UNIVERSITY OF CALIFORNIA

Los Angeles

Investigations on snoRNA processing in *Saccharomyces Cerevisiae*

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Biochemistry, Molecular and Structural Biology

by

Samuel Martin DeMario

2023

ABSTRACT OF THE DISSERTATION

Investigations on snoRNA processing in *Saccharomyces Cerevisiae*

by

Samuel Martin DeMario

Doctor of Philosophy in Biochemistry, Molecular and Structural Biology

University of California, Los Angeles, 2023

Professor Guillaume Chanfreau, Chair

snoRNAs are small non-coding RNAs which guide post-transcriptional modification of cellular RNAs. Here we present work investigating the processing of these critical RNAs. Many snoRNAs are encoded within the introns of coding genes. We show that in the absence of splicing, highly stable aberrant transcripts are generated. During the course of this work, we employed long-read RNA sequencing (ONT). ONT seq is an exciting new tool for studying RNA isoforms but visualization tools are scarce. Here we present our long-read sequencing visualization tool, NanoBlot and suggest future development directions. Finally, The TRAMP complex is a trimeric protein complex with semi-modular protein subunits. It is responsible for degradation and processing of a number of RNAs including excised introns and snoRNAs. We present work investigating the substrate specificity of 2 TRAMP confirmations.

The dissertation of Samuel Martin DeMario is approved.

Margot Elizabeth Quinlan

Roy Wollman

Xinshu Xiao

Guillaume Chanfreau, Committee Chair

University of California, Los Angeles

2023

To Marylin DeMario, the original Dr. DeMario.

Thesis Structure

The first two chapters of this dissertation are primarily composed of works which have been previously published. The sections where a previous publication is to be included have a citation and a DOI number. At those points it is suggested you read the indicated publication. A brief summary of each chapter is provided below:

- Chapter 1 gives general background knowledge relevant to multiple sections of this thesis. Some of the information is redundant with background provided in published works.
- Chapter 2 is a published paper on a novel class of hybrid-mRNA/snoRNAs. (Liu et al. 2022a)
- Chapter 3 is a published paper on a novel R-package called "Nanoblot". (DeMario et al. 2023)
- Chapter 4 is unpublished work focused on the substrate specificity of the TRAMP complex.
- Chapter 5 is a concluding chapter.

Scientific Environment

All research was conducted in the laboratory of Guillaume Chanfreau at the University of California, Los Angeles. The research was funded by NIGMS; Grant R35 GM130370 to G.F.C. S.M.D. was partially funded by the Cellular and Molecular Biology Training Grant (Ruth L. Kirschstein National Research Service Award GM007185).

Acknowledgement

I'd like to thank my undergraduate mentees who contributed significantly to these works: Kevin Xu, Erica Li, Indya Weathers, Yanru Liu, Isabel Breuer, Nanilei Nawahine, and Michael Pung. I'd also like to thank Charles Wang who initially mentored me in the Chanfreau lab. Finally, I'd like to thank my advisor Guillaume Chanfreau; I struggle to imagine how one could have been a better mentor.

Vita

Samuel DeMario was born in Georgia. In 2014 he graduated from West Hall High School. He then attended the University of Georgia where he worked in the laboratories of David Blum and David Hall respectively. Despite getting a ‘C’ in Biochemistry 2, Sam received his Bachelor of Science in Biochemistry in 2018. He began studying at the University of California, Los Angeles in 2018. As of 2023, he still cannot spell ‘*Saccharomyces Cerevisiae’* without looking it up.

1: Introduction

Splicing

Some Eukaryotic mRNAs contain introns which are removed during mRNA maturation. Most introns require the spliceosome a large multisubunit complex with both RNAs and proteins. The majority of eukaryotic splicing occurs through 2 chemical steps. In step 1 the 5'exon is removed and the intron lariat is formed. In step 2 the 5'exon is ligated to the 3'exon and the intron lariat is removed. The lariat is then de-brached by the debranching complex(Mohanta and Chakrabarti 2021).

The spliceosome is a large ribonucleoprotein complex which catalyzes the splicing reaction. It is composed of more than 100 different proteins and 5 non-coding RNAs. Not all of these proteins are essential; some increase splicing efficiency while others control splicing fidelity(Will and Lührmann 2011).

Nuclear Exosome

The nuclear exosome is a large protein complex with a variety of important biological functions. It is highly conserved in eukaryotes and is involved in ribosomal RNA processing, snoRNA processing, gene regulation and RNA degradation(Allmang et al. 1999). It forms a barrel structure and has two distinct exonucleases. At the "top" of the barrel is RRP6 which has low processivity and at the "bottom" is DIS3 which is highly processive. In order to reach DIS3 transcripts must pass through the barrel of the nuclear exosome (Ogami and Suzuki 2021).

snoRNAs

snoRNAs (small nucleolar RNAs) are noncoding RNA which form into RNPs and guide modification of other cellular RNAs. They guide modification by sequence specific recognition of the target sequence by base pairing with the target sequence. snoRNAs primarily guide modification of pre-ribosomal RNAs but they have also been shown to have regulatory functions (Huang et al. 2017, 2022). They are divided into 2 main classes, C/D box and H/ACA box. C/D box snoRNAs guide 2'OH methylation while H/ACA guide pseudouridylation. These two classes are characterized by conserved secondary structures as well as conserved sequences (C, C', H, H', etc.). Each class assembles with a unique set of proteins in order to form a mature RNP. Mature RNPs have two complete sets of the snoRNP proteins and have two separate guide sequences which allows each snoRNP to modify two specific transcriptomic loci.

snoRNA Processing

snoRNAs are genomically encoded in three ways:

1. Independently transcribed
2. Transcribed as part of a polycistronic cluster
3. Intron encoded

The proportion of snoRNAs encoded in each of these ways varies significantly between organisms. In Humans the majority of snoRNAs are intron encoded(Kishore et al. 2013). In drosophila the majority of snoRNAs are polycistronically encoded. In yeast the majority of snoRNAs are independently transcribed.

The snoRNAs are processed at the 3'end by the nuclear exosome with the help of the TRAMP complex(Perumal and Reddy 2002). This happens through repeated rounds of oligoA addition and then degradation. Transcriptional termination of independently transcribed snoRNAs occurs via the Nab3-Nrd1-Sen1 (NNS) termination pathway which helps to recruit

TRAMP and the nuclear exosome (Chaves-Arquero and Pérez-Cañadillas 2023; Steinmetz et al. 2001; Porrua et al. 2012). The 5' ends of snoRNAs are processed via the 5' -> 3' exonucleases Xrn1p and Rat1p. Xrn1p and Rat1p do not have decapping activity, this means that the 5' ends of snoRNAs need to undergo endonucleolytic cleavage prior to 5'end processing; this can occur through either splicing or by the endonuclease Rnt1p (Lee et al. 2003).

snoRNP protein assembly is a complex, multistep process. The protein components initially assemble independently of the snoRNA and are loaded onto the snoRNA with the help of chaperone proteins. The assembly happens both in the cytoplasm and the nucleoplasm. While the assembly of the snoRNA with the snoRNP proteins is relatively well known the interplay between transcription and RNA processing with protein assembly machinery is largely unknown(Massenet et al. 2017).

RNA Degradation: NMD/NGD

mRNAs have a relatively short biological half-life, with some estimates suggesting a median half-life of 11 minutes(Marín-Navarro et al. 2011; Miller et al. 2011). This is due in part to the existence of specialized mRNA decay pathways. Two important pathways are the nonsense-mediated decay pathway (NMD) and the no-go-decay (NGD) pathway.

In the NMD pathway transcripts with premature termination codons are recognized and targeted for degradation(Peccarelli and Kebaara 2014). This includes unspliced pre-mRNAs which have erroneously begun translation, transcripts with a frameshift mutation, transcripts with point mutations that create stop codons, as well as some natural targets which are regulated by NMD(Guan et al. 2006). In S. cerevisiae, the NMD pathway requires 3 proteins which are conserved across eukaryotes: UPF1p, UPF2p, and UPF3p(He et al. 1997). In higher order eukaryotes, such as *C. elegans, and H. sapiens,* additional protein factors are required. Once a

premature stop codon is recognized it is decaped and degraded via the decapping complex and exonuclease Xrn1p(Cao and Parker 2003). The mechanism for how premature stop codons are differentiated from canonical stop codons is not fully understood. The most prevalent hypothesis is the *faux*-UTR model, in which the proximity of other proteins such as the polyA binding protein or the exon-junction complex in humans prevents targeting by NMD(Amrani et al. 2004).

In the NGD pathway ribosomes become stalled on transcript (Powers et al. 2020). This eventually leads to a ribosome collision event which creates a unique disome structure which can then be recognized by other proteins (Juszkiewicz et al. 2018; Ikeuchi et al. 2019). These collisions trigger ubiquitination of ribosomes and activate the ribosome quality control (RQC) pathway which degrades the stalled ribosomes(Juszkiewicz and Hegde 2017; Garzia et al. 2017). Recently the endonuclease Cue2p was shown to cleave RNAs at the site of ribosome collisions (D'Orazio et al. 2019). The ribosomes are then removed from the RNA by Dom34p and Hbs1p and degraded by the exosome and Xrn1p(Atkinson et al. 2008).

Next-generation sequencing approaches

Next generation sequencing (NGS) is a group of biochemistry techniques that allow for the determination of the primary sequence of a nucleic acid. NGS has high sequencing quality with errors occurring less than 0.1% of the time(Buermans and den Dunnen 2014).

In order to perform NGS the sample must first be turned into a library. There exist many ways to construct a library with each giving slightly different information(Shendure et al. 2017). There are 3 types of library prep relevant to this thesis: total RNA, mRNA and 3'-end seq.

In total RNA library preps the RNA is converted into cDNA using random hexamer primers, this results in all of the input RNAs being sequenced. The primary drawback to this technique is that there exist many highly abundant non-coding RNAs which will take up a substantial amount of the sequenced reads.

In mRNA-seq one endeavors to sequence only mRNAs and avoid sequencing the highly abundant non-coding RNAs. This can be done in a few different ways: The library can be reverse transcribed from RNA to DNA using an oligo(dT) primer which preferentially transcribes the polyA tails of mRNAs(Moll et al. 2014). Alternatively, mRNAs can be physically purified from a sample using oligo(dT) beads(Zhao et al. 2018).

Finally, in 3'end sequencing the goal is to find the 3'-ends of nucleic acid samples. This can be done by using a polyA polymerase to in-vitro polyadenylation a sample prior to sequencing using a polyT sequencing primer. This has been shown to effectively sequence short oligoA tails added by the TRAMP complex(Roy and Chanfreau 2020).

Nanopore sequencing

Nanopore sequencing (ONT) is a DNA sequencing technique which gives near full length transcripts. In short, a primer pre-assembled with a motor protein is ligated onto the end of target nucleic acids. Then, the sample is loaded onto a flowcell with nanopores suspended in a lipid membrane. The motor protein pushes the nucleic acid through the nanopore and the electrical resistance across the pore is continuously measured. The change in resistance is used to determine the primary sequence with ~80% accuracy, this is significantly lower fidelity than NGS but, due to the length of ONT reads (>1kb) it is typically possible to unambiguously align them to a transcriptome(Sahlin and Medvedev 2021).

ONT has many notable benefits over next generation sequencing but the most relevant to this work is its ability to unambiguously identify RNA isoforms(Oikonomopoulos et al. 2020; Zhao et al. 2019). Due to the longer reads afforded by ONT sequencing isoform usage can be quantitatively measured. However, visualization tools for ONT sequencing data are not well developed and have primarily focused on alternative splice isoform usage(De Paoli-Iseppi et al. 2021).

2: Aberrant snoRNA isoforms created from intron encoded snoRNAs

2.2: Spliceosome Mediated Decay Introduction

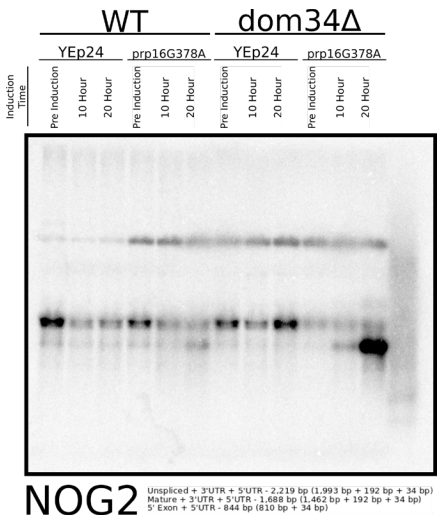
Spliceosome mediated decay (SMD) occurs when splicing is only partially complete before the spliceosome disassembles. In this case the mRNA is degraded likely through exonucleolytic cleavage via the nuclear exosome. To date only a few transcripts have been confirmed as SMD targets (Volanakis et al. 2013; Stepankiw et al. 2015). In order to study SMD we proposed to inhibit the second step of splicing in order to accumulate post-step one splicing products.

2.3: *Prp16* is an essential splicing factor with dominant negative mutants

Prp16 is an essential splicing factor which is required for the completion of the second chemical step but not the first. It is ATP-dependent. Two dominant dominant negative mutants, G378A and K379A, have been previously described (Schneider et al. 2002). The expression of these dominant negative mutants stalls the spliceosome in the post step one state trapping the target RNA.

The Chanfreau lab created an inducible plasmid form of Prp16dn. Prp16dn was under the control of the gal promoter. Cells were initially grown in YPRaffinose then shifted to YPGalactose for either 10 or 20 hours. This experiment was also conducted in *DOM34* knockouts, This was intended to stabilize the 5' exons which would typically be degraded by non-stop decay.

However, the induction of the dominant negative mutants resulted in the accumulation of a larger species than expected. This led us to investigate the extended species further.



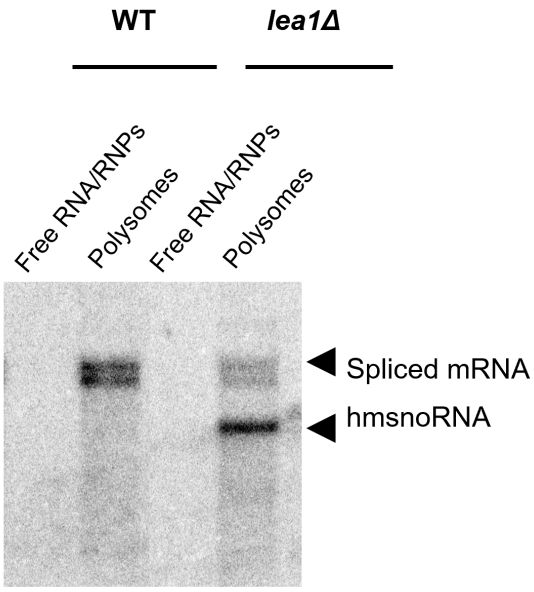
1.6: PNAS paper

The following section is an article previously published in PNAS on the hmsnoRNAs (Liu et al. 2022b).

1.7: Additional Experiments

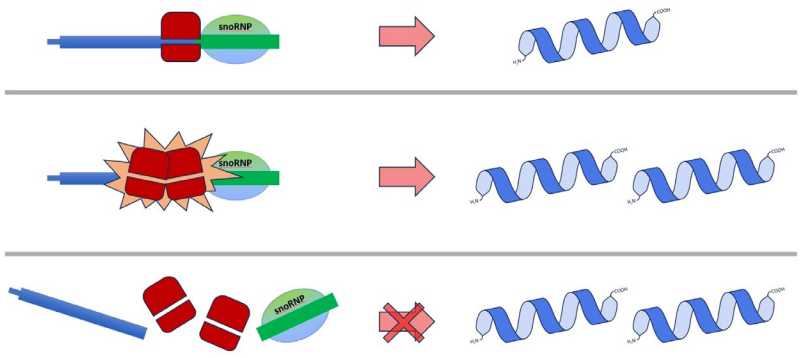
The hmsnoRNAs encode a reading frame and have a stop codon shortly into the intronic sequence. We became interested in the possibility that the hmsnoRNAs may be translated. In theory the hmsnoRNAs could create a truncated protein which would act as a regulatory peptide. To investigate this we conducted polysome fractionation to see if hmsnoRNAs fractionated with translating ribosomes. Fractionation was done on both WT and splicing mutant (lea1delta) cells. RNA was then isolated from each fraction and we performed Northern blots targeting the first exon of NOG2. We found that the hmosnoRNAs do fractionate with the

polysomes, suggesting they are potentially translated.

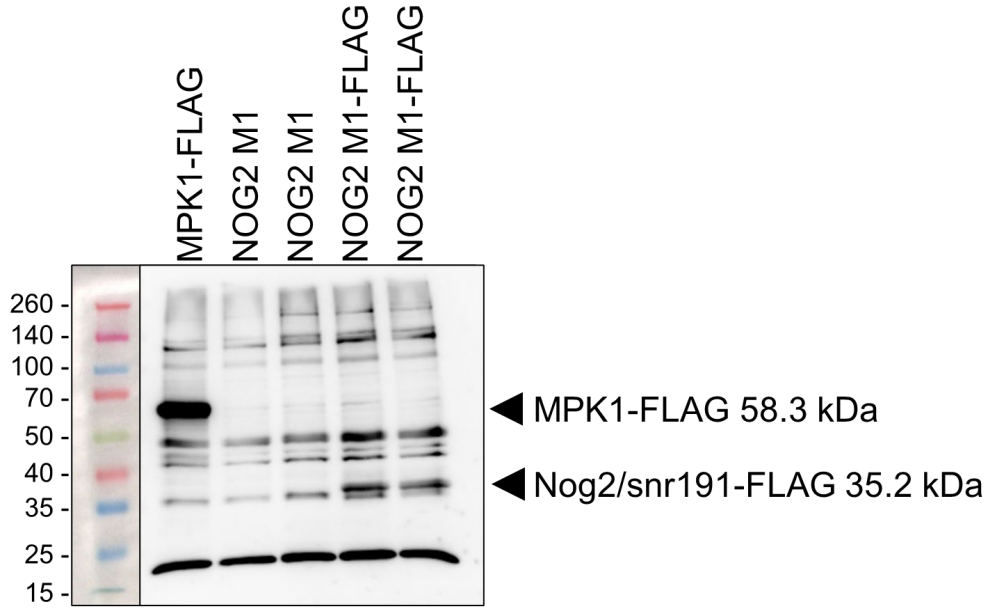


Northern Blot – Probe for Nog2 Exon1

The possibility of translating hmsnoRNAs was interesting to us as a synthetic biology tool. Transcripts with ribosomes collided on them are targeted for degradation by the No-Go-Decay pathway. Because the ribosome is unable to remove the snoRNAs from the hmsnoRNAs if the hmsnoRNA does not have a stop codon they become stalled and quickly degraded after the second ribosome collides with the first. This allows for tighter control of the protein produced by a single RNA transcript. We purpose this as a tool to control protein expression in mRNA based therapeutics.

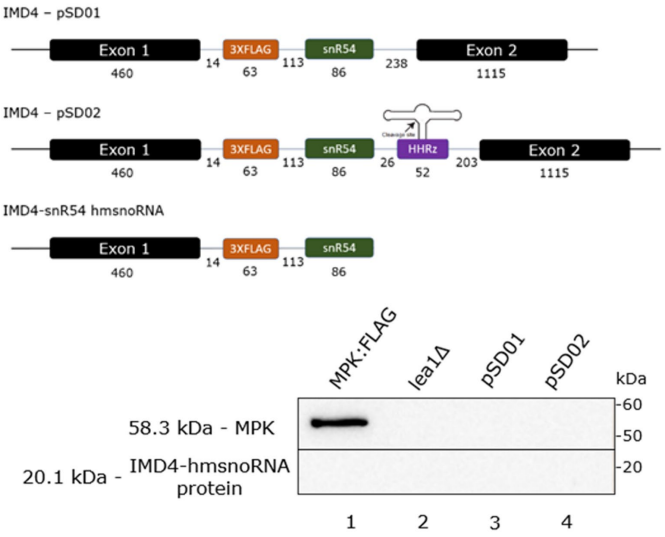


To test the possibility of hmsnoRNAs being translated we created a tagged version of the hmsnoRNA peptide. We used the M1 version of the NOG2/snR191 hmsnoRNA and added a 3XFLAG-tag to the end of the open reading frame.This resulted in the production of a FLAG-tagged peptide at the expected size. It is a possibility that the peptide is the result of translation of the "unspliced" isoform which still contains the intron downstream of the snoRNA, the 3' exon and the 3'UTR. To test this we have constructed a NOG2/hmsnoRNA construct with a hammerhead ribozyme sequence downstream of the snoRNA. At time of writing the construct has not been tested.



In addition to NOG2/snR191 hmsnoRNA we also created similar constructs for IMD4/snR54. The 5'ss were mutated resulting in the pSD01 plasmid. A hammerhead ribozyme (HHRz) was inserted in the intron downstream of snR54. We then transformed these constructs into yeast and performed a western blots against the FLAG peptide. Interestingly, none of the IMD4/snR54 hmsnoRNA constructs resulted in FLAG-tagged peptide. This is likely not due to lower transcription or unspliced RNA stability because ONT sequencing indicated that the level of IMD4 unspliced, spliced and hmsnoRNA were higher than in NOG2. This leads us to believe

that the class of the 3' snoRNA changes the translational efficiency of the hmsnoRNAs. snR191 is H/ACA and snR54 is C/D.



1.8: Future Directions: Biological Half Life

One possible future direction is to test the biological half-life and synthesis rates of the hmsnoRNA. We have consistently shown that in splicing mutants hmsnoRNAs accumulate to levels above the mature mRNA. However, we have never confirmed whether this is because of increased synthesis rate or increased biological halflife. One simple way to do this is using 4sU-seq. In 4sU-seq a modified nucleotide is added to the growth media and incorporated into newly synthesized transcripts. The modified nucleotide can then be biotinylated and purified allowing for separation of newly synthesized transcripts from transcripts synthesized before the spike in. By collecting a few timepoints an accurate synthesis rate and degradation rate can be calculated for each species(Boileau et al. 2021).

This is one notable consideration for using this technique on hmsnoRNAs. The readout for 4sU-seq is typically next-generation sequencing. However, as previously noted, next-

generation sequencing is unable to unambiguously detect hmsnoRNAs. There are a few possible strategies to circumvent this limitation. In-vitro poly-adenylation followed by 3'end sequencing could be used to investigate some of the hmsnoRNAs. In this case the 3'end read would be largely useless in determining whether a transcripts in a snoRNA or an hmsnoRNA but the 5' read would likely occur in the 5'exon allowing for unambiguous determination.

An alternative approach would be to use nano-pore sequencing. However, there are some considerations. First, multiple enzymatic steps are required to prepare the RNA for sequencing meaning that the results will be less quantitative. Second, the modification and biotin addition may prevent the transcript from going through the nanopore. There do exist commercial kits which use a reverse transcriptase to misincorporate guanidine across from modified uracil. This results in A U->C mismatch in the RT-products which can be read via next-gen sequencing. These transcripts should have no problem being read through the nanopore. However, this does not address the multiple enzymatic steps problem.

A final alternative approach is to use an alternative method such as a Northern blot to measure the half lives. In 4sU-seq the modified nucleotides are physically separated from the unmodified nucleotides. As a consequence the samples can be run independently on an agarose gel and individual Northern blots can be performed for each hmsnoRNA. This has the advantage of not requiring any enzymatic steps but also requires significantly more RNA to be collected. Additionally, each hmsnoRNA would need to be tested individually to greatly increase the time required to do the experiment. It is likely that each of the hmsnoRNAs will have different halflives and synthesis rates.

1.9: Future Directions: Genetically encoding internal polyA tail

Binding of PABPCp to the polyA tail of mRNAs has been shown to be important for efficient translation(Passmore and Coller 2022). In order to increase translational efficiency we

are interested in adding internal poly-A tails to our synthetic constructs. We plan to add oligo-A stretches to the intronic region upstream of the snoRNA to allow PABPCp to bind to the transcripts.

1.10: Future Directions: Engineering reporter system for hmsnoRNA translation

To facilitate further study of hmsnoRNA translation we propose the construction of a reporter system. This is an engineering challenge because translation of the unspliced form will always result in the creation of an identical peptide to the hmsnoRNA form. A dual reporter system could be employed to differentiate between the two; a second reporter gene can be included downstream of the first with an internal ribosome entry site. In this way there is an included control for translation of the unspliced form. As an example a GFP can be encoded in the 5'exon upstream of the snoRNA and a BFP can be encoded in a separate ORF downstream of the snoRNA. The ratio of GFP fluorescence to BFP fluorescence can be used to approximate the translation of the hmsnoRNA vs the translation of the unspliced form.

1.11: Future Directions: Study of hmsnoRNAs in other organisms

Finally, we propose to do further studies of hmsnoRNA in other organisms. All of the work we have done on hmsnoRNAs has been done in *Saccharomyces cerevisiae*. Since we have developed a protocol for sequencing hmsnoRNAs, study of other organisms should be relatively easy. When doing the sequencing on human cells we may need deeper sequencing depth because of the increased number of genes as well as the increased length of the average transcript. We may need to optimize the protocol in order to get more reads to relevant transcripts.

2: NanoBlot

2.1: Isoform visualization is important

RNA isoforms are versions of a specific genomic loci that vary in some structural characteristics. They may have differences in the transcriptional start sites, transcriptional termination sites, poly A tail length, or splicing. These variations can have profound effects on the functionality of the RNA. They can affect the primary structure of the protein product, change the translational efficiency of the RNA, change the localization of the RNA or change the stability of the RNA. Genes can have many isoforms and the abundance of each of these isoforms can change in different conditions. Changes in 3'UTR isoforms as well as alternatively spliced isoforms are common disease pathologies (Romo et al. 2017; Mills and Janitz 2012).

Isoforms can be called with a variety of softwares but that requires having specific previously annotated isoforms(Annaldasula et al. 2021). If your isoforms of interest are not discrete or if there exist many isoforms which need to be shown, visualization is challenging. To this end we developed an R-package based around ggplot for visualization of RNA isoforms without the need for discrete isoform annotations.

2.2: Visualizing Oxford nanopore sequencing data is inherently challenging

Visualizing Oxford Nanopore (ONT) sequencing data poses a formidable challenge due to its high information density. Similar to conventional sequencing techniques, ONT yields valuable insights into relative gene abundance but additionally it provides a myriad of additional information. ONT sequencing can reveal isoforms abundance as well as reveal new isoforms.

Furthermore, it unveils crucial details, including polyA tail lengths, transcriptional start and stop sites, as well as post-transcriptional modification data. This wealth of information presents a unique dilemma in the realm of data visualization, as there exists no universally applicable solution to create intuitive representations. In this chapter, we delve into the complexities of Oxford Nanopore sequencing data visualization, highlighting the need for tailored approaches to harness its full potential as well as introducing our software package for isoform visualization.

2.3: RNA Paper

The following section is an article previously published in RNA on NanoBlot (DeMario et al. 2023).

2.4: Future Development: Additional probe options

ONT data is incredibly information rich, in addition to full transcript lengths, it can be used to detect post transcriptional modifications as well as poly A tail length(Krawczyk et al. 2024). It would be a useful feature to be able to "probe" specific transcripts for specific modifications. This is in theory possible but there are some logistical limitations.

As ONT is still a relatively new technology, standards for how information should be encoded have not yet been widely adopted. In its current state NanoBlot only interacts with the BAM format, which has been a standard in next-generation sequencing for years. Information can be added to individual transcripts in a BAM file via the usage of arbitrary flags on an individual transcript.

While this strategy of adding tags to the BAM file could in theory be done in R the processing time and required resources would be considerable. It would be faster and less resource intensive to use C or one of the currently available tools for BAM manipulation (e.g. samtools).

2.5: Future Development: Bioconductor submission

Bioconductor is a software ecosystem that has become a standard in the bioinformatics field. It is primarily composed of R-packages although it does allow non-R code and usage of locally installed software. The current version of NanoBlot is mostly in compliance with Bioconductor standards with a few notable exceptions.

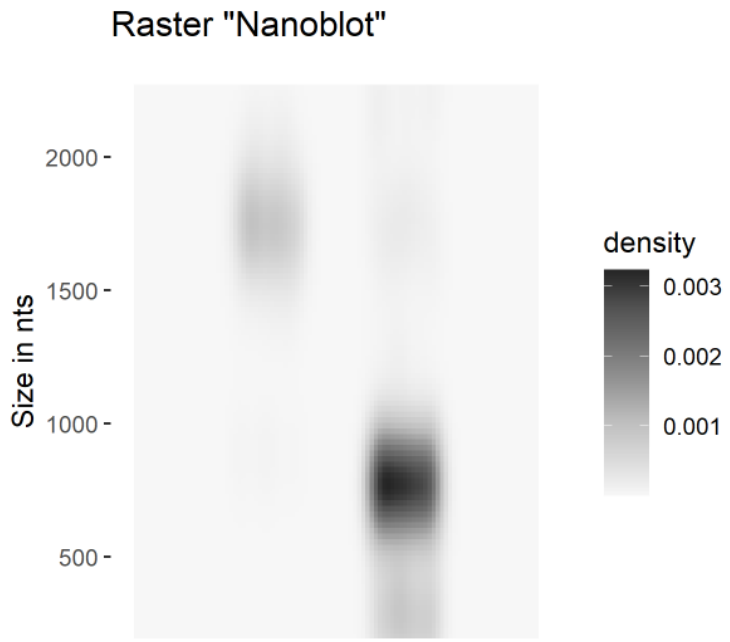
Packages on bioconductor are highly encouraged to use established formats for their data if an acceptable standard already exists. In the case of NanoBlots one relevant format is the gRange object, which specifies genomic regions. In its current state NanoBlot requires a user provided BED file. The rationale for this is that BED files are common and easy to edit/append data to via the use of common softwares (i.e. excel/google sheets). Internally, NanoBlot converts these BED files into gRange objects. Bioconductor's review will likely suggest that we use gRange objects directly.

The current implementation of NanoBlot's RT-PCR and RACE features uses bedtools to trim the reads based on a specific set of genomic coordinates. Bioconductor would likely suggest the removal of the bedtools requirement. At the time of writing the author is unaware of any Bioconductor packages which have an acceptable substitute for trimming of BAM files. An R-based implementation of this feature is theoretically possible but due to the potential size of the subsetted BAM file inputs and R's relatively inefficient string processing it is a suboptimal solution. Bioconductor allows C++ code via the "Rcpp" package. Due to C++ high efficiency in string processing I think the addition of a C++ function for trimming BAM reads is the best solution.

2.6: Future Development: Raster scanning

Early versions of NanoBlot had the option to output a NanoBlot as a raster scan (Figure). This feature was removed from the final release for two reasons. First, an unknown bug would occasionally cause a fatal error. The error particularly frequently occured when producing plots with a low number of samples or less reads overall. It could be circumvented by duplicating the input data.

The second reason that the raster scan option was removed was because the resulting blots were too close to real Northern blot images. While we did want interpretation of NanoBlots to be similar to Northern blots



Size estimation

3: TRAMP complex

3.1: Overview of the TRAMP complex

The TRAMP complex is a protein composed of 3 subunits MTR4p, TRF4/5p and AIR1/2p. It has been proposed that TRAMP is required to unwind secondary structures during RNA degradation(Jia et al. 2012). MTR4p is an ATP dependent RNA helicase which associates with the nuclear exosome and increases its activity. It has been shown to have a slight substrate preference for polyA regions (Jia et al. 2011). TRF4p and TRF5p are RNA polymerases which add short oligoA tails to the ends of transcripts. The role of Air1p and Air2p are not well understood. It has been shown that Air1p and Air2p show different relative affinities for Trf4 and Trf5. In biological isolates Air1p has been in complex with both Trf4p and Trf5p while Air2p shows a notable bias toward Trf4p (Delan-Forino et al. 2020). In order to investigate the role of the Air proteins in TRAMP complex activity we generated deletions of Air1 and Air2 as well as double mutants with Rrp6 deleted. We also constructed a triple mutant where all 3 genes were inactivated. 2 sequencing approaches were used.

3.2: Sequencing approaches

2 sequencing approaches were used to investigate TRAMP complex activity. The first approach was to use 3'end-sequencing(Fahmi et al. 2022). In 3'end-sequencing a NV(poly-T) primer is used during the reverse transcription step creating a library where one of the reads starts at the first nucleotide after the polyA tail.

There are a few important limitations of 3'end sequencing. First, internal polyA rich regions are often erroneously called as true 3'ends. If a particular genomic region has a high concentration of A and is abundantly transcribed the polyT primer can hybridize and allow

reverse transcription. This can be partially mitigated by flagging any reads that are immediately upstream of genomically rich A sequences. However, this is not a perfect solution as true polyA sites in yeast often follow a "YAAA" sequence leading to false negatives(Heidmann et al. 1992).

A second consideration is that the exact polyA site cannot be determined if the polyA site is A rich. This is an inherent problem in studying polyA sites and is impossible to differentiate. However, this problem frequently comes up in the analysis of the TRAMP mutants. Because the TRAMP complex is semi-processive, processing intermediates have oligo-A tails along a wide distribution. This increases the likelihood that the oligo A sequence occurs next to a genomically encoded A region causing ambiguity.

We also employed Nanopore sequencing as an alternative approach. We used two library preparation strategies, direct-RNA sequencing and cDNA sequencing. Direct-RNA sequencing allows us to potentially learn about post-translational modifications of the TRAMP targets. Using the cDNA library allowed us to multiple the samples. Both kits are based around polyA split-end-ligation. Because of this, they are both able to make estimates of the polyA tail length. It is with noting that the sequencing depth of ONT sequencing is much lower than 3'end sequencing making investigation of low-abundance TRAMP targets impossible with Nanopore sequence.

3.3: Genome-wide Analysis of poly(A) peaks detected in the rrp6 mutant and Air1/Air2 dependencies.

We previously analyzed poly(A) peaks accumulating in rrp6delta mutants, or in strains in which the exosome has been inactivated using the anchor away strategy (Roy et al. Genome Research 2016). Poly(A) peaks that accumulate specifically in the absence of Rrp6p or of nuclear exosome activity are indicative of sites of exosome activity which were previously polyadenylated by TRAMP. In order to analyze which poly(A) peaks are promoted by the activity

of Air1p or Air2p, we generated double mutants lacking Rrp6p and Air1p, or Rrp6p and Air2p. In addition, to determine which pA peaks are specifically promoted by TRAMP activity, we generated a triple mutant lacking Rrp6p, Air1p and Air2p. We generated poly(A)-based 3'-end sequencing data sets from the rrp6Δ, rrp6Δair1Δ, rrp6Δair2Δ, and rrp6Δair1Δair2Δ mutants and analyzed the results.

In agreement with previous research, deletion of AIR1p had a minor effect on the polyadenylation landscape of the cells, suggesting that its function can be mostly fulfilled by Air2p (Fig 1A). Deletion of AIR2 showed a more significant effect on poly-adenylation (avg. fold change vs rrp6: 1.24 +- 0.88). Removing both AIR1 and AIR2 from RRP6 deletion strains reveals a major change in the polyadenylation landscape of the cells. The majority of the points showed an increase in relative polyadenylation (3,289/5,210) with an average fold change of 1.44 +- 1.45. It is possible that the general stabilization we see is due to loss of some highly poly-adenylated species (snRNAs, rRNA)

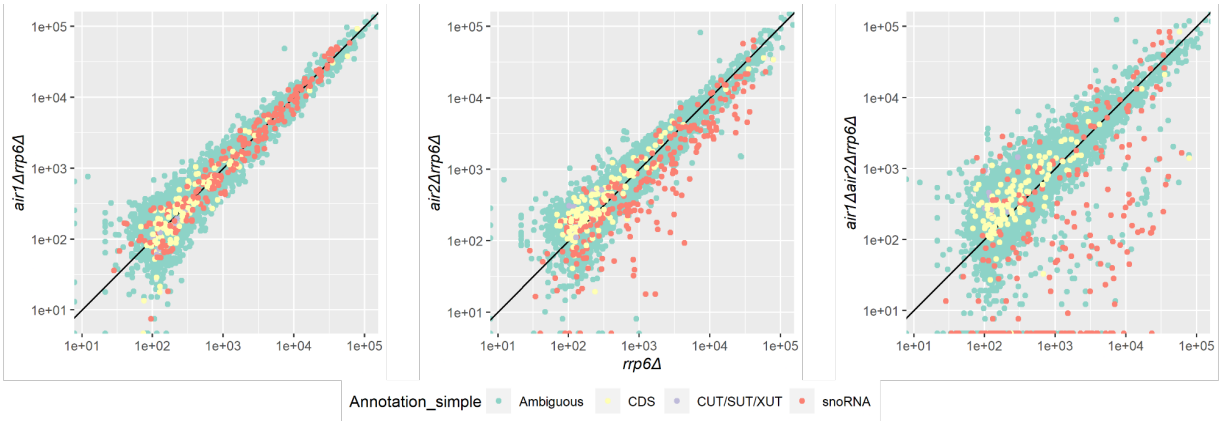
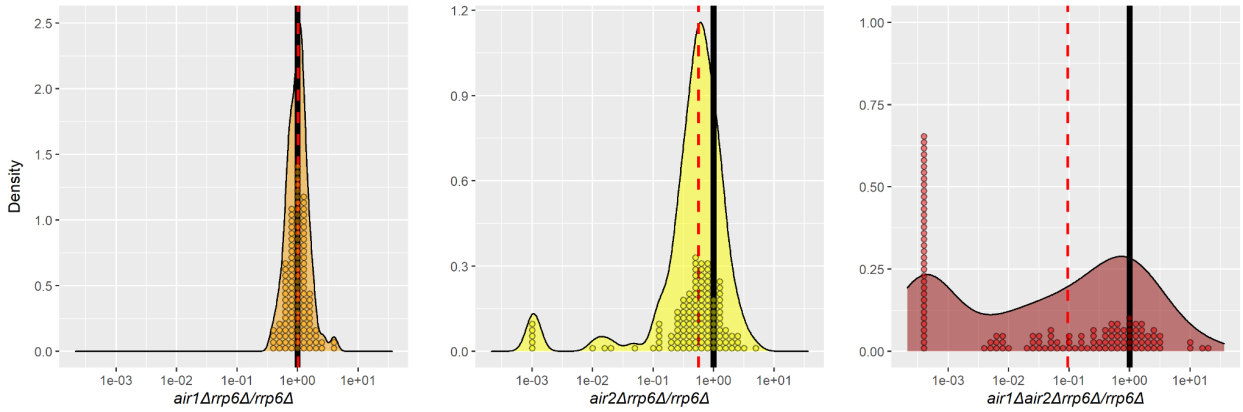


Figure 1. A) poly(A)-based 3'-end sequencing data for rrp6Δ, rrp6Δair1Δ, rrp6Δair2Δ, and rrp6Δair1Δair2Δ mutants. Individual reads were collapsed to the first non-A residue. Clusters were generated by grouping adjacent reads with allowed gaps of 5nts. The clusters were then annotated using previously reported annotation files. avg. fold change rrp6Δair1Δ vs rrp6Δ: 1.12 +- 0.842.

3.4: snoRNAs Processing

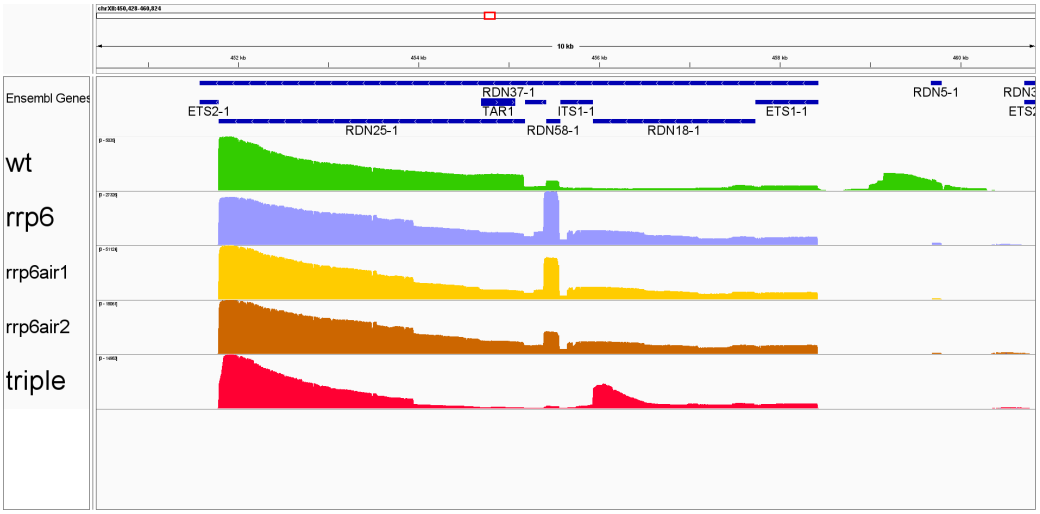


Analysis of polyadenylation downstream of snoRNA. Points within the last 10nts of and 500nts downstream of annotated snoRNAs were analyzed. The red dashed line shows the mean excluding the 3 highest and lowest points from each sample.

Next, we analyzed the peaks appearing downstream of snRNAs. The ***air1Δrrp6Δ*** again showed the lowest impact on polyadenylation downstream of snRNA (Mean value: 1.020 +- 0.547). Both the ***air1Δair2Δrrp6Δ*** and ***air2Δrrp6Δ*** mutants showed a decrease in downstream polyadenylation relative to the ***rrp6Δ*** single mutant. The effect was slightly stronger in the ***air1Δair2Δrrp6Δ*** (Mean value: 0.57 +- 0.67) vs the ***air2Δrrp6Δ*** (Mean value: 0.095 +- 2.5). This may be due to partial compensation of Air1p loss by Air2p.

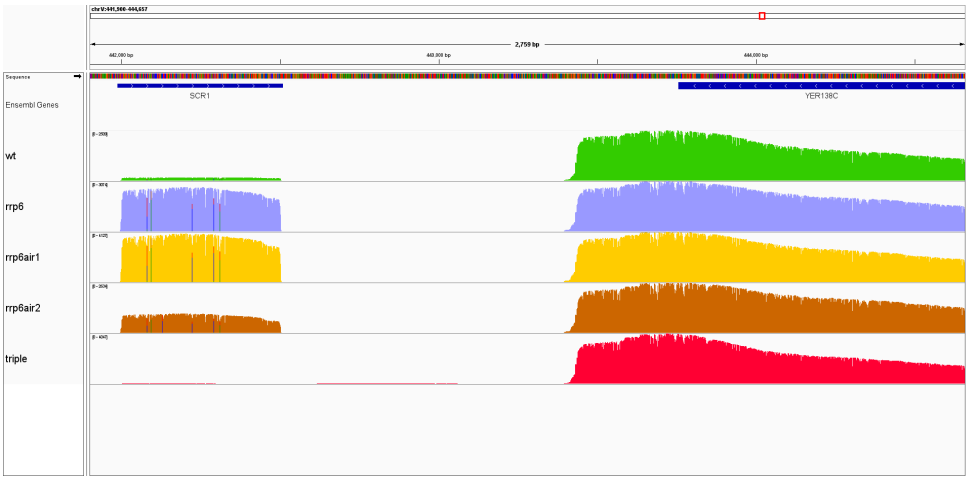
The most stabilized peaks in the ***air1Δair2Δrrp6Δ*** mutants correspond to a region ~450nts downstream of snR34 (Antisense to its neighboring gene, STE23), ~450nts downstream of snR10 (Antisense to its neighboring gene, MMS2), ~200nts downstream of snR65 (Antisense to RPS14A's 3' UTR) and ~100nts downstream of snR71 (Antisense to its neighboring gene, LIN1). Likely representing unprocessed pre-snoRNAs accumulating in the triple mutant.

3.5: TRAMP and Ribosomal RNA



Due to redundancy in the ribosomal RNA sequences read could not be uniquely mapped to a specific RDN loci. Analysis of ribosomal RNA showed a role for TRAMP in ribosomal RNA processing. In the triple mutants the 5.8s subunits lose almost all poly-adenylation at their 3' resulting in them not appearing in data. Additionally, in the triple mutants the 18s subunits show significant poly-adenylation at their 3' ends. Whether this is a degradation product or a processing defect is unknown.

3.6: TRAMP appears to play a role in scr1 Degridation

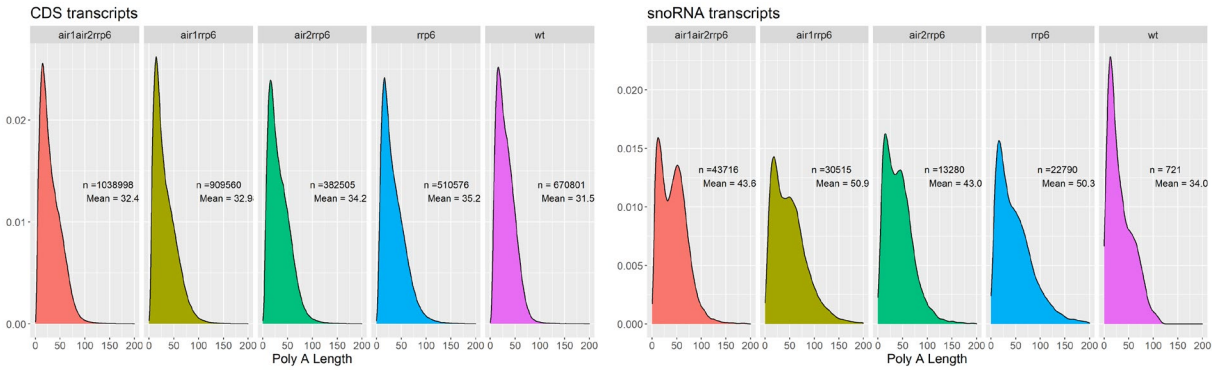


Analysis of other ncRNAs via direct Oxford Nanopore direct RNA sequencing suggests TRAMP plays a role in processing scr1. Previous research has shown that in *rrp44* cells scr1 transcripts accumulation in the nucleus (Leung et al. 2014). Our data suggest that TRAMP is responsible for poly-adenylation of the scr1 transcript. scr1 poly-adenylation seems to be partially dependent on AIR2 as deletion of air2 results in significantly lower poly-adenylatied scr1 transcripts compared to rrp6 or rrp6air1. Interestingly, a second much less abundant transcript is seen approximately 100nts downstream of scr1. It is largely disconnected from premature scr1 and has a distinct 5'end.

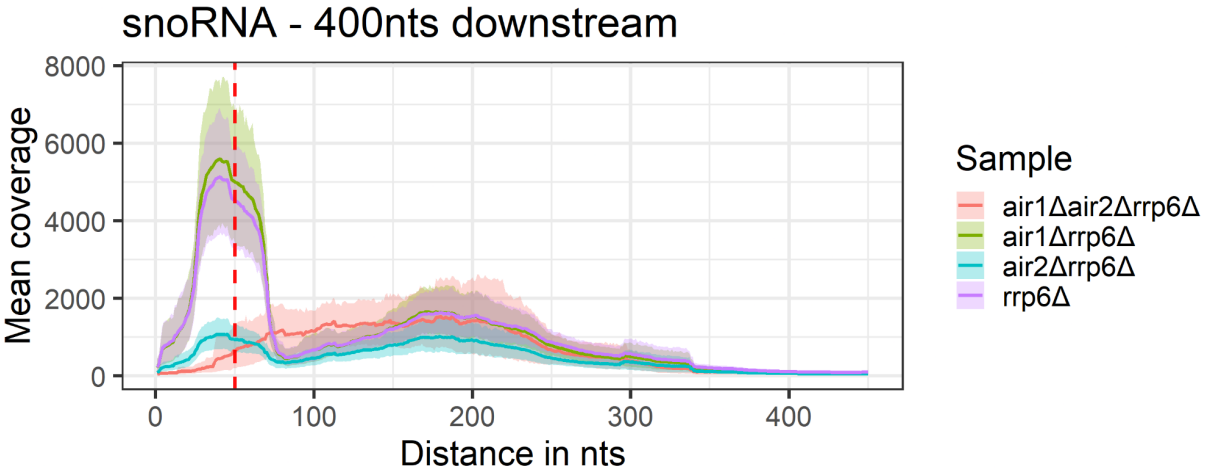
3.1: PolyA estimates

PolyA tail lengths were estimated for all transcripts using the direct RNA sequencing data |. The estimates were then subsetted to only reads which overlapped with coding sequences or snoRNAs. The WT samples had an average coding sequence polyA tail length of 31.5nts. As expected deletion of *RRP6* resulted in a notable increase in the polyA tail length to 35.2nts. Interestingly deletion of *AIR2* in addition to *RRP6* results in only a modest reduction in average CDS polyA tail length where as deletion of AIR1p and RRP6p resulted in a larger reduction similar to the one seen in the triple AIR1, AIR2 and RRP6 deltion. (Figure X) This suggests that there could be a role for AIR1 in maintenance of polyA tails.

When the reads were subsetted to only the snoRNA transcripts the results were more in line with previous observations about the redundancy of Air1 with Air2. deletion of RRP6 results in increased average polyA tail length for snoRNAs. Deletion of Air1 in conjunction with Rrp6 resulted in very little to no change in the polyA tail length. In contrast deletion of Air2 and Rrp6 as well as deletion of Air1, Air2, and Rrp6 resulted in a notable decrease in the polyA tail length compared to the single Rrp6 deletion. This suggests a requirement for Air2 in TRAMP targeting of pre-snoRNAs.



3.5: Metagene analysis of snoRNAs



Metagene analysis of region 400nts downstream of non-polycistronic snRNAs. The dashed red line indicates the 3'end of the snRNAs.

Metagene analysis was performed on the regions downstream of snRNAs. Polycistronic encoded snRNAs were removed to simplify analysis. The *air1Δair2Δrrp6Δ* triple mutants showed significantly less polyadenylation at the mature end of the snoRNAs, showing a role in late snoRNA processing. Additionally, both the *air1Δrrp6Δ* and *rrp6Δ* show a strong polyA peak at the mature 3'end of the snRNAs, implying a role in degradation. Interestingly, the triple mutant showed a much more subtle effect on snRNA 3' ends.

3.7: Future Direction: Individual gene/transcript polyA tail estimation

PolyA tail length calculations are computationally time consuming to generate. We have developed a workflow for generating polyA tail lengths for reads which map to specific genomic regions. For future research efforts it would be worthwhile to optimize the code for generating polyA tail lengths for a specific transcript. It currently works by extracting the read IDs of transcripts aligning to a specified position and then storing them as a list. Then, because polyA tail length estimation requires analysis of the raw signals, the read entries in the HDF5 files need to be extracted. Due to the large size of the FAST5 files, this step is extremely slow. Then, TailfindR can be run(Krause et al. 2019). This is also a notably time consuming step, although the computation time required is dependent on the number of reads analyzed.

5: Conclusions

5.1: Aberrant snoRNA isoform created from intron encoded snoRNAs

Splicing inactivation by a variety of techniques leads to the creation of a hybrid mRNA-snoRNA isoform. The hybrid mRNA-snoRNAs are generated by exonucleolytic degradation by the nuclear exosome. These isoforms occur for all intron encoded snoRNAs in the yeast genome including both C/D box snoRNAs and H/ACA box snoRNAs. We developed a long-read sequencing technique for the detection of these isoforms. Finally, we show that expression of mutant snoRNAs which lack the ACA motif does not result in hmsnoRNA formation showing that

the stability of hmsnoRNAs is dependent on snoRNP assembly. Possible future directions include adapting hmsnoRNAs to synthetic biology problems in order to control protein expression or stabilize RNAs as well as testing the biological halflives of hmsnoRNAs.

5.2: NanoBlot

While long-read sequencing technology has advanced significantly in recent years tools for visualization have lagged behind. Most currently available tools focus on calling discrete isoforms and therefore are not applicable if the isoforms being investigated have continuous lengths. To address this we developed NanoBlot, a tool for isoform visualization of long-read sequencing data. We attempted to create a representation which is similar to a classical Western/Northern blot to allow for easy interpretation. Future development directions for NanoBlot include additional probe options, code optimization,submission to Bioconductor and additional visualization options.

5.3: Substrate specificity of the TRAMP complex

We sought to investigate the substrate specificity of the TRAMP complex using complimentary sequencing approaches. Via 3'end-sequencing we were able to confirm that most processing of snoRNAs is dependent on Air2p with Air1p being largely redundant. We also confirm that the TRAMP complex plays a role in ribosomal RNA processing specifically in polyadenylation of the 5.8s ribosomal RNA. Finally, we use ONT sequencing to estimate the polyA tail lengths of a number of groups of cellular RNAs. We find evidence that there could be a role for AIR1p in maintenance of mRNA polyA tails.

Allmang C, Kufel J, Chanfreau G, Mitchell P, Petfalski E, Tollervey D. 1999. Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J* **18**: 5399–5410.

Amrani N, Ganesan R, Kervestin S, Mangus DA, Ghosh S, Jacobson A. 2004. A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature* **432**: 112–118.

Annaldasula S, Gajos M, Mayer A. 2021. IsoTV: processing and visualizing functional features of translated transcript isoforms. *Bioinformatics* **37**: 3070–3072.

Atkinson GC, Baldauf SL, Hauryliuk V. 2008. Evolution of nonstop, no-go and nonsense-mediated mRNA decay and their termination factor-derived components. *BMC Evol Biol* **8**: 1–18.

Boileau E, Altmüller J, Naarmann-de Vries IS, Dieterich C. 2021. A comparison of metabolic labeling and statistical methods to infer genome-wide dynamics of RNA turnover. *Brief Bioinform* **22**.

Buermans HPJ, den Dunnen JT. 2014. Next generation sequencing technology: Advances and applications. *Biochim Biophys Acta* **1842**: 1932–1941.

Cao D, Parker R. 2003. Computational Modeling and Experimental Analysis of Nonsense-Mediated Decay in Yeast. *Cell* **113**: 533–545.

Chaves-Arquero B, Pérez-Cañadillas JM. 2023. The Nrd1-Nab3-Sen1 transcription termination complex from a structural perspective. *Biochem Soc Trans* **51**: 1257–1269.

De Paoli-Iseppi R, Gleeson J, Clark MB. 2021. Isoform Age - Splice Isoform Profiling Using Long-Read Technologies. *Front Mol Biosci* **8**: 711733.

Delan-Forino C, Spanos C, Rappsilber J, Tollervey D. 2020. Substrate specificity of the TRAMP nuclear surveillance complexes. *Nat Commun* **11**: 3122.

DeMario S, Xu K, He K, Chanfreau GF. 2023. NanoBlot: An R-Package for Visualization of RNA Isoforms from Long Read RNA-sequencing Data. *RNA N Y N* rna.079505.122.

D'Orazio KN, Wu CC-C, Sinha N, Loll-Krippleber R, Brown GW, Green R. 2019. The

endonuclease Cue2 cleaves mRNAs at stalled ribosomes during No Go Decay. *Elife* **8**: e49117.

Fahmi NA, Ahmed KT, Chang J-W, Nassereddeen H, Fan D, Yong J, Zhang W. 2022. APA-Scan: detection and visualization of 3'-UTR alternative polyadenylation with RNA-seq and 3'-end-seq data. *BMC Bioinformatics* **23**: 396.

Garzia A, Jafarnejad SM, Meyer C, Chapat C, Gogakos T, Morozov P, Amiri M, Shapiro M, Molina H, Tuschl T. 2017. The E3 ubiquitin ligase and RNA-binding protein ZNF598 orchestrates ribosome quality control of premature polyadenylated mRNAs. *Nat Commun* **8**: 16056.

Guan Q, Zheng W, Tang S, Liu X, Zinkel RA, Tsui K-W, Yandell BS, Culbertson MR. 2006. Impact of nonsense-mediated mRNA decay on the global expression profile of budding yeast. *PLoS Genet* **2**: e203.

He F, Brown AH, Jacobson A. 1997. Upf1p, Nmd2p, and Upf3p are interacting components of the yeast nonsense-mediated mRNA decay pathway. *Mol Cell Biol* **17**: 1580–1594.

Heidmann S, Obermaier B, Vogel K, Domdey H. 1992. Identification of Pre-mRNA Polyadenylation Sites in Saccharomyces cerevisiae. *Mol Cell Biol* **12**: 4215–4229.

Huang C, Shi J, Guo Y, Huang W, Huang S, Ming S, Wu X, Zhang R, Ding J, Zhao W, et al. 2017. A snoRNA modulates mRNA 3' end processing and regulates the expression of a subset of mRNAs. *Nucleic Acids Res* **45**: 8647–8660.

Huang W, Sun Y-M, Pan Q, Fang K, Chen X-T, Zeng Z-C, Chen T-Q, Zhu S-X, Huang L-B, Luo X-Q, et al. 2022. The snoRNA-like lncRNA LNC-SNO49AB drives leukemia by activating the RNA-editing enzyme ADAR1. *Cell Discov* **8**: 117.

Ikeuchi K, Tesina P, Matsuo Y, Sugiyama T, Cheng J, Saeki Y, Tanaka K, Becker T, Beckmann R, Inada T. 2019. Collided ribosomes form a unique structural interface to induce Hel2-driven quality control pathways. *EMBO J* **38**: e100276.

Jia H, Wang X, Anderson JT, Jankowsky E. 2012. RNA unwinding by the Trf4/Air2/Mtr4

polyadenylation (TRAMP) complex. *Proc Natl Acad Sci U S A* **109**: 7292–7297.

Jia H, Wang X, Liu F, Guenther U-P, Srinivasan S, Anderson JT, Jankowsky E. 2011. The RNA helicase Mtr4p modulates polyadenylation in the TRAMP complex. *Cell* **145**: 890–901.

Juszkiewicz S, Chandrasekaran V, Lin Z, Kraatz S, Ramakrishnan V, Hegde RS. 2018. ZNF598 is a quality control sensor of collided ribosomes. *Mol Cell* **72**: 469–481.

Juszkiewicz S, Hegde RS. 2017. Initiation of quality control during poly (A) translation requires site-specific ribosome ubiquitination. *Mol Cell* **65**: 743–750.

Kishore S, Gruber AR, Jedlinski DJ, Syed AP, Jorjani H, Zavolan M. 2013. Insights into snoRNA biogenesis and processing from PAR-CLIP of snoRNA core proteins and small RNA sequencing. *Genome Biol* **14**: R45.

Krause M, Niazi AM, Labun K, Torres Cleuren YN, Müller FS, Valen E. 2019. tailfindr: alignment-free poly(A) length measurement for Oxford Nanopore RNA and DNA sequencing. *RNA N Y N* **25**: 1229–1241.

Krawczyk PS, Tudek A, Mroczek S, Dziembowski A. 2024. Transcriptome-Wide Analysis of mRNA Adenylation Status in Yeast Using Nanopore Sequencing. *Methods Mol Biol Clifton NJ* **2723**: 193–214.

Lee CY, Lee A, Chanfreau G. 2003. The roles of endonucleolytic cleavage and exonucleolytic digestion in the 5'-end processing of S. cerevisiae box C/D snoRNAs. *RNA N Y N* **9**: 1362–1370.

Leung E, Schneider C, Yan F, Mohi-El-Din H, Kudla G, Tuck A, Wlotzka W, Doronina VA, Bartley R, Watkins NJ, et al. 2014. Integrity of SRP RNA is ensured by La and the nuclear RNA quality control machinery. *Nucleic Acids Res* **42**: 10698–10710.

Liu Y, DeMario S, He K, Gibbs MR, Barr KW, Chanfreau GF. 2022a. Splicing inactivation generates hybrid mRNA-snoRNA transcripts targeted by cytoplasmic RNA decay. *Proc Natl Acad Sci* **119**: e2202473119.

Liu Y, DeMario S, He K, Gibbs MR, Barr KW, Chanfreau GF. 2022b. Splicing inactivation

generates hybrid mRNA-snoRNA transcripts targeted by cytoplasmic RNA decay. *PNAS* **119**.

Marín-Navarro J, Jauhiainen A, Moreno J, Alepuz P, Pérez-Ortín JE, Sunnerhagen P. 2011. Global estimation of mRNA stability in yeast. *Methods Mol Biol Clifton NJ* **734**: 3–23.

Massenet S, Bertrand E, Verheggen C. 2017. Assembly and trafficking of box C/D and H/ACA snoRNPs. *RNA Biol* **14**: 680–692.

Miller C, Schwalb B, Maier K, Schulz D, Dümcke S, Zacher B, Mayer A, Sydow J, Marcinowski L, Dölken L, et al. 2011. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol Syst Biol* **7**: 458.

Mills JD, Janitz M. 2012. Alternative splicing of mRNA in the molecular pathology of neurodegenerative diseases. *Neurobiol Aging* **33**: 1012. e11-1012. e24.

Mohanta A, Chakrabarti K. 2021. Dbr1 functions in mRNA processing, intron turnover and human diseases. *Biochimie* **180**: 134–142.

Moll P, Ante M, Seitz A, Reda T. 2014. QuantSeq 3′ mRNA sequencing for RNA quantification. *Nat Methods* **11**: i–iii.

Ogami K, Suzuki HI. 2021. Nuclear RNA Exosome and Pervasive Transcription: Dual Sculptors of Genome Function. *Int J Mol Sci* **22**.

Oikonomopoulos S, Bayega A, Fahiminiya S, Djambazian H, Berube P, Ragoussis J. 2020. Methodologies for Transcript Profiling Using Long-Read Technologies. *Front Genet* **11**: 606.

Passmore LA, Coller J. 2022. Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. *Nat Rev Mol Cell Biol* **23**: 93–106.

Peccarelli M, Kebaara BW. 2014. Regulation of natural mRNAs by the nonsense-mediated mRNA decay pathway. *Eukaryot Cell* **13**: 1126–1135.

Perumal K, Reddy R. 2002. The 3’ end formation in small RNAs. *Gene Expr* **10**: 59–78.

Porrua O, Hobor F, Boulay J, Kubicek K, D’Aubenton-Carafa Y, Gudipati RK, Stefl R, Libri D.

2012. In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. *EMBO J* **31**: 3935–3948.

Powers KT, Szeto J-YA, Schaffitzel C. 2020. New insights into no-go, non-stop and nonsense-mediated mRNA decay complexes. *Curr Opin Struct Biol* **65**: 110–118.

Romo L, Ashar-Patel A, Pfister E, Aronin N. 2017. Alterations in mRNA 3′ UTR isoform abundance accompany gene expression changes in human Huntington's disease brains. *Cell Rep* **20**: 3057–3070.

Roy KR, Chanfreau GF. 2020. Robust mapping of polyadenylated and non-polyadenylated RNA 3' ends at nucleotide resolution by 3'-end sequencing. *Methods San Diego Calif* **176**: 4–13.

Sahlin K, Medvedev P. 2021. Error correction enables use of Oxford Nanopore technology for reference-free transcriptome analysis. *Nat Commun* **12**: 2.

Schneider S, Hotz H-R, Schwer B. 2002. Characterization of dominant-negative mutants of the DEAH-box splicing factors Prp22 and Prp16. *J Biol Chem* **277**: 15452–15458.

Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, Waterston RH. 2017. DNA sequencing at 40: past, present and future. *Nature* **550**: 345–353.

Steinmetz EJ, Conrad NK, Brow DA, Corden JL. 2001. RNA-binding protein Nrd1 directs poly(A)-independent 3′-end formation of RNA polymerase II transcripts. *Nature* **413**: 327–331.

Stepankiw N, Raghavan M, Fogarty EA, Grimson A, Pleiss JA. 2015. Widespread alternative and aberrant splicing revealed by lariat sequencing. *Nucleic Acids Res* **43**: 8488–8501.

Volanakis A, Passoni M, Hector RD, Shah S, Kilchert C, Granneman S, Vasiljeva L. 2013. Spliceosome-mediated decay (SMD) regulates expression of nonintronic genes in budding yeast. *Genes Dev* **27**: 2025–2038.

Will CL, Lührmann R. 2011. Spliceosome structure and function. *Cold Spring Harb Perspect Biol* **3**.

Zhao L, Zhang H, Kohnen MV, Prasad KVSK, Gu L, Reddy ASN. 2019. Analysis of Transcriptome and Epitranscriptome in Plants Using PacBio Iso-Seq and Nanopore-Based Direct RNA Sequencing. *Front Genet* **10**: 253.

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. *Sci Rep* **8**: 4781.