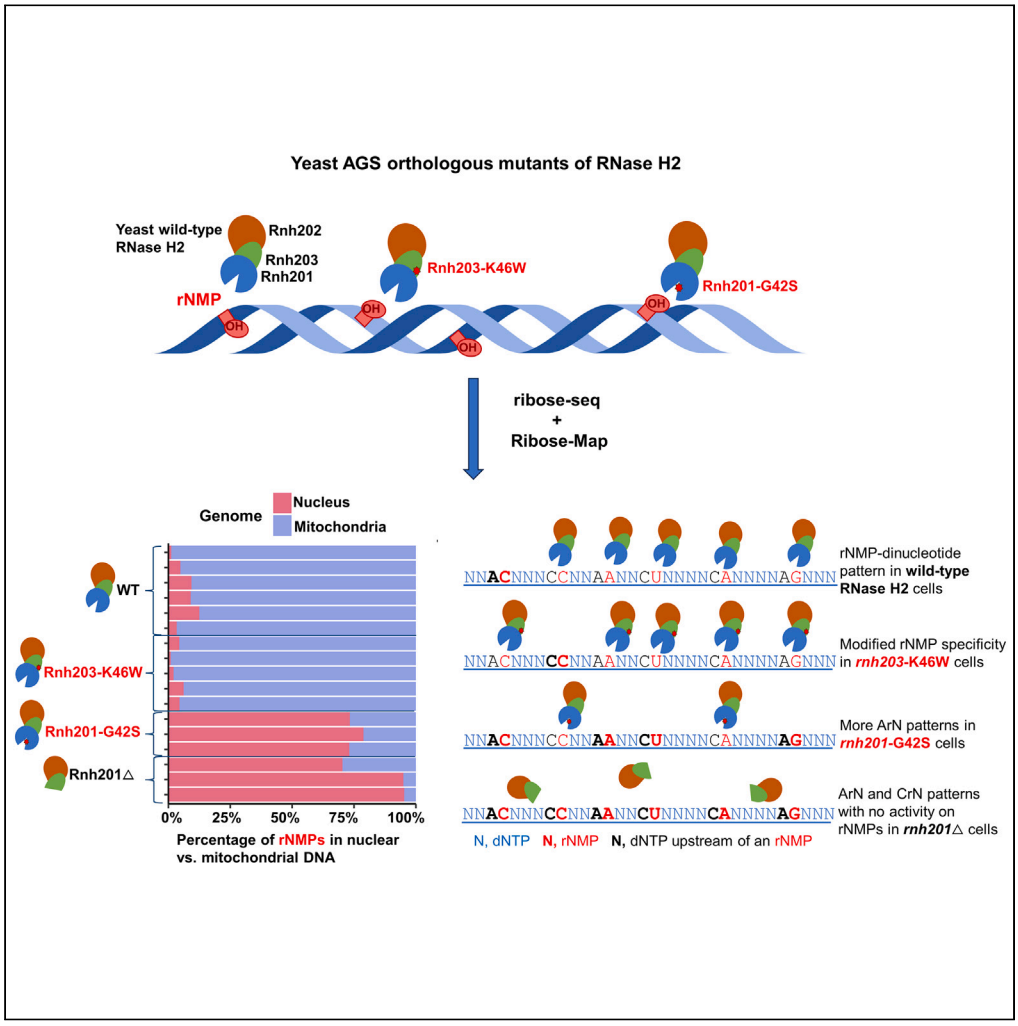
iScience

CellPress
OPEN ACCESS

Article

Distinct features of ribonucleotides within genomic DNA in Aicardi-Goutières syndrome ortholog mutants of *Saccharomyces cerevisiae*



Deepali L. Kundnani, Taehwan Yang, Alli L. Gombolay, ..., Zeel H. Mehta, Celine Mouawad, Francesca Storici

storici@gatech.edu

Highlights

Yeast AGS-mutant *rnh201*-G42S shows lower ribonucleotide activity

Yeast AGS-mutant *rnh203*-K46W has altered specificity on ribonucleotides

rC rises in AGS yeast mutant nuclear and mitochondrial DNA

In the G42S ortholog, rNMPs are preceded by dA; in R69W, rNMPs are preceded by dC

Kundnani et al., iScience *27*, 110012
June 21, 2024
https://doi.org/10.1016/j.isci.2024.110012

iScience

CellPress
OPEN ACCESS

Article

Distinct features of ribonucleotides within genomic DNA in Aicardi-Goutières syndrome ortholog mutants of *Saccharomyces cerevisiae*

Deepali L. Kundnani,[1] Taehwan Yang,[1] Alli L. Gombolay,[1,2] Kuntal Mukherjee,[1] Gary Newnam,[1] Chance Meers,[1,3] Ishika Verma,[1] Kirti Chhatlani,[1] Zeel H. Mehta,[1] Celine Mouawad,[1] and Francesca Storici[1,4,*]

SUMMARY

Ribonucleoside monophosphates (rNMPs) are abundantly found within genomic DNA of cells. The embedded rNMPs alter DNA properties and impact genome stability. Mutations in ribonuclease (RNase) H2, a key enzyme for rNMP removal, are associated with the Aicardi-Goutières syndrome (AGS), a severe neurological disorder. Here, we engineered orthologs of the human RNASEH2A-G37S and RNASEH2C-R69W AGS mutations in yeast *Saccharomyces cerevisiae*: *rnh201*-G42S and *rnh203*-K46W. Using the ribose-seq technique and the Ribose-Map bioinformatics toolkit, we unveiled rNMP abundance, composition, hotspots, and sequence context in these AGS-ortholog mutants. We found a high rNMP presence in the nuclear genome of *rnh201*-G42S-mutant cells, and an elevated rCMP content in both mutants, reflecting preferential cleavage of RNase H2 at rGMP. We discovered unique rNMP patterns in each mutant, showing differential activity of the AGS mutants on the leading or lagging replication strands. This study guides future research on rNMP characteristics in human genomes with AGS mutations.

INTRODUCTION

Aicardi-Goutières syndrome (AGS) is a rare genetic disorder that predominantly affects the brain, skin, and immune system.[1] AGS is typically diagnosed in infancy or early childhood and causes a range of neurological symptoms including developmental delays, intellectual disabilities, seizures, and movement abnormalities.[2] AGS is characterized by the calcification of the basal ganglia, chronic cerebrospinal fluid (CSF) lymphocytosis, negative serological investigations for common prenatal infections and raised levels of interferon alpha (IFN-a) in CSF and serum.[3] Patients with AGS have mutations in nucleic-acid metabolism regulators such as Three Prime Repair Exonuclease 1 (TREX1), SAM And HD Domain Containing Deoxynucleoside Triphosphate Triphosphohydrolase 1 (SAMHD1), Ribonuclease (RNase) H2, Adenosine Deaminase Acting on RNA 1 (ADAR1), and Interferon Induced with Helicase C Domain 1 (IFIH1).[3] More than 50% of patients with AGS exhibit mutations in one of the three subunits of the RNase H2-complex.[4]

The presence of ribonucleotides in DNA happens in the form of ribonucleoside monophosphates (rNMPs) and is mainly caused by the DNA polymerase incorporation of ribonucleotide triphosphates (rNTPs) or potentially by the ligation of RNA primers during DNA-lagging strand synthesis.[4,5] RNase H2 is a critical enzyme complex involved in ribonucleotide excision repair (RER) during DNA replication and repair. Failure to efficiently remove the embedded rNMPs in DNA can contribute to disease, compromising the integrity of the genome.[6,7] Moreover, rNMPs can hinder the interaction of proteins with DNA during DNA replication and transcription, potentially disrupting normal cellular functions.[8] Therefore, the efficient removal of rNMPs by RNase H2 is vital for maintaining genomic stability and preventing the accumulation of DNA damage.[9] Recently, the field has witnessed a surge of interest in the study of embedded rNMPs, with select groups developing sequencing techniques to capture rNMPs in genomic DNA.[10] One of the pioneering techniques, published in 2015, is *ribose-seq*,[11,12] which employs alkaline hydrolysis and a special spliced-RNA ligase to capture embedded rNMPs along with their upstream DNA. This method facilitates the generation of high-throughput sequencing libraries without bias toward any specific genomic sequence context.

Human RNase H2 consists of 3 subunits (A, B, and C), with RNase H2A as its catalytic subunit. In yeast *Saccharomyces cerevisiae*, the orthologous proteins to human RNase H2 subunits A, B and C are Rnh201, Rnh202, and Rnh203, respectively.[13–15] Studies in yeast have revealed that mutations in these genes lead to genomic instability, DNA damage, and replication defects.[16–18] Defective RNase H2 function in yeast increases embedded-rNMP presence, leading to replication stress, DNA breaks, and activation of damage response pathways, revealing the detrimental effects of rNMP presence in DNA.[16,19] Previous studies reporting rNMPs embedded in the nuclear genome of *S. cerevisiae* cells having the *rnh201*Δ genotype show preference for deoxyadenosine immediately upstream of the most abundant rCMPs and rGMPs.[20]

[1]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA
[2]Bacterial Special Pathogens Branch, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA
[3]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA
[4]Lead contact
*Correspondence: storici@gatech.edu
https://doi.org/10.1016/j.isci.2024.110012

RNase H2 Protein subunit alignment for *H.sapiens* and *S.cerevisiae*

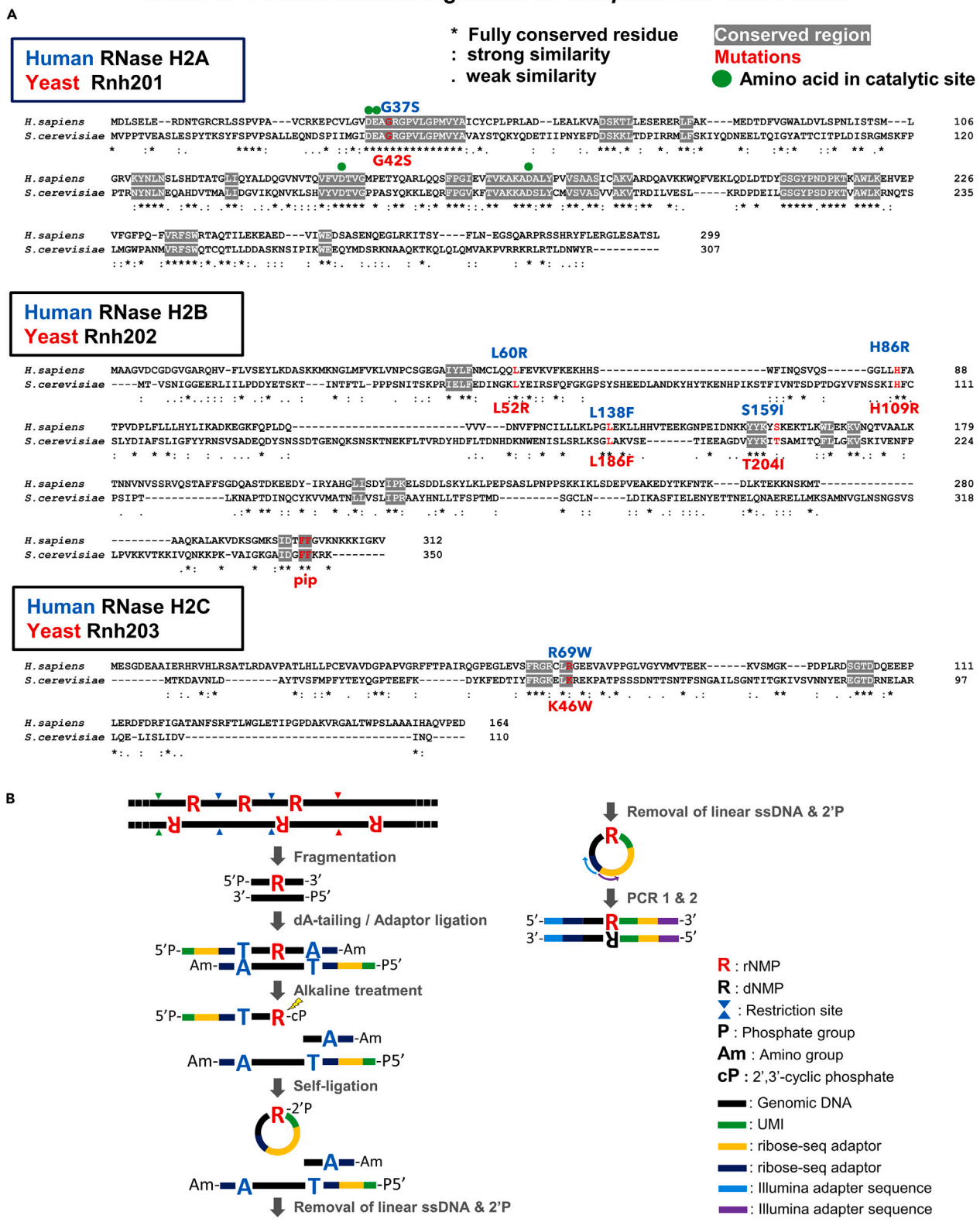


Figure 1. Schemes of human and yeast RNase H2 subunit amino acid sequence alignment with common AGS mutations and ribose-seq technique

(A) Alignment of amino acid sequences of RNase H2 catalytic subunit H2A, and accessory subunits H2B and H2C from *H. sapiens* and *S. cerevisiae*.[25] The gray highlights show the most conserved parts of the protein complex.[25] Amino acids highlighted in red show AGS mutations in human RNase H2 subunits and the corresponding mutations in the *S. cerevisiae* ortholog enzyme subunits within fully conserved or similar amino acids.[26] The amino acids with green circle on top represent those present in the catalytic site of the enzyme subunit A.[13] The mutations characterized in this study found in the conserved regions of RNase H2 subunits are RNase H2A-G37S that corresponds to Rnh201-G42S in *S. cerevisiae*, and RNase H2C-R69W that corresponds to Rnh203-K46W in *S. cerevisiae*.

(B) Schematic of the ribose-seq protocol. Genomic DNA is fragmented, dA-tailed, and ligated to a molecular barcode–containing sequencing adaptor. Alkali treatment denatures the DNA and cleaves at rNMP sites, exposing 2′,3′-cyclic phosphate and 2′-phosphate termini, which are recognized and ligated to the 5′-phosphate ends of the same molecules. Linear, unligated fragments are degraded using T5 exonuclease. Circular DNA molecules are PCR-amplified and sequenced at an Illumina platform.

Frequencies and patterns of rNMP embedment around early firing autonomously replicating sequences (ARSs) in *rnh201*Δ yeast strains show distinct preferences for deoxyadenosine upstream of the rNMPs embedded in the leading strand, and deoxycytidine upstream of the rNMPs embedded in the lagging strand.[21] Moreover, several reports have shown that RNase H2 is active on nuclear but not on yeast and human mitochondrial DNA[20,22–24]; thus, mutations in RNase H2 genes are mainly expected to impact the abundance and features of rNMPs embedded in nuclear DNA both in yeast and human cells.

Among the mutations reported in the human RNase H2-enzyme complex, the G37S mutation occurring close to the catalytic site residues of the RNase H2 subunit A (highlighted with green circles in Figure 1A), has been extensively studied.[14] The G37S mutation in RNase H2A almost abolishes the catalytic activity of RNase H2, invokes the cGAS–STING innate immune-sensing pathway, and is perinatally lethal in mice.[27,28] The RNase H2A subunit is better conserved among eukaryotic species in contrast to RNase H2B and RNase H2C subunits.[25,29–31] In the RNase H2 complex, the RNase H2C subunit is close to the catalytic site of subunit A.[14] Among AGS mutations in RNase H2B and H2C, the R69W mutation in RNase H2C has reduced activity *in vitro* on RNA-DNA hybrids and on a single rNMP embedded in DNA, highlighting its role in enzymatic function.[29] In patients with AGS having mutations in RNase H2A or RNase H2C, the degree of disability and interferon score, and the severity of symptoms are elevated, as compared to patients having mutations in RNase H2B.[32,33]

Because the AGS mutants, RNASEH2A-G37S and RNASEH2C-R69W are in conserved regions between the human and the yeast subunits, we engineered the corresponding yeast orthologs mutants, Rnh201-G42S and Rnh203-K46W, respectively, to characterize their impact on the rNMPs embedded in yeast genomic DNA. To capture the genomic rNMPs and the DNA upstream of the embedded rNMPs, we utilized the *ribose-seq* technique[11,12] (Figure 1B) and performed the sequence analysis using the Ribose-Map bioinformatics toolkit[34–36] to obtain the coordinates of rNMPs in the yeast nuclear and mitochondrial genomes. We determined embedded rNMP features such as rNMP abundance rate, composition, hotspots, and sequence preferences around the site of rNMP embedment in the DNA of cells carrying the AGS-orthologous mutations. Moreover, we compared the rNMP features of the AGS ortholog mutants with those found in the DNA of wild-type and *rnh201*Δ mutant cells. Due to the conserved functions of the RNase H2 complex in both human and yeast cells, we posit that the results of this study can contribute to a deeper understanding of the biology associated with the RNASEH2A-G37S and RNASEH2C-R69W mutations in human cells. Furthermore, our findings may provide insights into comprehending the effects of these AGS ortholog mutations on the genome of patients with AGS.

RESULTS

Generation of *S. cerevisiae* strains carrying the Aicardi-Goutières syndrome yeast ortholog mutations *rnh201*-G42S or *rnh203*-K46W, DNA-seq analysis, and ribose-seq libraries preparation

With the goal to study embedded-rNMP patterns in AGS mutants, we engineered two AGS orthologous mutations occurring in conserved domains of human and yeast RNase H2 subunits. Two AGS mutations commonly found in human patients are RNASEH2A-G37S and RNASEH2C-R69W. The yeast ortholog mutations are *rnh201*-G42S and *rnh203*-K46W, respectively (Figure 1A). Conserved regions of amino acid sequences between yeast and human RNase H2 subunits were identified from previously published work that compares sequence alignments across several species.[25] RNASEH2A-G37S is within the catalytic subunit, whereas RNASEH2C-K69W is in a non-catalytic subunit but it is still very close to the Subunit A-catalytic site within the RNase H2 complex.[14,37] The yeast ortholog mutations *rnh201*-G42S and *rnh203*-K46W were created in *S. cerevisiae* in the BY4742 background (*MATα his3*Δ1 *leu2*Δ0 *lys2*Δ0 *ura3*Δ0) using the *delitto perfetto* technique.[38] We built 14 ribose-seq libraries of rNMPs embedded in the genomic DNA of yeast strains carrying wild-type RNase H2, *rnh201*-G42S, *rnh203*-K46W, or *rnh201*Δ genotype (Figure 1B). In addition to these libraries, in the analyses, we also included two wild-type and one *rnh201*Δ ribose-seq libraries previously published derived from the same BY4742 backgrounds.[20] Multiple sets of restriction enzymes were utilized to fragment DNA during the preparation of various ribose-seq libraries. This approach aimed to explore diverse genomic DNA regions and account for variations in rNMP sequence context or rNMP patterns linked to the fragmentation method used (Table S1) and amplified using PCR with specific primers (Table S2). The different combinations of restriction enzymes (REs) used to fragment the yeast genomic DNA were (i) RE1: DraI, EcoRV, and SspI; (ii) RE2: AluI, DraI, EcoRV, and SspI; and (iii) RE3: RsaI and HaeIII. The sequencing data of the ribose-seq libraries were analyzed using the Ribose-Map bioinformatics toolkit[34–36] to first obtain the rNMP coordinates in the sacCer2[39] genome.

Genomic DNA from all genotypes utilized in this study was extracted and used for ribose-seq library preparation, and in part also for direct DNA sequencing. The DNA-seq data were used to observe mutation spectra to evaluate the genomic stability of the AGS-mutant cells compared to wild-type and *rnh201*Δ cells and for rNMP-count normalization in hotspot analysis (See STAR Methods). The yeast AGS-orthologous mutants Rnh201-G42S and Rnh203-K46W were previously shown to stimulate double-strand break (DSB) repair via cDNA in

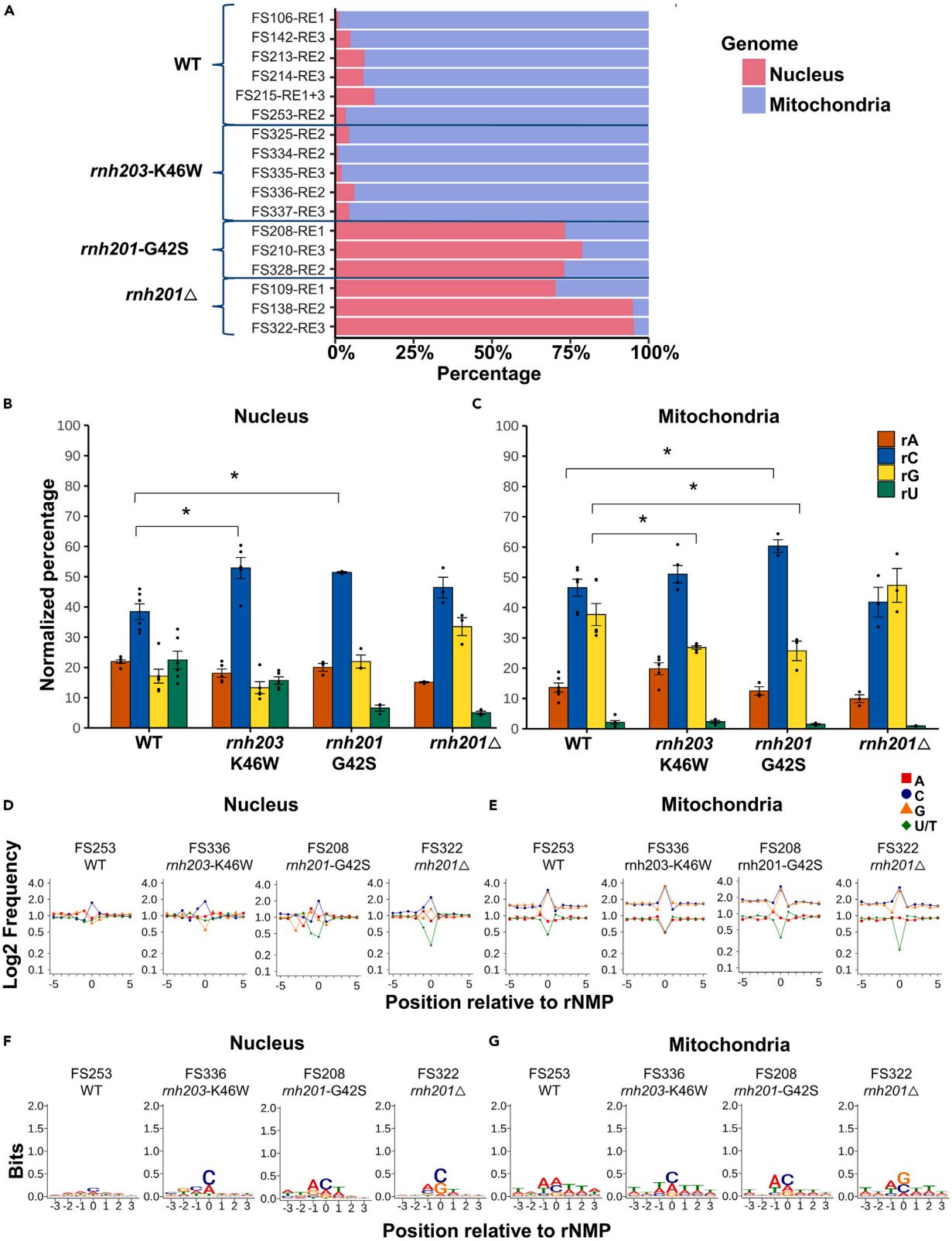

A
WT
rnh203-K46W
rnh201-G42S
rnh201△
FS106-RE1
FS142-RE3
FS213-RE2
FS214-RE3
FS215-RE1+3
FS253-RE2
FS325-RE2
FS334-RE2
FS335-RE3
FS336-RE2
FS337-RE3
FS208-RE1
FS210-RE3
FS328-RE2
FS109-RE1
FS138-RE2
FS322-RE3
0%
25%
50%
75%
100%
Percentage
Genome
Nucleus
Mitochondria
B
Nucleus
C
Mitochondria
Normalized percentage
rA
rC
rG
rU
WT
rnh203 K46W
rnh201 G42S
rnh201△
D
Nucleus
E
Mitochondria
A
C
G
U/T
FS253 WT
FS336 rnh203-K46W
FS208 rnh201-G42S
FS322 rnh201△
Log2 Frequency
Position relative to rNMP
F
Nucleus
G
Mitochondria
Bits
Position relative to rNMP

Figure 2. rNMP-embedment rates and patterns in the nuclear and mitochondrial genome of *S cerevisiae* AGS-ortholog mutants

(A) Horizontal stacked-bar plots showing rNMP-abundance rates as percentage of rNMPs found in the nuclear (red bars) and the mitochondrial (blue bars) genome of ribose-seq libraries (Table S1) of *S. cerevisiae* strain BY4742 with the indicated genotypes (wildtype, *rnh203*-K46W, *rnh201*-G42S, and *rnh201*Δ).

(B) Bar graphs corresponding to the mean value and standard error of normalized percentages of rNMPs found in nuclear and mitochondrial genome. Statistical comparisons made using Mann-Whitney *U* test between AGS mutant and wild type libraries are shown by a bar aligned with the rNMP base and *p* value of <0.05 denoted by single asterisk (*).

(C) DNA of rNMP libraries for the indicated genotypes. The percentages are normalized to the base composition of the reference genome, such that the expected percentage for each rNMP base is 25%.[20] Percentages without the normalization of rNMPs are shown in Figure S2 and are listed in Table S2.

(D and E) (D) Sequence context +/− 5 nucleotides (nt) from the site of rNMP presence, which is indicated by the 0 position, for all rNMPs found in nuclear and (E) mitochondrial DNA, respectively. The frequency of each nucleotide is normalized to the frequency of the corresponding nucleotide present in the nuclear or mitochondrial reference genome. The plots shown are for one sample library of each genotype with the library and genotype indicated on top of the plots. The library name and genotype for the displayed data are indicated on top of each plot. Red square, A; blue circle, C; orange triangle, G; and green rhombus, U. The sequence plots generated from all rNMP libraries of this study are presented in Figures S3–S6.

(F and G) (F) Sequence logo plots flanking +/− 3 nt from the rNMP position (0) of top 1 percentile locations with the highest rNMP counts observed in nuclear and (G) mitochondrial DNA. The *y* axis shows the level of sequence conservation, represented in bits. The library name and genotype for the displayed data are indicated on top of each plot. The Sequence logo plots for all the libraries are present in Figure S7.

S. cerevisiae cells when combined with each other or along RNase H1-null mutation and lowered survival frequency after the induction of a DSB as compared to wild-type RNase H2 cells.[40] The *rnh201*-G24S mutation also has elevated recombination and dinucleotide slippage rates in yeast cells.[26] Here, we examined germline and somatic mutational patterns of insertions and deletions. In the germline mutation spectra, we do not see any major differences in the mutational patterns between the genotypes (Figure S1A). However, we note that the wild-type library has higher percentage of dT deletion (dark orange panel) as compared to insertions (dark green panel); whereas AGS mutants and *rnh201*Δ libraries show similar percentages of dT insertions and deletions (Figure S1A). In somatic mutation spectra (Figure S1B), we note markedly higher dT insertions (dark green panel) in AGS mutants and *rnh201*Δ as compared to wild-type libraries. In addition, we see higher dT insertion at 5+ repeat units in *rnh201*-G42S and *rnh201*Δ as compared to wild-type and *rnh203*-K46W libraries (dark orange panel) suggesting higher events of genetic modifications in *rnh201*-G42S and *rnh201*Δ mutant compared to wild-type and *rnh203*-K46W cells. The somatic mutation spectra suggest that both the Rnh201-G24S and the Rnh203-K46W yeast ortholog mutants do have varying levels of impact on genome stability in yeast cells.

The ribose-seq libraries were prepared as described in previous studies.[11,12,24] Briefly, after genomic DNA fragmentation, the genomic fragments are dA-tailed and ligated to a molecular barcode–containing sequencing adaptor. Alkali treatment denatures the DNA and cleaves at rNMP sites, exposing 2′,3′-cyclic phosphate or 2′-phosphate termini, which are recognized and ligated to the 5′-phosphate ends of the same molecules by the *Arabidopsis thaliana* tRNA ligase, generating single-strand circles. Linear, unligated fragments are degraded. Circular DNA molecules are PCR-amplified and sequenced using an Illumina sequencing platform (Figure 1B). Successively, the sequencing reads are analyzed using the Ribose-Map bioinformatics tool to align the reads to the SacCer2 reference genome[39] and to obtain the genomic coordinates of the rNMPs at single nucleotide resolution, the sequence context, and hotspots of the rNMPs.[34–36]

Increased ribonucleoside monophosphate frequency in the nuclear genome of *rnh201*-G42S mutant cells

Previous work using alkaline gel electrophoresis of total genomic DNA, combined nuclear and mitochondrial DNA, derived from *rnh201*-G42S mutant-yeast cells revealed higher sensitivity to alkali compared to wild-type cells, suggesting increased level of rNMPs in the genome.[26] By leveraging ribose-seq and Ribose-Map techniques, it becomes possible to map rNMPs embedded in genomic DNA at single-nucleotide resolution. To study the frequency of rNMPs embedded in the yeast nuclear genome for each genotype, we independently counted the rNMPs present in nuclear and mitochondrial DNA in each individual ribose-seq library (Table S1). Since absolute rNMP counts in the libraries do not account for variations in their construction and sequencing yield, we determined the percentage of rNMPs in the nuclear and mitochondrial DNA from the total rNMP count obtained for each library. Because RNase H2 has been repeatedly shown not to be active on rNMPs embedded in mitochondrial DNA[20,22–24] rNMPs found in yeast mitochondrial DNA constitute the gross majority of the genomic rNMPs. We used this knowledge to estimate the increase in rNMPs in nuclear DNA due to the AGS-ortholog mutants by comparing nuclear and mitochondrial rNMP percentages in each library among the wild-type, *rnh203*-K46W, *rnh201*-G42S, and *rnh201*Δ genotype (Figure 2A). Wild-type libraries have less than 15% rNMPs in the nucleus from the total rNMPs in each library (Figure 2A, Table S1). On the contrary, *rnh201*Δ cells have >70% rNMPs in the libraries contributed from the nucleus. Libraries of the *rnh201*-G42S genotype show high rNMP embedded in the nucleus, almost to the level of those of the *rnh201*Δ genotype libraries, suggesting impaired catalytic activity of RNase H2. Differently, the percentages of rNMPs found in the nuclear DNA of *rnh203*-K46W libraries are low, like those detected in the wild-type libraries, suggesting little to no effect of the *rnh203*-K46W mutation on the catalytic activity of the RNase H2 complex.

Biased rC levels in *rnh201*-G42S and *rnh203*-K46W mutant genotypes

We normalized the rNMP-base composition with respect to the nucleotide base composition found in the reference genome. Bar plots of these normalized percentages of each rNMP base in the nucleus (Figure 2B) show higher rC percentages for *rnh201*-G42S and *rnh203*-K46W genotypes in comparison to the normalized percentage of rC in the wild-type genotype. The rNMP composition without normalization (Figure S2A) also indicates high rC percentages in AGS mutants amongst all other rNMP bases (rA, rG and rU). In the nuclear DNA of *rnh201*-G42S libraries,

alongside the increase in rC, there is a decrease in rU like that seen in *rnh201*Δ libraries (Figures 2B and S2A, Table S3). The biased rC level in nuclear DNA of *rnh201*-G42S and *rnh203*-K46W libraries is unique in comparison to the rNMP composition found in nuclear DNA of *rnh201*Δ libraries, as *rnh201*Δ libraries have high rC as well as rG level in nuclear DNA. Normalized rNMP-base percentages in the mitochondrial genome (Figure 2C) for all genotypes have biased rC and rG normalized percentages as compared to rA and rU, consistent with results previously seen in yeast ribose-seq libraries.[11] Interestingly, the nucleotide rC does appear to be predominantly higher than rG in mitochondrial DNA of the *rnh201*-G42S and *rnh203*-K46W libraries as compared to wild-type and *rnh201*Δ libraries (Figure 2C). The rNMP composition without normalization in the mitochondrial genome also shows that rC is significantly higher than rG in the *rnh203*-K46W and *rnh201*-G42S compared to the wild-type and the *rnh201*Δ libraries (Figure S2B, Table S3).

Specific sequence context upstream of ribonucleoside monophosphate sites in nuclear DNA of *S. cerevisiae* ortholog *rnh203*-K46W mutant libraries

To study the sequence context and nucleotide preference of rNMP sites in the yeast AGS ortholog genotypes, we calculated the deoxynucleotide frequencies around +/− 100 bases (zoomed out) and +/− 5 bases (zoomed in) near the unique sites of rNMP embedment for nuclear and mitochondrial genome in every library. After normalization to the mononucleotide frequencies observed in the reference genome, frequencies of deoxyribonucleotide monophosphates (dNMPs) around rNMP sites in the nuclear and mitochondrial genome show a preference (higher frequency) for dNMPs with base G or C in the mitochondria as compared to the nucleus, in which frequencies of A, C, G and T are almost similar (Figures 2D, 2E, and S3–S6). This feature of rNMPs in the yeast mitochondrial genome is in line with what has been observed before with other budding yeast strains[20] suggesting rNMP presence in GC clusters in the yeast mitochondrial genome.[41] Thus, despite the low percentage of dCMPs and dGMPs in mtDNA, rCs and rGs were often surrounded by dCMPs and dGMPs.

The sequence context near +/− 5 bases of the rNMP-embedment sites (0^{th} position) in nuclear DNA of the *rnh203*-K46W genotype shows the preference for dCMP one base upstream (position −1) (Figures 2D and S3). The dCMP preference at position −1 from the rNMP in *rnh203*-K46W libraries is particularly evident when the rNMP has base C or U (Figures S3 and S4). Differently, in *rnh201*-G42S libraries, we see preference for dAMP one base upstream (position −1) of the rNMPs (Figures 2D and S4), particularly when the rNMP has base A or G (Figures S3 and S4). The rNMP frequency observed in the 0 position of the sequence plots is consistent with the normalized rNMP percentages for the respective genotypes (Figures 2B and 2C). Unlike for the nuclear genome, there is no observable difference in the frequencies of dNMPs around the rNMP site among the different genotypes for the mitochondrial genome. As previously observed for rNMPs in yeast mitochondrial DNA,[20] rCMP and rGMP sites are surrounded by dCMPs and dGMPs, respectively, within GC-rich regions (Figures 2E, S5, and S6).

Unique ribonucleoside monophosphates-hotspot composition and consensus sequence in Aicardi-Goutières syndrome-orthologous mutants

To analyze the sequence context around the highly frequent rNMP-embedment sites in the yeast AGS-ortholog libraries, we obtained the hotspots of the rNMP counts present (1 percentile of highly abundant single nucleotide rNMP locations) in nuclear and mitochondrial DNA. Using the Ribose-Map Hotspot Module, we obtained the genomic sequence of these abundant rNMP sites 3-nucleotides upstream and downstream of the rNMPs on the same strand as the rNMP, and visualized the results using the sequence logo plots (Figures 2F, 2G, and S7). We find that rCMP (0 position) is the dominant rNMP in the nucleus for all RNase H2-mutant genotypes tested (*rnh203*-K46W, *rnh201*-G42S and *rnh201*Δ), whereas there is no dominant rNMP seen in hotspots from wild-type nuclear libraries (Figure 2F). The rCMP hotspots are preferentially preceded by dCMP in *rnh203*-K46W libraries (−1 position), whereas in *rnh201*-G42S and *rnh201*Δ libraries rCMP hotspots are preceded by dAMP and followed by dTMP (Figures 2F and S7A). Sequence-logo plots of rNMPs in mitochondrial DNA show that the rNMP position is more often an rCMP in the AGS mutant libraries (Figures 2G and S7B). Hotspots of rNMPs in mitochondrial DNA of all genotypes studied show an evident preference for dAMP at position −1 and a fit within the consensus TNArSWTW sequence previously identified[20] (Figures 2G and S7B).

To observe the composition and distribution of hotspots across each genotype, we analyzed recurring rNMPs, present in at least 2 libraries for each genotype and selected the most abundant rNMP sites based on the average enrichment factor (see STAR Methods) calculated for each unique rNMP embedment site in that genotype. The count of the rNMPs in each embedment site of each library was normalized based on the relative base coverage, the ratio of reads mapped per base with respect to average reads per base in the genome, obtained from the DNA-seq data to exclude false positive hotspots due to the falsely amplified alignment of reads in low complexity regions, such as simple repeat sequences. These normalized counts for each rNMP embedment site in each library were then converted to enrichment factor, a relative rNMP-count measure for each location with respect to total rNMP count in each library genome, further used to calculate the average-rNMP enrichment within libraries in each genotype[24] (see STAR Methods). We selected the top 75 sites in the nuclear genome and the top 25 sites in the mitochondrial genome having the highest average rNMP enrichment factor (common hotspots) to be visually represented with their base (Figure 3). The hotspots present in at least 80% of libraries with the same genotype are marked with an asterisk (Figure 3). The base composition of the hotspots observed in each genotype is consistent with the composition of all rNMPs present in nuclear and mitochondria in each genotype (Figures 2B and 2C). The normalized composition of common hotspots represented in Figure 3 are also shown (Figures S8A and S8B). Interestingly, the top-75 hotspots in the nuclear genome of AGS mutant (*rnh203*-K46W and *rnh201*-G42S) libraries have higher rC than wild-type and *rnh201*Δ libraries (Figures 3A–3D and S8A and Table S4E), We also see that *rnh203*-K46W libraries have a similar level of rA as wild-type libraries, but a lower rG such as *rnh201*-G42S as compared to wild-type and *rnh201*Δ libraries. The coordinates, ribonucleotide base, and genes located at hotspot locations are also indicated (Tables S4 and S5). These common hotspots validate

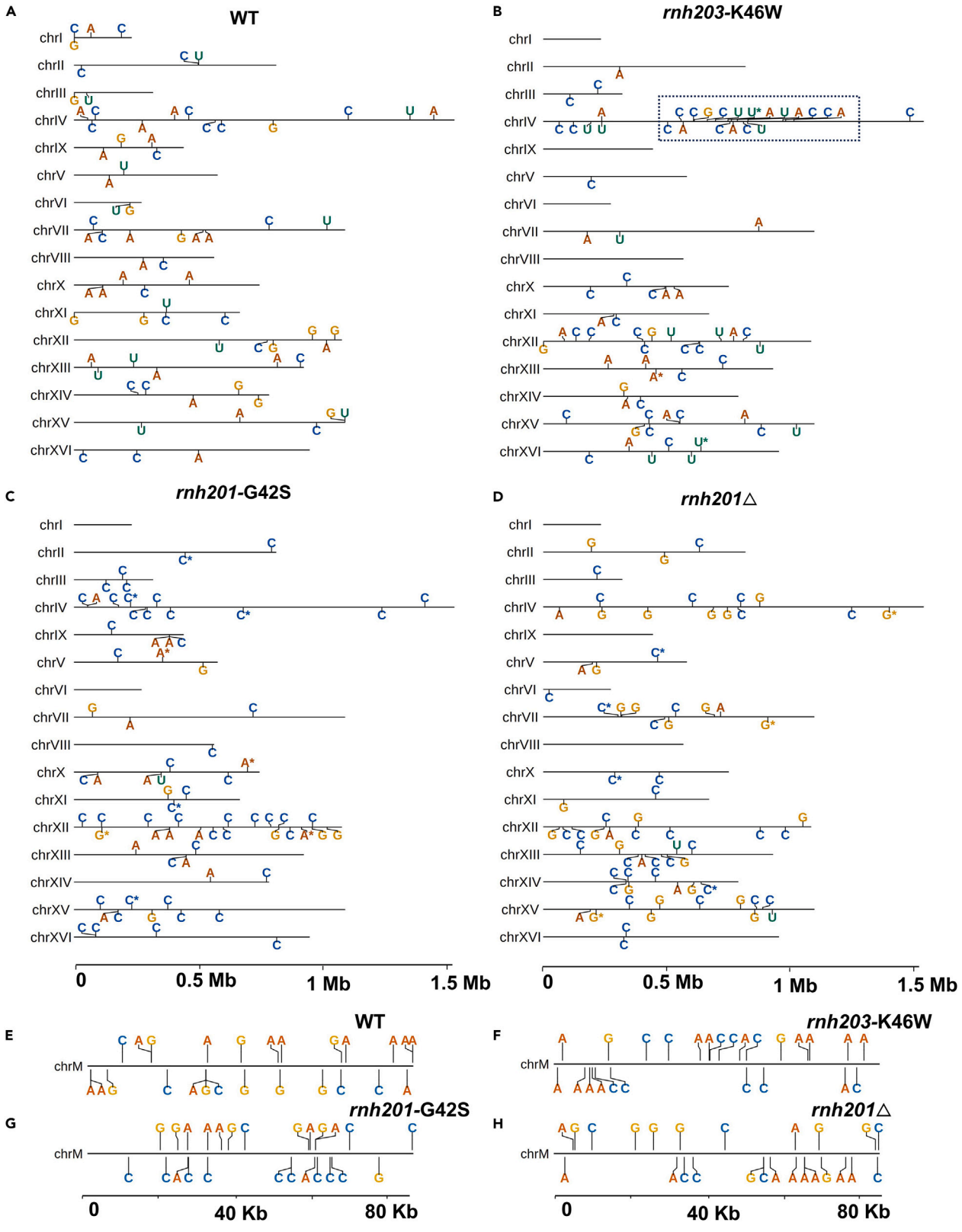

A
WT
B
rnh203-K46W
C
rnh201-G42S
D
rnh201△
chrI
chrII
chrIII
chrIV
chrIX
chrV
chrVI
chrVII
chrVIII
chrX
chrXI
chrXII
chrXIII
chrXIV
chrXV
chrXVI
0
0.5 Mb
1 Mb
1.5 Mb
E
WT
F
rnh203-K46W
G
rnh201-G42S
H
rnh201△
chrM
0
40 Kb
80 Kb

Figure 3. Most abundant common hotspots in nuclear and mitochondrial DNA of the Rnh203-K46W and Rnh201-G42S AGS ortholog mutants

Genome-mapped locations of the 75 most abundant common rNMP hotspots in nuclear DNA of ribose-seq libraries of (A) wild-type, (B) *rnh203*-K46W, (C) *rnh201*-G42S, and (D) *rnh201*Δ genotype. Genome mapped locations of the 25 most abundant common rNMP hotspots in mitochondrial DNA of ribose-seq libraries of (E) wildtype, (F) *rnh203*-K46W, (G) *rnh201*-G42S, and (H) *rnh201*Δ genotype. The represented annotations for rA, rC, rG, and rU are shown in red, blue, yellow, and green color letters, respectively. The common hotspots are selected based on occurrence in at least 2 libraries in each genotype and the highest rNMP Enrichment Factor (ratio of rNMP counts to average rNMP per base in genome). rNMP hotspots present in 80% or more libraries are marked with an asterisk (*). Chromosome locations +/− 3 nt of flanking sequence, enrichment factor, mapped genes, and composition of common hotspots is provided in Table S4 for the common hotspots in the nuclear genome and in Table S5 for the common hotspots in the mitochondrial genome. Common hotspot clusters identified on Chr IV in *rnh203*-K46W libraries are indicated by a dashed frame. Sequence logo plots for top 75 (nucleus) and top 25 (mitochondria) with top 2 and 5 percentile hotspots of unique rNMP locations (both nucleus and mitochondria) in each genotype are shown in Figure S8.

the rNMP-compositional changes we observe in the different genotypes (Figure 2B) whilst also suggesting unique locations of rNMP retention also seen in Figure 3 making the hotspot composition in both AGS mutants unique in comparison with the wild-type and *rnh201*Δ libraries (Figures 3D and S8A). The common hotspots in nuclear DNA of RNase H2 wild-type cells do not appear to cluster around a specific location, whereas we detect more than 15 rNMP-hotspot sites in a cluster on Chr IV between 0.5 Mb and 1 Mb in *rnh203*-K46W libraries (Figures 3A and 3B). To confirm the different sequence contexts for the hotspots observed in the different genotypes, we also looked at +/− 3-bp flanking consensus around top 75 common hotspots with different percentile thresholds of top 2% and 5% of common hotspots in each genotype consensus (Figures S8C and S8D). In nuclear DNA of wildtype cells, we do not detect a strong consensus in the rNMP position (0 position). Instead, the rC is predominant in all mutant genotypes, with rG also being abundant in *rnh201*Δ (Figure S8A). We also mapped the top 75 common hotspots in all genotypes to respective genes to find that more than 69 hotspots out of 75 hotspots map to protein coding genes in *rnh201*-G42S whereas in *rnh203*-K46W only 51 hotspots map to protein coding genes (Tables S4B and S4C). Increase in hotspot mapping to protein coding genes is also seen in *rnh201*Δ (58 hotspots, Table S4D) vs. wild-type libraries (49 hotspots, Table S4D). We conducted a similar analysis of common hotspots in mitochondria and visualized the top 25 common hotspots (Figures 3E–3H). The top-25 rNMP sites of the common hotspots in mitochondria show that the rC site is more abundant in *rnh203*-K46W and *rnh201*-G42S libraries compared to wild-type and *rnh201*Δ libraries (Figures S8B and S8D and Table S5E). By analyzing the consensus sequence of the top-2 and 5 percentiles of rNMP hotspots (Figure S8D), we detect more rA and rC in mitochondria of all genotypes, with rC being higher in *rnh201*-G42S, and rG being abundant only in *rnh201*Δ (Figures S8B and S8D).

Sequence-specific preferences in *S. cerevisiae* orthologs of RNase H2C-R69W and RNase H2A-G37S mutants

To further investigate and validate the sequence-context conservation of the rNMPs and their upstream and downstream dNMPs in the *S. cerevisiae* orthologs of human RNase H2C-R69W and RNase H2A-G37S mutants, we obtained heatmaps of the normalized frequencies of each rNMP mono- and di-nucleotide combination in nuclear and mitochondrial DNA (Figures 4A–4D). We used one-tailed Mann-Whitney *U* test to determine the significant preference or increase of each combination (Tables S6 and S7) and two tailed Mann-Whitney *U* test to determine significantly different frequencies between any of the RNase H2-mutant libraries and the wild-type libraries (Tables S8 and S9). Preferred dinucleotides (Tables S6 and S7) with significant differences in the pairwise comparisons (Tables S8 and S9) compared to wild-type libraries with *p* value <0.05 are highlighted with yellow arrows for *rnh203*-K46W libraries, green arrows for *rnh201*-G42S libraries and gray arrows for *rnh201*Δ libraries. The mononucleotide-heatmap analysis of rNMPs in nuclear DNA shows that the normalized frequency of rC increases significantly (*p* value <0.05 in Table S6) in both AGS-mutant genotypes, *rnh203*-K46W and *rnh201*-G42S (green and yellow arrow in Figure 4A); whereas, in the mitochondrial DNA, rC is preferred in only *rnh201*-G42S (Figure 4B, Table S7). The mononucleotide-heatmap analysis of rNMPs in nuclear DNA shows that the normalized frequency of rG is preferred in comparison to rA and rU (*p* value <0.05 in Table S6) in *rnh201*Δ (gray arrow in Figure 4A) but not in *rnh201*-G42S. Moreover, in mitochondrial DNA, while rG is still preferred in comparison to rA and rU in both the AGS-mutant libraries (*p* value <0.05 in Table S7), the frequency of rG decreases in both *rnh201*-G42S and *rnh203*-K46W as compared to wild-type and *rnh201*Δ libraries (Figure 4B, *p* value <0.05 in Table S9).

In the dinucleotide-heatmap analysis with the rNMP (R) and its upstream dNMP (N), NR combinations in the nucleus, CrC has the strongest frequency in *rnh203*-K46W as compared to wild-type libraries ((yellow arrow in Figure 4C, *p* value <0.05 in Tables S6 and S8), in which we observe preference for ArC. For the *rnh201*-G42S libraries, we see NR-dinucleotide combinations with preference for dAMP: ArA, ArC, and ArG, and for dGMP; GrA, GrC, and GrG (Figure 4C and *p* value <0.05 in Table S6). Among these dinucleotide preferences in the *rnh201*-G42S libraries ArA, ArG, and GrC frequencies are significantly different (*p* value <0.05 in Table S8) from wild-type libraries (green arrows in Figure 4C, *p* value <0.05 in Table S8). We also detect a strong CrU preference in *rnh201*-G42S libraries as well as in *rnh201*Δ libraries (green and gray arrow in Figure 4C, *p* value <0.05 in Table S6). The patterns for ArA, ArC, ArG and CrU in *rnh201*-G42S are consistent with the patterns observed in *rnh201*Δ libraries, whereas *rnh201*-G42S libraries prefer GrA, GrC, and GrG as compared to *rnh201*Δ libraries. On the other hand, *rnh201*Δ libraries have preference for CrA and CrC, as compared to *rnh201*-G42S libraries (Figure 4C, *p* value <0.05 in Table S6). Dinucleotide preferences in the mitochondrial genome (Figure 4D) are very similar for all genotypes, with a markedly conserved preference for CrA, ArC, ArG, and CrU, partially like those observed in the nuclear DNA of *rnh201*Δ libraries (Figures 4C and 4D, *p* value <0.05 in Table S7). However, there are some differences observed in the frequencies between genotypes, i.e., a lower frequency of ArC and ArG in *rnh203*-K46W libraries as compared to wild-type libraries (Figure 4D, *p* value <0.05 in Table S9).

In the heatmap analysis of the rNMP with its downstream dNMP, dinucleotide (RN), we detect weak preferences in both nuclear and mitochondrial DNA (highlighted using arrows in Figure S9 and *p* value <0.05 in Tables S6 and S7). The dinucleotide rUG is preferred only in nuclear

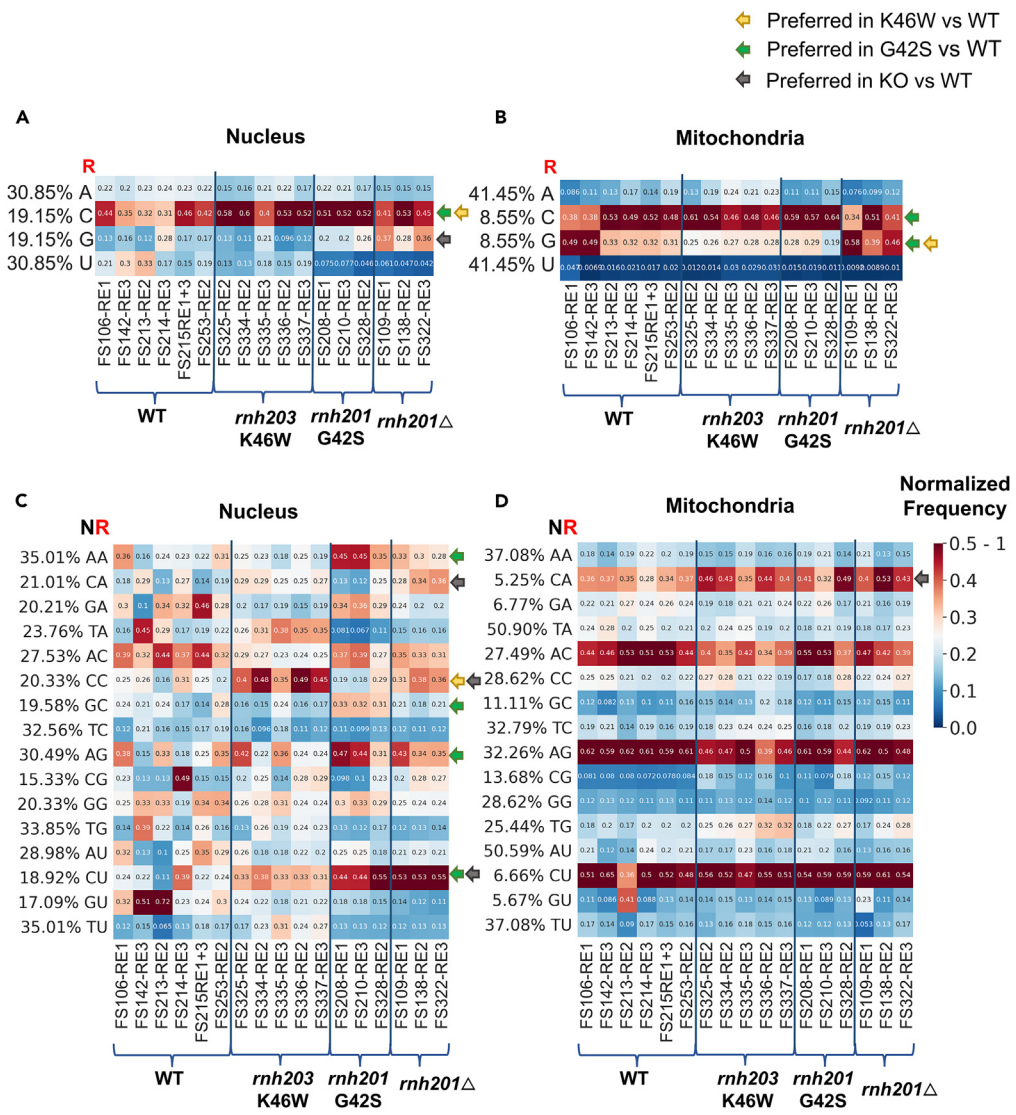


Figure 4. rNMP-base composition and dinucleotide patterns identified in the *rnh203*-K46W and *rnh201*-G42S AGS ortholog mutant libraries

Heatmap of normalized frequency of each type of rNMP (rA, rC, rG, and rU) in nuclear (A) and mitochondrial (B) DNA for all the ribose-seq libraries of this study. Heatmap analyses with normalized frequency of (C) nuclear and (D) mitochondrial NR dinucleotides (rA, rC, rG, and rU with the upstream deoxyribonucleotide with base A, C, G, or T) for all the ribose-seq libraries of this study. The formulas used to calculate these normalized frequencies are shown and explained in STAR Methods. Each column of the heatmap shows results of a specific ribose-seq library. Each library name is indicated underneath each column of the heatmap with its corresponding restriction enzyme (RE) set used. The ribose-seq libraries of the same genotype are also grouped together by brackets and separated by thick vertical blue lines. Each row shows results obtained for rNMP-mononucleotide or dinucleotide combination. The actual percentages of rNMP (rA, rC, rG, and rU) or dinucleotides of fixed base A, C, G, or T for the indicated base combinations (AA, CA, GA, and TA; AC, CC, GC, and TC; AG, CG, GG, and TG; and AT, CT, GT, and TT) present in the nuclear and mitochondrial genome of *S. cerevisiae* are shown to the left of the corresponding heatmaps. The observed % of rNMPs or dinucleotides with rNMPs with base A, C, G, or U were divided by the actual % of each rNMP or dinucleotide with fixed base A, C, G, or T in nuclear or mitochondrial DNA. The bar to the right shows how different frequency values are represented as different colors: white for 0.25; light red to red for 0.25 to 0.5–1, and dark blue to light blue for 0.25 to 0. Yellow, green, and gray arrows indicate the nucleotide frequency significantly preferred for *rnh203*-K46W, *rnh201*-G42S, and *rnh201*Δ mutant libraries, respectively, in comparison to wild-type libraries. Significantly different frequencies with p value of <0.05 are highlighted based on one tailed Mann Whitney-Wilcoxon U test between nucleotide frequencies in the AGS mutant genotypes vs. the expected value of 0.25 (Tables S6 and S7), as well as two tailed Mann-Whitney Wilcoxon U test AGS-mutant libraries vs. wild-type libraries (Tables S8 and S9). Dinucleotide heatmaps for the most frequent dNMP found downstream of the rNMPs in the nuclear and mitochondrial DNA are presented in Figure S9.

DNA of *rnh203*-K46W libraries but is not significantly higher than in wild-type libraries (p value <0.05 in Tables S6 and S8). The dinucleotides rAT, rCA, rCT, and rGT are preferred in the nuclear DNA of *rnh201*-G42S libraries significantly compared to wild-type libraries (highlighted using green arrows in Figure S9, p value <0.05 in Tables S6 and S8). The frequencies of dinucleotides rAT and rGT in *rnh201*-G42S libraries are

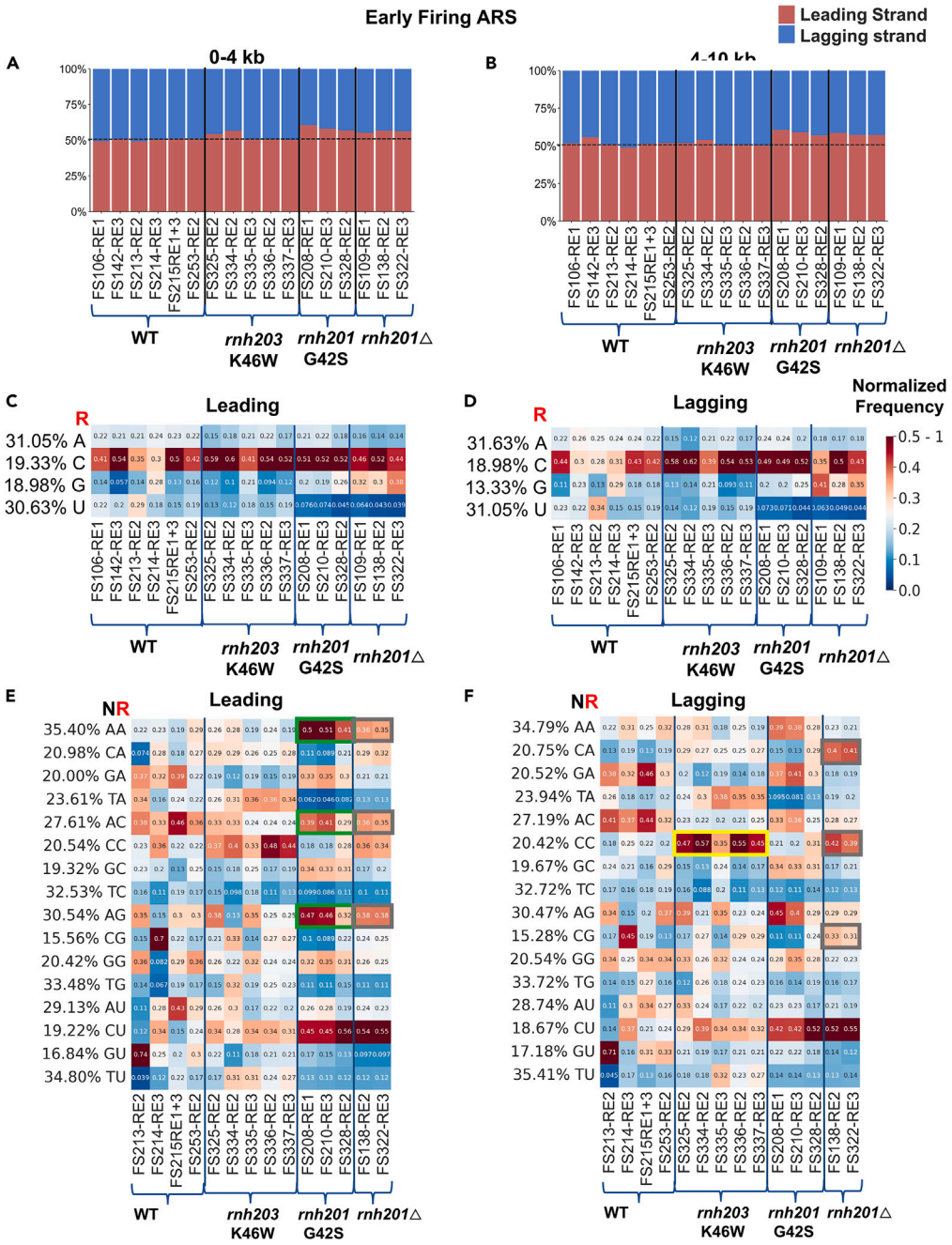


Figure 5. The *rnh203*-K46W and *rnh201*-G42S AGS ortholog mutant libraries display bias distribution, composition, and NR-dinucleotide patterns on the leading and lagging strand in yeast nuclear DNA

Stacked-percentage bar plots of rNMP counts in (A) 0–4,000 nt and (B) 4,000–10,000 nt windows on the leading (red bars) and lagging (blue bars) strand regions adjacent to early firing ARSs. Mono nucleotide heatmaps analysis in 4,000–10,000 nt windows in the leading (C) and lagging (D) strands with frequencies normalized to the dNMP content in the 4000–10,000-nt window around the early-firing ARSs on the leading or lagging strand, respectively. Dinucleotide NR (rA, rC, rG, and rU with the upstream deoxyribonucleotide with base A, C, G, or T) heatmap analysis to reveal preferences on the (E) leading and (F) lagging strand for wild-type, *rnh201*Δ, and AGS mutant libraries 4,000–10,000 nt from the early-firing ARSs. The formulas used to calculate these normalized frequencies are shown and explained in the STAR Methods. Each column of the heatmap shows results of a specific ribose-seq library. Libraries with more than 400 rNMPs in the leading and lagging strand are displayed in the heatmaps. Each library name is indicated underneath each column of the heatmap with its corresponding restriction enzyme (RE) set used. The ribose-seq libraries of the same genotype are also grouped together by brackets and separated by thick vertical blue lines. Each row shows results obtained for rNMP or dinucleotide combination. The actual percentages of rNMP (rA, rC, rG, and rU)

Figure 5. *Continued*

dinucleotides of fixed base A, C, G, or T for the indicated base combinations (AA, CA, GA, and TA; AC, CC, GC, and TC; AG, CG, GG, and TG; and AT, CT, GT, and TT) present in the 4,000–10,000-nt windows around early-firing ARSs on the leading and lagging strands are shown to the left of the heatmaps. The observed % of rNMPs and dinucleotides with rNMPs with base A, C, G, or U were divided by the actual % of each rNMP and dinucleotide with fixed base A, C, G, or T in 4,000–10,000-nt windows around ARSs on the leading and lagging strands. The bar to the right shows how different frequency values are represented as different colors: white for 0.25; light red to red for 0.25 to 0.5–1, and dark blue to light blue for 0.25 to 0. Yellow, green, and gray boxes indicate dinucleotide preferences found in the lagging and leading strands for *rnh203*-K46W, *rnh201*-G42S, and *rnh201*Δ mutant libraries.

like those in *rnh201*Δ libraries being preferred significantly compared to wild-type libraries (highlighted using gray arrows in Figure S9, p value <0.05 in Tables S6 and S8). Although not significant compared to wild-type libraries, the *rnh203*-K46W libraries have dominant rA and rU followed by dGMP in nuclear DNA (Figure S9A).

S. cerevisiae orthologs of RNase H2C-R69W and RNase H2A-G37S mutants show specific dinucleotide patterns on the leading and lagging strand around Aicardi-Goutières syndrome sequences of the yeast genome

To examine whether the yeast AGS-ortholog mutant libraries display similar or different strand bias patterns on the leading and lagging strands, we analyzed rNMP distribution and nucleotide patterns on the leading and lagging strands within windows of 0–4 kb and 4–10 kb from ARSs in the genome of wild-type, *rnh203*-K46W, *rnh201*-G42S, and *rnh201*Δ strains of this study (Figures 5 and S10). As previously observed, we found that as compared to wild-type, the *rnh201*Δ libraries tend to have a slight increased level of rNMPs on the leading strand.[21] Notably, such preference for the leading strand is even more evident in *rnh201*-G42S libraries in both 0–4 kb and 4–10 kb flanking regions of the ARSs (Figures 5A and 5B). In the mononucleotide-heatmap analysis of rNMPs present on the leading vs. lagging strands in the 4-10-kb region around early-firing ARSs, we do not detect strand-biased composition for the AGS mutant libraries, as seen for wild-type and *rnh201*Δ libraries of wild-type or mutant DNA-polymerase (Pol) genes.[21,28] We previously found that rNMPs have markedly distinct dinucleotide frequencies on the leading and lagging strands 4-10-kb upstream and downstream of early-firing ARSs in rNMP libraries of *rnh201*Δ cells expressing wild-type or mutant forms of Pol δ or Pol ε.[21] Leading and lagging strand-dinucleotide patterns show signatures of incorporation due to the activity of DNA Pol ε and Pol δ on these strands, respectively. In the *rnh201*Δ libraries, the normalized dinucleotide frequencies of ArA, ArC, and ArG are higher on the leading strand, while those for CrA, CrC, and CrG are higher on the lagging strand[21] (see gray frames for *rnh201*Δ in Figures 5E and 5F). In the heatmap analysis of dinucleotide preferences in nuclear DNA, we found that for *rnh203*-K46W libraries the preferred dinucleotide CrC has a higher frequency on the lagging strand as compared to the leading strand in the 4–10 kb region from early ARSs (Figures 5E and 5F; see yellow frames). This pattern of CrC being higher on the lagging vs. the leading strand, likely due to rNMP incorporation by Pol δ activity, is also seen in *rnh201*Δ libraries, but it is barely present in *rnh201*-G42S libraries (Figures 5E and 5F). For *rnh201*-G42S libraries, we detect a preference for ArA, ArC, and ArG dinucleotide patterns in the 4-10-kb region of the leading strand (green frames) as compared to the lagging strand, as also detected in the *rnh201*Δ libraries, but much stronger than in *rnh201*Δ libraries, showing incorporation patterns due to Pol ε activity (Figures 5E and 5F).

DISCUSSION

Rnh201-G42S (RNase H2A-G37S ortholog) but not Rnh203-K46W (RNase H2C-R69W ortholog) has reduced ribonucleotide excision repair activity

The G37S mutation in human RNase H2A is a common mutation in individuals with AGS, located in a conserved region of the protein, nearby the catalytic site.[37] This specific mutation affects the catalytic activity of the RNase H2 enzyme, impairing its ability to cleave rNMPs[42] and process RNA/DNA hybrids effectively.[31,43] In our analyses, the RNase H2A-G37S mutant ortholog in yeast, Rnh201-G42S, shows lack of catalytic activity because of the highly increased rNMP percentage in nuclear DNA relative to the rNMP presence in mitochondrial DNA, as observed for *rnh201*Δ libraries (Figure 2A). The R69W mutation in the human RNase H2C subunit affects the RNase H2-complex stability,[14,30] and it is located very close to the catalytic site of RNase H2A within the structure of the RNase H2 complex. The yeast ortholog mutant of RNase H2C-R69W is Rnh203-K46W. The ribose-seq libraries of *rnh203*-K46W, unlike those of *rnh201*-G42S and *rnh201*Δ cells, have less than 10% of rNMPs in the nuclear genome with most rNMPs found in the mitochondrial genome such as wild-type libraries. Because RNase H2 is not active on mitochondrial DNA[8,22,23,25] these results indicate that the RNase H2-catalytic activity is mostly intact in the *rnh203*-K46W but compromised in the *rnh201*-G42S genotype. The results obtained with the analysis of rNMP percentage in nuclear vs. mitochondrial DNA in the *rnh201*-G42S libraries correlate with an increased alkali sensitivity of yeast genome carrying this mutation.[26] The abundance of rNMPs in nuclear DNA of rnh201-G42S libraries is overall less than that observed in rnh201Δ libraries (Figure 2A), suggesting that the Rnh201-G42S mutant still retains some rNMP-incision activity. Another study looking at the enzymatic activity of RNase H2-G42S mutant found up to 13% activity of the mutant compared to wild-type RNase H2.[44] There are numerous other studies that infer rNMP abundance in the genome of cells by sensitivity test for stress-inducing agents[42,45] or nick translation reaction[46] that provide the information of rNMP levels. By using the ribose-seq technique to directly capture rNMP sites in genomic DNA and paired analysis with the Ribose-Map toolkit to map them at single-nucleotide resolution, our study enables to obtain specific location, composition of rNMPs, distribution, sequence context, and patterns of rNMP sites in yeast AGS-ortholog mutants, which can help to understand the consequences of the orthologous mutants in humans. In future studies it will be possible to apply the same techniques to directly capture and study rNMPs embedded in the DNA of AGS-mutant samples derived from human cell lines or patient samples.

Increased level of rC both in Rnh201-G42S and Rnh203-K46W and low rU in Rnh201-G42S mutant libraries reflects the preferential residual activity of the AGS ortholog mutants of RNase H2 on rG

Previous research investigating the composition of rNMPs in yeast DNA reported a general high rC frequency in the nuclear genome for all strains studied, including the BY4742 background, in wild-type and *rnh201*Δ cells.[20] Here, we found a consistent trend with an even higher level of rC in the nuclear genome of both AGS-ortholog mutants (Figure 2C). In line with results in yeast mitochondrial DNA, in which rG tends also to have a high frequency, particularly in the BY4742 strain,[20] *rnh201*Δ cells display a higher frequency of rG in nuclear DNA[20] (Figures 2B and 4A). Differently, in the two AGS-ortholog mutants studied here, the rG frequency is significantly lower than that of rC in nuclear DNA. These results highlight that RNase H2 prefers cleaving at rG followed by rC in yeast nuclear DNA. When RNase H2 activity is compromised, as observed in the *rnh203*-K46W and *rnh201*-G42S mutant libraries, rG becomes the preferred target. In fact, the reduced activity of the *rnh201*-G42S mutant is still sufficient to preferentially target rG in nuclear DNA. Markedly, also in mitochondrial DNA, differently from wild-type and *rnh201*Δ libraries, both *rnh201*-G42S and *rnh203*-K46W mutant libraries display a much higher frequency of rC compared to rG (Figure 2C). This observation suggests that, while RNase H2 is not active on rNMPs embedded in mitochondrial DNA, its activity on nuclear DNA may indirectly affect mitochondrial metabolism, for example affecting nucleotide pool composition. Interestingly, a recent study found that AGS mutations are associated with different levels of mitochondrial disfunction.[47]

Another rNMP-composition feature observed is the low level of rU in nuclear DNA of *rnh201*-G42S libraries as compared to wild-type libraries and almost as low as in the *rnh201*Δ libraries (Figures 2B and 4A). Even lower frequency of rU is observed in the mitochondrial genome (Figures 2C and 4B), in which there is no RNase H2 activity.[22,23] However, the level of nuclear rU in the *rnh203*-K46W libraries is not different from that observed in wild-type libraries. Overall, these results suggest that the RNase H2-complex activity is not very specific to rU. Most likely rU has a low level of incorporation by replicative DNA polymerases both in nuclear and mitochondrial genomes.

Ribonucleoside monophosphate hotspots in the yeast Aicardi-Goutières syndrome-ortholog mutants show unique composition and distribution as compared to wild-type and *rnh201*Δ samples

Hotspots, the top one percentile locations in the genome with the highest rNMP counts in each library (Figures 2C, 2D, and S7), and common hotspots, rNMP locations with the highest average rNMP enrichment recurring in at least two libraries of the same genotypes (Figures 3 and S8), show trends in composition mostly consistent with those of all genomic rNMPs in the corresponding genotypes. This observation aligns with previous findings in the human mitochondrial genome, where the top 20 common hotspots reflected rNMP-compositional differences across various human cell types. These hotspots appeared at distinct coordinates in each cell type compared to those found universally across all cell types, suggesting notable differences in embedment patterns among them.[24] Nuclear rNMP hotspots found in yeast wild-type libraries do not show preference for any rNMP (Figure 2C); whereas *rnh201*-G42S and *rnh203*-K46W libraries have a high level of rC hotspots, and *rnh201*Δ libraries have both rC and rG hotspots. A high level of rC in the top 75 common hotspots is found in the AGS mutants like in *rnh201*Δ mutant libraries (Figures 3A–3D and S8A). Moreover, in *rnh203*-K46W mutant libraries, the common hotspots with rU are more abundant than those with rG and are much more prevalent than in *rnh201*Δ and in *rnh201*-G42S libraries, as in wild-type cells (Figures 3A–3D and S8A). Both AGS-orthologous mutants have a low level of rG among the hotspots (Figures 3A–3D and S8A) and have a unique composition profile. The AGS-ortholog mutants have higher rC common hotspots than wild-type and *rnh201*Δ cells in both nuclear and mitochondrial DNA (Figures S8, S8A, and S8B). The *rnh203*-K46W cells have higher rA and lower rG hotspots than the *rnh201*-G42S cells in both nuclear and mitochondrial DNA (Figures S8A and S8B). Furthermore, the distribution of *rnh203*-K46W common hotspots reveals a cluster of rNMPs on chromosome IV within the 500,474 and 1009,648 loci (Figure 3B, Table S4). The different rNMP composition and distribution of the common hotspots in the different RNase H2-genotype libraries compared to those detected in the wild-type RNase H2 libraries reveal rNMP sites that are not processed by the mutant forms of RNase H2 allowing their accumulation. Therefore, the results underscore the preferential activity of RNase H2 for sites with rG in nuclear DNA.

Interestingly, the *rnh201*-G42S mutant exhibits a significant majority of common hotspots within protein-coding regions of the nucleus compared to the other genotypes (Table S4). Similarly, the *rnh201*Δ mutant shows an increased number of common hotspots in these regions compared to the wild-type and *rnh203*-K46W genotypes (Table S4), though not as extensively as the *rnh201*-G42S mutant. These results may suggest a differential function of RNase H2 in various genomic regions and could be associated with the observed bias in activity, particularly the lack of activity of the Rnh201-G42S mutant on the leading strand, as discussed later in discussion.

Both yeast Aicardi-Goutières syndrome-ortholog mutants Rnh201-G42S and the Rnh203-K46W show distinct ribonucleoside monophosphates patterns suggesting distinct activity on ribonucleoside monophosphates embedded in yeast nuclear DNA

The dinucleotide-heatmap analyses of nuclear rNMPs and their upstream dNMPs not only identify rNMP embedment sites, typically resulting from DNA polymerase activity, but also highlight sites of rNMP removal. The use of DNA polymerase mutants that incorporate abundant rNMPs in DNA in the *rnh201*Δ genotype reveals rNMP incorporation sites specific of the different polymerases.[21] In wild-type DNA polymerase genotype, sites of rNMP removal by the RER mechanism via RNase H2 are indicated by pattern differences between *rnh201*Δ and wild-type rNMP libraries. The inability of AGS-RNase H2 mutant to remove rNMPs is shown through differences between AGS mutant and wild-type libraries. Utilizing multiple restriction-enzyme combinations to fragment genomic DNA for ribose-seq library creation ensures that observed patterns across all libraries of a genotype are not artifacts of enzyme-fragmentation patterns. The dinucleotide heatmaps distinctly highlight sequence patterns in AGS-mutant libraries compared to wild-type and rnh201Δ libraries, differing from the conserved

dinucleotide-rNMP patterns seen in mitochondrial DNA analysis across all genotypes (Figures 4C and 4D). The *rnh203*-K46W libraries show a preference for dCMP upstream of rC (CrC), while *rnh201*-G42S libraries show a preference for ArA, GrC, ArG, and CrU dinucleotide patterns. The difference between dinucleotide patterns in nuclear DNA of the two AGS mutants suggest that the *rnh201*-G42S and *rnh203*-K46W mutations, beyond affecting the rate of cleavage at rNMP sites in DNA by RNase H2 enzyme[14,15] alter the enzyme specificity of cleavage, potentially modifying the DNA binding capacity of RNase H2. This observation is also supported by the fact that the ArA pattern is more frequent in the *rnh201*-G42S libraries than in the *rnh201*Δ libraries. Since *rnh201*Δ eliminates the catalytic activity of the RNase H2 enzyme, rNMP patterns seen in *rnh201*Δ reflect the rNMP patterns without RER. Thus, the CrA and CrC dinucleotides, which are frequent in *rnh201*Δ but not in *rnh201*-G42S libraries (Figure 4C), particularly on the lagging strand (Figures 5E and 5F) are the ones being preferentially removed in the *rnh201*-G42S cells. Vice versa, the ArA and ArC, and in part ArG dinucleotides, which are frequent in *rnh201*Δ and even more frequent in *rnh201*-G42S libraries but not in *rnh203*-K46W libraries (Figure 4C), particularly on the leading strand (Figures 5E and 5F) are the ones being preferentially removed in the *rnh203*-K46W cells. The dinucleotide patterns also reflect the type of rNMPs that are incorporated by DNA Pols. Our results confirmed a leading strand preference for dAMP upstream of rA, rC, rG, and in part rU, and a lagging stand preference for dCMP upstream of any rNMP in *rnh201*Δ cells, as previously found.[21] These strand bias preferences are consistent across different techniques used to capture rNMPs from *rnh201*Δ cells.[24] Pol δ, DNA polymerase of the lagging strand, incorporates fewer rNMPs compared to Pol ε, which primarily acts on the leading strand.[4,21,48] This is a key reason for the observed increase in rNMP embedment in the leading versus lagging strand in both *rnh201*Δ and *rnh201*-G42S libraries. It is a notable finding that *rnh201*-G42S libraries display an even higher leading-strand bias of rNMPs than *rnh201*Δ libraries (Figures 5A and 5B). Since a low fidelity DNA Pol ε mutant that incorporates more rNMPs shows bias for rNMPs preceded by dAMP on the leading strand, and vice versa, a DNA Pol δ mutant that incorporates more rNMPs shows bias for rNMPs preceded by dCMP on the lagging strand, the ArN-dinucleotide pattern mainly reflects Pol ε-incorporated rNMPs, and the CrN-dinucleotide pattern mainly reflects Pol δ-incorporated rNMPs.[21] Interestingly, the *rnh201*-G42S libraries show a preference for dAMP upstream of the rNMPs (Figure 4C), which is markedly accentuated on the leading strand (Figures 5C and 5D), indicating that the *rnh201*-G42S mutant of RNase H2 may not be able to cleave effectively leading-strand rNMPs, which are mainly incorporated by DNA Pol ε. On the contrary, the dinucleotide pattern CrC not only is stronger in the *rnh203*-K46W libraries than in any other genotype libraries (Figure 4C), but it is also accentuated on the lagging strand, which is synthesized mainly by Pol δ (Figures 5C and 5D). These results suggest that the Rnh203-K46W mutant does not efficiently process rNMPs introduced by Pol δ, especially rC, but it does process rNMPs introduced by Pol ε. Thus, both Rnh201-G42S and Rnh203-K46W represent separation-of-function mutants, revealing that the RNase H2-RER activity operates through distinct mechanisms on the leading vs. the lagging strand and/or toward rNMPs introduced by Pol ε vs. Pol δ. These findings may be related to the low complex stability of both the RNase H2A-G37S and the RNase H2C-R69W mutants.[15]

While we find a strong dNMP preferences upstream of the nuclear rNMP sites, the dNMP preference downstream of the nuclear rNMP sites is less conserved (Figure S9). However, we detected an apparent preference for the dGMP following rA and rU in libraries of *rnh203*-K46W cells. Such dinucleotide-downstream patterns may reflect an altered DNA binding capacity of the RNase H2 complex, potentially biased to specific sequence contexts without affecting the rate of rNMP removal. It has been reported that both the RNase H2A-G37S and the RNase H2C-R69W mutant enzymes have a reduce interaction with the proliferating cell nuclear antigen (PCNA).[15] Thus, it is also possible that the dinucleotide patterns observed in the Rnh203-K46W mutant libraries, e.g., CrC, rAG, and rUG, reflect poor interaction with PCNA. Whereas the reduced interaction with PCNA of the Rnh201-G42S mutant might be hindered in the *rnh201*-G42S libraries due to the low catalytic function of the enzyme.

Implications for Aicardi-Goutières syndrome and other diseases related to the processing of ribonucleoside monophosphates in DNA

There are numerous examples in which work in budding yeast has been relevant for better understanding the biology of human cells and/or human diseases. Particularly, human genetic defects associated with DNA replication and repair have often been studied directly in yeast because of the evolutionary conservation of genes and their functions. It is not a chance that yeast has been considered a "honorary mammal."[49] While the sequence similarity between the yeast and human RNase H2 subunits is low particularly in the non-catalytic subunits,[50] we chose to characterize the yeast *rnh201*-G42S and *rnh203*-K46W mutations, which are orthologs of the human mutations RNASEH2A-G37S and RNASEH2C-R69W mutations, respectively, and are in regions of the RNase H2 catalytic subunit A and the subunit C, respectively, that are conserved between yeast and human cells. Quantification of rNMPs in the combined nuclear and mitochondrial genome of the yeast AGS ortholog mutant Rnh201-G42S suggested increased level of rNMPs in the mutant.[26,42] Currently, there are no studies characterizing the features of rNMPs in the AGS or AGS-ortholog mutants. Our results show that the Rnh201-G42S mutant has significantly more rNMPs in nuclear DNA, relative to rNMPs in mitochondrial DNA, than wild-type Rnh201 and the Rnh203-K46W mutant, and almost to the level of those found in the Rnh201-null mutant. There is a good likelihood that such an effect is conserved in the corresponding human mutant RNaseH2A-G37S. Initial analysis of rNMP distribution in human DNA of RNASEH2A-defective cells carrying a low fidelity mutant of DNA Pol ε has revealed a leading-strand bias of rNMP incorporation in human nuclear DNA[51] in line with results obtained using corresponding mutants in budding yeast.[21] Another rNMP-processing feature observed in yeast *rnh201*-null cells, such as the presence of small deletions at short nucleotide repeats,[52] has also been observed in human RNASEH2A-defective cells.[53] These results support a significant conservation of rNMP features related to the RNase H2 activity between yeast and human cells. Therefore, we believe that the new features of rNMPs that we found in the yeast AGS ortholog mutants Rnh201-G42S and Rnh203-K46W have a significant value to better understand the effects and consequences of the corresponding RNase H2 mutants in human DNA.

Examining mutations in yeast AGS ortholog mutants of RNase H2 not only enhances our understanding of AGS mechanisms but also sheds light on genome maintenance processes and the effects of DNA repair deficiencies on human health. RNase H2 impairments and mutations are also observed in conditions such as systemic lupus erythematosus (SLE),[54,55] and Werner syndrome, indicating broader implications for autoimmune disorders and progeria.[56]

It is also interesting to note that the defects in RER are associated with squamous cell carcinoma in mice.[57] Moreover, defects in RNaseH2A and H2B subunits are reported to cause inflammatory response.[27,58] In a recent study, RNaseH2C has also been reported to be a novel metastasis susceptibility factor in breast cancer.[59] In a different context, RNaseH2A has been shown to be important for sustaining cancer proliferation,[60–65] and knockout of RNase H2 can have tumor suppressive effects.[58,66] By examining patterns of embedded rNMPs, we can gain insights into the diverse activities of the RNase H2 enzyme in removing rNMPs under different circumstances and in various cell types. Similar to how the RNASEH2A gene is recognized as a prognostic biomarker in many cancers,[62] rNMP features may also serve a predictive role, indicating the cellular state and progression of diseases, including cancer, AGS, and other immune-related conditions. With the multi-faceted roles of RNase H2 subunits in preventing human diseases, it is increasingly important to investigate the effects of different mutations in RNase H2-subunit genes and their relation to rNMP embedment in human diseases. This exploration opens new possibilities for disease prevention, early-stage detection, and interventions. Notably, this study not only identifies embedded rNMP features in faulty RNase H2 but also offers valuable guidance for future research on rNMP features in the broader human genome across various cell types.

Limitation of the study

Here, we have presented yeast models of two AGS mutations in RNase H2 in a pioneer work to understand the different rNMP characteristics found in the genomic DNA of two ortholog mutants of human RNaseH2, RNaseH2A-G37S and RNaseH2C-R69W. The analyses were performed in yeast ortholog mutants Rnh201-G42S and Rnh203-K46W corresponding to the AGS mutants RNase H2A-G37S and RNase H2C-R69W in human RNase H2, respectively. It would be interesting to explore patterns and hotspots of rNMPs embedded in the DNA of human cell lines or cells derived from patients carrying the same AGS mutations, as well as other AGS mutations. Complementary studies in human cells carrying AGS mutations would help to have a comprehensive understanding of the functions and consequences of the rNMPs embedded in the genome and their impact on DNA and RNA metabolism of the mutant cells. Due to the rare cases of patients with AGS and lack of knowledge on embedded rNMP features in human nuclear genome, yeast was used as the eukaryotic model to determine the feasibility of expanding the work to the human genome, while it is possible that not all features in AGS-orthologous mutants in yeast can be extrapolated to AGS mutants in humans.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Multiple sequencing alignment of RNase H2 protein sequences
  - Yeast strains and manipulations
  - Ribose-seq and library preparation
  - Preliminary data processing and alignment
  - Germline and somatic mutation analysis
  - Ribonucleotide percentages, composition, and sequence context analysis
  - Ribonucleotide count normalization and mapping common hotspots
  - Nucleotide-frequency analysis and visualization using heatmaps
  - Heatmap analysis around the ARS regions on the leading and lagging strands
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110012.

ACKNOWLEDGMENTS

We thank S. Biliya, N. Djeddar, and A. Bryksin from the Molecular Evolution Core for advice and support with high-throughput sequencing, and the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology for their research cyberinfrastructure resources and services.

We acknowledge S. Randhawa, Y. Lee and M. Sun for critically reading the manuscript; and all members of the Storici laboratory for assistance and feedback on this study.

We acknowledge funding from the National Institutes of Health, the National Institute of Environmental Health Sciences (R01ES026243 to F.S.), the Howard Hughes Medical Institute Faculty Scholar grant 55108574 (F.S.), the Mathers Foundation (AWD-002589 to F.S.), the W. M. Keck Foundation (to F.S.). to support this work, and the Tom and Marie Patton Distinguished Professor Award (to F.S.).

AUTHOR CONTRIBUTIONS

Conceptualization: F.S., D.L.K., and T.Y., and with help of A.L.G.; methodology: D.L.K. and T.Y. with help from A.L.G, K.M., G.N., C.M. (Chance Meers), I.V., K.C., Z.H.M., C.M. (Celine Mouawad), and F.S; investigation: D.L.K., A.L.G, T.Y., and F.S.; data curation: D.L.K. with help of T.Y. and A.L.G.; visualization, formal analysis, and validation: D.L.K.; interpretation: D.L.K. and F.S.; writing - original draft: D.L.K.; writing – review and editing: D.L.K., T.Y., and F.S.; funding acquisition, resources, and supervision: F.S.

DECLARATION OF INTERESTS

We have a patent related to this study: Storici, F., Hesselberth, J.R., and Koh, K. D. Methods to detect Ribonucleotides in deoxyribonucleic acids. GTRC-6522, 2013; U.S. 10,787,703 B1 Sep. *29*, 2020.
https://uspto.report/patent/grant/10,787,703.

Received: October 2, 2023
Revised: March 15, 2024
Accepted: May 14, 2024
Published: May 16, 2024

REFERENCES

1. Fazzi, E., Cattalini, M., Orcesi, S., Tincani, A., Andreoli, L., Balottin, U., De Simone, M., Fredi, M., Facchetti, F., Galli, J., et al. (2013). Aicardi–Goutieres syndrome, a rare neurological disease in children: A new autoimmune disorder? Autoimmun. Rev. *12*, 506–509. https://doi.org/10.1016/J.AUTREV.2012.08.012.
2. Al Mutairi, F., Alfadhel, M., Nashabat, M., El-Hattab, A.W., Ben-Omran, T., Hertecant, J., Eyaid, W., Ali, R., Alasmari, A., Kara, M., et al. (2018). Phenotypic and Molecular Spectrum of Aicardi-Goutières Syndrome: A Study of 24 Patients. Pediatr. Neurol. *78*, 35–40. https://doi.org/10.1016/J.PEDIATRNEUROL.2017.09.002.
3. Crow, Y.J., Massey, R.F., Innes, J.R., Pairaudeau, P.W., Rowland Hill, C.A., Woods, C.G., Ali, M., Livingston, J.H., Lebon, P., Nischall, K., et al. (2004). Congenital glaucoma and brain stem atrophy as features of Aicardi–Goutières syndrome. Am. J. Med. Genet. *129A*, 303–307. https://doi.org/10.1002/AJMG.A.30250.
4. Williams, J.S., Lujan, S.A., and Kunkel, T.A. (2016). Processing ribonucleotides incorporated during eukaryotic DNA replication. Nat. Rev. Mol. Cell Biol. *17*, 350–363. https://doi.org/10.1038/NRM.2016.37.
5. Łazowski, K., Faraz, M., Vaisman, A., Ashton, N.W., Jonczyk, P., Fijalkowska, I.J., Clausen, A.R., Woodgate, R., and Makiela-Dzbenska, K. (2023). Strand specificity of ribonucleotide excision repair in Escherichia coli. Nucleic Acids Res. *51*, 1766–1782. https://doi.org/10.1093/NAR/GKAD038.
6. Kellner, V., and Luke, B. (2020). Molecular and physiological consequences of faulty eukaryotic ribonucleotide excision repair. EMBO J. *39*, e102309. https://doi.org/10.15252/EMBJ.2019102309.
7. Sassa, A., Yasui, M., and Honma, M. (2019). Current perspectives on mechanisms of ribonucleotide incorporation and processing in mammalian DNA. Genes Environ. *41*, 3. https://doi.org/10.1186/s41021-019-0118-7.
8. Potenski, C.J., and Klein, H.L. (2014). How the misincorporation of ribonucleotides into genomic DNA can be both harmful and helpful to cells. Nucleic Acids Res. *42*, 10226–10234. https://doi.org/10.1093/NAR/GKU773.
9. Cerritelli, S.M., and El Hage, A. (2020). RNases H1 and H2: guardians of the stability of the nuclear genome when supply of dNTPs is limiting for DNA synthesis. Curr. Genet. *66*, 1073–1084. https://doi.org/10.1007/S00294-020-01086-8.
10. Zhou, Z.X., Williams, J.S., Lujan, S.A., and Kunkel, T.A. (2021). Ribonucleotide incorporation into DNA during DNA replication and its consequences. Crit. Rev. Biochem. Mol. Biol. *56*, 109–124. https://doi.org/10.1080/10409238.2020.1869175.
11. 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. Methods Mol. Biol. *2049*, 17–37. https://doi.org/10.1007/978-1-4939-9736-7_2/.
12. Koh, K.D., Balachander, S., Hesselberth, J.R., and Storici, F. (2015). Ribose-seq: global mapping of ribonucleotides embedded in genomic DNA. Nat. Methods *12*, 251–257. https://doi.org/10.1038/NMETH.3259.
13. Cerritelli, S.M., and Crouch, R.J. (2009). Ribonuclease H: the enzymes in Eukaryotes. FEBS J. *276*, 1494–1505. https://doi.org/10.1111/J.1742-4658.2009.06908.X.
14. Nishimura, T., Baba, M., Ogawa, S., Kojima, K., Takita, T., Crouch, R.J., and Yasukawa, K. (2019). Characterization of six recombinant human RNase H2 bearing Aicardi-Goutiéres syndrome causing mutations. J. Biochem. *166*, 537–545. https://doi.org/10.1093/JB/MVZ073.
15. Kind, B., Muster, B., Staroske, W., Herce, H.D., Sachse, R., Rapp, A., Schmidt, F., Koss, S., Cardoso, M.C., and Lee-Kirsch, M.A. (2014). Altered spatio-temporal dynamics of RNase H2 complex assembly at replication and repair sites in Aicardi–Goutières syndrome. Hum. Mol. Genet. *23*, 5950–5960. https://doi.org/10.1093/HMG/DDU319.
16. Sui, Y., Epstein, A., Dominska, M., Zheng, D.Q., Petes, T.D., and Klein, H.L. (2022). Ribodysgenesis: sudden genome instability in the yeast Saccharomyces cerevisiae arising from RNase H2 cleavage at genomic-embedded ribonucleotides. Nucleic Acids Res. *50*, 6890–6902. https://doi.org/10.1093/NAR/GKAC536.
17. Lockhart, A., Pires, V.B., Bento, F., Kellner, V., Luke-Glaser, S., Yakoub, G., Ulrich, H.D., and Luke, B. (2019). RNase H1 and H2 Are Differentially Regulated to Process RNA-DNA Hybrids. Cell Rep. *29*, 2890–2900.e5. https://doi.org/10.1016/J.CELREP.2019.10.108.
18. O'Connell, K., Jinks-Robertson, S., and Petes, T.D. (2015). Elevated Genome-wide instability in yeast mutants lacking RNase H activity. Genetics *201*, 963–975. https://doi.org/10.1534/GENETICS.115.182725.
19. Ashour, M.E., and Mosammaparast, N. (2021). Mechanisms of damage tolerance and repair during DNA replication. Nucleic Acids Res. *49*, 3033–3047. https://doi.org/10.1093/NAR/GKAB101.
20. 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., et al. (2020). Ribonucleotide incorporation in yeast genomic DNA shows preference for cytosine and guanosine preceded by deoxyadenosine. Nat. Commun. *11*, 2447. https://doi.org/10.1038/s41467-020-16152-5.
21. Xu, P., and Storici, F. (2021). Frequency and patterns of ribonucleotide incorporation around autonomously replicating sequences in yeast reveal the division of labor of replicative DNA polymerases. Nucleic Acids Res. *49*, 10542–10557. https://doi.org/10.1093/NAR/GKAB801.

22. Wanrooij, P.H., Engqvist, M.K.M., Forslund, J.M.E., Navarrete, C., Nilsson, A.K., Sedman, J., Wanrooij, S., Clausen, A.R., and Chabes, A. (2017). Ribonucleotides incorporated by the yeast mitochondrial DNA polymerase are not repaired. Proc. Natl. Acad. Sci. USA *114*, 12466–12471. https://doi.org/10.1073/PNAS.1713085114.
23. Berglund, A.K., Navarrete, C., Engqvist, M.K.M., Hoberg, E., Szilagyi, Z., Taylor, R.W., Gustafsson, C.M., Falkenberg, M., and Clausen, A.R. (2017). Nucleotide pools dictate the identity and frequency of ribonucleotide incorporation in mitochondrial DNA. PLoS Genet. *13*, e1006628. https://doi.org/10.1371/JOURNAL.PGEN.1006628.
24. Xu, P., Yang, T., Kundnani, D.L., Sun, M., Marsili, S., Gombolay, A.L., Jeon, Y., Newnam, G., Balachander, S., Bazzani, V., et al. (2024). Light-strand bias and enriched zones of embedded ribonucleotides are associated with DNA replication and transcription in the human-mitochondrial genome. Nucleic Acids Res. *52*, 1207–1225. https://doi.org/10.1093/NAR/GKAD1204.
25. El-Sayed, W.M.M., Gombolay, A.L., Xu, P., Yang, T., Jeon, Y., Balachander, S., Newnam, G., Tao, S., Bowen, N.E., Brůna, T., et al. (2021). Disproportionate presence of adenosine in mitochondrial and chloroplast DNA of Chlamydomonas reinhardtii. iScience *24*, 102005. https://doi.org/10.1016/J.ISCI.2020.102005.
26. Potenski, C.J., Epshtein, A., Bianco, C., and Klein, H.L. (2019). Genome instability consequences of RNase H2 Aicardi-Goutières syndrome alleles. DNA Repair *84*, 102614. https://doi.org/10.1016/J.DNAREP.2019.04.002.
27. Pokatayev, V., Hasin, N., Chon, H., Cerritelli, S.M., Sakhuja, K., Ward, J.M., Morris, H.D., Yan, N., and Crouch, R.J. (2016). RNase H2 catalytic core Aicardi-Goutières syndrome–related mutant invokes cGAS–STING innate immune-sensing pathway in mice. J. Exp. Med. *213*, 329–336. https://doi.org/10.1084/JEM.20151464.
28. Figiel, M., Chon, H., Cerritelli, S.M., Cybulska, M., Crouch, R.J., and Nowotny, M. (2011). The Structural and Biochemical Characterization of Human RNase H2 Complex Reveals the Molecular Basis for Substrate Recognition and Aicardi-Goutières Syndrome Defects. J. Biol. Chem. *286*, 10540–10550. https://doi.org/10.1074/JBC.M110.181974.
29. Chon, H., Vassilev, A., Depamphilis, M.L., Zhao, Y., Zhang, J., Burgers, P.M., Crouch, R.J., and Cerritelli, S.M. (2009). Contributions of the two accessory subunits, RNASEH2B and RNASEH2C, to the activity and properties of the human RNase H2 complex. Nucleic Acids Res. *37*, 96–110. https://doi.org/10.1093/NAR/GKN913.
30. Reijns, M.A.M., Bubeck, D., Gibson, L.C.D., Graham, S.C., Baillie, G.S., Jones, E.Y., and Jackson, A.P. (2011). The Structure of the Human RNase H2 Complex Defines Key Interaction Interfaces Relevant to Enzyme Function and Human Disease. J. Biol. Chem. *286*, 10530–10539. https://doi.org/10.1074/JBC.M110.177394.
31. Chon, H., Sparks, J.L., Rychlik, M., Nowotny, M., Burgers, P.M., Crouch, R.J., and Cerritelli, S.M. (2013). RNase H2 roles in genome integrity revealed by unlinking its activities. Nucleic Acids Res. *41*, 3130–3143. https://doi.org/10.1093/NAR/GKT027.
32. Crow, Y.J., Chase, D.S., Lowenstein Schmidt, J., Szynkiewicz, M., Forte, G.M.A., Gornall, H.L., Oojageer, A., Anderson, B., Pizzino, A., Helman, G., et al. (2015). Characterization of human disease phenotypes associated with mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and IFIH1. Am. J. Med. Genet. *167A*, 296–312. https://doi.org/10.1002/AJMG.A.36887.
33. Rice, G.I., Forte, G.M.A., Szynkiewicz, M., Chase, D.S., Aeby, A., Abdel-Hamid, M.S., Ackroyd, S., Allcock, R., Bailey, K.M., Balottin, U., et al. (2013). Assessment of interferon-related biomarkers in Aicardi-Goutières syndrome associated with mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, and ADAR: a case-control study. Lancet Neurol. *12*, 1159–1169. https://doi.org/10.1016/S1474-4422(13)70258-8.
34. Gombolay, A.L., and Storici, F. (2021). Ribose-Map: A bioinformatics toolkit for ribonucleotide sequencing experiments. Software Impacts *10*, 100136. https://doi.org/10.1016/J.SIMPA.2021.100136.
35. Gombolay, A.L., and Storici, F. (2021). Mapping ribonucleotides embedded in genomic DNA to single-nucleotide resolution using Ribose-Map. Nat. Protoc. *16*, 3625–3638. https://doi.org/10.1038/s41596-021-00553-x.
36. Gombolay, A.L., Vannberg, F.O., and Storici, F. (2019). Ribose-Map: a bioinformatics toolkit to map ribonucleotides embedded in genomic DNA. Nucleic Acids Res. *47*, e5. https://doi.org/10.1093/NAR/GKY874.
37. Li, L., Shen, W., Zheng, J., Lai, R., Tan, C., Tan, Y., and Jiang, Y. (2015). Mutation of the conserved GRG motif and decreasing activity of human RNase H2. Open Life Sci. *10*, 322–327. https://doi.org/10.1515/BIOL-2015-0033.
38. Storici, F., Lewis, L.K., and Resnick, M.A. (2001). In vivo site-directed mutagenesis using oligonucleotides. Nat. Biotechnol. *19*, 773–776. https://doi.org/10.1038/90837.
39. Cherry, J.M., Ball, C., Weng, S., Juvik, G., Schmidt, R., Adler, C., Dunn, B., Dwight, S., Riles, L., Mortimer, R.K., and Botstein, D. (1997). Genetic and physical maps of Saccharomyces cerevisiae. Nature *387*, 67–73. https://doi.org/10.1038/387s067.
40. Keskin, H., and Storici, F. (2015). Defects in RNase H2 Stimulate DNA Break Repair by RNA Reverse Transcribed into cDNA. MicroRNA *4*, 109–116. https://doi.org/10.2174/2211536604666150820120129.
41. de Zamaroczy, M., and Bernardi, G. (1986). The GC clusters of the mitochondrial genome of yeast and their evolutionary origin. Gene *41*, 1–22. https://doi.org/10.1016/0378-1119(86)90262-3.
42. Uehara, R., Cerritelli, S.M., Hasin, N., Sakhuja, K., London, M., Iranzo, J., Chon, H., Grinberg, A., and Crouch, R.J. (2018). Two RNase H2 Mutants with Differential rNMP Processing Activity Reveal a Threshold of Ribonucleotide Tolerance for Embryonic Development. Cell Rep. *25*, 1135–1145.e5. https://doi.org/10.1016/J.CELREP.2018.10.019.
43. Coffin, S.R., Hollis, T., and Perrino, F.W. (2011). Functional consequences of the RNase H2A subunit mutations that cause Aicardi-Goutières Syndrome. J. Biol. Chem. *286*, 16984–16991. https://doi.org/10.1074/jbc.M111.228833.
44. Rohman, M.S., Koga, Y., Takano, K., Chon, H., Crouch, R.J., and Kanaya, S. (2008). Effect of the disease-causing mutations identified in human RNase H2 on the activities and stabilities of yeast RNase H2 and archaeal RNase HII. FEBS J. *275*, 4836–4849. https://doi.org/10.1111/J.1742-4658.2008.06622.X.
45. Lazzaro, F., Novarina, D., Amara, F., Watt, D.L., Stone, J.E., Costanzo, V., Burgers, P.M., Kunkel, T.A., Plevani, P., and Muzi-Falconi, M. (2012). RNase H and Postreplication Repair Protect Cells from Ribonucleotides Incorporated in DNA. Mol. Cell *45*, 99–110. https://doi.org/10.1016/J.MOLCEL.2011.12.019.
46. Pizzi, S., Sertic, S., Orcesi, S., Cereda, C., Bianchi, M., Jackson, A.P., Lazzaro, F., Plevani, P., and Muzi-Falconi, M. (2015). Reduction of hRNase H2 activity in Aicardi–Goutières syndrome cells leads to replication stress and genome instability. Hum. Mol. Genet. *24*, 649–658. https://doi.org/10.1093/HMG/DDU485.
47. Dragoni, F., Garau, J., Sproviero, D., Orcesi, S., Varesio, C., De Siervi, S., Gagliardi, S., Cereda, C., and Pansarasa, O. (2022). Characterization of Mitochondrial Alterations in Aicardi–Goutières Patients Mutated in RNASEH2A and RNASEH2B Genes. Int. J. Mol. Sci. *23*, 14482. https://doi.org/10.3390/IJMS232214482.
48. Johnson, R.E., Klassen, R., Prakash, L., and Prakash, S. (2015). A Major Role of DNA Polymerase δ in Replication of Both the Leading and Lagging DNA Strands. Mol. Cell *59*, 163–175. https://doi.org/10.1016/J.MOLCEL.2015.05.038.
49. Resnick, M.A., and Cox, B.S. (2000). Yeast as an honorary mammal. Mutat. Res. *451*, 1–11. https://doi.org/10.1016/S0027-5107(00)00036-1.
50. Cerritelli, S.M., and Crouch, R.J. (2009). Ribonuclease H: The enzymes in eukaryotes. FEBS J. *276*, 1494–1505. https://doi.org/10.1111/j.1742-4658.2009.06908.x.
51. Koyanagi, E., Kakimoto, Y., Minamisawa, T., Yoshifuji, F., Natsume, T., Higashitani, A., Ogi, T., Carr, A.M., Kanemaki, M.T., and Daigaku, Y. (2022). Global landscape of replicative DNA polymerase usage in the human genome. Nat. Commun. *13*, 7221. https://doi.org/10.1038/s41467-022-34929-8.
52. Cho, J.E., Kim, N., Li, Y.C., and Jinks-Robertson, S. (2013). Two distinct mechanisms of Topoisomerase 1-dependent mutagenesis in yeast. DNA Repair *12*, 205–211. https://doi.org/10.1016/J.DNAREP.2012.12.004.
53. Reijns, M.A.M., Parry, D.A., Williams, T.C., Nadeu, F., Hindshaw, R.L., Rios Szwed, D.O., Nicholson, M.D., Carroll, P., Boyle, S., Royo, R., et al. (2022). Signatures of TOP1 transcription-associated mutagenesis in cancer and germline. Nature *602*, 623–631. https://doi.org/10.1038/s41586-022-04403-y.
54. Mackenzie, K.J., Carroll, P., Lettice, L., Tarnauskaitė, Ž., Reddy, K., Dix, F., Revuelta, A., Abbondati, E., Rigby, R.E., Rabe, B., et al. (2016). Ribonuclease H2 mutations induce a cGAS/STING-dependent innate immune response. EMBO J. *35*, 831–844. https://doi.org/10.15252/EMBJ.201593339.
55. Pendergraft, W.F., and Means, T.K. (2015). AGS, SLE, and RNASEH2 mutations: translating insights into therapeutic advances. J. Clin. Invest. *125*, 102–104. https://doi.org/10.1172/JCI78533.
56. Sugawara, S., Okada, R., Loo, T.M., Tanaka, H., Miyata, K., Chiba, M., Kawasaki, H., Katoh, K., Kaji, S., Maezawa, Y., et al. (2022). RNaseH2A downregulation drives inflammatory gene expression via genomic

DNA fragmentation in senescent and cancer cells. Commun. Biol. *5*, 1420. https://doi.org/10.1038/s42003-022-04369-7.

57. Hiller, B., Hoppe, A., Haase, C., Hiller, C., Schubert, N., Müller, W., Reijns, M.A.M., Jackson, A.P., Kunkel, T.A., Wenzel, J., et al. (2018). Ribonucleotide excision repair is essential to prevent squamous cell carcinoma of the skin. Cancer Res. *78*, 5917–5926. https://doi.org/10.1158/0008-5472.CAN-18-1099.
58. Aden, K., Bartsch, K., Dahl, J., Reijns, M.A.M., Esser, D., Sheibani-Tezerji, R., Sinha, A., Wottawa, F., Ito, G., Mishra, N., et al. (2019). Epithelial RNase H2 Maintains Genome Integrity and Prevents Intestinal Tumorigenesis in Mice. Gastroenterology *156*, 145–159.e19. https://doi.org/10.1053/J.GASTRO.2018.09.047.
59. Deasy, S.K., Uehara, R., Vodnala, S.K., Yang, H.H., Dass, R.A., Hu, Y., Lee, M.P., Crouch, R.J., and Hunter, K.W. (2019). Aicardi-Goutières syndrome gene Rnaseh2c is a metastasis susceptibility gene in breast cancer. PLoS Genet. *15*, e1008020. https://doi.org/10.1371/JOURNAL.PGEN.1008020.
60. Marsili, S., Tichon, A., Kundnani, D., and Storici, F. (2021). Gene Co-Expression Analysis of Human RNASEH2A Reveals Functional Networks Associated with DNA Replication, DNA Damage Response, and Cell Cycle Regulation. Biology *10*, 221. https://doi.org/10.3390/BIOLOGY10030221.
61. Flanagan, J.M., Funes, J.M., Henderson, S., Wild, L., Carey, N., and Boshoff, C. (2009). Genomics screen in transformed stem cells reveals RNASEH2A, PPAP2C, and ADARB1 as putative anticancer drug targets. Mol. Cancer Ther. *8*, 249–260. https://doi.org/10.1158/1535-7163.MCT-08-0636.
62. Zhang, L., Yang, Y., Cheng, L., Cheng, Y., Zhou, H.H., and Tan, Z.R. (2018). Identification of Common Genes Refers to Colorectal Carcinogenesis with Paired Cancer and Noncancer Samples. Dis. Markers *2018*, 3452739. https://doi.org/10.1155/2018/3452739.
63. Williams, K.A., Lee, M., Hu, Y., Andreas, J., Patel, S.J., Zhang, S., Chines, P., Elkahloun, A., Chandrasekharappa, S., Gutkind, J.S., et al. (2014). A Systems Genetics Approach Identifies CXCL14, ITGAX, and LPCAT2 as Novel Aggressive Prostate Cancer Susceptibility Genes. PLoS Genet. *10*, e1004809. https://doi.org/10.1371/JOURNAL.PGEN.1004809.
64. Xu, J., Liu, H., Yang, Y., Wang, X., Liu, P., Li, Y., Meyers, C., Banerjee, N.S., Wang, H.K., Cam, M., et al. (2019). Genome-wide profiling of cervical rna-binding proteins identifies human papillomavirus regulation of rnaseh2a expression by viral e7 and e2f1. mBio *10*, e02687-18. https://doi.org/10.1128/MBIO.02687-18.
65. Dai, B., Zhang, P., Zhang, Y., Pan, C., Meng, G., Xiao, X., Wu, Z., Jia, W., Zhang, J., and Zhang, L. (2016). RNaseH2A is involved in human gliomagenesis through the regulation of cell proliferation and apoptosis. Oncol. Rep. *36*, 173–180. https://doi.org/10.3892/OR.2016.4802.
66. Zhao, F., Liu, A., Gong, X., Chen, H., Wei, J., Chen, B., Chen, S., Yang, R., Fan, Y., and Mao, R. (2022). Hypoxia-induced RNASEH2A limits activation of cGAS-STING signaling in HCC and predicts poor prognosis. Tumori *108*, 63–76. https://doi.org/10.1177/03008916211026019.
67. Xu, P., and Storici, F. (2021). RibosePreferenceAnalysis: Analyzing the preference of rNMPs embedded in genomic DNA. Software Impacts *10*, 100149. https://doi.org/10.1016/J.SIMPA.2021.100149.
68. Siow, C.C., Nieduszynska, S.R., Müller, C.A., and Nieduszynski, C.A. (2012). OriDB, the DNA replication origin database updated and extended. Nucleic Acids Res. *40*, D682–D686. https://doi.org/10.1093/NAR/GKR1091.
69. Langmead, B., Wilks, C., Antonescu, V., and Charles, R. (2019). Scaling read aligners to hundreds of threads on general-purpose processors. Bioinformatics *35*, 421–432. https://doi.org/10.1093/BIOINFORMATICS/BTY648.
70. Depristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., Del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498. https://doi.org/10.1038/ng.806.
71. Bergstrom, E.N., Huang, M.N., Mahto, U., Barnes, M., Stratton, M.R., Rozen, S.G., and Alexandrov, L.B. (2019). SigProfilerMatrixGenerator: A tool for visualizing and exploring patterns of small mutational events. BMC Genom. *20*, 685. https://doi.org/10.1186/S12864-019-6041-2.
72. Stuckey, S., Mukherjee, K., and Storici, F. (2011). In Vivo Site-Specific Mutagenesis and Gene Collage Using the Delitto Perfetto System in Yeast Saccharomyces cerevisiae. Methods Mol. Biol. *745*, 173–191. https://doi.org/10.1007/978-1-61779-129-1_11.
73. Kalhorzadeh, P., Hu, Z., Cools, T., Amiard, S., Willing, E.M., De Winne, N., Gevaert, K., De Jaeger, G., Schneeberger, K., White, C.I., and De Veylder, L. (2014). Arabidopsis thaliana RNase H2 deficiency counteracts the needs for the WEE1 checkpoint kinase but triggers genome instability. Plant Cell *26*, 3680–3692. https://doi.org/10.1105/TPC.114.128108.
74. McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y.M., Buso, N., Cowley, A.P., and Lopez, R. (2013). Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res. *41*, W597–W600. https://doi.org/10.1093/NAR/GKT376.
75. Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL–EBI. Nucleic Acids Res. *38*, W695–W699. https://doi.org/10.1093/NAR/GKQ313.
76. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539. https://doi.org/10.1038/MSB.2011.75.
77. Shaban, N.M., Harvey, S., Perrino, F.W., and Hollis, T. (2010). The Structure of the Mammalian RNase H2 Complex Provides Insight into RNA·DNA Hybrid Processing to Prevent Immune Dysfunction. J. Biol. Chem. *285*, 3617–3624. https://doi.org/10.1074/JBC.M109.059048.
78. Baker Brachmann, C., Davies, A., Cost, G.J., Caputo, E., Li, J., Hieter, P., and Boeke, J.D. (1998). Designer Deletion Strains derived from Saccharomyces cerevisiae S288C: a Useful set of Strains and Plasmids for PCR-mediated Gene Disruption and Other Applications. Yeast *14*, 115–132. https://doi.org/10.1002/(SICI)1097-0061(19980130)14:2.
79. Keskin, H., Shen, Y., Huang, F., Patel, M., Yang, T., Ashley, K., Mazin, A.V., and Storici, F. (2014). Transcript-RNA-templated DNA recombination and repair. Nature *515*, 436–439. https://doi.org/10.1038/nature13682.
80. Storici, F., Bebenek, K., Kunkel, T.A., Gordenin, D.A., and Resnick, M.A. (2007). RNA-templated DNA repair. Nature *447*, 338–341. https://doi.org/10.1038/nature05720.
81. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Curr. Protoc. Bioinf. *43*, 11.10.1–11.10.33. https://doi.org/10.1002/0471250953.BI1110S43.
82. Wagih, O. (2017). ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics *33*, 3645–3647. https://doi.org/10.1093/BIOINFORMATICS/BTX469.
83. Saleem, M.A.M., Mendoza-Parra, M.A., Cholley, P.E., Blum, M., and Gronemeyer, H. (2017). Epimetheus - a multi-profile normalizer for epigenomic sequencing data. BMC Bioinf. *18*, 259. https://doi.org/10.1186/S12859-017-1655-3.
84. Füllgrabe, J., Gosal, W.S., Creed, P., Liu, S., Lumby, C.K., Morley, D.J., Ost, T.W.B., Vilella, A.J., Yu, S., Bignell, H., et al. (2023). Simultaneous sequencing of genetic and epigenetic bases in DNA. Nat. Biotechnol. *41*, 1457–1464. https://doi.org/10.1038/s41587-022-01652-0.
85. Yabuki, N., Terashima, H., and Kitada, K. (2002). Mapping of early firing origins on a replication profile of budding yeast. Gene Cell. *7*, 781–789. https://doi.org/10.1046/J.1365-2443.2002.00559.X.
86. Yousefi, R., and Rowicka, M. (2019). Stochasticity of replication forks' speeds plays a key role in the dynamics of DNA replication. PLoS Comput. Biol. *15*, e1007519. https://doi.org/10.1371/JOURNAL.PCBI.1007519.
87. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/BIOINFORMATICS/BTQ033.
88. Quinlan, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr. Protoc. Bioinf. *47*, 11.12.1–11.12.34. https://doi.org/10.1002/0471250953.BI1112S47.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw and analyzed data	This paper	GEO: GSE240399
Saccharomyces cerevisiae reference genome build SacCer2 and SacCer3	University of California Santa Cruz (UCSC) Genomes and *Saccharomyces* Genome Database (SGD), Cherry J et al., Nature, 1997[39]	https://genome.ucsc.edu/, RRID:SCR_005780 http://www.yeastgenome.org/, RRID:SCR_004694
Experimental models: Organisms/strains		
S. cerevisiae: Strain background: BY4742: *MATα his3Δ1 leu2Δ0 lys2Δ0 ura3* Δ0	Storici et al., Nat Biotech. 2001[38]	KK-1
S. cerevisiae: Strain background: BY4742 *rnh203*-K46W (CM1339)	This study	CM1339
S. cerevisiae: Strain background: BY4742 *rnh203*-K46W (CM1340)	This study	CM1340
S. cerevisiae: Strain background: BY4742 *rnh201*-G42S	This study	CM1333
S. cerevisiae: Strain background: BY4742 *rnh201Δ::hygMX4*	Balachander et al., Nat Commun 2019[11]	SB305
Oligonucleotides		
Adaptors. L1-L8 for ribose-seq, see Table S2	This study	N/A
Adaptor S for ribose-seq: GTTGCGACACGGATCTATCAACACT	This study	N/A
Primer PCR.1 in ribose-seq: GTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTTGATAGATCCGT GTCGCAAC	This study	N/A
Primer PCR.2 in ribose-seq: ACACTCTTTCCCTACACGAC	This study	N/A
Primers in the second PCR for ribose-seq, see Table S2 (PCR.701,702,705,712)	This study	N/A
Primers in the second PCR for ribose-seq, see Table S2 (PCR.501-508)	This study	N/A
Software and algorithms		
Ribose-map Bioinformatics Toolkit	Gombolay et al., Nat Protocols, 2021[35]	https://github.com/agombolay/ribose-map
Subtraction and mismatch removal	Xu et al., NAR, 2023[24]	https://doi.org/10.5281/zenodo.8121711 https://doi.org/10.5281/zenodo.8121723
Customized scripts for organelle abundance and composition visualizations	This Paper	https://doi.org/10.5281/zenodo.10455801
Custom R scripts for hotspot analysis	This Paper	https://doi.org/10.5281/zenodo.8152090
Mono and dinucleotide preference heatmaps and its statistics	Xu et al., Software Impacts, 2021[67]	https://github.com/xph9876/RibosePreferenceAnalysis https://doi.org/10.5281/zenodo.10457211
Other		
Autonomous replicating Sequences annotations, OriDB	OriDB, Siow C et al., NAR, 2012[68]	https://cerevisiae.oridb.org/; RRID:SCR_025365

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bowtie2	Langmead B et al. Bioinformatics, 2019[69]	https://bowtie-bio.sourceforge.net/bowtie2/index.shtml , RRID:SCR_016368
GATK (v.4.3.0)	Van der Auwera et al., Genomics in the Cloud, 2020[70]	https://gatk.broadinstitute.org/hc/en-us , RRID:SCR_001876
DNA-seq analysis pipeline	This paper	https://doi.org/10.5281/zenodo.10453018
SigProfilerMatrixGenerator(v 1.2.14 for sacCer3 genome)	Bergstrom EN et al., BMC Genomics, 2019[71]	https://osf.io/s93d5/wiki/home/ , RRID:SCR_023122

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Francesca Storici (storici@gatech.edu).

Materials availability

All unique/stable reagents generated in this study are available from the lead contact.

Data and code availability

- ribose-seq and DNA-seq data generated for this study have been deposited at NCBI's Gene Expression Omnibus repository under GEO:GSE240399 and are publicly available as of the date of publication. Accession numbers are also listed in the key resources table.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The *delitto perfetto* system[72] was used for creating mutant strain using the background BY4742 (strain). The experimental method for analysis of single base resolution of ribonucleotide embedment used is ribose-seq.[11,12] Analysis used for sequencing data analysis of ribose-seq is using Ribose-Map Toolkit[34–36] paired with modified shared or common hotspot analysis previously reported[24] and nucleotide heatmaps using RibosePreferenceAnalysis software.[67]

METHOD DETAILS

Multiple sequencing alignment of RNase H2 protein sequences

To observe the conserved domains between RNase H2 subunits of human and yeast, we used the amino acid sequences of RNase H2 catalytic subunit H2A, and accessory subunits H2B and H2C from *H. sapiens* and *S. cerevisiae*.[13,25,73] We performed multiple sequence alignment using Clustal Omega[74–76] to predict conserved domains between all three subunits of RNase H2 enzyme complex in human and yeast. We annotated the amino acids (Asp-35,Glu-35,Asp142, and Asp170) involved in the catalytic activity,[77] conserved domains in gray[25] and yeast ortholog mutations in red (*rnh201*-G42S, *rnh202*-L52R, *rnh202*-H109R, *rnh202*-L186F, *rnh202*-T204I, and *rnh203*-K46W) previously reported.[26]

Yeast strains and manipulations

We used haploid *Saccharomyces cerevisiae* strains from the BY4742 background. Strain BY4742 is a derivative of S288C, was used in the *S. cerevisiae* gene disruption project, and have *his3*Δ1 *leu2*Δ0 *lys2*Δ0 *ura3*Δ0 genotype with MATα mating type.[78] Standard genetic and molecular biology methods were used for growth, gene disruption, isolation of mutants, selection, genome engineering, colony PCR, and sequence analysis of genomic DNA.[72,79,80] Other genotypes were derived from KK-1[20] by using the *delitto perfetto* method[72] to generate *rnh201*-G42S (GGC->AGC codon change) and *rnh203*-K46W (AAG -> TGG codon change) mutations to result in yeast orthologs of AGS-ortholog mutants. The *rnh201*Δ genotype was derived by replacing *RNH201* via transformation with a PCR product containing *hygMX4* or *kanMX4* cassette flanked by 50 nucleotides of sequence homologous to regions upstream and downstream of *RNH201* open reading frame.[20] All mutations were confirmed by sequence analysis of PCR products obtained from amplification of a DNA region surrounding the specific mutation. The primers used for strain construction are listed in Table S2.

Ribose-seq and library preparation

The analyzed ribose-seq libraries were constructed, following the previous protocols with a few modifications.[18,22,23] Three or more ribose-seq libraries were prepared for each genotype of the yeast strains using up to three different sets of restriction enzymes (RE1: DraI, EcoRV, and SspI; RE2: AluI, DraI, EcoRV, and SspI; and RE3: HaeIII and RsaI). This strategy allowed us (i) to verify that the conclusions taken from our analyses of ribose-seq data are not influenced by a particular set of restriction enzymes used to fragment the DNA extracted from the different yeast genotypes, and (ii) to further confirm reproducibility of the results. All ribose-seq libraries have a specific barcode within the sequence of the Unique Molecular Identifier (UMI) to distinguish the libraries from each other in the sequencing run and eliminate PCR duplicates.[36] All commercially available enzymes in the ribose-seq protocol were used according to the manufacturer's instructions. In all DNA purification steps, nuclease free water was used to elute DNA. *S. cerevisiae* cells were inoculated in 150 mL of a YPD liquid medium containing yeast extract, peptone, and 2% (wt/vol) dextrose and incubated in a shaking incubator for 2 days to reach stationary phase. Genomic DNA was extracted using Qiagen Genomic DNA protocol "Preparation of Yeast Samples". Successively, 40 μg of yeast genomic DNA was fragmented using restriction enzymes to produce blunt-ended fragments with an average size of 450 base pairs (bp) in length. Multiple sets of restriction enzymes were used for different library preparations, as shown in Table S1. The fragmented DNA was then purified by QIAquick PCR Purification Kit (QIAGEN). The purified DNA fragments were tailed with dATP (New England Biolabs, NEB) by using Klenow Fragment (3′→5′ exo-) (NEB) for 30 min at 37°C and purified by QIAquick PCR Purification Kit. Following dA-tailing and purification, the DNA fragments were annealed with a partially double-stranded adapter (Adapter.L1∼L8 with Adapter.S, Table S2) by T4 DNA ligase (NEB) for overnight incubation at 16°C. After that, the adapter-ligated products were purified using HighPrep PCR (MAGBIO). The annealed fragments were treated with a final concentration of 0.3 M NaOH for 2 h at 55°C to denature the DNA strands, and to cleave at the rNMP sites resulting in 2′,3′-cyclic phosphate, 3′-phosphate, or 2′-phosphate termini. This was followed with neutralization using 2 M HCl and purification using HighPrep RNA Elite (MAGBIO). All the successive purification steps were performed using HighPrep RNA Elite (MAGBIO). The single-stranded DNA fragments were initially incubated at 95°C for 3 min and then on ice for 2 min to fully denature any secondary structures of DNA. incubated with a final concentration of 1 μM *Arabidopsis thaliana* tRNA ligase (AtRNL), 50 mM Tris-HCl pH 7.5, 40 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, and 300 μM ATP in a volume of 20 μL for 1 h at 30°C and followed by purification. AtRNL ligates the 2′-phosphate ends of rNMP-terminated ssDNA fragment to its opposite 5′-phosphate end, which results in a circular ssDNA.

The fragments were then treated with T5 Exonuclease (NEB) in a volume of 50 μL for 30 min at 37°C to degrade all remaining linear ssDNA fragments. After purification, the circular fragments were incubated with a final concentration of 1 μM 2′-phosphotransferase (Tpt1), 20 mM Tris-HCl pH 7.5, 5 mM MgCl2, 0.1 mM DTT, 0.4% Triton X-100, and 10 mM NAD^+ in a volume of 40 μL for 1 h at 30°C to remove the 2′-phosphate presenting at the ligation junction. After Tpt1 treatment and purification, the circular fragments were amplified by two rounds of polymerase chain reaction (PCR) with Q5-High Fidelity polymerase (NEB). Both PCR rounds begin with an initial denaturation at 98°C for 30 s. Denaturation at 98°C for 10 s, primer annealing at 65°C for 30 s, and DNA extension at 72°C for 30 s are then performed. These three steps are repeated for 6 cycles in the first PCR round, and for 11 cycles in the second PCR round. Successively, there is a final extension reaction at 72°C for 2 min for both PCRs.

The first round of PCR was performed to amplify and introduce the adapter sequences of Illumina TruSeq CD index primers. The primers PCR.1 and PCR.2 used for the first round were the same for all libraries. The second round of PCR was performed to attach specific indexes i7 and i5 for each library. The sequences of PCR primers and indexes can be found in Table S2. The amplified DNA fragments were then performed with a double-sided size selection to select specific sizes of DNA between 250 bp and 650 bp using HighPrep PCR (MAGBIO). The resulting ribose-seq libraries were quantified with Qubit and Bioanalyzer. The prepared libraries were sequenced on the Illumina HiSeq X ten and NextSeq 500 platforms at Admera Health. For this study, we built 4 libraries for analysis of wild-type RNase H2, five libraries for *rnh203*-K46W, three libraries for *rnh201*-G42S, and two libraries for *rnh201*Δ. For controls, we also included two wild-type RNase H2 (FS106 and FS142)[20] and one *rnh201*Δ libraries (FS138)[20] that were previously published.

Preliminary data processing and alignment

For the ribose-seq libraries, the sequencing reads consist of an eight-nucleotide UMI, a three-nucleotide molecular barcode, the tagged nucleotide (the nucleotide tagged during ribose-seq from which the position of the rNMP is determined), and the sequence directly downstream from the tagged nucleotide. The UMI corresponds to sequence position 1–6 and 10–11, the molecular barcode corresponds to position 7–9, the tagged nucleotide corresponds to position 12, and the tagged nucleotides downstream of the sequence corresponds to positions 13+ of the raw FASTQ sequences. The rNMP is the reverse complement of the tagged nucleotide. Before aligning the sequencing reads to the reference genome, the reads were trimmed based on sequencing quality and the custom ribose-seq adaptor sequence using cutadapt 1.16 (-q 15 -m 62 -a "AGTTGCGACACGGATCTATCA"). In addition, to ensure accurate alignment to the reference genome,[39] reads containing fewer than 50 nucleotides of genomic DNA (those nucleotides located downstream from the tagged nucleotide) after trimming were discarded. Following quality control, the Alignment and Coordinate Modules of the Ribose-Map toolkit were used to process and analyze the reads.[36] The Alignment Module de-multiplexed the trimmed reads by the appropriate molecular barcode, aligned the reads to the reference genome[39] (sacCer2) using Bowtie2,[69] and de-duplicated the aligned reads using UMI-tools, as also mentioned in previously published study about ribose-seq analysis on yeast strains.[20] Based on the alignment results, the Coordinate Module filtered the reads to retain only those with a mapping quality score of at least 30 (probability of misalignment <0.001) and calculated the chromosomal coordinates and per-nucleotide counts of rNMPs.

Due to the efficient removal of rNMPs by RNase H2, nuclear libraries of wild-type cells generally had a much lower number of reads compared to the mitochondrial libraries of the same cells, and thus had a higher number of background reads that originated from the capture of restriction enzyme ends likely by residual activity of T4 DNA ligase. To improve the quality of analysis and filter out any background reads or artifacts, we filtered out those reads that showed restriction enzyme cleavage sites at rNMP embedment sites, which removes any linear single strand DNA-only fragments being captured. We then also filter out rNMP mismatches between the nucleotide in reference genome versus sequencing reads. The final file generated is of BED format and all further analysis is performed this bed file after filtration of restriction enzyme site and mismatched rNMPs.

Germline and somatic mutation analysis

DNA extracted from the yeast cells for ribose-seq was also sequenced using Illumina HiSeq X Ten Platform. The fastq reads were trimmed for any residual Illumina adapters and filtered to keep reads above the quality threshold of 15 and length 50 bp. Alignment of trimmed reads was done on sacCer3 genome[39] using bowtie2[69] to generate SAM (Sequence Alignment Map) and indexed BAM (Binary Alignment Map) file, which was further used to get vcf (variant call format) files for germline and somatic mutations. For generating germline mutations, we used the pipeline previously published for yeast genome analysis[53] which used GATK's(v4.3.0) Haplotype caller[70,81] to obtain mutants and use a filtering criteria of QD < 2.0 and FS > 200 to obtain high quality germline variants. To obtain somatic mutations, we used GATK's(v4.3.0) Mutect2[70,81] to generate variants which were further filtered using MutectFilterCalls. After obtaining high quality germline and somatic mutations, we observed indel patterns in AGS mutant cell lines using SigProfilerMatrixGenerator[71] (v1.2.14 for SacCer3 genome) which also finds matching indel patterns in COSMIC (Catalogue of Somatic mutations in Cancer) database.

Ribonucleotide percentages, composition, and sequence context analysis

Since each library can have different coverage, we used percentage of rNMPs in nuclear and mitochondrial DNA to determine the rate of embedment in difference genotypes. We then visualized these percentages using horizontal bar graphs generated using customized R scripts.

The Composition Module of Ribose-Map was used to obtain raw and normalized percentages of each rNMP base (A, C, G, U/T) in nuclear and mitochondrial DNA of all genotypes. The normalized percentages are the raw rNMP counts normalized on the nucleotide counts in the corresponding (nuclear or mitochondrial) reference genome. The Sequence Module of Ribose-Map was used to generate nucleotide sequence context plots encoded using the R-ggplot package. The 1 % of the most abundant rNMP sites in the genome of each library were selected as hotspots, and the consensus sequence around these hotspots was visualized using customized R scripts and plotted using ggseqlogo.[82]

Ribonucleotide count normalization and mapping common hotspots

We addressed a concern that the high frequency of rNMP locations might be attributed to variations in sequence depth, suggesting they could be hotspots due to this factor.[83,84] To obtain the sequence depth we use fastq reads obtained after trimming step mentioned in Germline and somatic mutation analysis method. Alignment was performed using bowtie2[69] to sacCer2[39] genome to obtain an alignment file in SAM (Sequence Alignment Map) format. The SAM file was then sorted and indexed to a BAM (Binary Alignment Map) file.

Coverage per base pair (c_b) and mean genomic coverage per base pair (c_{gb}) as calculated using SAMtools. Relative base coverage (C_{R_b}) of each base (b) in the genome is calculated as follows:

$$C_{R_b} = \frac{C_b}{C_{gb}}, \text{where } C_{gb} = \frac{\sum_{b=1}^{n} C_b}{n} \text{ and } n \text{ is total number of base pairs in genome}$$

This relative coverage per base (C_{R_b}) is further used to calculate rNMPs per base coverage (R_{c_b}) by normalizing the rNMP counts per base pair (R_b) using the following formula:

$$R_{c_b} = \frac{R_b}{C_{R_b}}$$

We calculate R_{c_b} for each base pair in each Library using respective genotype coverage information. Further to obtain a relative measure of each site of embedment with respect to rNMPs observed in each library, we calculate Enrichment Factor, EF_{c_b} using rNMPs per base coverage R_{c_b} for each base in each library, also previously used for generating hotspots and enriched zones[24]

The enrichment factor is calculated relative to the length of genome and region of interest, region here being each base pair.

$$EF_{c_b} = \frac{R_{c_b}/L_b}{R_{c_g}/L_g}, \text{where } R_{c_g} = \frac{\sum_{b=1}^{n} R_{c_b}}{n} \text{ and } n = \text{total number of base pairs in genome}$$

$$L_b = 1 \text{ for } 1 \text{ base pair and } L_g \text{ is the length of genome in base pair.}$$

R_{c_b} rounded to the closest integer $\lfloor R_{c_b} \rceil$ in each Library is used to calculate the presence of rNMPs in each library of every genotype. Top common hotspot locations in each genotype are determined based on highest mean Enrichment Factor and recurrence of rNMPs in a genomic location in at least 2 libraries of each genotype.

Nucleotide-frequency analysis and visualization using heatmaps

To generate the mononucleotide heatmaps, in every nuclear and mitochondrial ribose-seq library, the frequency of each type of rNMP (FN: FA, FC, FG, or FU) was calculated as a ratio of percentage of each rNMP base divided by the percentage of each dNTP base in the reference genome. Normalized frequency for each rNMP base was calculated using the frequency of the respective rNMP base divided by the sum of frequencies of each rNMP base.[20,67] The sum of all normalized frequencies in the mononucleotide heatmaps sums up to 1.

$$F_N = \frac{R_N}{D_N}, NF_N = \frac{F_N}{F_A+F_C+F_G+F_U}$$

$$\text{where } R_N \text{ is percentage of ribonucleotides wher } N\epsilon\{A, C, G, U\}$$

$$\text{and } D_N \text{ is percentage of dinucleotides in reference genome where } N\epsilon\{A, C, G, T\}$$

For dinucleotide heatmaps, normalized frequency was calculated for each deoxyribonucleotide immediately upstream or downstream of rNMP embedment site for each base such that the sum of NR (upstream) or RN (downstream) frequencies for each R would sum upto 1.[20,67]

$$F_{NR} = \frac{R_{NR}}{D_{NN}}, NF_{NR} = \frac{F_{NR}}{F_{AR}+F_{CR}+F_{GR}+F_{TR}}$$

$$F_{RN} = \frac{R_{RN}}{D_{NN}}, NF_{RN} = \frac{F_{RN}}{F_{RA}+F_{RC}+F_{RG}+F_{RT}}$$

$$\text{where } R_{NR} \text{ is percentage of dinucleotide } NR, \text{ upstream nucleotide } N\epsilon\{A, C, G, T\} \text{ of ribonucleotides}, R\epsilon\{A, C, G, U\}$$

$$R_{RN} \text{ is percentage of dinucleotide } RN, \text{ downstream nucleotide } N\epsilon\{A, C, G, T\} \text{ of ribonucleotides}, R\epsilon\{A, C, G, U\} \text{ and}$$

$$D_{NN} \text{ is percentage of dinucleotides in reference genome where } N\epsilon\{A, C, G, T\}$$

Heatmap analysis around the ARS regions on the leading and lagging strands

ARS annotations of OriDB[68] were downloaded and transferred to the sacCer2 reference genome[39] with Liftover software. Only confirmed ARSs were included (n = 410.[68] Among them, 276 ARSs have a known firing time.[85] They were divided into two halves, early-firing ARSs ($T \leq 24.7$ min, n = 139) and late-firing ARSs ($T > 24.7$ min, n = 137). The upstream and downstream 15-kb regions are considered as the flanks of an ARS. The 5′-upstream flanks of ARSs for both the Watson and Crick strands correspond to the lagging strand, and the 3′-downstream flanks of ARSs correspond to the leading strand. If two ARSs are close to each other and the distance between them is smaller than 30 kb, the position of the collision point of the corresponding converging replication forks is calculated with their firing times and the average fork moving speed of 1.6 kb/min.[86] Bedtools[87,88] was used to get rNMPs overlapping in 4 -10 kb of leading and lagging strand from early firing ARS to generate dinucleotide frequency heatmaps for each rNMP and normalized on the dinucleotide frequencies found in respective strand.

QUANTIFICATION AND STATISTICAL ANALYSIS

We used a one-sided Mann-Whitney *U* test to compare frequency results of heatmap data with expected frequency of 0.25 for each rNMP base and dinucleotide combination, in nuclear and mitochondrial DNA of each genotype library. We also did two-sided Mann-Whitney U test to compare frequencies of AGS mutants or *rnh201*Δ mutant with wild-type frequencies. *P*-values of these tests are indicated in Tables S6, S7, S8, and S9.