

# **Lost in Transcription: Investigating rNMP incorporation patterns in genomic DNA of *S. cerevisiae* under varying metabolic pathways**

*Alex Kovensky, Kamala Natarajan*

*Storici Lab - Spring 2023*

## **Abstract**

The incorporation of ribonucleotide monophosphates (rNMPs) into DNA during replication and transcription can cause genomic instability and disease. In this study, we analyzed the influence of environmental factors on rNMP incorporation in varying strains of budding yeast grown on glucose and galactose carbon sources. We employed RNA-sequencing techniques and subsequent bioinformatics analyses, such as ribose-map, heatmap analysis, and RNA-seq, and omic integration. Our findings reveal that, while differences in growth environment impacted rNMP incorporation rates and patterns into genomic DNA, these changes were limited in scope. Higher variation in rNMP incorporation bias was observed in mutants grown in galactose than in glucose. Additionally, correlative multi-omics integration displayed a slight, but significant, preference for rNMP incorporation into gene coding regions over the TSS region ( $\pm 200\text{bp}$ ). Moreover, mitochondrial DNA showed not to vary in rNMP incorporation bias patterns regardless of mutant strain and environment type, implying the mechanisms involved in the differential energy pathways tested in *S. cerevisiae* are largely unrelated to incorporation bias. Overall, while our study underscores the influence of environmental factors on rNMP incorporation into genomic DNA, it is still unclear as to the quantitative and consistent impact of environmental factors on specific incorporation bias. This study further illustrates that rNMP incorporation is a multifaceted phenomenon influenced by various factors and necessitates additional research.

## **1.0 - Introduction**

The study of ribonucleotide monophosphate (rNMP) incorporation into DNA, often perceived as a form of contamination, is crucial in understanding the complex relationship between DNA and RNA. rNMPs, base-pairs originating from RNA, can be integrated into DNA during transcription. While certain genes, such as *rnh201*, have evolved to correct these errors prior to DNA replication, these mechanisms are not always perfect. The accumulation of rNMPs has been implicated in various health studies, with notable relevance to medical conditions such as Aicardi-Goutieres syndrome, which is linked to mutated RNaseH2 (Koh et al. 2015). DNA and RNA exhibit distinct structural organizations: DNA forms as a double-stranded molecule, whereas RNA is single-stranded with a diverse array of secondary structures and base-pairing possibilities. Despite the growing body of research on this topic, there remains a significant knowledge gap as most studies have only utilized glucose as a carbon source when studying budding yeast, a model organism for rNMP research. Although it utilizes glycolysis preferentially, budding yeast exhibits differential gene expression and metabolic patterns when grown on alternative carbon sources such as galactose. This study aims to investigate the impact

of the yeast's growth environment, particularly the utilization of both glucose and galactose, on rNMP accumulation in its genomic DNA. Our central research question pertains to how changes in growth environment (and associated metabolic differences) affect rNMP incorporation into genomic DNA, with the hypothesis that incorporation patterns would differ substantially in yeast cells grown in galactose.

## 2.0 Part I - Cell preparation

To address this question, we employed *Saccharomyces cerevisiae*, or budding yeast, as our model organism. *S. cerevisiae* is a eukaryote and shares similarities in cell structure with plants and animals, making it an appropriate model for studying cellular processes across a wide range of organisms. The yeast's small single cell size, short doubling time, and ease of culturing allow for rapid and cost-effective production and maintenance of multiple strains. The genetic manipulability of *S. cerevisiae*, including gene addition and deletion through homologous recombination techniques, further enhances its utility as a model organism. Its well-documented and regularly updated genome sequence, consisting of over 12 million base pairs and 6,275 genes organized on 16 chromosomes, provides a solid foundation for genetic studies (ScienceDirect). The ability to culture this yeast as a haploid simplifies the isolation of mutants and haploid-diploid hybrids. Moreover, extensive research on *S. cerevisiae*'s intracellular organelles, particularly mitochondria, has significantly advanced our understanding of eukaryotic organelles, including human mitochondrial function and disease.

*Figure 2.1* illustrates *Schizosaccharomyces pombe* and *S. cerevisiae* cells under the microscope. *S. pombe* is slightly longer than *S. cerevisiae*, as seen in the figure, due to its reproduction through fission rather than budding. Compared to *S. pombe*, *S. cerevisiae* has a more advanced genetic toolbox, including a larger collection of available mutants and more established techniques for genetic manipulation. *S. pombe* and *S. cerevisiae* diverged from a common ancestor approximately 400 million years ago; however, *S. cerevisiae* has a closer relationship to higher eukaryotes than *S. pombe*, including humans (Yanagida 2002). As a result, *S. cerevisiae* serves as a more suitable model for research on biological pathways and processes, such as investigating the incorporation of rNMPs into genomic DNA and other related cellular processes.

## 2.1 Methods

In order to compare the growth environment for *S. cerevisiae*, three growth mixtures were developed: YPD liquid-agar, YPD liquid (no agar), and YPGal liquid. These media were developed to be used in initial cell growth, as well as experimentation around identifying varied metabolic pathways (section 2.1.4). The YPD liquid and agar mixture consists of 1% yeast extract, 2% peptone, 2% dextrose, 2% agar and 1000ml water. It includes all amino acids and glucose required for yeast cells to ferment. YPD liquid consists of 1% yeast extract, 2% peptone and 3% glycerol and 900ml water. The YPGal liquid media consists of 1% yeast extract, 2% peptone and 2% galactose and 1000ml water for experimental groups 4, 5 ,6 and 9; 970 ml water

for group 7. This medium will allow the yeast cells to switch their metabolism from fermentation to respiration. Our group created a stock of YPD liquid-agar media for yeast cell fermentation; other stocks and mentioned mediums were developed by other groups.

Every group was assigned a designated strain-medium combination: our group was tasked with preparing *E134 rnh201 pol2-M644G* double mutant in YPD liquid and agar. Note that, while this would be developed and later prepared (section 3.0) for sequencing, later bioinformatic analysis would analyze a separate but similar strain (*E134 rnh201 pol2-4*) in YPD liquid and agar due to insufficient evidence of correct gene library preparation (section 5.2). Initial stock patches of *S. cerevisiae* were streaked on 01/12/2023 (*Fig 2.3a (1)*) from isolated colonies with genotype E134 WT (*Fig 2.2*).

In order to develop an initial *E134 rnh201* K.O. mutant, the G418 resistance gene was chosen to serve as an antibiotic marker. *rnh201* gene is replaced by G418 resistance gene (antibiotic marker). Mutants of *rnh201* will grow in G418 media, thus being selected against the wild-type to successfully contain knockout for our experimental process. Primer design for amplifying the marker will contain homologous sequence with both target and selective genes.

### 2.1.1 Cell transformation

For the transformation protocol, our *rnh201* and *pol-2* double mutant was inoculated in 5 ml of YPD at 30°C overnight. The SSD was heat-denatured at 100°C for 5 min and was immediately put on ice prior to transformation. The culture was transferred and spun several times in the centrifuge into a cell suspension after adding a solution of 0.1M LiAC-TE which made the cells permeable. 3 microliters of SSD was added to 20 microliters of PCR product along with LiAc-PEG-TE which was used to adhere the DNA to the cells such that the proximity will optimize the potential entry into the cells. SSD acted as carrier DNA, allowing targeting of the PCR product and protection from degradation. The transformations were incubated at 30°C for 30 minutes and heat-shocked at 42°C for 15 minutes to drive DNA into cells. The transformed cells were then collected by centrifugation and plated onto two YPD plates. These cells were allowed to incubate at 30°C for 2-3 days.

Two rounds of replica plating were used to select for our desired strain (E134; *MATα ade5-1 lys2-14A trp1-289 his7-2 leu2-3,112 ura3-52*), and especially select for the transformed G418 marker (*Fig 2.3a*), followed by selection of transformed colonies (*Fig 2.3b*). Note that one of the initial stock plates cracked, causing a potential source of contamination. Additionally,  $\alpha$  and  $\alpha$  mating types were not labeled on the final growth plate (YPD x MAT); while it was assumed that  $\alpha$  grew (as described by our selected strain, E134), this is considered a source of error.

PCR utilized to amplify the transformed sample, particularly *rnh201* upstream and downstream regions. 50  $\mu$ l NF H<sub>2</sub>O was first transferred into a 1.5ml tube, then samples were taken from

transformed colony plates (T1A, T1B, T2A); half was streaked on a new YPD plate to single colonies, with the other half resuspended in NF H<sub>2</sub>O for colony PCR. Each tube was incubated in the heat block for 5 minutes, then immediately put on ice to heat shock the samples. A master mix was created using EconoTaq PLUS GREEN 2X Master Mix (25 µl), 10nM forward primer (2.5 µl), 10nM reverse primer (2.5 µl), DNA (10 µl), and NF H<sub>2</sub>O (10 µl, 1/5 of initial 50 µl). These mixes were labeled M1-4 (*Table 2.1*). 40 µl of this master mix was aliquoted into each (x16) labeled PCR tube (WT, T1A-B, T2A; M1-4). RnH35.1 primer is used to create the knockout for the upstream *rnh201* gene (reaction 3 as per protocol) and *rnh201.2* primer is used to create the knockout for the downstream *rnh201* gene (reaction 4 as per protocol). Samples were purified using QIAquick column purification (section 3.1.12) with the MinElute PCR purification kit. Samples that indicated successful transformation of the G418 antibiotic marker (*Fig 2.2, Fig 2.5*) were sequenced using Sanger sequencing (*Fig 2.7*).

*Table 2.1: Mix guidelines for colony PCR.*

Mix Number (#)	Forward primer	Reverse primer
1	RnH35.1	RnH35.2
2	RnH35.3	RnH201.2
3	RnH35.1	K1
4	K2	RnH201.2

### 2.1.2 RNA extraction

RNA was extracted for later bioinformatic analysis using RNeasy extraction kits. The yeast cells were harvested in a 15 ml centrifuge tube by centrifuging at 1000 x g for 5 min at 4°C. The supernatant was decanted, and remaining media was removed by aspiration. The cells were resuspended with 100 µl of freshly prepared Buffer Y1 containing zymolyase. The mixture was incubated for 10–30 min at 30°C with gentle shaking to generate spheroplasts. 350 µl Buffer RLT was added and the mixture was vortexed vigorously to lyse the spheroplasts. 250 µl ethanol (96–100%) was added to the homogenized lysate and mixed well by pipetting. 700 µl of the sample was transferred to an RNeasy spin column that was placed in a 2 ml collection tube to wash the spin column membrane using 350 µl Buffer RW1. The 80 microliters of DNase I incubation mix was directly added to the RNeasy spin column membrane, and placed on the benchtop (20–30°C) for 15 min. 350 µl Buffer RW1, 500 µl Buffer RPE were added to the spin column and centrifuged. Finally, the RNeasy spin column was placed in a new 1.5 ml collection tube and had 30–50 µl RNase-free water added directly to the spin column membrane. The RNA was eluted upon final centrifugation. The final extracted RNA concentration from Nanodrop was 35.8 ng/µl.

### 2.1.3 Restriction digest

Samples also underwent restriction digestion in order to prepare for later restriction digestion reactions (section 3.0). 1.5ml tubes (3x) were prepared for use in restriction digestion. Each was labeled for double digestion (DD) or single digestion (XbaI and ScaI, respectively). Keeping all components on ice, each digestion mixture was created according to *Table 2.2*. The plasmid concentration used was 0.9 µg/µl, and the digestion enzymes were added last before mixing by micropipette and incubation in H<sub>2</sub>O at 37°C for 15 minutes.

*Table 2.2: Mix guidelines for restriction digestion.*

COMPONENT	Double digestion	Single digestion (XbaI)	Single digestion (ScaI)
DNA	1 µg	1 µg	1 µg
10X rCutSmart Buffer	5 µl	5 µl	5 µl
XbaI	1.0 µl	1.0 µl	
ScaI	1.0 µl		1.0 µl
H <sub>2</sub> O	to 50 µl	to 50 µl	to 50 µl

### 2.1.4 Fermentation

Cells from our isolated stock patches were allowed to grow in flasks containing either YPD (glucose) or YPGal (galactose) liquid medium, with the purpose being to qualitatively depict the difference between gene expression pathways- CO<sub>2</sub> would be produced in YPD but not YPGal. 100 ml of YPD liquid was transferred to a conical flask and a whole patch of WT cells was transferred to the liquid. Rubber balloons were attached to the top of each flask in order to capture any gasses released during incubation, then each mixture was incubated overnight. Notably, the first attempt at this experiment failed, as neither the YPD nor the YPGal balloon inflated. Another flask with sampled yeast cells and YPD liquid medium yielded results (*Fig 2.8*).

## 2.2 Results

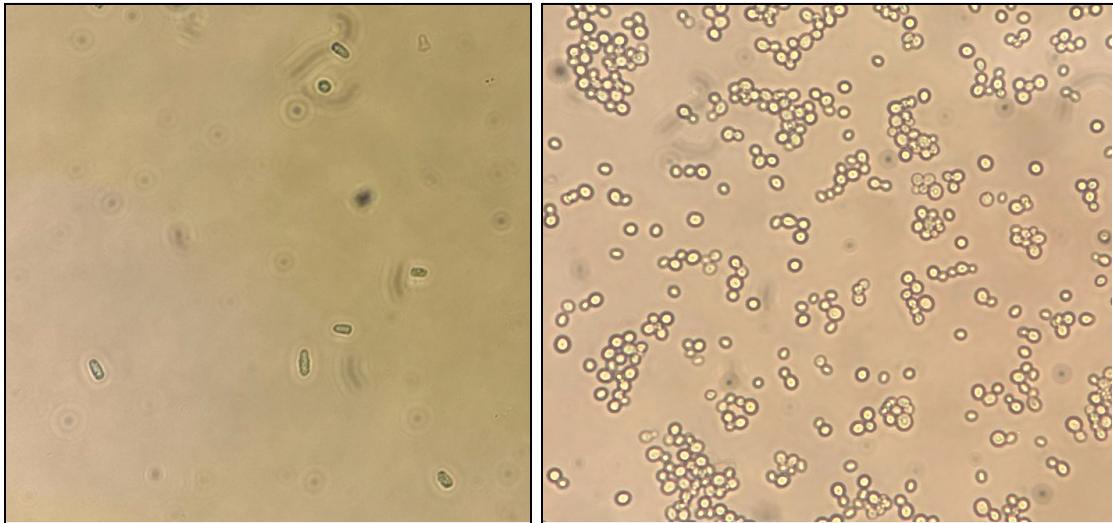
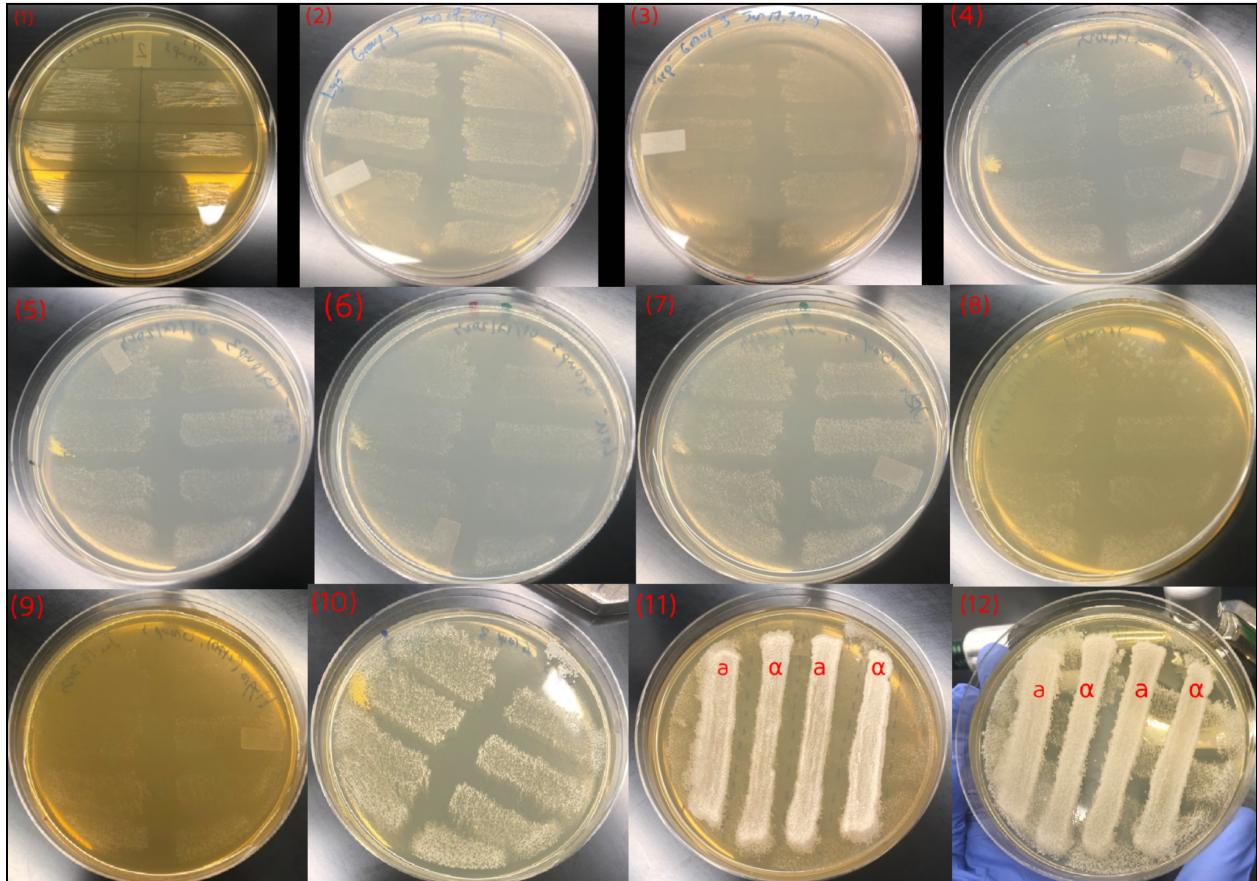


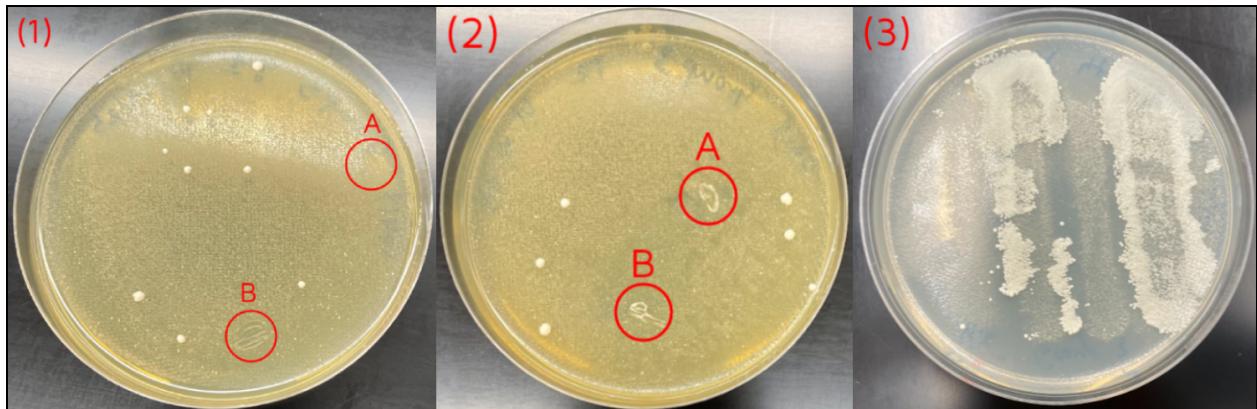
Figure 2.1: Microscopy results. The left image shows *S. pombe*, while the right shows *S. cerevisiae* (our model organism). *S. pombe* is slightly longer than *S. cerevisiae*, as it reproduces through fission rather than budding. *S. cerevisiae* was preferred for this study due to its ease of use in the lab (smaller genome, shorter reproduction time, precedence for use as model organism), as well as its known dual energy pathways.



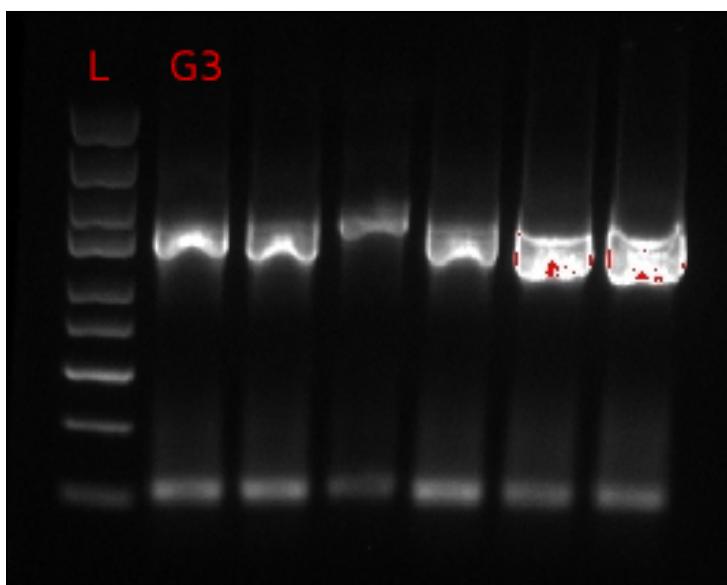
*Figure 2.2: Incubated cell streaking colonies.* Individual colonies are represented by isolated circles on the plate. Due to reproduction by budding, colonies of *S. cerevisiae* represent single gene strains. Several single colonies were sampled for production of cell stocks (Fig 2.3a (1)).



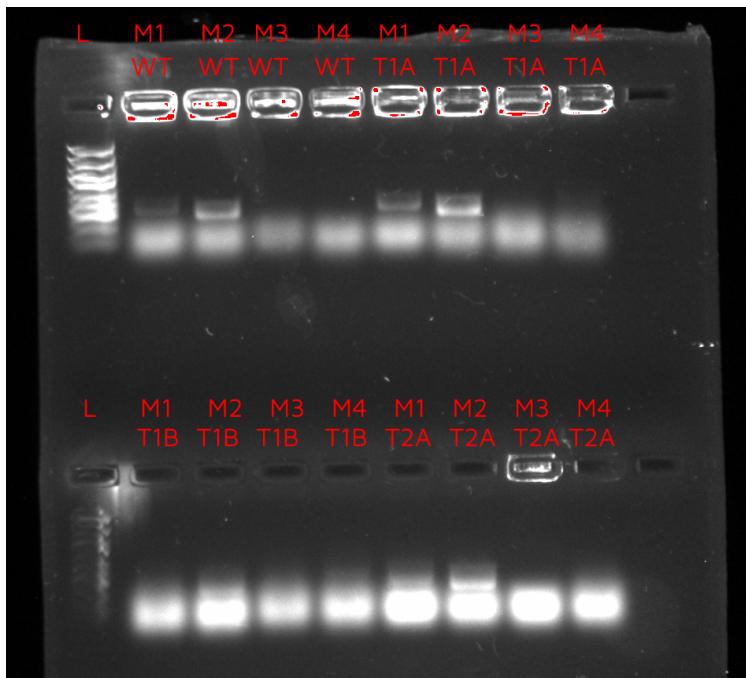
*Figure 2.3a: Plating results (pre-transformation).* Agar plates used for an initial round of replica plating, ordered as such: Patch plate 2 (1), Lys- (2), Trp- (3), His- (4), Ade- (5), Leu- (6), Ura- (7), G418 (8), Hygro (9), YPG (10), MAT (11), YPD (12). Note that “ $\alpha$ ” and “a” mating pairs are listed for MAT and YPD. Due to technical issues, the image of colony growth on plate 1 was lost; an image of the pre-growth plate is instead shown. Unknown colony growth began on plate 4, with an additional unknown colony present on plate 10 onward.



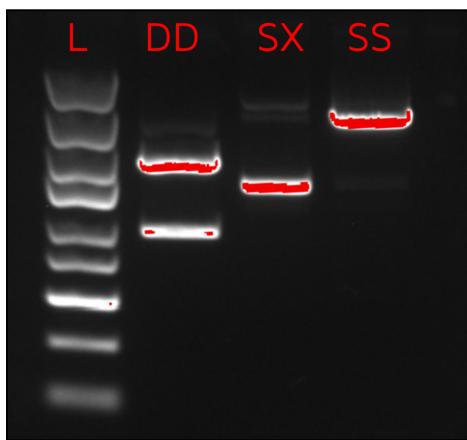
*Figure 2.3b: Plating results (post-transformation).* Final plates from a second round of replica plating after transformation of G418 resistance marker. Plates 1 and 2 depict G418 plates with single mutant *rnh201*-K.O. (G418-marked) colonies. Samples A and B were used for both stock preparation and colony PCR. Plate 3 depicts final colonies of presumed genotype E134 after a similar round of plating to *Figure 2.3a*. Note that “α” and “a” columns were erroneously left unlabeled.



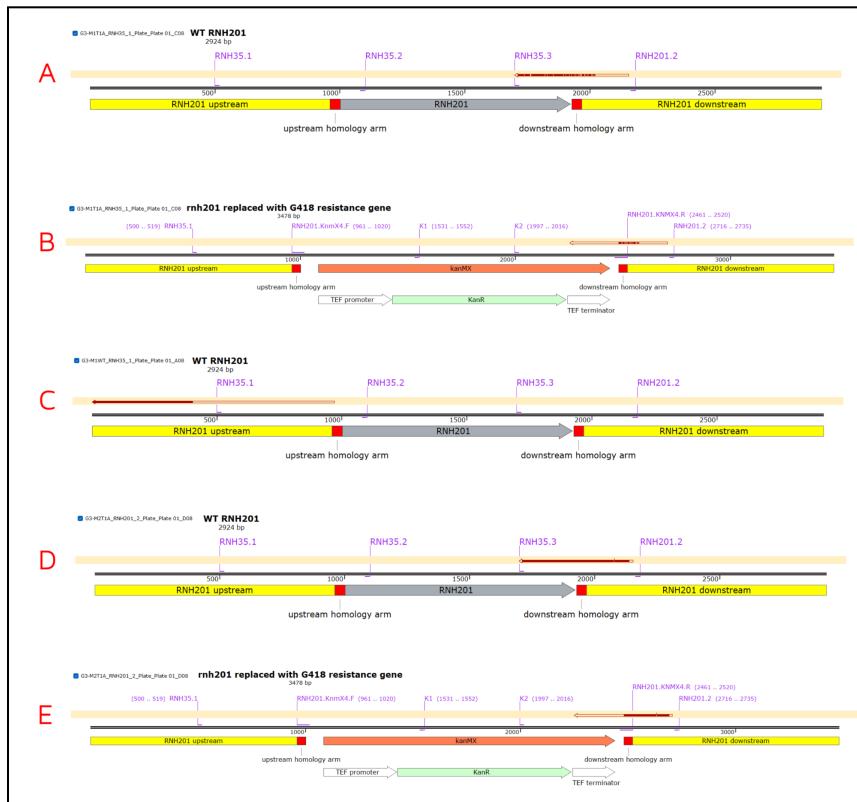
*Figure 2.4: Agarose gel for G418-marker PCR.* The L column indicates the gene ladder, and G3 indicates our G418-transformation PCR sample. G3 shows two bands (~800-1000bp), with the top presumably containing complete plasmids. The bottom band may indicate excess material, such as primer dimers.



*Figure 2.5: Agarose gel for colony PCR.* The L columns indicate the gene ladder used. M1-4 indicates reaction mixtures 1-4 (*Table 2.1*). T1A,B and T2A indicate individual colonies sampled from transformant plates T1 and T2 (*Figure 2.3b*). W1 and samples showed expected bands (~500bp) based on forward and reverse primers used. T1A showed clear bands (~500bp) but in opposite mixture treatments than expected. T1B and T2A samples appear to show bands (~500bp), but samples are too blurred to indicate expected reaction. Lower bands (<200bp) likely indicate primer dimers, or other small excess material)



*Figure 2.6: Gel result for restriction digestion.* Columns L, DD, SX, and SS represent the gene ladder, double digest (XbaI + ScaI), single digest 1 (XbaI), and single digest 2 (ScaI), respectively. DD depicts two bands, 2200bp and 750bp (two large pieces), and SX shows one large band around 1000bp. While SS shows similarity to SX, the size (~5000bp) is larger than expected (~1000bp).



*Figure 2.7: Sanger sequencing results.* Results for both WT and *rnh201* mutant strains. Based on the mixture-sample combination described in *Figure* (Fig 2.5), each sequence represents the following: M1-T1A (A), M1-T1A (B), M1-WT (C), M2-T1A (D), and M2-T1A (E). Sequences A, C, and D show alignment against the expected WT strain, while sequences B and E show alignment against the expected knockout sequence (E134).



*Figure 2.8: Fermentation experiment results.* Flasks show cells grown in YPD (left) vs YPGal (right).

### 3.0 Part II - Gene library preparation (Ribose-seq)

We utilized the ribose-seq protocol (*Figure 3.1*) in order to isolate and quantify rNMPs for genome library construction. Ribose-seq is one of very few rNMP-specific pipelines, and was developed in part in the Storici lab. Ribose-seq is designed to isolate embedded rNMPs within DNA through fragmentation of samples, exposure of rNMPs, and self-ligation around each rNMP fragment to allow for gene library construction and sequencing (Balachander et al. 2019).

#### 3.1 Materials & Methods

Ribose-seq protocol began with extraction of genomic DNA from isolated mutant colonies of *S. cerevisiae*, followed by manipulation by various processes to reduce genome size, allowing for isolation and cyclization of individual rNMPs within strings of DNA, along with a unique molecular identifier (UMI) and adaptor sequences for later sequencing and analysis.

##### 3.1.1 Extraction

Extraction utilized a modified DNeasy Blood & Tissue Kit protocol that consisted of cell wall lysis and degradation by 800 µl zymolyase in 4.8ml Buffer Y1 (incubated at 30°C for 1 hour), followed by spheroplast suspension in 1.44 ml Buffer ATL and both 160 µl proteinase K and 40 µl RNase A for denaturation and digestion of proteins and RNA (vortexed and incubated at 56°C H<sub>2</sub>O for 30 minutes). This ultimately left behind genomic DNA precipitated by 3.2 ml Buffer ALE, then purified through use of DNeasy Mini spin columns (8 x 600 µl) and 500 µl Buffers AW1, AW2, and 100 µl nuclease free water (NF H<sub>2</sub>O) (incubated at room temperature (RT) for 1 minute before centrifugation).

DNA concentration (ng/µl), purity, and amount (µg) were measured via Qubit and Nanodrop (*Table 3.1*). While the actual amount of DNA (9.79 µg) was lower than desired (20-40 µg), the highest yield of any group was 12.07 µg; the TAs increased concentration through evaporation, increasing sample concentration to that listed for Nanodrop (967.6 ng/µl).

*Table 3.1: DNA concentrations, purity, and amounts.* Concentration and total amount were measured using Qubit, while concentration and purity were measured using Nanodrop. Ideal purity was listed as 1.8, so sample purity is lower than ideal.

DNA measurement method	Concentration (ng/µl)	Purity	Total amount (µg)
Qubit	163.17	n/a	9.79
Nanodrop	967.6	1.52	n/a

### 3.1.2 Fragmentation & End Repair

Extracted DNA (4 x 16 µl) was fragmented via NEBNext dsDNA Fragmentase (4 x 2 µl) and NF H<sub>2</sub>O. Sample mixtures were vortexed before being incubated at 37°C for 40 minutes, then fragmentation was stopped by addition of 0.5M EDTA (4 x 5 µl). Samples were purified using bead purification protocol (refer to section describing bead purification protocol) with AMPure XP DNA beads, 400 µl 70% ethanol (EtOH) during wash steps, and final addition of 42 µl NF H<sub>2</sub>O. 40 µl supernatant was removed and run through agarose gel electrophoresis to ensure fragmentation occurred as intended (*Fig 3.3*).

After fragmentation, the samples (amounts from lab notebook, 2 samples) were mixed with 10X NEBNext End Repair Reaction Buffer (10 µl), End Repair Enzyme Mix (5 µl), and NF H<sub>2</sub>O (amounts from lab notebook) and incubated at 20°C for 30 minutes in order to initiate end repair, converting all sticky ends to blunt ends. The end repair product was purified using bead purification with HighPrep PCR DNA beads (section 3.1.11). Wash steps utilized 500 µl 80% EtOH, and 42 µl NF H<sub>2</sub>O was added after drying. 40 µl of the supernatant of the purified end repair product was removed from each tube and placed in 1.5 mL tubes for use in future steps.

### 3.1.3 dA-tailing

End-repaired gene products underwent dA-tailing to allow for adaptor ligation (*Fig 3.2*). A dA-tailing mixture (2 x 1.5mL tubes) was combined using 10X NEBuffer 2 (5µl), 5mM dATP (1 µl), Klenow Fragment (3'→5' exo-) (3 µl), end-repair DNA (40µl), and NF H<sub>2</sub>O (1µl), with the Klenow Fragment mixed in last by micropipette to initiate enzymatic activity. The tubes were spun down then incubated in water at 37°C for 30 minutes, then purified by column purification (section 3.1.12).

### 3.1.4 Adaptor ligation

The ribose-seq adaptor was prepared using 500mM Tris-Cl pH 7.5 (1 µl), 2.5M NaCl (1 µl), 50mM EDTA (1 µl), 50µM adaptor oligo L (25 µl), 500µM adaptor oligo S (12.5 µl) and NF H<sub>2</sub>O (9.5 µl). This mixture was incubated in boiling water for 5 minutes then gradually cooled down to 25°C for 1 hour. After incubation, the mixture was purified using a G-25 column, and stored at -20°C until used.

Prepared ribose-seq adaptor (12 µl) was added to two PCR tubes, along with 10X T4 DNA ligase buffer (5 µl), dA-tailed DNA (28µl). T4 DNA ligase was pipetted into the mixture to initiate enzymatic activity before being spun down briefly and incubated overnight at 16°C. These tubes were then incubated at 65°C for 10 minutes to inactivate the T4 DNA ligase then stored at -20°C until purified via bead purification (section 3.1.11). DNA beads were incubated at RT for 30 minutes before use. 90 µl beads were transferred into each 1.5mL tube, and 400 µl 80% EtOH. 32 µl warm NF H<sub>2</sub>O was added after drying, and 30 µl (2x) of the final purified product was transferred to PCR tubes for use in future steps.

### 3.1.5 Alkali treatment

Purified adaptor-ligated samples (2 x 30 µl) were set aside for alkali treatment, exposing the 2',3'-cyclic phosphate on the rNMP to allow for later cyclization by AtRNL. A mixture containing 2M NaOH (7.5 µl) and DNA (30 µl, 24µl + 6µl RT NF H<sub>2</sub>O) was mixed into PCR tubes (2x) via micropipette and spun down briefly before being incubated at 55°C for 1.5 hours. 2M HCl (8.5 µl) was added to neutralize each tube, then the pH was checked using pH test paper until the mixture was between a pH of 6-8. Either 2M NaOH (0.3 µl) or 2M HCl (0.3 µl) was added intermittently to the mixture and vortexed before tested using pH paper until within the desired range.

### 3.1.6 Self-ligation by AtRNL

Taking the purified gene product from the previous section, self-ligation was used to cyclize each isolated rNMP, preparing the gene product for sequencing. The 1.5ml tube containing the DNA was incubated on a heat block for 5 minutes before being placed on ice to heat shock the sample. PCR tubes (4x) were prepared and labeled with M and P1-3, respectively. After briefly spinning down the DNA, mixtures containing DNA (13 µl), 10X AtRNL buffer (2 µl), NF H<sub>2</sub>O (5 µl, M only) and 4 µM AtRNL (5 µl, P1-3 only) were added to the PCR tubes and mixed by micropipette. The tubes were briefly spun down before being incubated at 30°C for 1 hour. After incubation, all four samples (M, P1-3) were purified using RNA bead purification (section 3.1.11) to remove non-cyclized ssDNA and remaining reaction components. The P1-3 tubes were combined into a single 1.5ml tube (P), while the M tube was added to a fresh 1.5ml tube (M). After the RNA beads had incubated at RT for 30 minutes and vortexed, 90 µl RNA beads were added to the M tube, and 108 µl was added to the P tube; the samples were slowly pipetted to mix before being incubated at RT for 15 minutes. The total three wash steps used 400 µl 70% EtOH, and warm NF H<sub>2</sub>O was added to both tubes (32µl to M, 42µl to P). The purified products were transferred into fresh PCR tubes (M, P1-4), with 30µl transferred between M tubes and 10µl transfer from P to each PCR tube P1-4.

### 3.1.7 Degradation of linear ssDNA

In order to remove rNMP-barren linear ssDNA from the self-ligated gene product, each PCR tube was combined with 10X NEBuffer 4 (5 µl), NF H<sub>2</sub>O (14µl to M, 34µl to P1-4), and T5 Exonuclease (1 µl). Note that roughly double the amount of required NF H<sub>2</sub>O was added to tube P4 due to an error in managing each sample. Each tube was mixed by pipetting, spun down briefly, then incubated at 37°C for 30 minutes. To remove any degraded ssDNA components, as well as any remaining digestion reaction components, each sample was purified using RNA bead purification similar to the previous section (section 3.1.6). 1.5 ml tubes (5x) were prepared to match each PCR tube, with each mixture transferred to the associated 1.5 ml tube. 27 µl warm NF H<sub>2</sub>O was added to each tube once dried, and 25 µl of supernatant from each tube was transferred to fresh PCR tubes, labeled accordingly.

### 3.1.8 Removal of 2'-phosphate

All samples M and P1-4 underwent removal of 2'-phosphate from the cyclized rNMPs to avoid interference with primers during PCR. Each tube was combined with 10X Tpt1 Buffer (4 µl), 50mM NAD<sup>+</sup> (8 µl), and 15.7µM Tpt1 (3 µl), then spun down and incubated at 30°C for 1 hour. Each sample was purified using RNA bead purification (section 3.1.6). 10 µl RT NF H<sub>2</sub>O was added to the 1.5ml tube before RNA beads were added, and 22 µl warm NF H<sub>2</sub>O was only added to tubes M and P1. After resuspension of the beads, M was incubated at RT for 5 minutes, while P1 was transferred into P2, resuspended, and so on for P3 and P4. 20 µl of supernatant was transferred from each 1.5ml tube into labeled PCR tubes.

### 3.1.9 Amplification by PCR

Samples were amplified via two PCR steps (PCR1 & PCR2). Mixtures were prepared for PCR1 containing DNA from the previous step (20 µl), Q5 2x Master Mix (25 µl), 10µM P1 forward primer (2.5 µl), and 10µM reverse primer (2.5 µl). Samples were mixed by pipette, spun down, and placed in the PCR machine for PCR1. Each PCR product was placed into one of two 1.5 ml tubes (M or P, respectively) and purified using column purification (section 3.1.12). 32 µl NF H<sub>2</sub>O was added to each column before final incubation and centrifugation. Each purified product was transferred to fresh PCR tubes (either M or P, respectively).

After purification, each sample underwent PCR2 to replicate strands and incorporate i5/i7 indices for Illumina sequencing. A mixture of Q5 High-Fidelity 2x Master Mix (25 µl), 10µM D70X (2.5 µl), 10µM D50X (2.5 µl), and NF H<sub>2</sub>O (15 µl) was mixed into each tube before being spun down and placed in the PCR machine for PCR2. After removal from the machine, samples from M and P were compared using PAGE (*Fig 3.4*).

### 3.1.10 Size selection

Each sample (M, P) from PCR2 underwent two steps of size selection to isolate the cyclised DNA (220bp to 650bp) for Illumina sequencing. Size selection consisted of DNA bead purification (section 3.1.11), with 12 µl beads for step 1 and 40 µl beads for step 2. Wash steps used 400 µl 80% EtOH. 52 µl warm NF H<sub>2</sub>O was added after drying in step 1, followed by transfer of 50 µl purified product into new (2x) prepared 1.5ml tubes. 23 µl water NF H<sub>2</sub>O was added after drying in step 2, and 21 µl of supernatant was transferred into new (2x) prepared 1.5ml tubes. Final purified samples were sent to the bioanalyzer to evaluate success of sample preparation before being sent for Illumina sequencing (*Fig 3.5*).

### 3.1.11 Bead purification

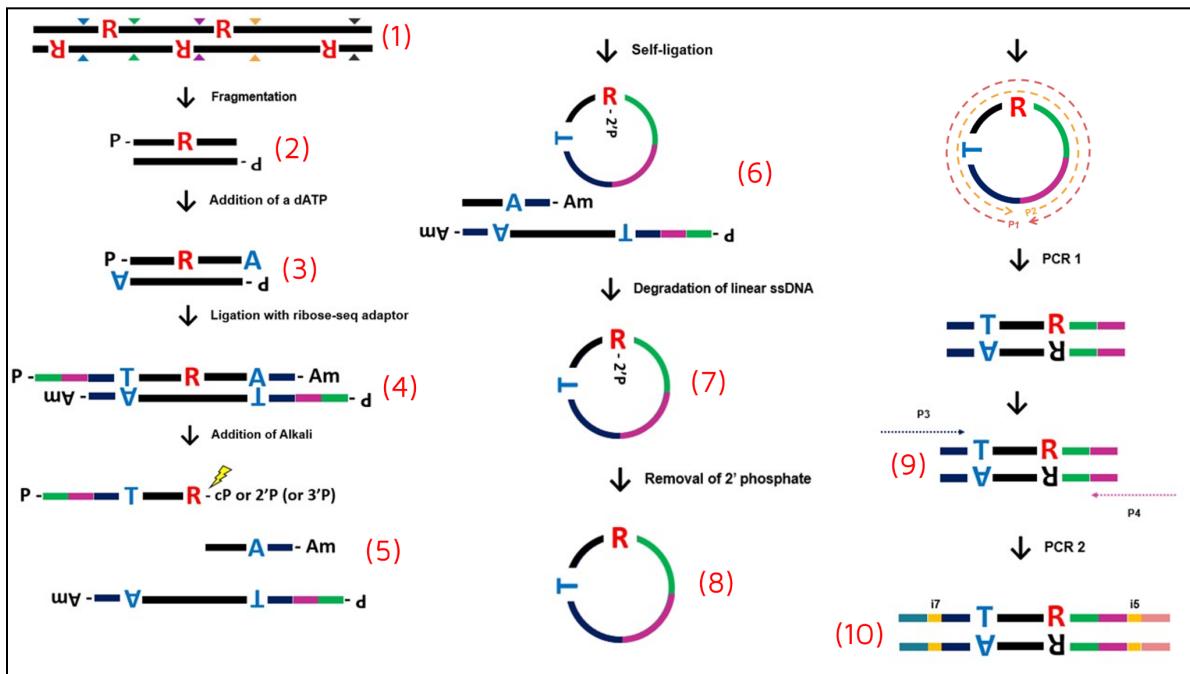
Sample tubes were diluted with 25 µl NF H<sub>2</sub>O and briefly spun down. Magnetic beads were vortexed before 90 µl were mixed into each sample tube by micropipette. Samples were incubated at RT for 5 minutes before incubating in a magnetic rack for 3 minutes. Samples

underwent two wash steps, with the supernatant first removed via P200 micropipette, followed by incubation at RT for 30 seconds with ethanol (EtOH) via P1000 micropipette. The final round of purification was followed by an additional supernatant removal step, then air-dried until just after the beads were no longer shiny in appearance (over-drying, i.e. cracking, was avoided to avoid DNA loss). Dried tubes were removed from the magnetic rack and mixed thoroughly with NF H<sub>2</sub>O by micropipette and spun down, then incubated at RT for 3 minutes, followed by incubation at RT for another three minutes in the magnetic rack to trap the beads. The purified product was then transferred out of the tubes.

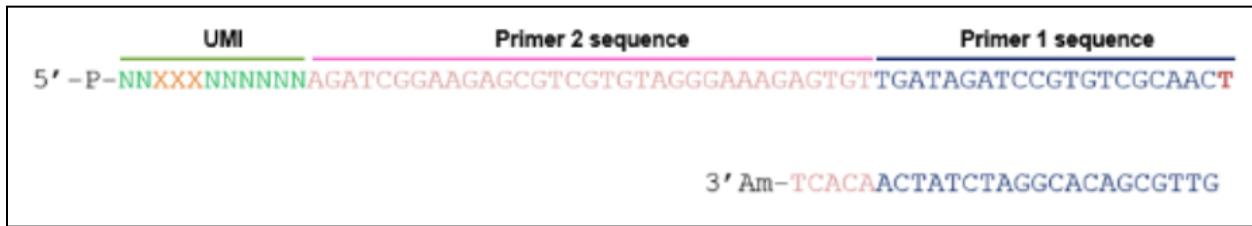
### 3.1.12 Column purification

Column purification used the QIAquick PCR Purification Kit. Each tube was treated with 250 µl Buffer PB to bind PCR products, then vortexed and briefly spun down. The contents of each tube was then transferred to a QIAquick spin column and centrifuged at 13,000 rpm for 1 minute; the flow-through was then discarded. 750 µl Buffer PE (wash) was added to each spin column and then centrifuged again, then centrifuged one additional time after discarding the flow-through to allow for removal of any residual wash buffer. The contents of each spin column was transferred to new 1.5mL tubes, then 30 µl NF H<sub>2</sub>O was added to each column and incubated at RT for 5 minutes before being centrifuged a final time for 1 minute.

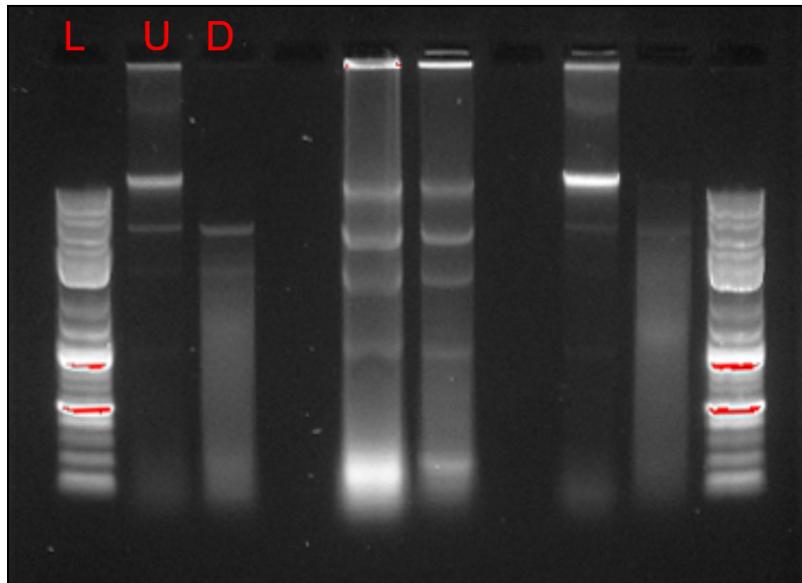
## 3.2 Results



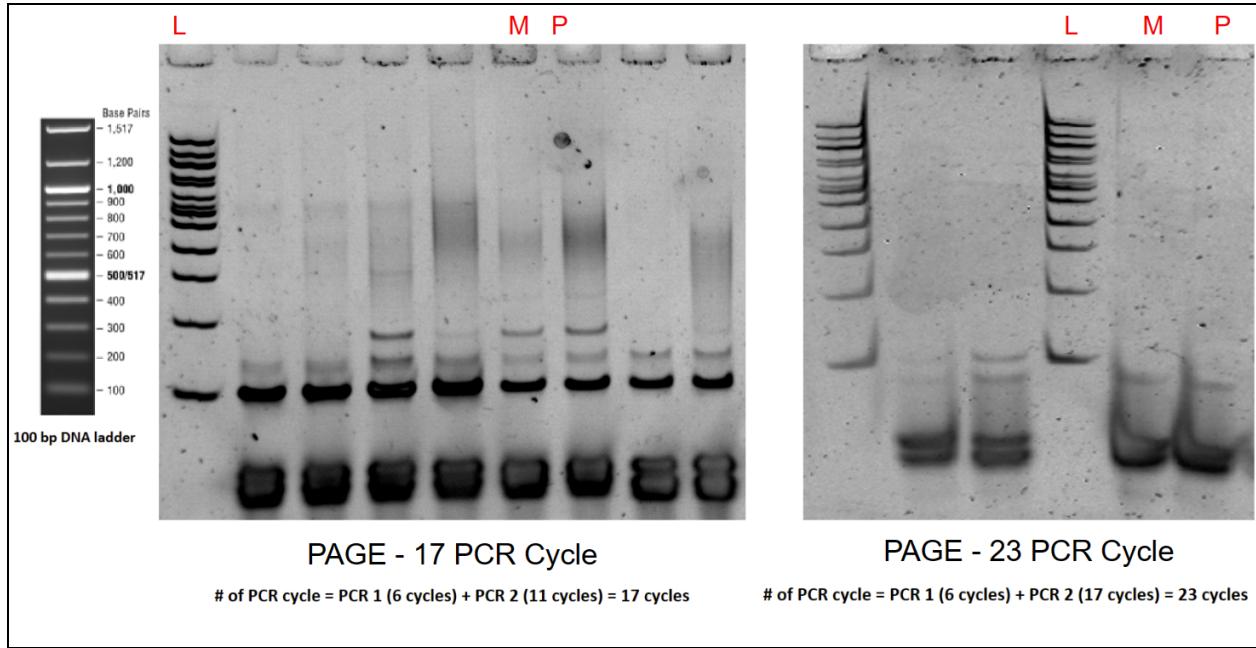
*Figure 3.1: Ribose-seq overview.* The gene product of each step is labeled as follows: DNA extraction (1), fragmentation and end repair (2), dA-tailing (3), adaptor ligation (4), alkali treatment (5), self-ligation by AtRNL (6), linear ssDNA degradation (7), 2'-phosphate removal (8), PCR1 (9), PCR2 (10), and size selection (unlisted).



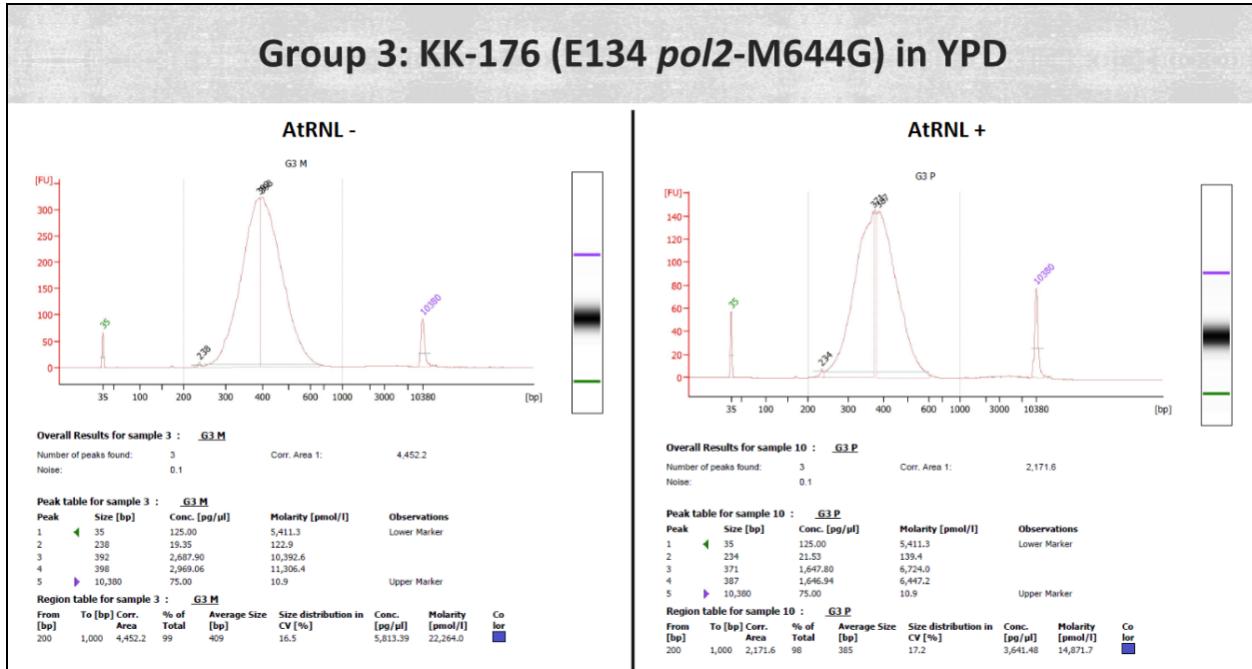
*Figure 3.2: Ribose-seq adaptor structure.* The ribose-seq adaptor consists of two primer sequences and a UMI (65 nt long & 25 nt short oligonucleotides, thus partially double-stranded). The UMI consists of 8 random (NNNNNNNN) and 3 specific (XXX, Barcode) nucleotides, and the Primer 1 & 2 sequences contain partial adaptor sequences to allow for recognition by the Illumina sequencing platform. The long strand contains a 3' overhang T at the end of Primer 1 on the long strand, and a 3' amino modifier (3' Am) on the opposite end of the short strand, preventing ligation of dNTPs.



*Figure 3.3: Fragmentation gel.* Agar gel in 1X TAE Buffer and loading dye. Comparing undigested (U) to digested (D) treatments and gene ladder (L). Undigested sample showed larger fragments around 5000 and 1000bp, while the digested sample showed a long streak between ~1500bp and 100bp.



*Figure 3.4: PAGE results for PCR.* PAGE stained with SYBR Gold, 1X TBE buffer and loading dye. AtRNL+ (P) and AtRNL- (M) samples were compared to a gene ladder (L). Constructed gene libraries represented by range between 220bp and 650bp. Bands below 200bp indicate primer dimers and other excess material. Two PAGEs were run after an initial batch of 11 cycles (left) and 17 cycles (right) of PCR 2 in order to compare sample clarity. Both M and P columns are similar for both gels, with ranges indicated between ~200 to 800 bp, although column P shows a higher density of genetic material than column M.



*Figure 3.5: Bioanalyzer results for KK-176 E134 pol2-M644G (G3). Bioanalyzer results show similar peaks for both treatments with (+) and without (-) AtRNL, with the main peak ranging between 200-700 bp for both. Average size 385bp for AtRNL+ and 409 bp for AtRNL-. These peaks fall within the expected range for AtRNL+, but no peak was expected for AtRNL-.*

## 4.0 Part III - Bioinformatics

Bioinformatic analysis was done using the UNIX environment within Georgia Tech's Partnership for an Advanced Computing Environment (PACE) Instructional Cluster Environment (ICE). Modules were installed and loaded via anaconda3, and additional analysis (section 4.1.4) utilized an installed R environment.

### 4.1 Bioinformatic analysis

Due to insufficient evidence of adequate gene library construction (section 5.2), the strain from section 3.1 was replaced with a pre-existing gene library (E134 *rnh201 pol2-4* in YPD); all bioinformatic analysis will therefore utilize this strain. A table listing the finalized samples, as well as associated UMIs and FS#s, is available in section 4.2 (*Table 4.1*).

*Table 4.1: Finalized dataset for bioinformatic analysis. Old/replacement datasets are grayed out.*

FS #	Sample	Enzymes	UMI (5'->3')	Builder	Barcode
FS379	KK-44 (E134 WT) in YPD	HaeIII, RsaI	CCG	Group1	CGG
FS162	E134 in YPD	HaeIII, RsaI	CTG	Group2	CAG
FS138	E134 rnh201 in YPD	HaeIII, RsaI	TGA	Group8	TCA
FS321	E134 rnh201 pol2-4	HaeIII, RsaI	CCG	Group3	CGG
FS383	TY-148 (E134 rnh201 pol2-M644G) in YPD	HaeIII, RsaI	TGT	Group7	ACA
FS382	TY-146 (E134 rnh201) in YPD	HaeIII, RsaI	GCT	Group6	AGC
FS380	KK-44 (E134 WT) in YPGal	HaeIII, RsaI	AGC	Group4	GCT
FS381	KK-176 (E134 pol2-M644G) in YPGal	HaeIII, RsaI	AGC	Group5	GCT
FS384	TY-148 (E134 rnh201 pol2-M644G) in YPGal	HaeIII, RsaI	CCG	Group9	CGG

## 4.2 Results

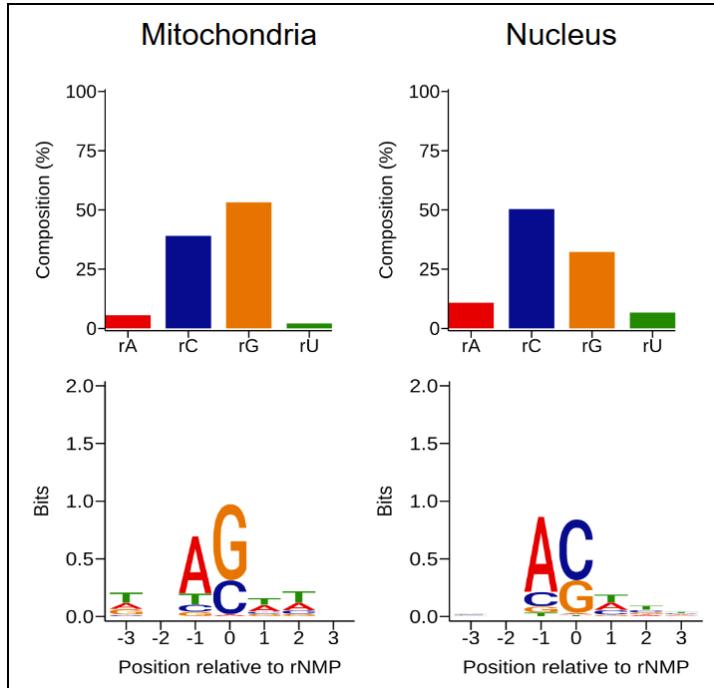
Results are listed according to the bioinformatic process involved. See associated discussion sections 5.3.0-3 for further analysis.

### 4.2.1 Ribose-map

Reads were initially trimmed using TrimGalore (ribosemap environment), generating .fastq files for use in ribose-map modules. A config file was supplied by the TA and adjusted accordingly. Ribose-map alignment and coordinate modules (*Table 4.2*) were first run sequentially, followed by ribose-map hotspot, sequence, and composition modules (*Figure 4.1*). Note that, due to an unresolved error in the code, no sequence plots were generated correctly for any group and are thus unlisted. Ribose-map results can be located at the following directory within PACE: /storage/home/hpaceice1/shared-classes/materials/bios4590b/ribosemap\_analysis/results/FS384\_KN.

*Table 4.2: Combined alignment (left) and coordinate (right) reads data.*

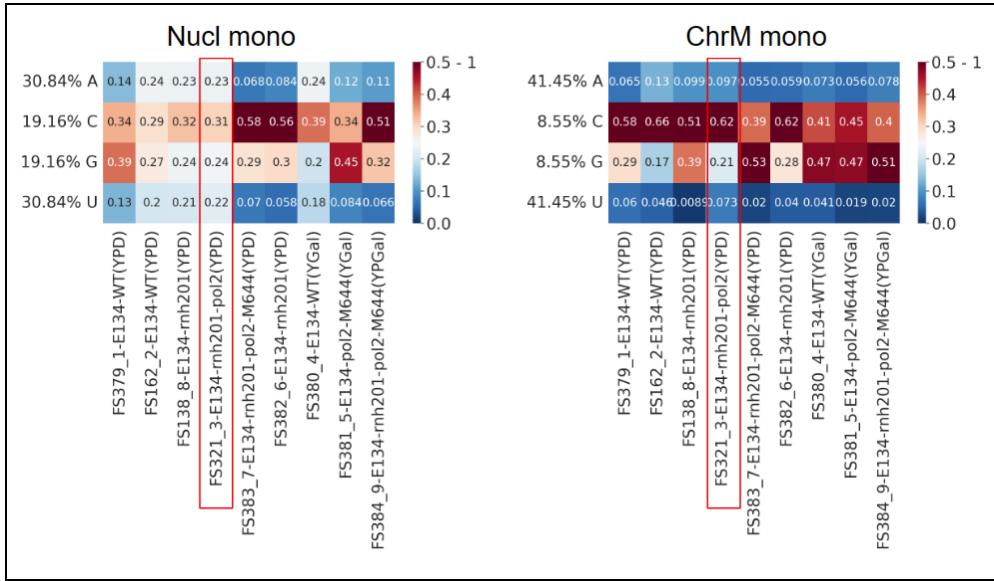
Raw reads	Alignment rate	Reads in nucleus	Reads in ChrM	Total reads	% rNMPs in Nucl/Total	% rNMPs in ChrM/Total
4768076	96.02%	937227	19034	956261	98.01%	1.99%



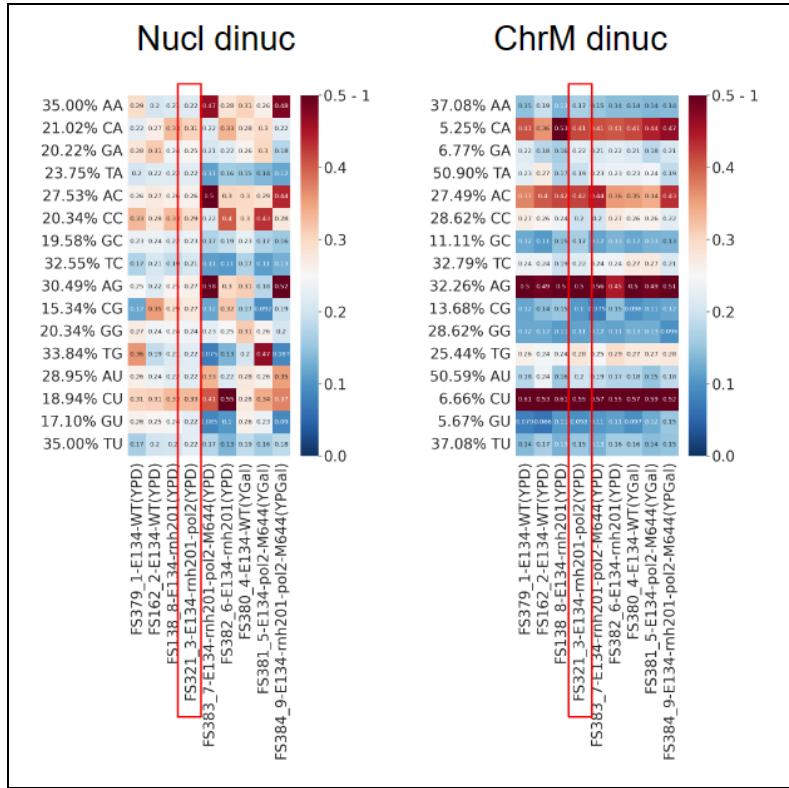
*Figure 4.1: Combined composition (top) and hotspot (bottom) plots.* Plots compare overall rNMP composition in mitochondrial (chrM) versus nuclear (nucl) DNA, as well as dNMP type bias up to three bp away from embedded rNMPs in both types of DNA.

#### 4.2.2 Heatmaps

Heatmaps were generated (misc environment, conda) using TA pre-generated pathways/script and ribose-map data. Mononucleotide, dinucleotide, and trinucleotide heatmaps were generated; only mononucleotide (*Figure 4.2*) and dinucleotide (*Figure 4.3*) were analyzed. See section 5.3.2 for in-depth analysis. Heatmaps are available in the following directory:  
[/storage/home/hpaceice1/shared-classes/materials/bios4590b/heatmap\\_analysis/results/plots\\_KN](http://storage/home/hpaceice1/shared-classes/materials/bios4590b/heatmap_analysis/results/plots_KN)



*Figure 4.2: Mononucleotide plots for nuclear (left) and mitochondrial (right) DNA. Values are normalized. The y-axis depicts percentage (%) ribonucleotide by type between all samples and the x-axis depicts individual strain-media combinations. The strain that our group used for bioinformatic processing is boxed in red. Blue cells indicate bias against a particular type of rNMP, while red cells indicate bias towards (enrichment of) a particular type of rNMP.*



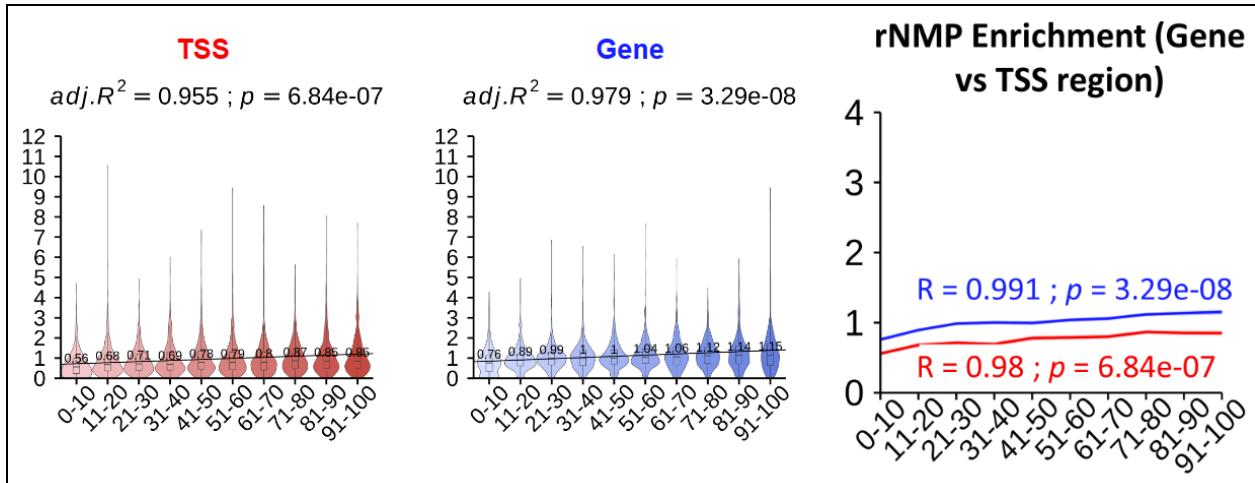
*Figure 4.3: Dinucleotide plots for nuclear (left) and mitochondrial (right) DNA. Values are normalized. The y-axis depicts the percentage (%) dNMP→rNMP combinations (dNMP preceding rNMP) between all samples, and the x-axis depicts individual strain-media combinations. The strain that our group used for bioinformatic processing is boxed in red. Blue cells indicate bias against a particular type of preceding dNMP, while red cells indicate bias towards (enrichment of) a particular preceding dNMP.*

#### 4.2.3 RNA-seq

RNA-seq and ribose-seq data analysis was used for isolated RNA (section 2.1.2). TA pre-generated files, directories, and pipelines were utilized and followed. Sequences were trimmed using TrimGalore (ribosemap, conda) then aligned using Bowtie2. Samtools modules were used to assign pairs for paired end reads (fixmate), sorting (sort), naming (addreplacerg) and indexing (index) before reads were counted using featureCounts. Overall alignment rate was found to be 98.16%.

#### 4.2.4 Omics integration

We also checked for multi-omics correlations (*Figure 4.4*) within the data in order to identify potential influences on cellular mechanisms behind rNMP incorporation bias. The association between transcription ( $\pm 200$  TSS) and rNMP incorporation (gene) was checked using RNA-seq data and ribose-seq data, respectively. rNMP enrichment factor (REF) was obtained by a pre-written bash script, and R was used to correlate RNA-seq and ribose-seq data.



*Figure 4.4: Correlation line plots for multi-omics analysis.* The above plots depict relative rNMP enrichment in the TSS region (left), gene-coding region (middle), and overall comparison between the two regions (right). The violin box plots (left, middle) have y-axes of rNMP enrichment factor (REF) and x-axes showing expression percentile, overall highlighting trends in rNMP enrichment throughout all samples.

## 5.0 Discussion

The present study aimed to examine the process of ribonucleotide incorporation into DNA and its effects on DNA replication and repair mechanisms, using a series of molecular biology techniques and bioinformatics analyses. Our experimental process was divided into three parts, each focusing on a specific aspect of the study. In Part I, we used a series of transformation steps to incorporate the G418 marker gene, as well as replica plating to isolate transformed colonies. Part II focused on DNA extraction and preparation using ribose-seq, resulting in a constructed gene library; agar gel/PAGE electrophoresis were used for qualitative analysis of samples before bioanalysis and sequencing. Finally, Part III delved into a comprehensive bioinformatic analysis of ribose-map, heatmap, and omics data to understand the patterns and preferences in ribonucleotide incorporation in both nuclear and mitochondrial DNA. In the following paragraphs, we discuss the main findings and implications of these results, shedding light on the intricacies of ribonucleotide incorporation in DNA and its potential impact on cellular processes.

## 5.1 Part I

In the replica plating process (*Fig 2.3*), plates 2-7 were used sequentially to first select for the desired E134 genotype, followed by G418 to test for transformant colonies, Hygro as an additional antibiotic selection plate, YPG and YPD for growth, and MAT for mating pairs. Pre-transformation plates (*Fig 2.3a*) demonstrated an expected lack of G418 colonies, while colonies appeared on this plate in *Figure 2.3b*. Mating pairs were not labeled post-transformation, but the growth of only half of the mating pairs compared to both mating pair

types in *Figure 2.3a* provided further evidence that transformation occurred. Gel electrophoresis results (*Fig 2.4*) reinforced the evidence of successful incorporation of the G418 marker gene.

In the colony PCR analysis (*Fig 2.5*), W1 samples (M1-4) were the only ones to display the expected gel results, with bands indicating that reactions with mixtures 1 and 2 resulted in longer strands. Although clear bands appeared for T1A M1-2, this was contrary to the expected reaction with M3-4; the wells may have been accidentally loaded in reverse order. Due to clear banding, both WT and T1A samples were further processed for subsequent DNA extraction. T1B and T2A samples appeared to have some interaction with M1-2, and potentially M3-4 for T1B; regardless of expectations, the gel was too blurred in these columns for any significant results, so these samples were not processed in further steps. Sanger sequencing of WT and T1A samples supported the initial transformation of the G418 gene.

Regarding the restriction digestion (*Fig 2.6*), while DD and SX treatments seem to have proceeded as expected, with two bands in DD and one in SX, the SS treatment likely did not occur as intended. Although SS shows one band, as did SX, this band is around 5000bp rather than the expected ~1000bp. The size of the band in SS is similar to that of an undigested full sample, suggesting there was likely an error related to the addition of Scal to that sample. Since DD indicates that Scal still retains enzymatic activity (i.e., two bands), this implies that there was likely insufficient Scal added to the SS treatment.

## 5.2 Part II

In *Table 3.1*, we observed issues with our samples relatively early on. Our DNA extraction yielded only around 9 µg, while we had hoped for a 20-40 µg yield. This suggests that our group, as well as others, may not have been as precise as necessary - as mentioned in section 3.1.1, no group managed to extract more than 13 µg of DNA. This lower-than-expected yield could be attributed to a lack of precision, an inefficient protocol, or both. Notably, there was a blockage during the elution of the DNEasy spin columns, which, based on the Nanodrop purity values below 1.8, was likely protein-based. A longer incubation period or higher concentrations of proteinase K might have improved purity and overall concentrations. Moreover, insufficient DNA may have impacted the success of subsequent steps, with lingering protein concentrations potentially limiting enzymatic activity.

Fortunately, it appears that enzymatic activity during fractionation (section 3.1.2) was not adversely affected by the presence of excess peptides. The key result from *Figure 3.3* shows that the digested sample (D) displayed a wide-ranging "smudge" across 1000 bp to <100 bp, indicating the presence of fragments of various sizes, as opposed to the undigested sample (U) with larger cell material present (~5000 to 1000 bp). Faintly-visible bands at the upper range of the digested sample may indicate the presence of excess peptides mentioned in *Table 3.1*. However, *Figure 3.3* overall suggests that fragmentation was not an issue.

Our PAGE results (*Fig 3.4*) displayed material in the ~200-800 bp range for both AtRNL+ (P) and AtRNL- (M) samples. We expected a streak of genetic material for P and a lack of material in that range for M, similar to the column to the right of the first P column. Although P shows a higher concentration of genetic material than M, the presence of material in that range for M indicates that something went wrong. While size selection appeared to occur as expected, the presence of a peak in AtRNL- was unexpected, as it represents treatment without AtRNL, leading to no cyclization of rNMP-embedded fragments and complete degradation by T5 exonuclease. This result is consistent with the PAGE gels (*Fig 3.4*), where both M and P columns display DNA streaks within the constructed library ranges (220 bp - 650 bp). The peak in AtRNL- indicates either unsuccessful degradation of linear ssDNA or contamination in the M sample during self-ligation. Unfortunately, this meant that the AtRNL+ samples were not worth sequencing (data would be non-significant or unclear), and bioinformatic analysis would instead utilize a similar strain (E134 *rnh201 pol2-4*) in YPD.

### 5.3 Part III

Discussions of bioinformatic results described throughout section 4.2 are split into three distinct portions, as is the case for the aforementioned results section.

#### 5.3.1 Ribose-map analysis

As ribose-map involves data generation by various modules, there is a decent amount of data to work with for analysis. Based on results from *Table 4.2*, it is crucial to note that most reads (>98%) appeared within nuclear DNA as opposed to mitochondrial DNA. While it would make sense to find a significant enrichment of rNMPs within nuclear DNA of the double-K.O. mutant strain examined, the significant lack of rNMPs in mitochondrial DNA implies that, at least for YPD, mutagenesis selecting for increased rNMP incorporation does not impact mitochondria. Considering the mechanism involved with energy production under fermentation, this may make sense (considering mitochondria is not important in fermentation compared to its role in respiration); however, a key note here is that *almost no* rNMPs were present, meaning it would make sense to look for similar trends in samples grown in YPGal.

Looking at *Figure 4.1*, there seems to be a general bias towards incorporation of rNMP-C and rNMP-G in both mitochondrial and nuclear DNA; there does appear to be precedence for this in the literature (Balachandar et al. 2020). Interestingly, preference for each ribonucleoside seems to be swapped between the two DNA sources, as rC is higher in the nucleus compared to a similarly-high rG in mitochondria. Looking down to the generated hotspot plots, there appears to be preference towards dNMP-A directly preceding rC or rG; there is precedence for this in the literature (Balachandar et al. 2020), and this pattern is further discussed in section 5.3.2. Interestingly, there appears to be a slight bias towards dNMP-T at various bp around a central rNMP - while this pattern is partially observed in *Figure 4.3* (as discussed in section 5.3.2), this

pattern also appears to apply up to two bp in mitochondria, and up to three bp in the nucleus. Interestingly, there appears to be a consistent lack of bias towards any particular dNMP at the -2bp position from the rNMP in both the mitochondrial and nuclear DNA.

### 5.3.2 Heatmap analysis

The generated mononucleotide heatmap (*Fig 4.2*) demonstrates a consistent pattern across all groups, including wild type grown in YPD and YPG. rC is integrated at higher rates than other ribonucleotides, and rA and rU are consistently lacking in all mutants and wildtypes cultivated in YPD and YPG. Interestingly, galactose appears to slightly boost the incorporation rate of rG along with rC. Ribonucleotide incorporation also appears less biased in most YPD samples, while there tends to be somewhat-inconsistent, but stronger bias towards particular rNMPs in YPGal. This may be due to yeast cells grown in galactose incorporating ribonucleotides into nuclear DNA at a faster pace because of the different carbon source used for energy production; compared to the glycolytic system used by *S. cerevisiae* on glucose, the galactose metabolism pathway produces more ATP and NADPH, which can promote enhanced nucleotide synthesis and ribonucleotide incorporation into nucleus DNA (Wu et al. 2014).

The difference in DNA replication machinery between mitochondria and the nucleus may account for the increased rate of ribonucleotide incorporation in mitochondrial DNA. DNA polymerases in the nucleus synthesize DNA using deoxyribonucleotides (dNTPs) as the basic unit, while a unique group of enzymes in mitochondria utilize both dNTPs and ribonucleotides (rNTPs) as substrates (Clayton, 1982). The special characteristics of mitochondrial DNA, such as its small size, circular structure, and proximity to the electron transport chain that generates reactive oxygen species (ROS), have been related to the employment of rNTPs during mitochondrial DNA replication (Lee, 2010).

Although rNMP incorporation occurred in both nuclear and mitochondrial DNA, the rate of incorporation has been found to typically be higher in nuclear DNA. This disparity can be attributed to several factors, such as the overall effect of rNMP incorporation being diminished by replication errors in mitochondrial DNA being spread over several copies (Rhee, 2018). The increased stability of the C-G base pair compared to the A-T base pair may help explain the preference for ribonucleotide C or G incorporation into DNA. The stability of the double-stranded DNA structure is crucial for the integrity and functionality of the genetic material, enhancing the binding of proteins to DNA and aiding in the prevention of mistakes in DNA replication and transcription.

Looking instead at the dinucleotide heatmap (*Fig 4.3*), some patterns immediately reappear compared to the mononucleotide heatmap. Similarly to the mononucleotide heatmap, there seems to be a strong bias towards certain rNMPs (and in this case, dNMP-rNMP pairings) in mitochondrial DNA compared to nuclear DNA. Notably, a pattern of preference for rNMP-C and

rNMP-G to be preceded by dNMP-A has already been recorded in the literature; considering the ribose-seq process interacted predominantly with nuclear DNA, this pattern seems to be retained in the largely-untouched mitochondrial DNA. Looking carefully at the nuclear dinucleotide plot, while the trend is not as clear as in mitochondrial DNA, this trend seems somewhat present in nuclear DNA as well (Balanchander et al. 2020). Also recorded in the literature, a large proportion of rNMPs seems to be preceded by dNMP-A (especially rNMP-U, with a pairing of over 50% between all samples) disproportionate to other dNMPs (El-Sayed et al. 2021). Compared to the general lack of bias towards any particular rNMP identified in the mononucleotide plot for our analyzed library, we also see a similar lack of bias for any particular dNMP-rNMP pair within our library, potentially implying an overall lack of incorporation bias for that particular strain.

While most rNMPs in the mitochondrial dna of the mononucleotide plot had a strong bias towards or against incorporation, not all dNMP-rNMP pairs appear as strongly biased: despite >50% TA or >50% AU in samples overall, these pairings were not strongly selected for or against. While AU pairings vary more between nuclear DNA samples, this pattern generally remains for TA (do note, however, that % instances of either is cut in half for both pairings). Possibly related mechanistically to the findings by Balachander et al. (2020), there also seems to be a distinct bias against preceding rNMPs with dNMP-T. Notably, this appears most frequently in some mutant strains in YPD and all strains in YPGal, perhaps indicating that something in the respiration pathway under YPGal has a similar effect to mutation on strains in YPD. Interestingly, while there does also seem to be bias against the pairing in mitochondrial DNA, this pattern is not quite as strong as in most nuclear DNA samples; this pattern is also consistent across samples in both YPD and YPGal, implying that energy pathway (and associated change in gene expression) has little impact on mitochondrial DNA in regards to rNMP incorporation (which is further supported by *Table 4.2*, as described in section 5.3.1).

### 5.3.3 Omics analysis (RNA-seq)

Looking at the RNA-seq (section 4.2.3) and multi-omics results (*Fig 4.4*), a few points stand out. Noting an overall alignment rate of 98.16% and the strong R<sup>2</sup> and p values for all plots of *Figure 4.4*, these values would imply that any visible trends would likely be significant. Although the difference between enrichment in the TSS and gene-coding regions is very slight (~0.2-0.3 REF), high confidence in both trends does indicate a significant increase in overall enrichment for the gene-coding region compared to the TSS ( $\pm 200$  bp). This general increase in enrichment could potentially be linked to the slight increase in overall rNMP diversity in nuclear DNA compared to mitochondrial DNA as seen in *Figure 4.1*, but further research may be required to prove or disprove this relationship.

## 6.0 Conclusions

Our overall findings suggest that, while differences in rNMP incorporation bias do seem to exist between samples in glucose versus galactose, these differences are not entirely consistent, necessitating further research. Although changes in incorporation bias are evident in the nucleus, mitochondrial DNA remained unaffected by alterations in rNMP incorporation rate and growth medium, signifying that gene expression-related rNMP incorporation mechanisms are primarily linked to the nucleus. This may be due to the unique characteristics of mitochondrial DNA, such as its separate replication machinery discussed in section 5.3.2.

The occurrence of unexpected results during gene library preparation, such as lower DNA yields and unexpected peaks in AtRNL- samples, highlights the need for further optimization of experimental procedures. Moreover, the observed trends in ribonucleotide incorporation patterns and potential influences of energy pathways on rNMP incorporation warrant further investigation. Future research should analyze additional samples under varying conditions, concentrating on unveiling the mechanisms involved in rNMP incorporation for both nuclear and mitochondrial DNA. Furthermore, the role of replication machinery in determining rNMP incorporation rates and preference should be explored in greater depth to better understand these underlying mechanisms and the consequences of these processes.

## References

- Balachander, S., Gombolay, A. L., Yang, T., Xu, P., Newnam, G., Keskin, H., El-Sayed, W. M. M., Bryksin, A. V., Tao, S., Bowen, N. E. D., Schinazi, R. F., Kim, B., Koh, K. D., Vannberg, F. O., and Storici, F. (2020). Ribonucleotide incorporation in yeast genomic DNA shows preference for cytosine and guanosine preceded by deoxyadenosine. *Nature Communications*, 11(2447), 1-14. <https://doi.org/10.1038/s41467-020-16152-5>
- Balachander, S., Yang, T., Newnam, G., El-Sayed, W. M. M., Koh, K. D., and Storici, F. (2019) Capture of Ribonucleotides in Yeast Genomic DNA Using Ribose-Seq. *Yeast Systems Biology*, 2049, 17-37. [https://doi.org/10.1007/978-1-4939-9736-7\\_2](https://doi.org/10.1007/978-1-4939-9736-7_2)
- Clayton, D. A. (1982). Replication of animal mitochondrial DNA. *Cell*, 28(4), 693-705. doi: 10.1016/0092-8674(82)90094-x
- El-Sayed, W. M. M., Gombolay, A. L., Xu, P., Yang, T., Jeon, Y., Balachandar, S., Newnam, G., Tao, S., Bowen, N. E., Bruna, T., Borodovsky, M., Schinazi, R. F., Kim, B., Chen, Y., and Storici, F. (2021) Disproportionate presence of adenosine in mitochondrial and chloroplast DNA of Chlamydomonas reinhardtii. *iScience*, 24, 1-14. <https://doi.org/10.1016/j.isci.2020.102005>

Koh, K. D., Balachander, S., Hesselberth, J. R., & Storici, F. (2015). Ribose-seq: global mapping of ribonucleotides embedded in genomic DNA. *Nature Methods*, 12(3), 251-262. doi: 10.1038/NMETH.3259

Lee, C. M., Sedivy, J. M., & Ideker, T. (2010). Carcinogen exposure and stress-induced proliferation share a common signaling pathway. *Annals of the New York Academy of Sciences*, 1203(1), 89-97. doi: 10.1111/j.1749-6632.2010.05617.x

Rhee, D. B., & Pinto, R. M. (2018). DNA replication errors and the rise of mitochondrial DNA variants. In *Mitochondrial DNA: Methods and Protocols*, 81-94. Humana Press, New York, NY. [https://doi.org/10.1007/978-1-4939-7758-3\\_7](https://doi.org/10.1007/978-1-4939-7758-3_7)

*Saccharomyces cerevisiae* *Saccharomyces cerevisiae* - an overview | ScienceDirect Topics. [online] <https://www.sciencedirect.com/topics/neuroscience/saccharomyces-cerevisiae> (Accessed April 28, 2023)

Wu, X., Li, L., Jiang, H., & Shen, G. (2014). Comparison of the effects of glucose and galactose on the metabolism of *Saccharomyces cerevisiae* using a bioreactor. *Journal of Industrial Microbiology & Biotechnology*, 41(1), 189-196. doi: 10.1007/s10295-013-1378-6

Yanagida, M. (2002). The model unicellular eukaryote, *Schizosaccharomyces pombe*. *Genome Biology*, 3(3), comment 2003.1. <https://doi.org/10.1186/gb-2002-3-3-comment2003>