mt-ATP6 regulatory axis rescues ATP synthase in the diabetic heart and offers a novel role for lncRNA Kcnq1ot1. *Am J Physiol Cell Physiol* 322: C482–C495, 2022. doi:10.1152/ajpcell.00446.2021.
- Shi Z, Yang WZ, Lin-Chao S, Chak KF, Yuan HS. Crystal structure of Escherichia coli PNPase: central channel residues are involved in processive RNA degradation. *RNA* 14: 2361–2371, 2008. doi:10.1261/rna.1244308.
- Stone CM, Butt LE, Bufton JC, Lourenco DC, Gowers DM, Pickford AR, Cox PA, Vincent HA, Callaghan AJ. Inhibition of homologous phosphorolytic ribonucleases by citrate may represent an evolutionarily conserved communicative link between RNA degradation and central metabolism. *Nucleic Acids Res* 45: 4655–4666, 2017. doi:10.1093/nar/gkx114.
- Hui MP, Foley PL, Belasco JG. Messenger RNA degradation in bacterial cells. *Annu Rev Genet* 48: 537–559, 2014. doi:10.1146/annurev-genet-120213-092340.
- Briani F, Carzaniga T, Dehò G. Regulation and functions of bacterial PNPase. *Wiley Interdiscip Rev RNA* 7: 241–258, 2016. doi:10.1002/wrna.1328.

13. Zangrossi S, Briani F, Ghisotti D, Regonesi ME, Tortora P, Deho G. Transcriptional and post-transcriptional control of polynucleotide phosphorylase during cold acclimation in *Escherichia coli*. *Mol Microbiol* 36: 1470–1480, 2000. doi:10.1046/j.1365-2958.2000.01971.x.
14. Pobre V, Arraiano CM. Next generation sequencing analysis reveals that the ribonucleases RNase II, RNase R and PNPase affect bacterial motility and biofilm formation in *E. coli*. *BMC Genomics* 16: 72, 2015. doi:10.1186/s12864-015-1237-6.
15. Saramago M, Barria C, Dos Santos RF, Silva IJ, Pobre V, Domingues S, Andrade JM, Viegas SC, Arraiano CM. The role of RNases in the regulation of small RNAs. *Curr Opin Microbiol* 18: 105–115, 2014. doi:10.1016/j.mib.2014.02.009.
16. Sarkar D, Park ES, Emdad L, Randolph AR, Valerie K, Fisher PB. Defining the domains of human polynucleotide phosphorylase (hPNPaseOLD-35) mediating cellular senescence. *Mol Cell Biol* 25: 7333–7343, 2005. doi:10.1128/MCB.25.16.7333-7343.2005.
17. Portnoy V, Palnizky G, Yehudai-Resheff S, Glaser F, Schuster G. Analysis of the human polynucleotide phosphorylase (PNPase) reveals differences in RNA binding and response to phosphate compared to its bacterial and chloroplast counterparts. *RNA* 14: 297–309, 2008. doi:10.1261/rna.698108.
18. Liu P, Huang J, Zheng Q, Xie L, Lu X, Jin J, Wang G. Mammalian mitochondrial RNAs are degraded in the mitochondrial intermembrane space by RNASET2. *Protein Cell* 8: 735–749, 2017. doi:10.1007/s13238-017-0448-9.
19. Wang G, Chen HW, Oktay Y, Zhang J, Allen EL, Smith GM, Fan KC, Hong JS, French SW, McCaffery JM, Lightowlers RN, Morse HC 3rd, Koehler CM, Teitell MA. PNPase regulates RNA import into mitochondria. *Cell* 142: 456–467, 2010. doi:10.1016/j.cell.2010.06.035.
20. Stickney LM, Hankins JS, Miao X, Mackie GA. Function of the conserved S1 and KH domains in polynucleotide phosphorylase. *J Bacteriol* 187: 7214–7221, 2005. doi:10.1128/JB.187.21.7214-7221.2005.
21. Goltzrosch B, Lin CL, Li CL, Yang WZ, Chu LY, Agrawal S, Yuan HS. Crystal structure of dimeric human PNPase reveals why disease-linked mutants suffer from low RNA import and degradation activities. *Nucleic Acids Res* 46: 8630–8640, 2018. doi:10.1093/nar/gky642.
22. Lin CL, Wang Y, Yang WZ, Hsiao YY, Yuan HS. Crystal structure of human polynucleotide phosphorylase: insights into its domain function in RNA binding and degradation. *Nucleic Acids Res* 40: 4146–4157, 2012. doi:10.1093/nar/gkr1281.
23. Wong AG, McBurney KL, Thompson KJ, Stickney LM, Mackie GA. S1 and KH domains of polynucleotide phosphorylase determine the efficiency of RNA binding and autoregulation. *J Bacteriol* 195: 2021–2031, 2013. doi:10.1128/JB.00062-13.
24. Shepherd DL, Hathaway QA, Pinti MV, Nichols CE, Durr AJ, Sreekumar S, Hughes KM, Stine SM, Martinez I, Hollander JM. Exploring the mitochondrial microRNA import pathway through polynucleotide phosphorylase (PNPase). *J Mol Cell Cardiol* 110: 15–25, 2017. doi:10.1016/j.yjmcc.2017.06.012.
25. Shimada E, Ahsan FM, Nili M, Huang D, Atamdede S, TeSlaa T, Case D, Yu X, Gregory BD, Perrin BJ, Koehler CM, Teitell MA. PNPase knockout results in mtDNA loss and an altered metabolic gene expression program. *PLoS One* 13: e0200925, 2018. doi:10.1371/journal.pone.0200925.
26. Chen HW, Rainey RN, Balatoni CE, Dawson DW, Troke JJ, Wasiak S, Hong JS, McBride HM, Koehler CM, Teitell MA, French SW. Mammalian polynucleotide phosphorylase is an intermembrane space RNase that maintains mitochondrial homeostasis. *Mol Cell Biol* 26: 8475–8487, 2006. doi:10.1128/MCB.01002-06.
27. Croston TL, Thapa D, Holden AA, Tveter KJ, Lewis SE, Shepherd DL, Nichols CE, Long DM, Olfert IM, Jagannathan R, Hollander JM. Functional deficiencies of subsarcolemmal mitochondria in the type 2 diabetic human heart. *Am J Physiol Regul Integr Comp Physiol* 307: H54–H65, 2014. doi:10.1152/ajpheart.00845.2013.
28. Palmer JW, Tandler B, Hoppel CL. Biochemical properties of subsarcolemmal and interfibrillar mitochondria isolated from rat cardiac muscle. *J Biol Chem* 252: 8731–8739, 1977. doi:10.1016/S0021-9258(19)75283-1.
29. Baseler WA, Dabkowski ER, Williamson CL, Croston TL, Thapa D, Powell MJ, Razunguzwa TT, Hollander JM. Proteomic alterations of distinct mitochondrial subpopulations in the type 1 diabetic heart: contribution of protein import dysfunction. *Am J Physiol Regul Integr Comp Physiol* 300: R186–R200, 2011. doi:10.1152/ajpregu.00423.2010.
30. Baseler WA, Dabkowski ER, Jagannathan R, Thapa D, Nichols CE, Shepherd DL, Croston TL, Powell M, Razunguzwa TT, Lewis SE, Schnell DM, Hollander JM. Reversal of mitochondrial proteomic loss in Type 1 diabetic heart with overexpression of phospholipid hydroperoxide glutathione peroxidase. *Am J Physiol Regul Integr Comp Physiol* 304: R553–R565, 2013. doi:10.1152/ajpregu.00249.2012.
31. Dabkowski ER, Baseler WA, Williamson CL, Powell M, Razunguzwa TT, Frisbee JC, Hollander JM. Mitochondrial dysfunction in the type 2 diabetic heart is associated with alterations in spatially distinct mitochondrial proteomes. *Am J Physiol Regul Integr Comp Physiol* 299: H529–H540, 2010. doi:10.1152/ajpheart.00267.2010.
32. Jagannathan R, Thapa D, Nichols CE, Shepherd DL, Stricker JC, Croston TL, Baseler WA, Lewis SE, Martinez I, Hollander JM. Translational regulation of the mitochondrial genome following redistribution of mitochondrial microRNA in the diabetic heart. *Circ Cardiovasc Genet* 8: 785–802, 2015. doi:10.1161/CIRCGENETICS.115.001067.
33. Claycomb WC, Lanson NA Jr, Stallworth BS, Egeland DB, Delcarpio JB, Bahinski A, Izzo NJ Jr, HL-1 cells: a cardiac muscle cell line that contracts and retains phenotypic characteristics of the adult cardiomyocyte. *Proc Natl Acad Sci U S A* 95: 2979–2984, 1998. doi:10.1073/pnas.95.6.2979.
34. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72: 248–254, 1976. doi:10.1006/abio.1976.9999.
35. Thapa D, Nichols CE, Lewis SE, Shepherd DL, Jagannathan R, Croston TL, Tveter KJ, Holden AA, Baseler WA, Hollander JM. Transgenic overexpression of mitofilin attenuates diabetes mellitus-associated cardiac and mitochondrial dysfunction. *J Mol Cell Cardiol* 79: 212–223, 2015. doi:10.1016/j.yjmcc.2014.11.008.
36. Hathaway QA, Durr AJ, Shepherd DL, Pinti MV, Brandebura AN, Nichols CE, Kunovac A, Goldsmith WT, Friend SA, Abukabba AB, Fink GK, Nurkiewicz TR, Hollander JM. miRNA-378a as a key regulator of cardiovascular health following engineered nanomaterial inhalation exposure. *Nanotoxicology* 13: 644–663, 2019. doi:10.1080/17435390.2019.1570372.
37. Hathaway QA, Nichols CE, Shepherd DL, Stapleton PA, McLaughlin SL, Stricker JC, Rellick SL, Pinti MV, Abukabba AB, McBride CR, Yi J, Stine SM, Nurkiewicz TR, Hollander JM. Maternal-engineered nanomaterial exposure disrupts progeny cardiac function and bioenergetics. *Am J Physiol Heart Circ Physiol* 312: H446–H458, 2017. doi:10.1152/ajpheart.00634.2016.
38. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(delta delta C(T)) method. *Methods* 25: 402–408, 2001. doi:10.1006/meth.2001.1262.
39. Untergrasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* 35: W71–74, 2007. doi:10.1093/nar/gkm306.
40. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12: 357–360, 2015. doi:10.1038/nmeth.3317.
41. Perteau M, Kim D, Perteau GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* 11: 1650–1667, 2016. doi:10.1038/nprot.2016.095.
42. Dong C, Shen S, Keleş S. AdaliftOver: high-resolution identification of orthologous regulatory elements with Adaptive liftOver. *Bioinformatics* 39: btad149, 2023. doi:10.1093/bioinformatics/btad149.
43. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550, 2014. doi:10.1186/s13059-014-0550-8.
44. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28: 511–515, 2010. doi:10.1038/nbt.1621.
45. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools. *GigaScience* 10: giab008, 2021. doi:10.1093/gigascience/giab008.
46. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer, 2009.
47. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43: e47, 2015. doi:10.1093/nar/gkv007.

48. McDermaid A, Monier B, Zhao J, Liu B, Ma Q. Interpretation of differential gene expression results of RNA-seq data: review and integration. *Brief Bioinform* 20: 2044–2054, 2019. doi:10.1093/bib/bby067.
49. Blighe K, Rana S, Lewis S. EnhancedVolcano: publication-ready volcano plots with enhanced colouring and labeling (Online). 2018. <https://github.com/kevinblighe/EnhancedVolcano> [Accessed 14 Nov 2019].
50. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498–2504, 2003. doi:10.1101/gr.1239303.
51. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* 17: 10–12, 2011. doi:10.14806/ej.17.1.200.
52. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10: R25, 2009. doi:10.1186/gb-2009-10-3-r25.
53. Hunter JD. Matplotlib: a 2D graphics environment. *Comp Sci Eng* 9: 90–95, 2007. doi:10.1109/MCSE.2007.55.
54. Alasoo K. wiggleplotr: make read coverage plots from BigWig files (Online). Bioconductor. R package version 1.28.0. 2024 <https://bioconductor.org/packages/release/bioc/html/wiggleplotr.html> [Accessed 7 Mar 2024].
55. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* 25: 1841–1842, 2009. doi:10.1093/bioinformatics/btp328.
56. Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. ViennaRNA Package 2.0. *Algorithms Mol Biol* 6: 26, 2011. doi:10.1186/1748-7188-6-26.
57. Han S, Liang Y, Ma Q, Xu Y, Zhang Y, Du W, Wang C, Li Y. LncFinder: an integrated platform for long non-coding RNA identification utilizing sequence intrinsic composition, structural information and physicochemical property. *Brief Bioinform* 20: 2009–2027, 2018. doi:10.1093/bib/bby065.
58. Charif D, Lobry JR. SeqinR 1.0-2: a contributed package to the {R} project for statistical computing devoted to biological sequences retrieval and analysis. In: *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations*, edited by Bastolla U, Porto M, Roman HE, Vendruscolo M. New York: Springer Verlag, 2007, p. 207–232.
59. Stothard P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28: 1102, 1104, 2000. doi:10.2144/00286ir01.
60. Levin A, Lis M, Ponty Y, O'Donnell CW, Devadas S, Berger B, Waldspuhl J. A global sampling approach to designing and reengineering RNA secondary structures. *Nucleic Acids Res* 40: 10041–10052, 2012. doi:10.1093/nar/gks768.
61. Liaw A, Wiener M. Classification and regression by randomForest. *R News* 2: 18–22, 2002.
62. Evans JS, Murphy MA, Holden ZA, Cushman SA. Modeling species distribution and change using Random Forests. In: *Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications*, edited by Drew CA, Wiersma YF, Huettmann F. New York: Springer, 2011, p. 139–159.
63. Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41: 207–234, 2005. doi:10.1016/j.pep.2005.01.016.
64. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nat Methods* 19: 679–682, 2022. doi:10.1038/s41592-022-01488-1.
65. Schrodinger L. The PyMOL Molecular Graphics System, Version~1.8, edited by Holder T. New York: PyMOL, 2015.
66. Bandiera S, Ruberg S, Girard M, Cagnard N, Hanein S, Chretien D, Munnich A, Lyonnet S, Henrion-Caude A. Nuclear outsourcing of RNA interference components to human mitochondria. *PLoS One* 6: e20746, 2011. doi:10.1371/journal.pone.0020746.
67. Hollander JM, Lin KM, Scott BT, Dillmann WH. Overexpression of PHGPx and HSP60/10 protects against ischemia/reoxygenation injury. *Free Radic Biol Med* 35: 742–751, 2003. doi:10.1016/s0891-5849(03)00400-3.
68. Cabili MN, Dunagin MC, McClanahan PD, Biaesch A, Padovan-Merhar O, Regev A, Rinn JL, Raj A. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol* 16: 20, 2015. doi:10.1186/s13059-015-0586-4.
69. Li Z, Liu L, Jiang S, L Q, Feng C, Du Q, Zou D, Xiao J, Zhang Z, Ma L. LncExpDB: an expression database of human long non-coding RNAs. *Nucleic Acids Res* 49: D962–D968, 2021. doi:10.1093/nar/gkaa850.
70. Zhao Y, Liu S, Zhou L, Li X, Meng Y, Li Y, Li L, Jiao B, Bai L, Yu Y, Zhang S, Li W, Hoffman AR, Hu JF, Cui J. Aberrant shuttling of long noncoding RNAs during the mitochondria-nuclear crosstalk in hepatocellular carcinoma cells. *Am J Cancer Res* 9: 999–1008, 2019.
71. Mohammad G, Kowluru RA. Nuclear genome-encoded long non-coding rnas and mitochondrial damage in diabetic retinopathy. *Cells* 10: 3271, 2021. doi:10.3390/cells10123271.
72. Zhao Y, Zhou L, Li H, Sun T, Wen X, Li X, Meng Y, Li Y, Liu M, Liu S, Kim S, Xiao J, Li L, Zhang S, Li W, Cohen P, Hoffman AR, Hu JF, Cui J. Nuclear-encoded lncRNA MALAT1 epigenetically controls metabolic reprogramming in HCC Cells through the mitophagy pathway. *Mol Ther Nucleic Acids* 23: 264–276, 2021. doi:10.1016/j.omtn.2020.09.040.
73. Zhan J, Jin K, Xie R, Fan J, Tang Y, Chen C, Li H, Wang DW. AGO2 protects against diabetic cardiomyopathy by activating mitochondrial gene translation. *Circulation* 149: 1102–1120, 2024. doi:10.1161/CIRCULATIONAHA.123.065546.
74. Leucci E, Patella F, Waage J, Holmström K, Lindow M, Porse B, Kauppinen S, Lund AH. microRNA-9 targets the long non-coding RNA MALAT1 for degradation in the nucleus. *Sci Rep* 3: 2535, 2013. doi:10.1038/srep02535.
75. Yoon JH, Abdelmohsen K, Srikantan S, Yang X, Martindale JL, De S, Huarte M, Zhan M, Becker KG, Gorospe M. LincRNA-p21 suppresses target mRNA translation. *Mol Cell* 47: 648–655, 2012 [Erratum in *Mol Cell* 50: 303, 2013]. doi:10.1016/j.molcel.2012.06.027.
76. Lin C, Miles WO. Beyond CLIP: advances and opportunities to measure RBP-RNA and RNA-RNA interactions. *Nucleic Acids Res* 47: 5490–5501, 2019. doi:10.1093/nar/gkz295.
77. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME Suite: tools for motif discovery and searching. *Nucleic Acids Res* 37: W202–W208, 2009. doi:10.1093/nar/gkp335.
78. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38: 576–589, 2010. doi:10.1016/j.molcel.2010.05.004.
79. Gudenäs BL, Wang L. Prediction of lncRNA subcellular localization with deep learning from sequence features. *Sci Rep* 8: 16385, 2018. doi:10.1038/s41598-018-34708-w.
80. McIntyre GJ, Yu Y, Lomas M, Fanning GC. The effects of stem length and core placement on shRNA activity. *BMC Mol Biol* 12: 34, 2011. doi:10.1186/1471-2199-12-34.
81. Miyagawa R, Tano K, Mizuno R, Nakamura Y, Ijiri K, Rakwal R, Shibato J, Masuo Y, Mayeda A, Hirose T, Akimitsu N. Identification of cis- and trans-acting factors involved in the localization of MALAT-1 noncoding RNA to nuclear speckles. *RNA* 18: 738–751, 2012. doi:10.1261/rna.028639.111.
82. Zhang X, Hamblin MH, Yin KJ. The long noncoding RNA Malat1: its physiological and pathophysiological functions. *RNA Biol* 14: 1705–1714, 2017. doi:10.1080/15476286.2017.1358347.
83. Jeandard D, Smirnova A, Tarassov I, Barrey E, Smirnov A, Entelis N. Import of non-coding RNAs into human mitochondria: a critical review and emerging approaches. *Cells* 8: 286, 2019. doi:10.3390/cells8030286.
84. Wang G, Shimada E, Zhang J, Hong JS, Smith GM, Teitell MA, Koehler CM. Correcting human mitochondrial mutations with targeted RNA import. *Proc Natl Acad Sci USA* 109: 4840–4845, 2012. doi:10.1073/pnas.1116792109.
85. Niazi F, Valadkhan S. Computational analysis of functional long non-coding RNAs reveals lack of peptide-coding capacity and parallels with 3' UTRs. *RNA* 18: 825–843, 2012. doi:10.1261/rna.029520.111.
86. Zubradt M, Gupta P, Persad S, Lambowitz AM, Weissman JS, Rouskin S. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat Methods* 14: 75–82, 2017. doi:10.1038/nmeth.4057.
87. Ahmad S, Sarai A. Analysis of electric moments of RNA-binding proteins: implications for mechanism and prediction. *BMC Struct Biol* 11: 8, 2011. doi:10.1186/1472-6807-11-8.
88. Borrayo E, et al. Genomic signal processing methods for computation of alignment-free distances from DNA sequences. *Plos One* 9: 11, 2014.