nature portfolio

Peer Review File

Yeast EndoG prevents genome instability by degrading extranuclear DNA species

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

In this study, Yu et al. use a sequencing-based approach to characterize the insertion of cytoplasmic DNA fragments within the nuclear genome in budding yeast. They provide a convincing validation of this novel method (called Break-Ins), and demonstrate its accuracy in identifying simple or complex insertions at an inducible DSB site, in a mutant context previously demonstrated to enhance such events (Fig. 1a-c; Extended Data Fig. 1-2). They further use Break-Ins to demonstrate that transfer of mitochondrial (mt) and transposon DNA to the nuclear genome increases during chronological (stationary phase) and replicative aging (Fig. 1d-e; Fig. 2). Importantly, they identify the Nuc1/EndoG nuclease as an important regulator of these processes (Fig. 2), by degrading long mtDNA species (Fig. 3; Extended Data Fig. 5) and Ty retrotransposon cDNA intermediates (Fig. 4; Extended Data Fig.3-4), while promoting the insertion of short mtDNA fragments, likely representing incomplete degradation products (Fig. 2; Extended Data Fig. 3-4). Overall, the featured data are of high quality, appear carefully controlled for confounding factors (e.g. survival during stationary phase; preferred size for NHEJ substrates) and provide detailed insights into the mechanisms by which cytoplasmic DNA species could jeopardize genome stability, a phenomenon relevant in several physiological and pathological situations in other models, including human.

My main comment comes from the authors' conclusion that their observations relate to *cytoplasmic* DNA degradation, i.e. to Nuc1-mediated DNA degradation in the cytoplasm (as notably stated in the title and the summary). As pointed by the authors, while Nuc1/EndoG is a mitochondrial protein, it had been shown to relocate out of the mitochondria in two studies (references 32-33). However, both studies are actually compatible with Nuc1 relocalization to the nucleus rather than into the cytoplasm. In reference 32, imaging assays show that stress situations trigger Nuc1 relocalization from the mitochondria to the nucleus. In reference 33, a fraction of Nuc1 is detected outside mitochondria during meiosis, but solely based on biochemical fractionation assays, with cytosolic fractions likely contaminated by nuclear components. In this line, all the observations featured by Yu et al. could be explained by Nuc1 degrading retrotransposon or mtDNA species in the nucleus, thus controlling their nuclear availability rather than their cytoplasmic abundance or their transfer into the nucleus. To address this point, the authors need to perform the following experiments:
- They should assess where Nuc1 localizes in their experimental situations, e.g. by performing microscopy analysis of Nuc1-GFP cells (as in ref. 32), notably during stationary phase. Alternatively, fractionation experiments, if carefully controlled, could be considered.
- They should probe whether Nuc1 nuclear localization is required for the observed phenotypes, by looking at genomic insertion frequencies and TRP1-mtDNA transfer to the nucleus in the absence of Kap123, the karyopherin that reportedly controls the nuclear import of Nuc1 (ref. 32).

Minor comments:
- Analysis of the features of donor sequences suggests that inserted DNA fragments tend to originate more frequently from R-loop-prone regions in stationary phase cells (Extended Data Fig.

3e). However, the R-loop maps used for this analysis were obtained from exponentially-growing cells (ref. 68). This limitation should be indicated.
- The authors argue that Nuc1 activity mostly prevents the more frequent nuclear transfer of very long mtDNA, as compared to its role in stable short NUMTs insertions. However, as I understand it, 1% of nuc1Δ cells have transferred the long TRP1 mtDNA fragment to the nucleus at 8 days of stationary phase (Fig. 3b; text p10), while 1% of wt cells show short mtDNA insertions at the HO break at the same timepoint (Fig 2b; Suppl. Table 1). Could the authors clarify how they estimate a 1000-fold difference between these two measurements, as indicated? It would also be useful to plot the proportion of each type of donor DNA sequences, during stationary phase, using pie charts as in Extended Data Fig. 2a.
- As for Ty1 cDNAs, which are reverse transcription products, mtDNA could be transferred to the nucleus through an RNA intermediate, especially since EndoG can degrade RNA and since donor products often arise from DNA:RNA hybrid regions. Could the authors discuss this possibility? Were mitochondrial R-loop maps taken into consideration when looking at the features of mtDNA donor sequences (Extended Data Fig. 3e)?
- The authors state that "In nuc1Δ cells, long mtDNA is transferred to the nucleus, yet this DNA is not frequently inserted" (p12). I understand that this low frequency refers to the 1 out of 180 TRP+ spores carrying a stable TRP1 insertion. Could the authors provide the frequency of mitotically-stable TRP1 integrations among TRP+ cells during stationary phase (Fig. 3b data)?
- In the discussion, the authors indicate that "circular DNA devoid of centromeres shortens the lifespan of the cells" (p14). However, nuc1Δ cells, although displaying increased circular mtDNA levels, do not exhibit a shorter lifespan (Extended Data Fig. 4e). This is probably due to the small (1%) fraction of cells carrying such molecules - can the authors clarify this point?
- Extended Data Figure 1 legend: what does "grigosequencing" mean?

Reviewer #2 (Remarks to the Author):

In the presented work Yang Yu, Xin Wang and coworkers have been using a yeast model where cytoplasmic DNA can be captured into a specific nuclear locus to study the role of DNA2 and Nuc1 in preventing such insertions. The approach is quite ingenious and is further complemented by another model to track the mtDNA incorporation into nucleus using a nutrient selection marker. However, the involvement of Nuc1 (EndoG) in mtDNA degradation is well established, so apart for the methodological innovation, the novelty regarding mechanistic insights into mtDNA maintenance or NUMT generation are unclear. Also the genomic stability stated in the title, in the yeasts has not really been addressed.

As a general note: The introduction and results headings are missing? The methods section is after the figure legends, which is probably not following the journal standard? Also, the much of text explaining the results belongs to discussion. If the journal permits, these could be combined. The readability of the manuscript would also benefit from restructuring and reduction of redundancies.

Page 4: "[...] showed a smear of products above the band [...]"
Why is the "normally repaired MATa fragment" still present, if there should be a selection to lose

this site due to the HO induction? What percentage of these have lost the HO site due to small indel or a point mutation? Why does the double-strand break not lead to erosion of the ends prior NHEJ and result also in shorter PCR products? Does a PCR of a single colony produce one band or two bands? In Fig 1C control PCRs without the HO induction should be also presented.

"[...] were represented by a lower read number, likely representing events more difficult to sequence."
What are events that are more difficult to sequence?

Page 5: "Indeed, a dramatic increase in insertions was observed [...]"
How is this normalized? Can differences in PCR efficiency be ruled out? Same for the Fig 2: How is the insertion frequency normalized to account for differences in the amount of the sequenced DNA. If the "normally repaired" MATa product was used, please indicate this.

Page 7: "Unexpectedly, NUMTs were reduced by [...]
Considering that mitochondrial DNA is large (86 kb as mentioned later in the text) can it be that your PCR method is not detecting long insertions? EndoG will chop up the DNA to smaller fragments, increasing the likelihood of short insertions. The decay of the released mtDNA to shorter fragments in the ageing cells could then explain the increased insertion rate. You could try to see what happens in YME1 deficient cells, which should have increased mtDNA release into the cytosol. This is actually addressed later using a different method, but the authors do not tell what they consider as "long" mtDNA insertion? Also regarding the title, is there evidence of increased genomic instability in these yeasts?

Page 10: What makes the mtDNA circular in the nucleus? This is also mentioned later, but its relevance to the study is unclear.

Page 12: "In nuc1D cells, long mtDNA is transferred to the nucleus, yet this DNA is not frequently inserted at HO breaks or anywhere else in the genome."

Same comment as for the previous. How do you know it is not inserted but that your methods are not picking the long insertions up? You might be able to see these by analyzing the MATa locus of individual clones (ones not producing a short PCR product) using a Southern blot.

"[...] we tested 400 to >1000 NHEJ products individually by PCR. "
This is a bit vague. Also, the PCR biases towards short products, so it can still be that only a minority of the insertions are detected.

Page 15: "it prevents the far more frequent nuclear transfer of very long mtDNA that also occasionally integrates into the genome."
So long mtDNA transfer (to nucleus?) is more frequent, but integration is not (if it is occasional, figure 3 legend states "rarely integrates")? What is this based on? The discussion is somewhat confusing.

Page 25, middle section of the text: MATa -locus should be in capitals. There are also other small issues, such as using symbol font instead of proper ascii greek letters.

Reviewer #3 (Remarks to the Author):

This manuscript analyzes insertions that are introduced into HO-induced double-stranded breaks in budding yeast using analysis of reads generated by Illumina MiSeq sequencing. The authors demonstrate an increase in the ligation of insertions in stationary phase cells and reveal a role for the EndoG/Nuc1 nuclease in suppressing the introduction of Ty fragments but increasing the introduction of mtDNA fragments into these junctions. In contrast, Nuc1 prevents the introduction of larger mtDNA fragments as measured by the formation of circular mtDNA molecules, suggesting that the small mtDNA fragments arise from incomplete mtDNA degradation.

This is a very interesting and straightforward manuscript providing evidence for the role of Nuc1 in preventing aberrant NHEJ in stationary phase cells. The authors do an exemplary job of performing important controls, such as the requirement of the products on the NHEJ DNA ligase 4, and appropriate filtering of the MiSeq reads. Many of their observations, such as the propensity of multi-insertions and the length biases of the insertions, are rather unexpected. I believe this manuscript will be of high interest and will be appropriate for publication when the following comment is addressed.

A key missing experiment in this manuscript is the analysis of a nuc1 spt3 double mutant (nuc1 to cause increased Ty insertion, spt3 to suppress Ty transcription) to show that the Ty inserts do arise from the cDNA and not due to Ty-specific fragments formed in the genomic DNA. The current draft assumes that the Ty DNA inserts arise from reverse transcription and cDNA formation in the cytoplasm, but this assumption is never tested.

Reviewer #3 (Remarks on code availability):

The code is a combination of shell scripts and perl code to link together the output of other programs (e.g. blast/bwa/pear/bedtools).

The iDSBins and iDSBindel “packages” are clearly more mature (last update was older) and seem more polished. The comments seem quite legible and I can follow along reasonably well (though the SCAR program names in the iDSBins markdown seem to refer to some older code that no longer exists).

The LargeInsertionFeature “package” isn’t as polished (a lot of hard-coded paths that you would have to change if you were running it on your machine and there’s a shell script variable that I’m not sure where it gets defined), but certainly better than lots of code generated by researchers that I’ve seen.

They do provide example data to run their programs on, which is a big plus.

I haven't tried to run their pipelines, though. And subtle bugs would be really hard to identify. Given the fact that the authors can identify lots of these events, my guess is that bugs in the pipeline (if any) would be subtle. The authors would have caught the really bad problems.

At a higher level of review, I think the overall strategy of how they are attempting to identify insertions in the induced DSB seems reasonable, and I have no concerns with their software.

We thank the reviewers for their excellent suggestions, which have significantly improved the manuscript. We have conducted several additional experiments to address the reviewers' concerns. New/altered text in the manuscript is highlighted in grey color. The main message of the manuscript remains unchanged. The major points of the revised manuscript are:

A. Kap123 karyopherin-mediated transfer of Nuc1 to the nucleus does not regulate insertions of mtDNA or retrotransposon cDNA.

B. Small indels were analyzed as requested by the reviewers, and we found that the pattern of indels differs in the stationary phase, indicating an increased role of polymerase Pol4 in non-dividing cells. This important new result is shown in the new main figure 2.

C. Ty's cDNA insertions are dependent on Spt3.

Responses to reviewer comments:

Reviewer #1:

My main comment comes from the authors' conclusion that their observations relate to *cytoplasmic* DNA degradation, i.e. to Nuc1-mediated DNA degradation in the cytoplasm (as notably stated in the title and the summary). As pointed by the authors, while Nuc1/EndoG is a mitochondrial protein, it had been shown to relocate out of the mitochondria in two studies (references 32-33). However, both studies are actually compatible with Nuc1 relocalization to the nucleus rather than into the cytoplasm. In reference 32, imaging assays show that stress situations trigger Nuc1 relocalization from the mitochondria to the nucleus. In reference 33, a fraction of Nuc1 is detected outside mitochondria during meiosis, but solely based on biochemical fractionation assays, with cytosolic fractions likely contaminated by nuclear components. In this line, all the observations featured by Yu et al. could be explained by Nuc1 degrading retrotransposon or mtDNA species in the nucleus, thus controlling their nuclear availability rather than their cytoplasmic abundance or their transfer into the nucleus. To address this point, the authors need to perform the following experiments:

They should assess where Nuc1 localizes in their experimental situations, e.g. by performing microscopy analysis of Nuc1-GFP cells (as in ref. 32), notably during stationary phase. Alternatively, fractionation experiments, if carefully controlled, could be considered.
- They should probe whether Nuc1 nuclear localization is required for the observed phenotypes, by looking at genomic insertion frequencies and TRP1-mtDNA transfer to the nucleus in the absence of Kap123, the karyopherin that reportedly controls the nuclear import of Nuc1 (ref. 32).

The reviewer asked where Nuc1 nuclease degrades Ty cDNA or mtDNA and suggested that Nuc1 could degrade cDNA and mtDNA in the nucleus. We will address this question starting with a few notes. First, insertions of nuclear DNA fragments are not dependent on Nuc1. Therefore, the suggested Nuc1 activity in the nucleus would need to be selective toward mtDNA and cDNA, and not the nuclear genome. Second, the presence of Nuc1 in the nucleus would likely lead to cell death due to the degradation of chromosomes, and our analysis of templated insertions was done in cells that recovered and formed colonies. Third, Ty cDNA replication occurs within VLP (virus-like particles) in the cytoplasm. As neither VLPs nor the transposition of Ty1 has ever been reported in mitochondria, we can exclude that Nuc1 degrades cDNA in mitochondria. Fourth, as noted by the reviewer, we cited a manuscript that demonstrated a small fraction of Nuc1 in the

cytoplasm that viral RNA. These are the major reasons we originally proposed that Nuc1 can degrade DNA outside of mitochondria and the nucleus.

The reviewer suggested performing fluorescence microscopy to follow Nuc1-GFP localization. We tagged Nuc1 with GFP; however, we found that yeast stationary phase cells, which have no GFP-tagged protein, dramatically increase autofluorescence. This autofluorescence was observed across the entire spectrum of wavelengths, excluding the possibility of using other fluorescence tags. As recommended by literature listing common causes of autofluorescence, we tried several types of media and different strain backgrounds, but it was not helpful. This autofluorescence phenomenon has been previously observed in stationary cells in bacteria, yeast, and humans [1]. The figure below shows an example of such an image. In summary, fluorescence microscopy to examine Nuc1-GFP localization in stationary phase cells is not possible. To test whether Nuc1 level increases in stationary phase cells we tagged Nuc1 with FLAG tag and testes the level of protein in growing and old stationary phase cells but have not observed any change. This negative result is also shown for reviewer below but not included in a revised manuscript.

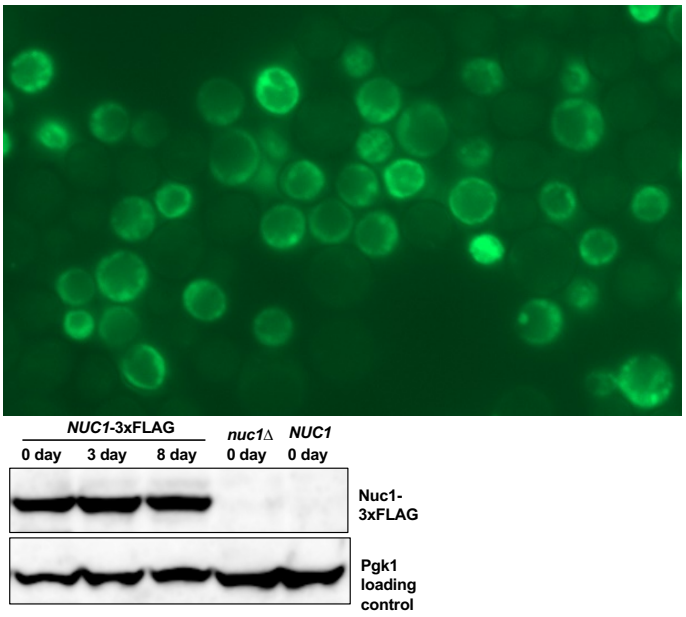

Figure for reviewer: Autofluorescence of yeast cells in 3 days stationary phase (top). These cells do not express GFP. Western blot analysis of Nuc1 protein in stationary phase cells (bottom).

The reviewer asked us to test the possible role of Kap123 because the Kap123 karyopherin was previously shown to mediate the transfer of Nuc1 to the nucleus. If Nuc1 transfer to the nucleus was required for the degradation of mtDNA and cDNA, we would expect that the *kap123*Δ mutant phenocopies *nuc1*Δ and is epistatic with the *nuc1*Δ mutant. However, the results demonstrate that *kap123*Δ has distinct phenotypes from *nuc1*Δ with respect to NUMTs, insertions of cDNA fragments, or the transfer of long mtDNA to the nucleus.

A. The *kap123*Δ mutant shows a decrease in Ty cDNA insertions compared to the wild type, which is opposite to the *nuc1*Δ mutant that has a 10-fold increase compared to the wild type. Moreover, the elimination of Kap123 in *nuc1*Δ cells decreased the level of Ty cDNA insertions by 26-fold when compared to *nuc1*Δ alone. These results indicate that Kap123 is important in transposition itself, as previously reported. Therefore, these data can neither support nor exclude the role of Kap123 in the transfer of Nuc1 to the nucleus, as in the *kap123*Δ mutant, the Ty DNA substrate for Nuc1 is likely not formed or formed at a low level. We note that the specific role of Kap123 in transposition is not known and is beyond the scope of this work.

B. The *kap123*Δ mutant shows a 3-fold decrease in NUMTs, which is much less than the 13-fold decrease observed in *nuc1*Δ cells. Moreover, the double mutant showed a further decrease in NUMTs compared to each single mutant, with a 74-fold decrease compared to the wild type. These results demonstrate that Kap123 plays a role in NUMTs formation, but this role is independent of Nuc1. The exact role of Kap123 in NUMT formation is not known but considering the broad impact of karyopherins in the transport of proteins from the cytoplasm to the nucleus, it may be related to one of many nuclear proteins needed for the capture of DNA fragments at DSBs. Accordingly, we found a decrease in insertions of all types of DNA, including nuclear genome insertions. See discussion of Kap123 related results below.

C. The *kap123*Δ mutant, unlike *nuc1*Δ cells, does not increase the transfer of long mtDNA; rather, it decreases it. The double mutant *kap123*Δ *nuc1*Δ is comparable to *nuc1*Δ cells. Because in *kap123*Δ cells there is less transfer of long *TRP1*-mtDNA compared to the wild type, and this decrease depends on Nuc1, it suggests that Nuc1 activity toward mtDNA is higher in *kap123*Δ potentially because there is less transfer of Nuc1 to the nucleus and therefore more complete degradation of mtDNA outside of the nucleus. See discussion of Kap123 related results below.

In conclusion, the role of Kap123 in the transfer of Nuc1 to the nucleus is not necessary for the degradation of long mtDNA. However, Kap123 may play a direct or indirect role in the formation of NUMTs and in general in all types of longer insertions. Additionally, we confirm previous findings implicating Kap123 in transposition. These new data are presented in new **Supplementary Figure 7**. We state in the paper (page 11-12):

"It was previously shown that Kap123 karyopherin mediates the transfer of Nuc1 to the nucleus in stressed cells [2]. Here, we tested the possible role of Kap123-mediated transfer of Nuc1 to the nucleus in the insertions of cDNA or mtDNA and the transfer of long mtDNA to the nucleus, all of which are controlled by Nuc1. Unlike in *nuc1*Δ mutant cells, *kap1*23Δ cells almost entirely lost insertions of retrotransposon cDNA. Additionally, the elimination of *KAP123* dramatically reduced insertions of retrotransposon cDNA in *nuc1*Δ cells. These results likely reflect the known role of Kap123 in promoting transposition [3,4]. Also, *kap123*Δ reduced NUMTs formation by 3-fold compared to the wild type, which is less than the reduction observed in *nuc1*Δ single mutants. The *kap123*Δ *nuc1*Δ double mutant showed much lower levels of insertions compared to each single mutant suggesting independent role of Kap123 and Nuc1 in promoting NUMTs. Insertions from the nuclear genome decreased in *kap123*Δ as well. Finally, *kap123*Δ, unlike *nuc1*Δ, showed no increase but rather a decrease in long mtDNA transfer to the nucleus, and this decrease was entirely dependent on Nuc1 (**Supplementary Fig. 7**). These results suggest that less transfer of Nuc1 to the nucleus in *kap123*Δ may help degrade mtDNA, or alternatively, that Kap123 plays a

positive role in capturing all types of DNA by transporting protein(s) needed for efficient DSB repair by NHEJ. We conclude that the degradation of extranuclear DNA by Nuc1 likely occurs before its transfer to the nucleus.

Considering the valid concern raised by the reviewer that we do not know where mtDNA/cDNA are degraded, we decided to eliminate the suggestion that Nuc1 degrades cytoplasmic DNA from the title and abstract. We do not know if mtDNA is degraded in the mitochondria or after its release from the mitochondria. A more precise way to refer to these DNA species is as "extranuclear," which indicates the origin of the DNA captured at DSB rather than where it is degraded. The title of the manuscript was changed to: "*Yeast EndoG Prevents Genome Instability by Degrading Extranuclear DNA Species*" We also adjusted the model to indicate the source of the DNA and the outcome of mtDNA transfer, and we no longer claim that mtDNA is degraded in the cytoplasm.

Minor comments:
- Analysis of the features of donor sequences suggests that inserted DNA fragments tend to originate more frequently from R-loop-prone regions in stationary phase cells (Extended Data Fig. 3e). However, the R-loop maps used for this analysis were obtained from exponentially-growing cells (ref. 68). This limitation should be indicated.

We agree and have noted this in the results section (page 6):

"*Inserted fragments also originated from R-loop-prone regions that were previously mapped in growing cells*."

- The authors argue that Nuc1 activity mostly prevents the more frequent nuclear transfer of very long mtDNA, as compared to its role in stable short NUMTs insertions. However, as I understand it, 1% of nuc1Δ cells have transferred the long TRP1 mtDNA fragment to the nucleus at 8 days of stationary phase (Fig. 3b; text p10), while 1% of wt cells show short mtDNA insertions at the HO break at the same timepoint (Fig 2b; Suppl. Table 1). Could the authors clarify how they estimate a 1000-fold difference between these two measurements, as indicated?

The NUMTs frequency shown in all figures is calculated among cells that repaired DSB by imprecise NHEJ, which typically corresponds to about 0.1% of all cells plated, while the long *TRP1*-mtDNA frequency is calculated among all cells plated. If both frequencies are calculated among all cells plated, it results in a ~250-fold difference. To provide a more precise comparison, we performed additional plating and estimated NHEJ efficiency in stationary phase cells. We found an increased rate of NHEJ in aging cells, which corrected our initial estimation of the difference between long mtDNA transfer and NUMTs. As end joining slightly increases in aging, the difference between the wild-type level of NUMTs and the number of long *TRP1*-mtDNA in *nuc1*Δ cells is approximately 250-fold and not 1000-folds. We corrected the number in results and discussion sections. We stated in results section (original manuscript) on page 6:

"Among NHEJ products that have altered the HO cleavage site (~0.2-0.4% of cells plated), about 1% carried NUMTs which corresponds to about 2-4 NUMTs per 10^5 cells plated."

It would also be useful to plot the proportion of each type of donor DNA sequences, during stationary phase, using pie charts as in Extended Data Fig. 2a.

The pie chart in Fig. 2a shows the distribution of different donor sequences in stationary phase at 8 days. Additionally, we added pie charts to **Fig. 5b**.

- As for Ty1 cDNAs, which are reverse transcription products, mtDNA could be transferred to the nucleus through an RNA intermediate, especially since EndoG can degrade RNA and since donor products often arise from DNA:RNA hybrid regions. Could the authors discuss this possibility?

NUMTs do not originate from mitochondrial RNA reverse transcribed to cDNA because we have not observed a single event where a spliced sequence was inserted at DSBs among tens of thousands of NUMTs. We noted in the original manuscript on page 8 that such insertions were not observed. However, RNA:DNA hybrids could possibly be inserted. These types of hybrids could originate from Ty replication intermediates, mtDNA, or from genomic locations. We are not aware of studies demonstrating the ability of NHEJ to ligate dsDNA ends with DNA:RNA hybrids. As we do not have evidence supporting or excluding it, we simply state:

"We propose that Nuc1 nuclease limits cDNA insertions by degrading incomplete Ty1 replication intermediates in stationary phase cells. Besides dsDNA and ssDNA that can be inserted at DSBs, it is possible that DNA:RNA hybrids are also captured by NHEJ."

- Were mitochondrial R-loop maps taken into consideration when looking at the features of mtDNA donor sequences (Extended Data Fig. 3e)?

No, we excluded mitochondrial DNA from the analysis. **Supplementary Figure 3e** shows the analysis of insertions and features of the nuclear genome only. In the methods section, we state:

"Distance analysis was based on edge distance between an insertion and its closest genome feature. Insertions coming from Ty retrotransposons, mtDNA, rDNA, MATa, or 2µ plasmid were excluded from these analyses."

In the revised manuscript, we added this information to the figure legend.

- The authors state that "In *nuc1*Δ cells, long mtDNA is transferred to the nucleus, yet this DNA is not frequently inserted" (p12). I understand that this low frequency refers to the 1 out of 180 TRP+ spores carrying a stable TRP1 insertion. Could the authors provide the frequency of mitotically-stable TRP1 integrations among TRP+ cells during stationary phase (Fig. 3b data)?

We have tested approximately 300 independent wild-type and *nuc1Δ* cells carrying *TRP1*-mtDNA in the nucleus for the stability of the Trp1+ marker. None of them carried stable *TRP1*-mtDNA. We added this information to the results section. We note that it is not possible to test thousands or millions of colonies carrying circular or integrated *TRP1*-mtDNA for stability as there is no counterselection for Trp1. Stability can only be tested one by one by streaking colonies on nonselective media followed by replica plating on Trp- plates.

Among 180 independent Trp+ spores tested, only one showed a stable Trp1 marker. To ensure that this single stable Trp+ colony corresponds to *TRP1* gene integration, we separated the chromosomes of this strain using CHEF and probed it with the *TRP1* sequence. We observed integration of TRP1 within one of the chromosomes. This is shown in the new panel c of **Supplemental Figure 6**.

- In the discussion, the authors indicate that "circular DNA devoid of centromeres shortens the lifespan of the cells" (p14). However, nuc1Δ cells, although displaying increased circular mtDNA levels, do not exhibit a shorter lifespan (Extended Data Fig. 4e). This is probably due to the small (1%) fraction of cells carrying such molecules - can the authors clarify this point?

The reviewer is correct that *nuc1*Δ cells age at a normal rate, and at the population level, nuclear mtDNA circles are unlikely to have a significant impact since only a small fraction of cells carry these circles. We clarified this issue by stating in the discussion:

"While *nuc1*Δ cells show increased levels of nuclear circular mtDNA, these circles are present in a small fraction of cells and therefore do not impact the population lifespan."

- Extended Data Figure 1 legend: what does "grigosequencing" mean?

We thank the reviewer for carefully reading the manuscript. The mistake has been corrected.

Reviewer #2 (Remarks to the Author):

In the presented work Yang Yu, Xin Wang and coworkers have been using a yeast model where cytoplasmic DNA can be captured into a specific nuclear locus to study the role of DNA2 and Nuc1 in preventing such insertions. The approach is quite ingenious and is further complemented by another model to track the mtDNA incorporation into nucleus using a nutrient selection marker.

We thank the reviewer for complimenting our methods to study the transfer of mtDNA to the nucleus.

However, the involvement of Nuc1 (EndoG) in mtDNA degradation is well established, so apart for the methodological innovation, the novelty regarding mechanistic insights into mtDNA maintenance or NUMT generation are unclear. Also the genomic stability stated in the title, in the yeasts has not really been addressed.

The role of EndoG in the degradation of mtDNA has indeed been described, for example, in response to VDAC oligomerization that results in the release of mtDNA from mitochondria. However, no published work has demonstrated that EndoG nuclease protects nuclear genome stability by degrading mtDNA species. Additionally, this work shows that EndoG degrades retrotransposon cDNA, which has not been previously shown to our knowledge. One form of genome instability that increases in the absence of EndoG is the insertions of retrotransposon cDNA at DSBs, and the second is the transfer of large mtDNA fragments to the nucleus, which can affect up to ~1% of cells. These occur during starvation or meiosis. This is also the first report

to establish the important role of yeast EndoG in the formation of NUMTs. We propose that NUMTs are byproducts of EndoG activity degrading large mtDNA. In general, we provide the first evidence in any organism that the degradation of extranuclear DNA fragments is important for the stability of the nuclear genome.

We note that any insertion of mtDNA or cDNA, which affects the integrity of the genome, is a form of genome instability. Similarly, extrachromosomal circular DNA, which is very common in cancer cells, is also a form of genome instability.

As a general note: The introduction and results headings are missing? The methods section is after the figure legends, which is probably not following the journal standard? Also, the much of text explaining the results belongs to discussion. If the journal permits, these could be combined. The readability of the manuscript would also benefit from restructuring and reduction of redundancies.

We made significant changes in the revised manuscript to better align with the style of *Nature Communications*. For the discussion, we aimed to make it clearer for the readers by adding some conclusions directly in the results section and leaving the discussion chapter for the overall main points of the manuscript. *Nature Communications* suggests having a succinct discussion.

Page 4: "[…] showed a smear of products above the band […]"
Why is the "normally repaired MATa fragment" still present, if there should be a selection to lose this site due to the HO induction? What percentage of these have lost the HO site due to small indel or a point mutation?

All these questions are excellent points. In the revised work, we put more effort into explaining all results, not just the longer (>10 bp) templated insertions. We focused on NUMTs and other insertions because these events had not been studied systematically before. Small insertions/deletions (indels) have been studied for over 25 years by many laboratories, so we did not discuss them extensively. In the revised manuscript, we cited some of this published work and better described NHEJ efficiency and indels frequency. The revised Supplemental Table 1 shows the frequency of long insertions and indels, which are the major outcomes of NHEJ. Additionally, we carefully analyzed NHEJ efficiency and the most common indels during chronological aging in yeast, which to our knowledge had not been examined before. We added several interesting points to our work:

A. The efficiency of error-prone NHEJ increases in chronologically aging cells (new **Fig. 2 a**).

B. The most common indels pattern change during chronological aging (new **Fig. 2 b,c**). Among indels the relative proportion of insertions increases when compared to cycling cells. This suggests that the role of Pol4 polymerase that responsible for insertions (among indels) increases in repair by NHEJ in stationary phase cells.

We state in the first results chapters:
"Most common sequence changes during DSB repair by NHEJ are small insertions and deletions of a few nucleotides (indels) that were previously characterized in detail (e.g. [5-7])."
After stating the frequency of long >10 bp insertions, we also say:

"The remaining events were small indels (Supplementary Table 1)."

Finally, we describe new results on small indels analysis:

"We also noted an increase of error-prone NHEJ and altered distribution of NHEJ junctions in chronologically aged cells. The proportion of "+CA" or "+ACA" nucleotide insertions that are mediated by Pol4[8,9] increased at the expense of "-ACA" mediated by Pol2 (**Fig. 2),** suggesting an increased role of Pol4-mediated error-prone NHEJ in stationary phase cells. In young *pol4*D cells, nearly all junctions are represented by "-ACA" deletions while in *pol4*D old stationary phase cells, additional longer deletions are common."

Why does the double-strand break not lead to erosion of the ends prior NHEJ and result also in shorter PCR products?

Primers used for amplicon sequencing that amplify the *MAT***a** sequence are positioned just outside of the *MAT***a** cleavage site, ensuring they pick up repaired sequences as long as there were no deletions that match the primers. We have also used primers located further outside of *MAT***a** and obtained nearly the same results as described in the original manuscript (**Supplementary Fig. 2b**). Longer deletions are not a common outcome of NHEJ in yeast. These deletions are generated by the alternative end-joining pathway (Alt-EJ), which is an inefficient pathway in yeast.

The most common sequence changes during DSB repair by NHEJ are small indels. Therefore, the primers used here amplify these events, which form the majority of the PCR product. As these typically involve just 1-3 nucleotide differences, the PCR product size on the agarose gel appears the same as the one without any break induction and repair. Longer insertions, which are the focus of this work, represent only up to a few percent of all events.

Does a PCR of a single colony produce one band or two bands? In Fig 1C control PCRs without the HO induction should be also presented.

A single colony will have only one band as it represents a single NHEJ product. We and others have analyzed thousands of individual colonies in previously published work. We have added a no-cut control to **Figure 1c**.

"[…] were represented by a lower read number, likely representing events more difficult to sequence." What are events that are more difficult to sequence?

To name a few possibilities: First, PCR does not amplify all sequences equally, although we used a high-performance hot-start polymerase, KAPA. Second, some repair sequences can be present in lower copy numbers due to lower cell numbers (not all colonies representing individual NHEJ products are the same size). Third, sequencing clusters during Illumina sequencing do not form equally well for all sequences. Please see additional comments on sequencing bias below.

Page 5: "Indeed, a dramatic increase in insertions was observed […]"

How is this normalized?

It is calculated per colony number pooled together for amplicon sequencing, and each colony represents an individual NHEJ product. This is explained in methods section.

Can differences in PCR efficiency be ruled out? Same for the Fig 2: How is the insertion frequency normalized to account for differences in the amount of the sequenced DNA. If the "normally repaired" MATa product was used, please indicate this.

All samples are processed the same way in wild type and mutants, so overall differences in the number of insertions are not dependent on PCR efficiency. As explained above, it is possible that some events are not amplified by PCR or are amplified less than others. To account for the fact that we do not sequence all templated insertions, we changed the figure description to

"Frequency of insertion among sequenced NHEJ products."

We note that insertion events represented by a high sequencing read number or just a single read are scored the same way as a single event. Therefore, read number does not affect frequency. We eliminated all clonal events by setting very stringent deduplication parameters as described in the methods section.

We assume that the reviewer refers to "normally repaired" events as the most common NHEJ products – short indels. As explained above, nearly all products of DSB repair were sequenced, including indels and long insertions. The exceptions are rare events associated with longer deletions that eliminate sequences specific to primers and possibly very long insertions over several kb.

Page 7: "Unexpectedly, NUMTs were reduced by […]
Considering that mitochondrial DNA is large (86 kb as mentioned later in the text) can it be that your PCR method is not detecting long insertions? EndoG will chop up the DNA to smaller fragments, increasing the likelihood of short insertions. The decay of the released mtDNA to shorter fragments in the ageing cells could then explain the increased insertion rate.

The Illumina MiSeq platform has a read length of 600 bp, so we cannot sequence longer events. PCR could amplify insertions of up to a few kb in length, but these are not fully sequenced by MiSeq. In cases where sequencing reads represented fragments from the same locus, separated by up to 3 kb, and in the opposite direction, we considered it a longer insertion. We found very few such events. Second, we transformed DNA of different lengths (separately) into cells that were plated on galactose-containing plates to induce a DSB. Individual colonies, and therefore individual NHEJ products, were tested for insertion. Insertions were most frequent for transformed DNA fragments of 60 to 200 bp. Longer fragments were inserted rarely. Therefore, we think that NHEJ has a preference for capturing shorter fragments around 100 bp. In the future, other methods can be applied to sequence rare longer insertions.

The interesting possibility raised by this reviewer is added to the results section (page 11):

"It's also possible that in *nuc1*Δ cells, very long mtDNA is occasionally inserted at DSBs, but these insertions cannot be sequenced using the MiSeq platform."

You could try to see what happens in YME1 deficient cells, which should have increased mtDNA release into the cytosol. This is actually addressed later using a different method, but the authors do not tell what they consider as "long" mtDNA insertion?

As suggested by the reviewer, we created a *yme1*Δ mutant and performed Break-Ins analysis. We confirmed that the *yme1*Δ mutant has a higher rate of TRP1-mtDNA transfer to the nucleus, and we found that this mutant has no NUMTs in growing cells. However, at 8 days of the stationary phase, a 3-fold higher frequency of NUMTs is observed compared to WT (see below). We decided not to include this side observation in the manuscript that focuses entirely on Nuc1. More work is required to understand the relationship between Nuc1 and Yme1 with respect to mtDNA transfer to the nucleus.

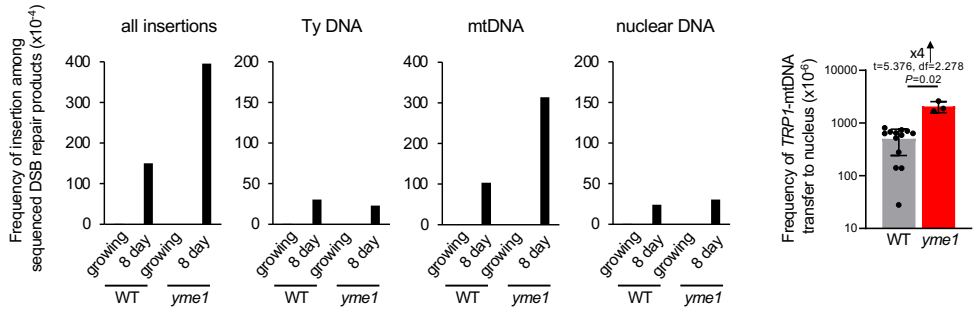

Figure for Reviewer: Break-Ins analysis (left) and *TRP1*-mtDNA transfer in *yme1*Δ (right).

Finally, we have not used the words "long mtDNA insertion." We followed long TRP1-mtDNA transfer, and by "long" we mean the size of the TRP1 gene (~1.5 kb) plus some mtDNA flanking sequences transferred to the nucleus. These fragments are rarely inserted; we found only 1 out of ~500 Trp+ colonies that carried mtDNA-TRP1 integrated into the genome.

Also, regarding the title, is there evidence of increased genomic instability in these yeasts?

As described above, by "genome instability" we mean any change in sequence or copy number within the nuclear genome. This manuscript focuses on such events – insertions or the gain of new circular DNA in the nucleus. The overall number of insertions and circular DNA increases in *nuc1*Δ cells. Perhaps the reviewer was asking about subsequent genome instability caused by the presence of circular mtDNA in the nucleus or by integrated and highly repetitive mtDNA. These questions were not addressed here.

Page 10: What makes the mtDNA circular in the nucleus? This is also mentioned later, but its relevance to the study is unclear.

We think that one of the DNA repair pathways is responsible for circle formation, similar to many previous reports on circle formation in yeast and humans. We state:

"These nuclear mtDNA circles are likely formed by ligation of linear DNA ends by NHEJ or other DSB repair pathways as was previously demonstrated for other types of circles (e.g. [10])."

Page 12: "In nuc1D cells, long mtDNA is transferred to the nucleus, yet this DNA is not frequently inserted at HO breaks or anywhere else in the genome." Same comment as for the previous. How do you know it is not inserted but that your methods are not picking the long insertions up? You might be able to see these by analyzing the MATa locus of individual clones (ones not producing a short PCR product) using a Southern blot.

We know these long TRP1-mtDNA fragments are not inserted because the TRP+ marker is not stable. After streaking on nonselective media plates, the Trp1 marker is lost in more than 99% of cases. As stated above, only 1 in ~500 TRP1-mtDNA fragments was integrated into the genome.

"[…] we tested 400 to >1000 NHEJ products individually by PCR. "
This is a bit vague. Also, the PCR biases towards short products, so it can still be that only a minority of the insertions are detected.

Here, instead of pooling colonies for MiSeq sequencing, we performed colony PCR to check the individual NHEJ products. All tested colonies picked randomly generated PCR products, so there was no bias.

Page 15: "it prevents the far more frequent nuclear transfer of very long mtDNA that also occasionally integrates into the genome."
So long mtDNA transfer (to nucleus?) is more frequent, but integration is not (if it is occasional, figure 3 legend states "rarely integrates")? What is this based on? The discussion is somewhat confusing.

We found that only 1 out of 180 tested Trp^+ colonies carried stable, integrated TRP1 in the genome among mtDNA transferred during meiosis, and 0 out of ~300 tested in mitotic cells. To ensure that the single stable Trp+ colony represented integration, we separated the chromosomes of this strain using CHEF, probed the Southern blot with a TRP1-specific DNA probe, and found that TRP1 was integrated within one of the chromosomes (New panel c in **Supplementary Figure 6c**).

Page 25, middle section of the text: MATa -locus should be in capitals. There are also other small issues, such as using symbol font instead of proper ascii greek letters.

We thank the reviewer for carefully reading the manuscript. The mistakes have been corrected.

Reviewer #3 (Remarks to the Author):

This manuscript analyzes insertions that are introduced into HO-induced double-stranded breaks in budding yeast using analysis of reads generated by Illumina MiSeq sequencing. The authors demonstrate an increase in the ligation of insertions in stationary phase cells and reveal a role for the EndoG/Nuc1 nuclease in suppressing the introduction of Ty fragments but increasing the introduction of mtDNA fragments into these junctions. In contrast, Nuc1 prevents the introduction

of larger mtDNA fragments as measured by the formation of circular mtDNA molecules, suggesting that the small mtDNA fragments arise from incomplete mtDNA degradation.

This is a very interesting and straightforward manuscript providing evidence for the role of Nuc1 in preventing aberrant NHEJ in stationary phase cells. The authors do an exemplary job of performing important controls, such as the requirement of the products on the NHEJ DNA ligase 4, and appropriate filtering of the MiSeq reads. Many of their observations, such as the propensity of multi-insertions and the length biases of the insertions, are rather unexpected. I believe this manuscript will be of high interest and will be appropriate for publication when the following comment is addressed.

We thank reviewer for complimenting our work.

A key missing experiment in this manuscript is the analysis of a nuc1 spt3 double mutant (nuc1 to cause increased Ty insertion, spt3 to suppress Ty transcription) to show that the Ty inserts do arise from the cDNA and not due to Ty-specific fragments formed in the genomic DNA. The current draft assumes that the Ty DNA inserts arise from reverse transcription and cDNA formation in the cytoplasm, but this assumption is never tested.

This is an excellent comment. As suggested by the reviewer, we constructed both single *spt3*Δ and *spt3*Δ *nuc1*Δ double mutants. As expected, both *spt3*Δ and *spt3*Δ *nuc1*Δ had a dramatic decrease in insertions from Ty cDNA, while *spt3*Δ showed no major change of NUMTs formation (new panel b in **Fig. 5**). We also noted that templated insertions from the nuclear genome are decreased in *spt3*Δ cells, which can't be simply explained based on known better survival of *spt3*Δ cells because *sch9*Δ cells that survive much longer as well do not show a decreased level of insertions (**Supp Fig 3f**). We found that *spt3*Δ cells had many templated insertions from the nuclear genome at day 33 in the stationary phase suggesting that genome instability is delayed in these cells. However, we decided not to include day 33 result as this would require more work and is outside the scope of this work, which focuses on yeast EndoG.

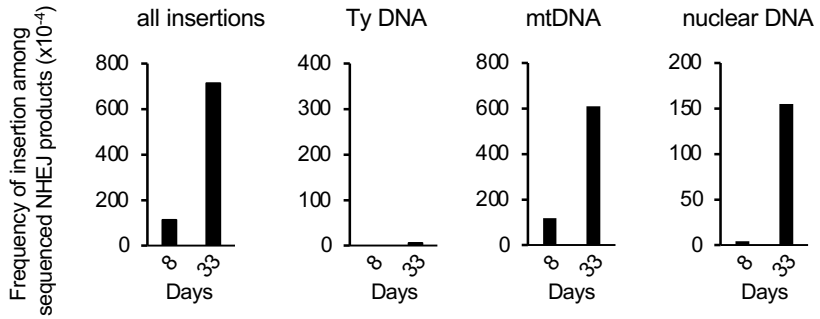

Figure for reviewer. Break-Ins analysis of *spt3*Δ cells in 33 days of chronological aging.

Reviewer #3 (Remarks on code availability):

The code is a combination of shell scripts and perl code to link together the output of other programs (e.g. blast/bwa/pear/bedtools).

The iDSBins and iDSBindel "packages" are clearly more mature (last update was older) and seem more polished. The comments seem quite legible and I can follow along reasonably well (though the SCAR program names in the iDSBins markdown seem to refer to some older code that no longer exists).

The LargeInsertionFeature "package" isn't as polished (a lot of hard-coded paths that you would have to change if you were running it on your machine and there's a shell script variable that I'm not sure where it gets defined), but certainly better than lots of code generated by researchers that I've seen.

We thank the reviewer for the suggestion. We have removed the SCAR name in the iDSBins markdown. The LargeInsertionFeature package has been thoroughly polished. The reviewer can find it here: LargeInsertionFeature : https://github.com/gucascau/LargeInsertionFeature.git

They do provide example data to run their programs on, which is a big plus.

I haven't tried to run their pipelines, though. And subtle bugs would be really hard to identify. Given the fact that the authors can identify lots of these events, my guess is that bugs in the pipeline (if any) would be subtle. The authors would have caught the really bad problems.

At a higher level of review, I think the overall strategy of how they are attempting to identify insertions in the induced DSB seems reasonable, and I have no concerns with their software.

1 Surre, J. *et al.* Strong increase in the autofluorescence of cells signals struggle for survival. *Sci Rep* **8**, 12088, doi:10.1038/s41598-018-30623-2 (2018).

2 Buttner, S. *et al.* Endonuclease G regulates budding yeast life and death. *Mol Cell* **25**, 233-246, doi:10.1016/j.molcel.2006.12.021 (2007).

3 Dakshinamurthy, A., Nyswaner, K. M., Farabaugh, P. J. & Garfinkel, D. J. BUD22 affects Ty1 retrotransposition and ribosome biogenesis in Saccharomyces cerevisiae. *Genetics* **185**, 1193-1205, doi:10.1534/genetics.110.119115 (2010).

4 Suresh, S. *et al.* Ribosomal protein and biogenesis factors affect multiple steps during movement of the Saccharomyces cerevisiae Ty1 retrotransposon. *Mob DNA* **6**, 22, doi:10.1186/s13100-015-0053-5 (2015).

5 Moore, J. K. & Haber, J. E. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. *Mol Cell Biol* **16**, 2164-2173 (1996).

6 Emerson, C. H. *et al.* Ku DNA End-Binding Activity Promotes Repair Fidelity and Influences End-Processing During Nonhomologous End-Joining in Saccharomyces cerevisiae. *Genetics* **209**, 115-128, doi:10.1534/genetics.117.300672 (2018).

7 Liang, Z., Sunder, S., Nallasivam, S. & Wilson, T. E. Overhang polarity of chromosomal double-strand breaks impacts kinetics and fidelity of yeast non-homologous end joining. *Nucleic Acids Res* **44**, 2769-2781, doi:10.1093/nar/gkw013 (2016).

8 Tseng, S. F., Gabriel, A. & Teng, S. C. Proofreading activity of DNA polymerase Pol2 mediates 3'-end processing during nonhomologous end joining in yeast. *PLoS Genet* **4**, e1000060, doi:10.1371/journal.pgen.1000060 (2008).

9 Wilson, T. E. & Lieber, M. R. Efficient processing of DNA ends during yeast nonhomologous end joining. Evidence for a DNA polymerase beta (Pol4)-dependent pathway. *J Biol Chem* **274**, 23599-23609 (1999).

10 Cortes-Ciriano, I. *et al.* Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat Genet* **52**, 331-341, doi:10.1038/s41588-019-0576-7 (2020).

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

In this revised version, the authors have carefully responded to my different comments, either by providing additional data, or by modifying the text. The manuscript is now suitable for publication.

Reviewer #2 (Remarks to the Author):

The authors have addressed my previous concerns. The YME1 results are intriguing, but I agree that expanding the work based on this observation is beyond the current scope.

I have only a couple of small suggestions.

1) The new highlighted sentence in the summary needs revision. As a suggestion: “In nondividing stationary phase cells, error-prone Pol4-mediated nonhomologous end-joining increases, leading to common insertions of nuclear mtDNA (NUMTs) and retrotransposon cDNA.“
2) Remove spaces before the reference numbers.
3) I note that some greek letters, such as delta in the nuc1Δ are still in symbol font
4) There are still some grammatical errors, missing commas etc. Please see e.g. the last sentence of the discussion as an example.

Reviewer #3 (Remarks to the Author):

The authors have done an excellent job of addressing all of my comments. I believe that the updated manuscript is appropriate for publication.

Reviewer #3 (Remarks on code availability):

The code was always available. The code quality for the updated module is definitely improved.

Responses to reviewers

We thank all reviewers for their time to re-review our manuscript. Few remaining questions from reviewer #2 are addressed below.

Reviewer #1 (Remarks to the Author):

In this revised version, the authors have carefully responded to my different comments, either by providing additional data, or by modifying the text. The manuscript is now suitable for publication.

Reviewer #2 (Remarks to the Author):

The authors have addressed my previous concerns. The YME1 results are intriguing, but I agree that expanding the work based on this observation is beyond the current scope.

I have only a couple of small suggestions.

1) The new highlighted sentence in the summary needs revision. As a suggestion: “In nondividing stationary phase cells, error-prone Pol4-mediated nonhomologous end-joining increases, leading to common insertions of nuclear mtDNA (NUMTs) and retrotransposon cDNA.“

We changed this phrase to:

In non-dividing stationary phase cells, Pol4-mediated non-homologous end-joining increases, resulting in frequent insertions of 1-3 nucleotides, and insertions of mtDNA (NUMTs) or retrotransposon cDNA.

2) Remove spaces before the reference numbers.

Spaces were removed.

3) I note that some greek letters, such as delta in the nuc1Δ are still in symbol font

The sign “delta” written in “Symbol” font was replaced with Times New Roman font.

4) There are still some grammatical errors, missing commas etc. Please see e.g. the last sentence of the discussion as an example.

We corrected the last sentence of the discussion. Whole manuscript was edited.

Reviewer #3 (Remarks to the Author):

The authors have done an excellent job of addressing all of my comments. I believe that the

updated manuscript is appropriate for publication.

Reviewer #3 (Remarks on code availability):

The code was always available. The code quality for the updated module is definitely improved.

4) There are still some grammatical errors, missing commas etc. Please see e.g. the last sentence of the discussion as an example.