pCLASH: Mapping snoRNA-guided 2'-O-Methylation Targets on pre-mRNA

Alex Choi

GCS Senior Thesis
Duke University
Spring 2021

Principal Investigator: Christopher Holley, MD, PhD
Research Mentor: Brittany Elliott, PhD

GCS Faculty Advisor: Charles Gersbach, PhD
GCS Topic: Engineer Better Medicines

Table of Contents

Abstract

Small nucleolar RNAs (snoRNAs) are noncoding RNAs canonically known to guide post-transcriptional modifications on ribosomal RNAs. It has been previously shown that box C/D snoRNAs are capable of guiding 2'-O-methylations (Nm) on mRNAs, with these modifications regulating physiologic gene expression and protein translation. To date, the comprehensive, transcriptome-wide set of mRNA targets of snoRNAs has not been identified. Here, we demonstrate pCLASH, a method to map the global set of pre-mRNA targets of Nm modification by snoRNAs through crosslinking, ligation, and sequencing of hybrids. Particularly novel aspects of this method include the use of phenol-toluol extraction (PTex) to isolate crosslinked RNA-protein along with the combination of nuclei isolation and rRNA depletion prior to immunoprecipitation. Over 2900 hybrids were obtained from pCLASH, with 166 of them being snoRNA-mRNA hybrids. Of note, the snoRNA-mRNA hybrids/Total Reads ratio using pCLASH was several orders of magnitude higher than previously published methods to identify snoRNA-mRNA interactions. Interestingly, numerous mitochondrial targets were revealed, meriting future explorations regarding the role of snoRNA-guided Nm on mitochondrial function and oxidative phosphorylation.

Introduction

A. Overview of Small Nucleolar RNAs (snoRNAs)

Small nucleolar RNAs, or snoRNAs, are canonically known to guide post-transcriptional modifications on other RNAs, with ribosomal RNAs (rRNAs) and small nuclear RNAs (snRNAs) being the primary targets. snoRNAs are typically 60-300 nt long and can be classified into two classes: C/D box snoRNAs and H/ACA box snoRNAs[1].

Our research group focuses on studying C/D box snoRNAs, which get their name from two conserved sequence elements: box C (RUGAUGA, where R is a purine) and box D (CUGA) found at the 5' and 3' ends of the snoRNA, respectively. These elements are replicated in the C' and D' boxes located in the central region of the snoRNA, as shown in Figure 1. The snoRNA forms a characteristic kink-turn structure that allows for the assembly of the snoRNA ribonucleoprotein (snoRNP) complex. This complex consists of four core proteins: fibrillarin (FBL), NOP56, NOP58, and 15.5kD. Fibrillarin is the methyltransferase that catalyzes 2'-O-Methylation (Nm), or the transfer of the methyl group from S-adenosyl-methionine (SAM) onto the 2'-hydroxyl group on the ribose of the target RNA, as depicted in Figure 2. NOP56, NOP58, and 15.5kD protect the snoRNA from degradation and direct its localization to the nucleolus[2].

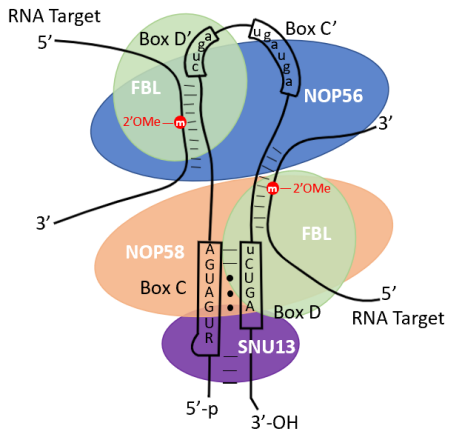


Figure 1. C/D Box snoRNA Structure

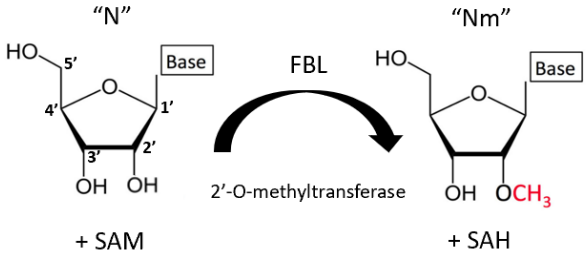


Figure 2. Transfer of the methyl group onto the target base by Fibrillarin (FBL)

The snoRNA binds to its target RNA through a short (7-21 nucleotide) antisense element, and the methyltransferase fibrillarin catalyzes the Nm modification on the nucleotide of the target RNA 5 nucleotides upstream of the D/D' box, as indicated in Figure 1.

B. Functional Role of snoRNAs on mRNA Targets

SnoRNAs canonically guide modifications on rRNAs, having a functional role in the maturation and stabilization of rRNA[3]. However, much less is known regarding the presence and biological role of snoRNAs on mRNA targets. Recently, our lab has shown that box C/D snoRNAs can guide 2'-O-methylations on mRNAs, with a functional role of regulating the expression and translation of the target mRNA. We demonstrated that in the coding sequence of peroxidasin (pxdn) mRNA, internal Nm modification results in increased mRNA levels but decreased peroxidase activity due to inhibition of translation as shown in Figure 3. This revealed a novel mechanism of regulatory control for the Nm modification on mRNA, influencing mRNA abundance and protein levels in vitro and in vivo[4]. Since there are hundreds of C/D box snoRNAs with unexplored functions, snoRNA-guided Nm of mRNA could be a widespread post-transcriptional mechanism to regulate physiologic gene expression in vivo. The discovery of such novel mRNA targets of Nm by snoRNAs could open up exciting avenues for future research.

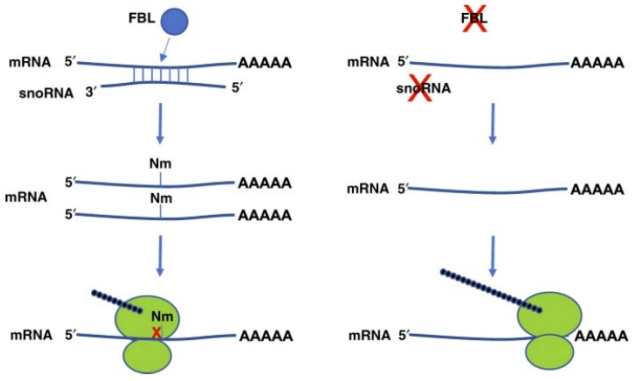


Figure 3. Nm increases mRNA abundance but inhibits translation in Pxdn mRNA (Elliott et al. 2019)

C. Existing Techniques to Identify snoRNA Targets

There have been several methods developed in recent years to identify snoRNA target sites. Crosslinking and immunoprecipitation (CLIP) relies on UV crosslinking to introduce a covalent bond between the protein of interest and any target RNAs. This covalent linkage protects the RNA-protein complex from RNase digestion as well as during stringent purification steps after IP. Recovered RNA segments can then be reverse transcribed, PCR amplified, and put through next-generation sequencing methods to identify novel snoRNA targets at nucleotide resolution[5]. There have been improvements to this CLIP method, including photoactivatable-ribonucleoside-enhanced CLIP (PAR-CLIP) for increased crosslinking efficiency and enhanced CLIP (eCLIP) for improved library generation efficiency and specificity[6,7]. While these CLIP methods have revealed new targets of snoRNA modification in physiologic settings, the vast majority of obtained targets were ribosomal RNA targets, as these constitute the majority of snoRNA-target RNA interactions[5].

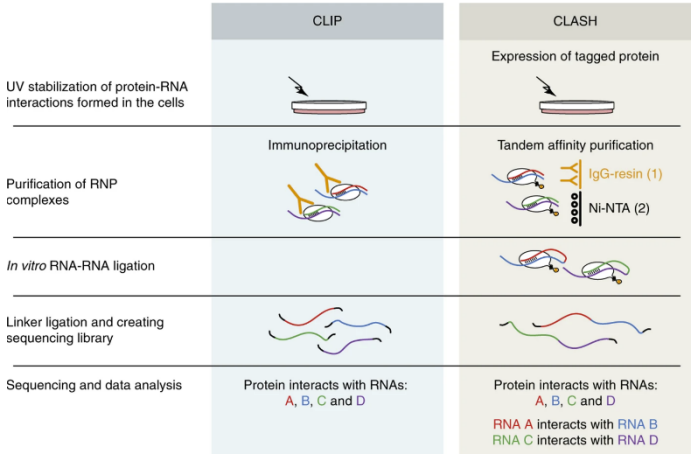


Figure 4. Comparison of CLIP and CLASH Methods (Helwak and Tollervey 2019)

Crosslinking, ligation, and sequencing of hybrids (CLASH) is an alternative method to detect transcriptome-wide RNA-protein interactions. CLASH is a modification of the traditional CLIP protocol, with an additional proximity ligation step of the RNAs as shown in Figure 4 such that RNA-RNA interactions

mediated by an RNA-binding protein can be identified[8,9]. A noteworthy variant of CLASH is quick CLASH (qCLASH), which streamlines the traditional CLASH process by eliminating the radiolabeling and SDS-PAGE steps known to result in large amounts of sample loss[10]. The success of CLASH in identifying mRNA-miRNA interactions provides a strong premise for its application in mRNA-snoRNA interactions[10,11].

D. Method Development to Identify novel mRNA Targets of snoRNAs

Ribosomal RNA constitutes approximately 80% of all RNA in the cell. By contrast, mRNA only makes up roughly 4% of all RNA[12]. Consequently, when targeting snoRNPs, CLIP or CLASH without rRNA depletion leads to the vast majority of obtained reads being rRNAs, which masks other snoRNA targets including mRNAs[5,13]. Since we are interested in mRNA targets of snoRNAs, the aim of this project is to develop a workflow that will detect snoRNA-mRNA interactions in a high-throughput manner.

Several methods to sequentially enrich for mRNAs have been utilized in this project. Phenol-toluol extraction (PTex) is a method to purify crosslinked RNA-protein complexes using a biphasic extraction depicted in Figure 5[14]. This allows for depletion of DNA, lipids, and non-crosslinked RNA and protein for downstream processing of only the RNA-protein complexes of interest.

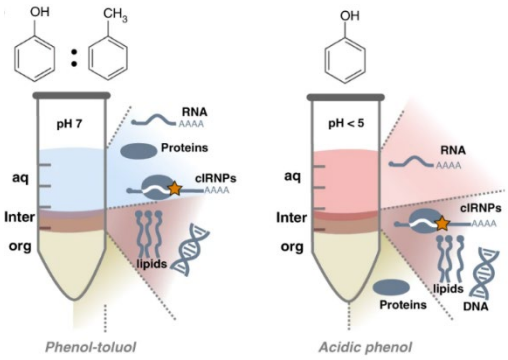


Figure 5. PTex Extraction of Crosslinked RNA-Protein (Urdaneta et al. 2019)

One method to enrich for mRNA is poly(A) selection, where oligo (dT) primers are used to hybridize to the 3' polyadenylated tails of mature mRNAs. An alternative

method to increase the proportion of mRNAs detected is to deplete the sample of rRNAs. This can be done through hybridization capture of rRNAs using rRNA-binding probes and subsequent removal using magnetic beads. While poly(A) selection is often cheaper and has been shown to have increased quantification of protein-coding genes, rRNA depletion provides greater coverage of intronic regions and has better performance for degraded RNAs[15]. Both mRNA enrichment and rRNA depletion were attempted in this project through the mCLASH and pCLASH protocols, respectively.

E. Potential Application on Therapeutics and Drug Development

In the field of epitranscriptomics studying RNA modifications, N6-methyladenosine (m6A) is perhaps the most widely studied modification, being the most prevalent post-transcriptional modification on eukaryotic mRNA[16]. In 2012, researchers discovered that the methyl group can be added by a “writer” enzyme called METTL3 and removed by “eraser” enzymes FTO or ALKBH5[17,18]. The connection to disease arose in 2014, when researchers showed that without the METTL3 enzyme, human embryonic stem cells grew uncontrollably and did not undergo normal differentiation[19]. In June 2016, Storm Therapeutics became the first start-up dedicated to creating a therapeutic based on RNA modifications. A year later, researchers substantiated the link between m6A and cancer, demonstrating that METTL3 was an essential gene for growth of acute myeloid leukemia cells and that reduced METTL3 resulted in normal differentiation into noncancerous cells[20,21]. Most recently in October 2020, Storm announced STC-15, the first-in-class drug candidate targeting METTL3 being developed for Phase I human clinical trials to treat acute myeloid leukemia[22].

Similar therapeutic applications could be on the horizon for snoRNA-guided Nm. SnoRNA dysregulation has been implicated in various cancers, and snoRNAs have been characterized as essential actors in lung cancer tumorigenesis and progression[23,24]. Dr. Holley has shown that Rpl13a snoRNAs are critical mediators of oxidative stress, and increased expression of several oxidative stress-related snoRNAs have been linked to cartilage aging and osteoarthritis[25,26]. The first snoRNA-associated disease was Prader-Willi Syndrome, with patients demonstrating deletions of the SNORD116 gene cluster, and research is still being conducted on transitions to clinical treatment[27].

One of the remaining technical hurdles for the field of epitranscriptomics is that methods to map and detect RNA modifications remain crude. This project has the potential to provide insight into more accurate and efficient mapping of RNA modifications on mRNA transcripts in vivo. The discovery of novel mRNA targets of snoRNA-guided Nm could spur further research linking additional diseases to snoRNAs and deeper explorations into specific modifications meriting clinical drug targeting as has been done with m6A.

Materials and Methods

The diagram of the full mCLASH and pCLASH workflows can be seen in Supplementary Figures 1 and 2 in the Appendix, respectively.

Cell Collection and UV Crosslinking

40 X 150 mm plates of HEK293T cells were cultured to 80% confluence. After aspiration of media and washing with Dulbecco's phosphate-buffered saline (DPBS), plates were UV crosslinked with 254 nm UV light at 450 mJ/cm^2. Cells were collected in DPBS and centrifuged at 500*xg*. Supernatants were discarded and cell pellets were processed for nuclear isolation (pCLASH) or were stored at -80°C.

Nuclear Fractionation using FR-iCLIP [pCLASH only]

Nuclear fractionation was performed immediately after UV crosslinking according to the Fractionation iCLIP protocol created by Brugiolo et al., which is designed specifically for UV crosslinked cells[28]. The cell pellets from 40 x 150 mm plates of WT HEK293T cells were washed in PBS and centrifuged at 180*xg* for 5 minutes at 4°C in 4-6 50 mL conical tubes. Upon removal of the supernatant, the pellet was gently resuspended in 30-40 mL Hypotonic Buffer per tube. After incubation on ice for 15 minutes and subsequent centrifugation at 425*xg* for 10 minutes at 4°C, cell pellets were resuspended in 30-40 mL Lysis Buffer 0.3%. Following incubation on ice for 10 minutes and centrifugation at 950xg for 10 minutes at 4°C, cell pellets were resuspended in 30-40 mL Lysis Buffer 0.5%. After incubation on ice for 10 minutes and centrifugation at 950xg for 10 minutes at 4°C, the supernatant was discarded to yield the pellet containing the nuclear fraction.

Phenol-Toluol Extraction (PTex)

UV-crosslinked RNA-protein was extracted using phenol-toluol according to Urdaneta et al. 2019[14]. Cell pellets were resuspended in 600 µL DPBS and added to a set of tubes each containing 300 µL phenol (pH 8), 300 µL toluol, and 200 µL bromochloropropane (BCP). After mixing (1 min, 1400 rpm, RT) and centrifugation (5 min, 16,100*xg*, 4°C), the aqueous layer consisting of crosslinked RNPs, RNA, and protein was extracted by pipetting. This aqueous layer was added to tubes with 300 µL Solution D (denaturing solution), 600 µL phenol (pH 8), and 200 µL BCP. After mixing and centrifugation as before, a syringe was used to remove the upper three quarters of aqueous phase (free RNA) and the lower three quarters of the organic phase (free proteins), leaving the interphase with crosslinked RNPs. This step was repeated with 400 µL water, 200 µL ethanol, 400 µL phenol (pH 8), and 200 µL BCP added to the resulting interphase. After mixing and centrifugation as before, a fresh syringe was used to isolate the interphase. 1.5 mL of 100% ethanol was added to each tube of interphase for precipitation. 2 µL of GlycoBlue coprecipitant was also added for improved yield after precipitation. RNA was precipitated at -20°C overnight. The RNA-protein complexes were then pelleted by centrifuging at 18,100*xg* at 4°C for 30 min. After decanting the supernatant, pellets were dried for approximately 10 minutes and resuspended in approximately 30 µL Tris buffer per plate of cells through heating at 50°C for 5 min. RNA quality analysis was performed using a Bioanalyzer (Agilent) and concentrations measured using a Qubit (Thermo Fisher).

mRNA Enrichment using Poly(A) Selection [mCLASH only]

For mCLASH, 600 µL of NEBNext Oligo $d(T)_{25}$ beads were used to bind ~1500 µg of crosslinked RNA-protein and washed twice in 1 mL RNA binding buffer. After pulse vortexing and removal of supernatant, beads were resuspended in 600 µL RNA binding buffer. 10 µL SUPERaseIN and 600 µL of the RNA sample were added and the resulting tube heated for 5 minutes at 65°C. After cooling on ice for 2 minutes and incubation at room temperature for 10 minutes, the supernatant was separated from the beads bound to poly(A) RNA. The beads were then washed three times with 600 µL Wash Buffer to remove unbound RNA. The poly(A) RNA-protein was eluted from the beads in 200 µL Tris Buffer and the yield assessed using the Qubit. This poly(A) selection process was repeated once for a total of two rounds of mRNA enrichment. The resulting RNA was visualized and quantified on a Bioanalyzer (Agilent) to ensure adequate rRNA depletion. Typical yields were about 10 µg of 2X (A) mRNA-protein from 1500 µg of total RNA-protein.

RiboZero pre-rRNA Subtraction [pCLASH only]

For elimination of pre-rRNA in pCLASH, 3 reactions of RiboZero (Illumina) were used on ~88 µg of PTex nuclei, based on the maximum volume that 3 RiboZero reactions could hold. Ethanol precipitation was used to de-salt and clean up the pre-rRNA depleted sample. In addition, a “second pass” was performed in which the RiboZero protocol was carried out again on bead after previously bound pre-rRNA from the first pass had been eluted off by incubating the beads at 80°C for 2 minutes. Since the 45S rRNA precursor is too large (13.3 kb) to be visualized on the Bioanalyzer, RT-qPCR was utilized to quantify pre-rRNA depletion.

mCLASH/pCLASH Protocol

mCLASH/pCLASH was performed as detailed in a protocol devised by Brittany Elliott (see Supp Figs 1,2 in Appendix). The mRNA enriched/rRNA-depleted sample was fragmented by adding 1 µL of a 1:25 dilution of RNAse I for 5 minutes at 37°C and immunoprecipitated using 10 µg of IgG or NOP56 antibody overnight. After stringent washing (6 different buffers of varying salt stringency, 4 rounds of washing for each buffer), the 5’ end of the RNA was phosphorylated using T4 Polynucleotide Kinase (T4 PNK) and ATP. A subsequent intermolecular ligation was performed using T4 RNA ligase 1, with the primary product being ligation of the 3’ end of the snoRNA to the 5’ end of the target RNA. Upon dephosphorylation of the RNA using Antarctic Phosphatase, the miRCat-33 3’ linker was ligated using T4 RNA ligase 2 truncated KQ for specific ligation of the pre-adenylated 3’ linker onto the free -OH at the 3’ end of the target RNA. The NOP56 protein was removed using Proteinase K, and the sample precipitated using ethanol and sodium acetate. 5’ phosphorylation was done using T4 PNK and the 5’ linker subsequently ligated to the 5’ end of the target RNA using T4 RNA ligase 1. The RNA was then reverse transcribed into cDNA using SuperScript III as the reverse transcriptase. Free RNAs were degraded through incubation with 100 mM NaOH at 98°C for 20 minutes. The cDNA was then amplified using PCR for 19 cycles using the Q5 High-Fidelity 2X Master Mix (NEB) and forward/reverse primers recognizing the 3’ and 5’ linkers. After ethanol precipitation of the sample overnight, gel purification using an E-Gel 4% Agarose gel (Invitrogen) was performed to eliminate empty linkers and primer dimers. The Zymoclean Gel DNA Recovery Kit (Zymo) was used to extract DNA from the gel fragment, and NucAway Spin Columns (Thermo

Fisher) were used to de-salt the sample. The samples were run on an Agilent 2100 Bioanalyzer using a High Sensitivity DNA Chip to analyze library sizing and quality prior to sequencing.

MiniSeq Sequencing

1 nM of each sample in a total volume of 7.5 µL was sequenced using the MiniSeq system (Illumina). The 1 nM library was denatured using 0.1N NaOH. Pre-chilled Hybridization Buffer was added to the library to achieve a concentration of 1.8 pM in a 500 µL volume. A PhiX control was also denatured using 0.1N NaOH and diluted to the same concentration of 1.8 pM in 500 µL volume. 100 µL of diluted PhiX and 400 µL of diluted library were combined and loaded onto the MiniSeq. The presence of unique molecular identifiers (UMIs) on the primers used for each sample allowed for multiplex sequencing, as these barcodes could be used for sample identification prior to data analysis. Sequencing data was analyzed using the Hyb pipeline as described by Travis et al[29].

NovaSeq Sequencing

Upon initial sequencing with the MiniSeq, the purified library was sent for sequencing on the Illumina NovaSeq 6000 for increased sequencing depth. Library preparation was performed similarly to the procedure for the MiniSeq, with the primary difference being a library pool concentration of 4 nM rather than the 1 nM used for the MiniSeq platform.

Results

PTex yields crosslinked RNA-protein

As an initial test to determine whether RNA could be visualized after PTex, 12 plates of HEK293T cells were crosslinked at 450 mJ/cm^2. Processing the combined sample using PTex yielded 450 µg of RNA, and the presence of RNA was visualized on an agarose gel shown in Figure 6. The clear 28S and 18S bands demonstrate the presence of RNA after extraction using PTex.

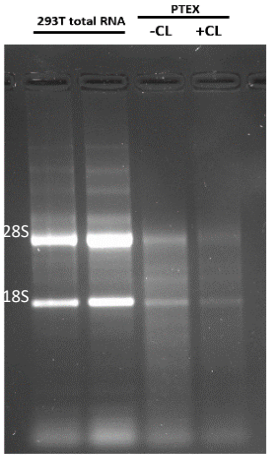


Figure 6. PTex -/+ CL Samples vs. 293T Total RNA on Agarose Gel

To determine the presence of protein after PTex, a western blot was performed for GAPDH – a stable, constitutively expressed RNA-binding protein acting as an ideal housekeeping protein. The input control lanes consisted of ~15 µg of HEK293T cell pellets in each sample that had either been UV-crosslinked at 450 mJ/cm^2 (+CL) or not crosslinked (-CL). The PTex samples were obtained from a combined set of 12 plates of 15 cm HEK293T plates for each sample that had either been UV-crosslinked (+CL) or not (-CL) prior to PTex extraction. As shown in Figure 7, the GAPDH protein corresponding to about 37 kDa was highly abundant in both input lanes. Significantly more GAPDH was detected in the PTex +CL than the PTex -CL lane, as expected with successful crosslinking and isolation of RNA-protein using PTex.

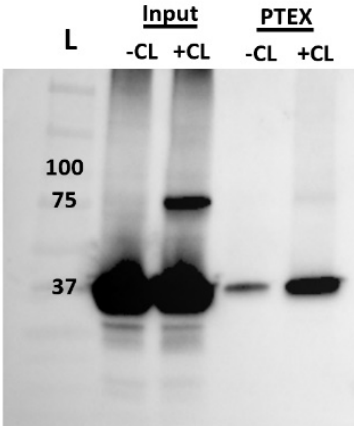


Figure 7. GAPDH Blot of PTex +/- CL and +/- RNase Treatment with input controls (L=ladder)

Poly(A) Selection enriches mRNA from PTex RNA-Protein

An Agilent 2100 Bioanalyzer was used to assess ribosomal RNA levels after successive rounds of poly(A) selection. As shown in Figure 8, in the total cell lysate RNA as well as the total RNA-protein after PTex, the 18S and 28S bands of rRNA were clearly visible. After 1X poly(A), the sample still had traces of 18S and 28S rRNA. It was after a second round of poly(A) selection that the RNA was visibly depleted of rRNA and adequately enriched for mRNA as indicated by the diverse bands without rRNA traces. This is confirmed in the Bioanalyzer electropherogram to the right, demonstrating that the peaks corresponding to the 18S and 28S rRNA have notably disappeared after two rounds of poly(A) selection.

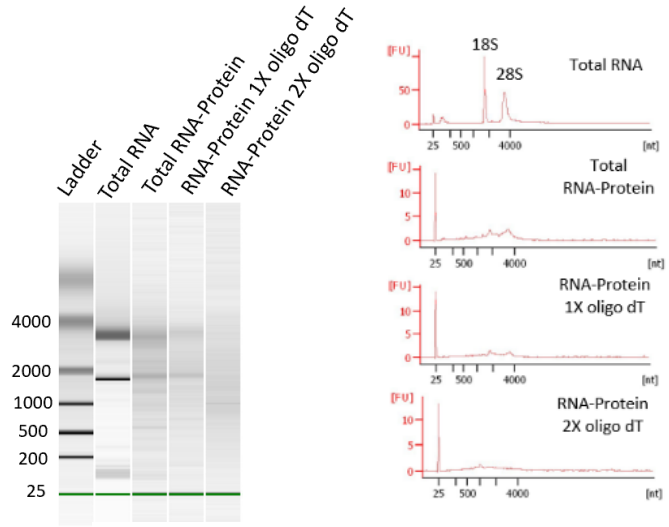


Figure 8. Bioanalyzer gel and electropherogram of PTex RNA after successive poly(A) selection

NOP56 can be immunoprecipitated from Poly(A)-selected RNA-Protein

Immunoprecipitation using a previously tested NOP56 antibody overnight followed by western blot revealed that a NOP56 band could be visualized in the +CL PTex 2X poly(A) sample (Figure 9). This is confirmed by the corresponding NOP56 bands clearly seen in the -CL and +CL input controls. The bands above the NOP56 band were initially thought to be possible crosslinked products of

Figure 9. NOP56 IP of +CL PTex 2X poly(A) sample with input controls

NOP56 bound to larger RNAs but were later confirmed to be background noise from the beads and NOP56 antibody.

RNA is detected in NOP56-IP Poly(A)-selected RNA-Protein

In addition to NOP56 protein, we also sought to confirm that RNA was detectable in the PTex 2X poly(A) NOP56 IP samples. As shown by the agarose gel in Figure 10, RNA is detectable in both the total +CL PTex sample and the +CL PTex 2X poly(A) sample after immunoprecipitating for NOP56. It can be seen that in the mRNA enriched sample, the 18S rRNA band is not detectable unlike in the total PTex sample, confirming rRNA depletion.

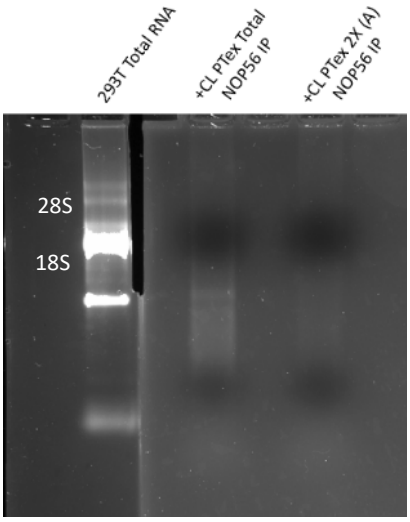


Figure 10. Agarose gel of +CL PTex Total and +CL PTex 2x poly(A) after NOP56 IP

PTex RNA-Protein is fragmented to a target size range

In order to investigate the effects of various fragmentation conditions on RNA sizing and quantity, three different trials of fragmentation using RNase I on +CL PTex samples were compared on the Bioanalyzer (Figure 11). Successful fragmentation can be evidenced by the depletion of longer RNA (>1000 nt) and the collapsing of signal onto shorter RNA (<200 nt). Based on the fragment sizes and RNA concentrations, fragmenting for between 2.5-5 minutes seemed to have the best balance between desired target fragmentation (size: 82-87 nt) and RNA yield.

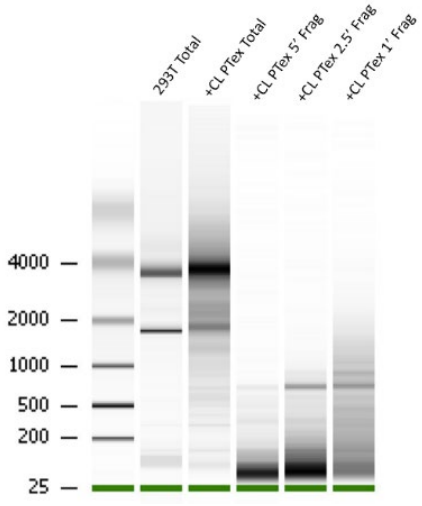


Figure 11. Bioanalyzer gel of +CL PTex samples under various fragmentation conditions

Libraries can be generated from fragmented samples

In addition to the wild-type 293T cells immunoprecipitated with NOP56, an additional WT 293T sample immunoprecipitated with IgG was prepared to serve as a negative control, as well as a SNORD32a/SNORD51 knockout (DKO12) sample immunoprecipitated with NOP56 to get an early idea of whether knocking out those snoRNAs could affect the snoRNA-mRNA reads obtained. 40 plates worth of 150 mm plates for each of the three samples (WT IgG, WT NOP56, DKO12 NOP56) were UV crosslinked at 0.45 J/cm^2, PTex extracted, and poly(A) enriched. The limiting sample was the DKO12 sample yielding 3.5 µg after poly(A) enrichment, so 3.5 µg of each sample was processed using the mCLASH protocol.

RNA quality control analysis results of the three sample libraries using the Agilent 2100 Bioanalyzer are depicted in Figure 12. Of note, the gel purification process was repeated until undesired products <200 bp (primer dimers and empty adapter ligations) were sufficiently removed, as can be seen by the lack of peaks below 200 bp. Both NOP56 samples (WT NOP56, DKO12 NOP56) showed peaks around 310 bp whereas the WT IgG negative control peaked at around 355 bp.

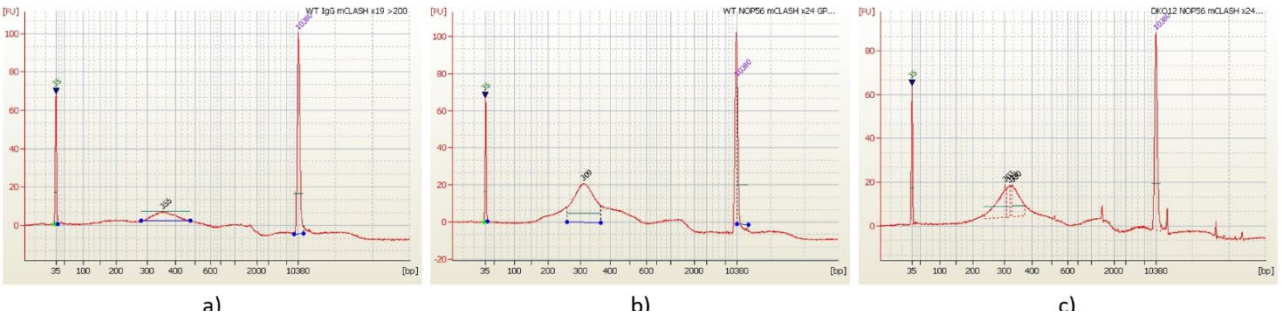


Figure 12. Bioanalyzer Electropherogram Images of a) WT IgG, b) WT NOP56, c) DKO12 NOP56

snoRNA-mRNA hybrids are detectable after sequencing

Sequencing of the three libraries was performed using the MiniSeq platform, with the obtained hybrid counts by interaction type summarized in Figure 13. Interestingly, the WT IgG negative control yielded a significantly lower number of total hybrids, being approximately two orders of magnitude less than the total number of hybrids obtained from the other NOP56 samples. Moreover, there were no snoRNA-mRNA or snoRNA-rRNA hybrids found at all in the IgG sample, which was promising. Since the proximity ligation to produce chimeras only occurs when RNAs are closely held together by protein, this result seemed to indicate that the obtained snoRNA-mRNA hybrids from the other two NOP56 IP samples are real interactions of mRNAs specifically with the NOP56 protein. However, even after taking additional steps to eliminate non-specific reads, there were still very few snoRNA-mRNA interactions and considerably more mRNA-mRNA and rRNA-rRNA interactions in both the WT NOP56 and DKO12 NOP56 samples.

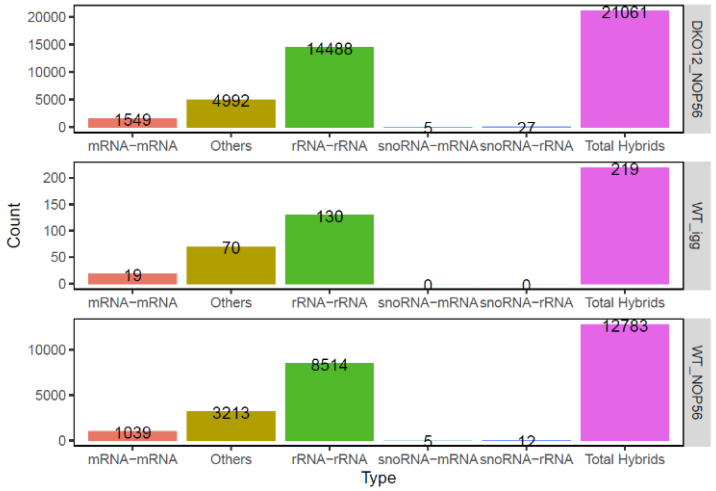


Figure 13. Summary of Interaction Types among All Obtained Hybrids after 2nd mCLASH Run

Transition to pCLASH

After obtaining a significantly lower number of snoRNA-mRNA hits than expected using mCLASH, a few hypotheses were developed as to why this could be occurring. One possibility was over-fragmentation, as the few snoRNAs that did appear as hits were highly abundant and significantly longer (>200 nt) than most snoRNAs, as typical box C/D snoRNAs have an average length of 80 nt[30]. Thus, it could be that the current

fragmentation conditions are over-digesting the smaller snoRNA fragments. Another hypothesis was that a significant number of snoRNA interactions are with pre-mRNA localized in the nucleus. Since it is known that snoRNAs are canonically found in the nucleolus and that 2'-O-methylations are guided to pre-rRNAs prior to processing and export, a similar phenomenon could be happening to pre-mRNAs. If so, the current workflow would be selecting against these interactions as poly(A) specifically enriches for mature mRNAs with polyadenylated tails, which would primarily be localized in the cytoplasm after export from the nucleus. These two hypotheses were tested by transitioning to pCLASH aimed at detecting snoRNA-pre-mRNA interactions. Instead of enriching for mRNA, nuclei were isolated from crosslinked cells to eliminate the bulk of rRNA within mature ribosomes in the cytoplasm, and the Ribo-Zero Kit (Illumina) was be employed to clear the majority of pre-rRNA in the nuclei. The fragmentation conditions were further adjusted by further diluting the RNase I (1:50 dilution instead of 1:25).

Nuclear Fractionation yields successful isolation of nuclei

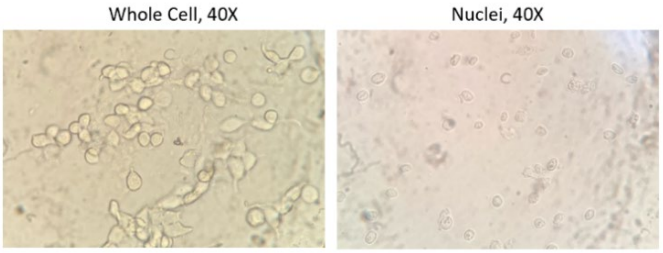


Figure 14. Comparison of Cell Morphology on Microscope for Whole Cells vs. Nuclear Fraction (40X Magnification)

The FR-iCLIP method of nuclear fractionation resulted in successful isolation of nuclei from our UV crosslinked HEK293T cells. As shown by the microscope image in Figure 14, the nuclei in the nuclear fraction looked significantly smaller than cells in the whole cell fraction. In addition, the nuclear pellet looked whiter by eye compared to a yellower whole cell fraction, which is a commonly observed result of successful nuclear isolation[31]. PTex of this nuclear fraction obtained from 40 X 150 mm plates of 293T cells yielded 168 µg of total PTex RNA. For reference, typical PTex of 40 plates

usually yields around 900 µg of total RNA. Thus, nuclear isolation resulted in approximately 19% of total yield, which is close to the 14% of RNA estimated to be contained in the nucleus compared to the total RNA within an entire cell[32].

rRNA Subtraction using Ribo-Zero is able to deplete pre-rRNA

After 2X RiboZero, 8.6 µg of RNA was yielded, translating to a ~10% yield from the original 88 µg of PTex nuclei. As shown in Figure 15, comparison of the PTex nuclei sample and both 1X and 2X RiboZero samples show marked decreases in the 28S and 18S bands compared to a 293T Total PTex sample. However, the Bioanalyzer was not able to characterize pre-rRNA quantity as its size of 13.3 kb was too large to be visualized.

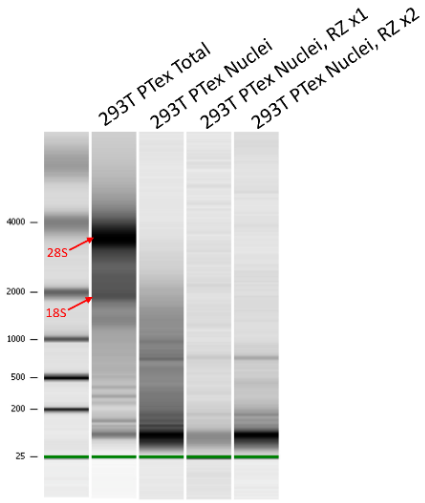


Figure 15. Bioanalyzer Gel of 293T Total PTex, 293T PTex Nuclei, and 293T PTex Nuclei after 1X and 2X RiboZero

Instead, qPCR was used to quantify pre-rRNA levels, with raw qPCR results of the four samples (PTex Total, PTex Nuclei, PTex Nuclei 1X RiboZero, PTex Nuclei 2X RiboZero) shown in Figure 16. The red lines indicate samples run using the 18S primer and the yellow lines are samples run using the 28S primer. The figure demonstrates an increase in approximately 1-2 CT cycles following nuclear isolation, and a significant increase in about 9-10 cycles after 1X RiboZero. The 2X RiboZero samples did not seem to show a significant change compared to the 1X RiboZero samples.

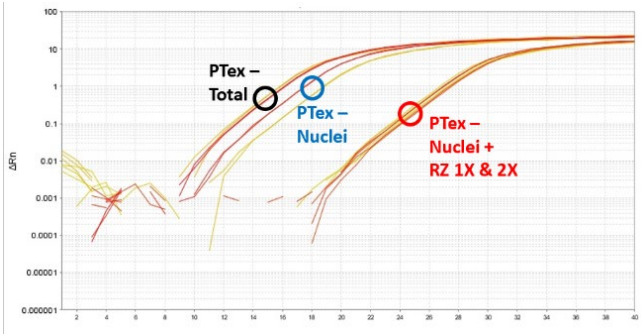


Figure 16. qPCR Result of 293T Total PTex, 293T PTex Nuclei, and 293T PTex Nuclei after 1X and 2X RiboZero

Translating the CT data into quantitative percentages through ΔΔCT and RQ calculations, there was approximately a 60-87% reduction in pre-rRNA after nuclear isolation, and >99.9% reduction after just 1X RiboZero as depicted in Figure 17. This indicated that 1X RiboZero was sufficient to deplete more than 99% of pre-rRNA in our PTex nuclei sample.

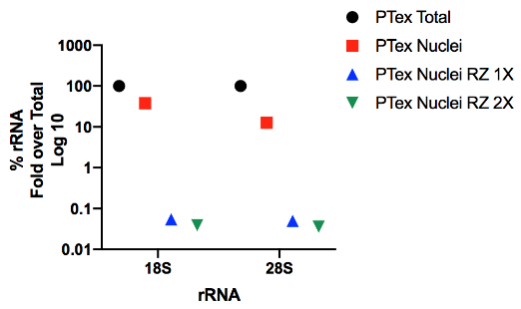


Figure 17. % rRNA depletion after 1X and 2X RiboZero

Successful Library Generation after pCLASH

10.8 µg of UV crosslinked, nuclei-isolated, PTex extracted, 2x RiboZero rRNA-depleted RNA-protein was put through the pCLASH workflow to generate libraries. 19 cycles of amplification were used to amplify the cDNA, and two rounds of gel purification were employed to remove primer dimers as well as empty adapter ligations from the library. The BioAnalyzer trace shown in Figure 18 demonstrates the high quality of the purified library, with a clear, single peak at ~280 nt without any other noticeable peaks indicating potential contaminating side products.

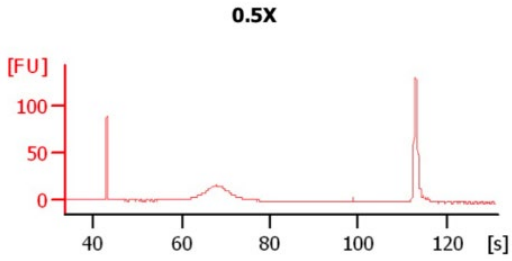


Figure 18. BioAnalyzer Quality Control Plot of Purified pCLASH Library

MiniSeq Sequencing reveals adequate library quality

The purified library was sequenced on the MiniSeq platform to ensure adequate library quality prior to sequencing on the NovaSeq. As shown in Figure 19, 451 total hybrids were obtained, 19 of which were snoRNA-mRNA hybrids. While this was still not a

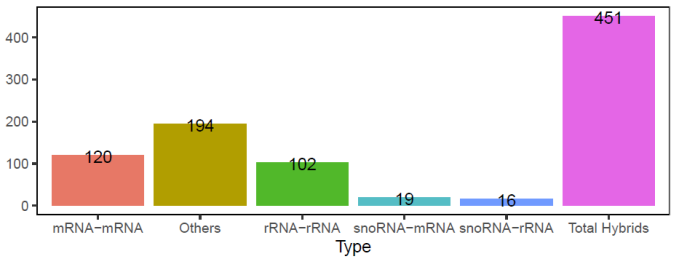


Figure 19. Hybrids Obtained from Illumina MiniSeq Sequencing Categorized by Interaction Type

considerable number of snoRNA-mRNA hybrids captured, it was still an improvement over the 5 snoRNA-mRNA hybrids obtained from mCLASH. Moreover, the ratio of snoRNA-mRNA hybrids to the total number of hybrids was much improved from mCLASH, raising optimism for the deeper sequencing run on the NovaSeq platform.

NovaSeq results of pCLASH returns significantly more snoRNA-mRNA hits

Sequencing the purified library using the Illumina NovaSeq 6000 yielded a total of 2960 hybrids, as shown in Figure 20. Interestingly, pCLASH captured considerably more snoRNAs within hybrids, with over 200 snoRNA hybrids compared to less than 20 in the mCLASH NOP56 run. This is even more remarkable considering the mCLASH run had substantially more total hybrids at over 12,000. Additionally, the number of rRNA-rRNA hybrids was significantly reduced (695) compared to the mCLASH run (8514), which attests to the success of nuclei isolation and RiboZero rRNA depletion. It is particularly noteworthy that 166 snoRNA-mRNA hybrids were obtained, which is a significant increase over the 5 snoRNA-mRNA hybrids obtained from mCLASH. Moreover, the ratio of snoRNA-mRNA to total reads is considerably higher with this pCLASH run, as evidenced by Figure 21. pCLASH has a snoRNA-mRNA/Total Reads ratio an order of magnitude higher than mCLASH, as well as several orders of magnitude greater than previously published CLASH methods.

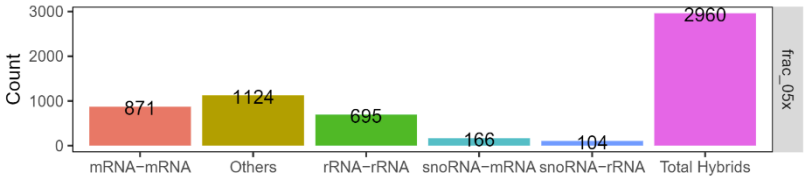


Figure 20. Hybrids Obtained from Illumina NovaSeq Sequencing Categorized by Interaction Type

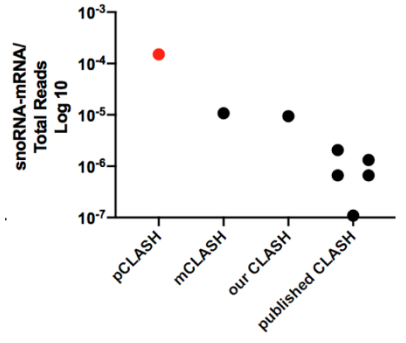


Figure 21. Comparison of snoRNA-mRNA Hybrids relative to Total Reads for various CLASH methods

pCLASH reveals mitochondrial targets of snoRNAs

Closer inspection of the obtained hybrids still revealed that the ribosomal RNA genes still accounted for the highest number of reads as shown in Figure 22, demonstrating the sheer quantity of rRNA and their interactions with NOP56 and snoRNAs. SNORD3A was the most common snoRNA found in hybrids, which is expected considering it is the most abundant box C/D snoRNA[33]. Interestingly, numerous mRNAs associated with snoRNAs were found to be encoded in mitochondrial genes, highlighted in yellow. This is particularly exciting as there has been a previously identified physiological link between snoRNAs and mitochondria, particularly with Rpl13a snoRNAs regulating ROS and oxidative stress[25]. The potential role of 2'-O-methylations mediated by snoRNAs on mitochondrial gene targets and the effect on mitochondrial function could be an exciting direction to take for future research.

	found in hyb	#reads
1	RNA28SN5	1982
2	SNORD3A-201	235
3	RNA18SN5	209
4	RNA5S1	206
5	MTRNR2L8-001	171
6	WDR74-001	158
7	ZNF354B-001	113
8	MTRNR2L12-001	75
9	MTRNR2L1-001	41
10	PTCH2-201	35
11	MT-CO1-201	28
12	MT-CO2-201	27
13	ST8SIA1-001	25
14	MT-ATP6-201	23
15	TMEM107-004	22
16	EEF2-001	22
17	AVIL-004	17
18	MT-CO3-201	17
19	PTMA-003	16
20	RPS6-001	15

Figure 22. Genes Obtained within Hybrids Sorted by Count

Discussion

pCLASH successfully yielded over 160 snoRNA-mRNA hybrids, with a significantly improved ratio of snoRNA-mRNA hybrids compared to the total number of obtained hybrids. The fact that numerous mitochondrial targets were obtained is particularly interesting, and several preliminary experiments have been conducted to begin validating some of these targets. For one, an RTL-P assay using cells with siRNA knockdowns of the Nm methyltransferase fibrillarin (FBL) has yielded promising preliminary results. This assay relies on quantitative PCR with reverse transcription at a lower concentration of dNTPs sensitive enough to detect Nm modifications on mRNA; early results indicate that FBL KD cells may have higher read-through when assaying for several mitochondrial genes identified in pCLASH, hinting at the presence of methylations. The RTL-P assay will be repeated using cells with a higher percentage of FBL KD for additional validation; this assay can also be attempted on cells with complete knockouts of several snoRNAs of interest to further validate the presence of Nm on specific mRNA transcripts.

In addition to RTL-P, the obtained mitochondrial targets will be cross-referenced using other methods currently employed in the lab to strengthen the case for a select few targets that can be explored in greater depth. Two such methods are computational methods that can bioinformatically predict putative target sites of snoRNAs: PLEXY and snoGlobe. Several of the mitochondrial hits have returned positive results when checked using PLEXY, particularly with the Rpl13a snoRNAs. Another is a biochemical method currently being developed by Dr. Yinzhou Zhu in the lab called RibOxi-Seq, a technique capable of detecting 2’-O-methylations on mRNAs based on differential

responses to oxidation depending on methylation status. Lastly, proteomics can be used to identify differentially expressed proteins either on cell lines knocked out for particular snoRNAs or mouse models with genetic knockouts of snoRNAs. If multiple lines of evidence point to the same mitochondrial gene target, this would provide strong evidence to further explore that gene and the mechanisms by which it may affect human physiology.

In terms of the functional effects of these potential modifications to mitochondrial targets, several mitochondrial assays could be employed to begin to answer this question. One method could be the Seahorse assay, where different chemicals are used to inhibit each mitochondrial complex to provide insight into the cause of mitochondrial dysfunction. The Seahorse assay could be performed on either genetically modified cell lines, or on isolated mitochondria from wild-type and snoRNA KO hearts from mice. Another functional assay could be immunoblotting for the mitochondrial complexes with an antibody cocktail, with each antibody targeting an individual complex. Both of these approaches could provide further insight into the specific aspect of mitochondrial function and oxidative phosphorylation that the modification could be influencing.

Our group has also explored improving this existing pCLASH method to be quicker and more efficient. We have attempted to conduct pCLASH at a larger scale to potentially obtain more hybrids and uncover additional mRNA targets; however, one recurrent issue was the inability to completely exclude empty adapter ligations (~150-170 bp) from the actual desired library (~200-400 bp). Even after multiple gel purifications, this undesired side product was extremely difficult to eliminate, with the

consecutive gel purifications resulting in significant loss of product. Thus, a novel technique that could be incorporated to the existing pCLASH workflow to circumvent this problem is the idea of template switching to generate the library. This method, summarized by the flowchart in Figure 23, relies on a special reverse transcriptase capable of switching the template it utilizes to transcribe the DNA at the 5' end of the RNA template, after it has synthesized the first strand[34]. This novel method of generating indexed cDNA libraries eliminates the need to add adapters at all, completely avoiding the formation of troublesome empty adapter side products. Preliminary experiments using this template switching method seem very promising, with much cleaner libraries generated in a much shorter period of time. This could be an exciting new alteration to our current pCLASH workflow that could significantly improve the efficiency and purity of the generated libraries.

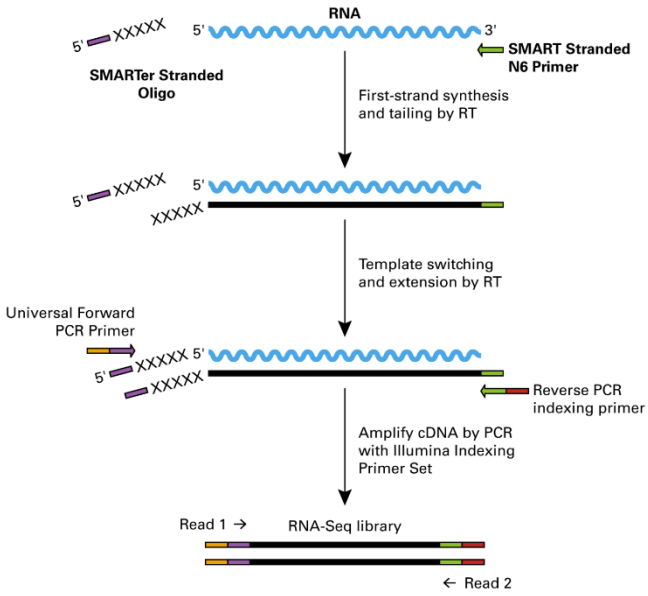


Figure 23. Flowchart of SMARTer Stranded RNA-Seq Kit Library Generation (Takara Bio)

Acknowledgements

I would like to thank Dr. Brittany Elliott for being such an amazing mentor and individual who cared not only about my laboratory skills but my well-being as a person. You always made me look forward to coming to lab every day, and I knew you would always be there to give me any help that I needed. Without your guidance and support, I would be nowhere near where I am today.

I am deeply grateful to Dr. Christopher Holley, who accepted me with open arms the summer after freshman year and has been a trusted advisor and mentor ever since. I greatly appreciate how you always made time to meet with me not only for research-related questions but also regarding other class projects, career plans, and life as a physician. I truly feel lucky that I had such a kind, caring PI during my time at Duke.

I would also like to thank fellow members of the Holley Lab including Dr. Yinzhou Zhu, Dr. Hala Abou Assi, Dr. Hsiang-Ting Ho, and Xinhe Yin for their valuable assistance and support in the lab.

References

1. Liang, J., Wen, J., Huang, Z., Chen, X.-, Zhang, B.-, & Chu, L. (2019). Small Nucleolar RNAs: Insight Into Their Function in Cancer. *Frontiers in Oncology*, *9*, 587.

2. Bratkovič, T. ž., Božič, J., & Rogelj, B. (2019). Functional diversity of small nucleolar RNAs. *Nucleic Acids Research*, *48*(4), 1627–1651.

3. Reichow SL, Hamma T, Ferre-D'Amare AR, Varani G. The structure and function of small nucleolar ribonucleoproteins. *Nucleic Acids Res*. (2007) 35:1452–64.

4. Elliott, B. A., Ho, H. T., Ranganathan, S. V., Vangaveti, S., Ilkayeva, O., Abou Assi, H., Choi, A. K., Agris, P. F., & Holley, C. L. (2019). Modification of messenger RNA by 2′-O-methylation regulates gene expression in vivo. *Nature Communications*, *10*(1), 3401.

5. Gumienny, R. *et al.* High-throughput identification of C/D box snoRNA targets with CLIP and RiboMeth-seq. *Nucleic Acids Res* **45**, 2341-2353 (2016).

6. Bieniasz, P. D., & Kutluay, S. B. (2018). CLIP-related methodologies and their application to retrovirology. *Retrovirology*, *15*(1), 35.

7. Van Nostrand, E. L., Pratt, G. A., Shishkin, A. A., Gelboin-Burkhart, C., Fang, M. Y., Sundararaman, B., Blue, S. M., Nguyen, T. B., Surka, C., Elkins, K., Stanton, R., Rigo, F., Guttman, M., & Yeo, G. W. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nature methods*, *13*(6), 508–514.

8. Helwak, A., Tollervey, D. Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH). *Nat Protoc* **9,** 711–728 (2014).

9. Lin, C., & Miles, W. O. (2019). Beyond CLIP: advances and opportunities to measure RBP–RNA and RNA–RNA interactions. *Nucleic Acids Research*, *47*(11), 5490–5501.

10. Gay, L. A., Sethuraman, S., Thomas, M., Turner, P. C. & Renne, R. Modified Cross-Linking, Ligation, and Sequencing of Hybrids (qCLASH) Identifies Kaposi's Sarcoma-Associated Herpesvirus MicroRNA Targets in Endothelial Cells. *J. Virol.* **92**, (2018).

11. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding. *Cell* **153**, 654–665 (2013).

12. Wu, J., Xiao, J., Zhang, Z., Wang, X., Hu, S., & Yu, J. (2014). Ribogenomics: the Science and Knowledge of RNA. *Genomics, Proteomics & Bioinformatics*, *12*(2), 57–63.

13. Dudnakova, T., Dunn-Davies, H., Peters, R., & Tollervey, D. (2018). Mapping targets for small nucleolar RNAs in yeast. *Wellcome Open Research*, *3*, 120.

14. Urdaneta, E. C., Vieira-Vieira, C. H., Hick, T., Wessels, H.-H., Figini, D., Moschall, R., Medenbach, J., Ohler, U., Granneman, S., Selbach, M., & Beckmann, B. M. (2019). Purification of cross-linked RNA-protein complexes by phenol-toluol extraction. *Nature Communications*, *10*(1), 990.

15. Zhao, S., Zhang, Y., Gamini, R., Zhang, B., & von Schack, D. (2018). Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: polyA+ selection versus rRNA depletion. *Scientific Reports*, *8*(1), 4781.

16. Zhang, C., Fu, J., & Zhou, Y. (2019). A Review in Research Progress Concerning m6A Methylation and Immunoregulation. *Frontiers in immunology*, *10*, 922.

17. Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y. G., & He, C. (2011). N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature chemical biology*, *7*(12), 885–887.

18. Zheng, G., Dahl, J. A., Niu, Y., Fedorcsak, P., Huang, C. M., Li, C. J., Vågbø, C. B., Shi, Y., Wang, W. L., Song, S. H., Lu, Z., Bosmans, R. P., Dai, Q., Hao, Y. J., Yang, X., Zhao, W. M., Tong, W. M., Wang, X. J., Bogdan, F., Furu, K., … He, C. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Molecular cell*, *49*(1), 18–29.

19. Batista, P. J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., Bouley, D. M., Lujan, E., Haddad, B., Daneshvar, K., Carter, A. C., Flynn, R. A., Zhou, C., Lim, K. S., Dedon, P., Wernig, M., Mullen, A. C., Xing, Y., Giallourakis, C. C., & Chang, H. Y. (2014). m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell stem cell*, *15*(6), 707–719.

20. Vu, L., Pickering, B., Cheng, Y. *et al.* The N^6-methyladenosine (m^6A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. *Nat Med* **23,** 1369–1376 (2017).

21. Barbieri, I., Tzelepis, K., Pandolfini, L. *et al.* Promoter-bound METTL3 maintains myeloid leukaemia by m^6A-dependent translation control. *Nature* **552,** 126–131 (2017).

22. Blundy, K. (2020, October 22). *Storm Therapeutics selects first-in-class clinical candidate targeting METTL3*. STORM Therapeutics.

23. Zhang, D., Zhou, J., Gao, J., Wu, R. Y., Huang, Y. L., Jin, Q. W., Chen, J. S., Tang, W. Z., & Yan, L. H. (2019). Targeting snoRNAs as an emerging method of therapeutic development for cancer. *American journal of cancer research*, *9*(8), 1504–1516.

24. Mourksi, N. E., Morin, C., Fenouil, T., Diaz, J. J., & Marcel, V. (2020). snoRNAs Offer Novel Insight and Promising Perspectives for Lung Cancer Understanding and Management. *Cells*, *9*(3), 541.

25. Michel, C. I., Holley, C. L., Scruggs, B. S., Sidhu, R., Brookheart, R. T., Listenberger, L. L., Behlke, M. A., Ory, D. S., & Schaffer, J. E. (2011). Small nucleolar

RNAs U32a, U33, and U35a are critical mediators of metabolic stress. *Cell metabolism*, *14*(1), 33–44.

26. Peffers, M.J., Chabronova, A., Balaskas, P. *et al.* SnoRNA signatures in cartilage ageing and osteoarthritis. *Sci Rep* **10,** 10641 (2020).

27. Bieth, E., Eddiry, S., Gaston, V. *et al.* Highly restricted deletion of the SNORD116 region is implicated in Prader–Willi Syndrome. *Eur J Hum Genet* **23,** 252–255 (2015).

28. Brugiolo, M., Botti, V., Liu, N., Müller-McNicoll, M., & Neugebauer, K. M. Fractionation iCLIP detects persistent SR protein binding to conserved, retained introns in chromatin, nucleoplasm and cytoplasm. *Nucleic Acids Research*, ***45***(18), 10452–10465 (2017).

29. Travis, A. J., Moody, J., Helwak, A., Tollervey, D., & Kudla, G. Hyb: A bioinformatics pipeline for the analysis of CLASH (crosslinking, ligation and sequencing of hybrids) data. *Methods*, ***65***(3), 263–273 (2014).

30. Marz, M., Gruber, A. R., Höner zu Siederdissen, C., Amman, F., Badelt, S., Bartschat, S., Bernhart, S. H., Beyer, W., Kehr, S., Lorenz, R., Tanzer, A., Yusuf, D., Tafer, H., Hofacker, I. L., & Stadler, P. F. Animal snoRNAs and scaRNAs with exceptional structures. *RNA Biology*, ***8***(6), 938–946 (2011).

31. Wilson, K. L., & Sonnenberg, A. (2016a). *Intermediate Filament Associated Proteins (Volume 569) (Methods in Enzymology, Volume 569)* (1st ed.). Academic Press.

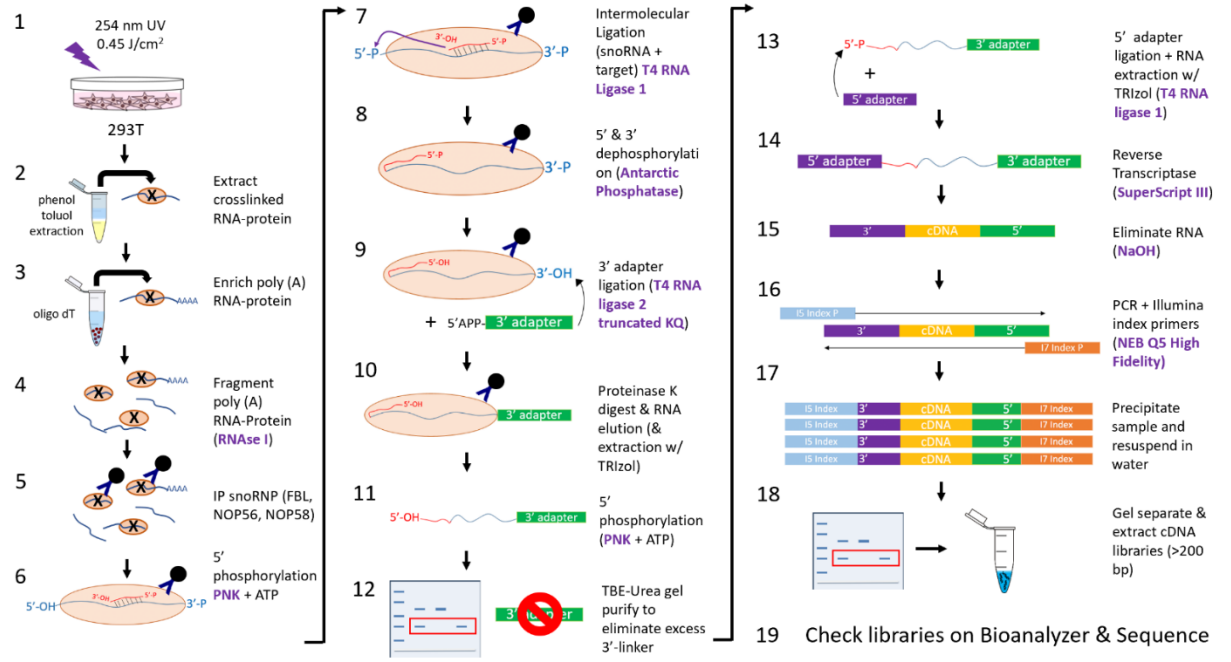
32. *RNA Content of a Typical Human Cell*. (n.d.). Gene Quantification. http://gene-quantification.org/what-is-RNA.pdf

33. Stamm, S., & Lodmell, J. S. (2019). C/D box snoRNAs in viral infections: RNA viruses use old dogs for new tricks. *Non-Coding RNA Research*, ***4***(2), 46–53.

34. Takara Bio USA, Inc. (2021, February). *SMARTer® Stranded RNA-Seq Kit User Manual*. Takara Bio.

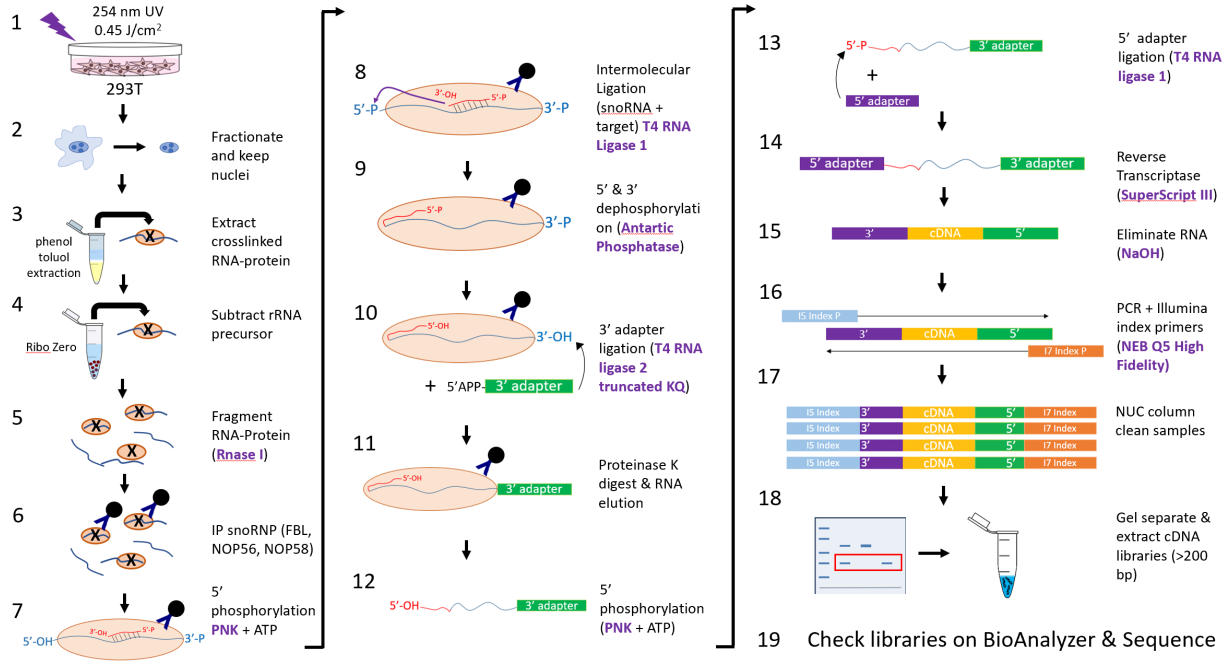
Appendix

Supplementary Figure 1. mCLASH Workflow Diagram



"mCLASH"

Supplementary Figure 2. pCLASH Workflow Diagram



"Precursor CLASH"

pCLASH - Version 4