# Round #1

## Recommender (Anna-Sophie Fiston-Lavier)

**Paper needs some clarification**

In this article, Teoli et al analyzed the variation in the expression level of transposable element (TE) insertions in RNAseq datasets from individuals with various numbers of sex chromosomes in order to test the toxic effect of Y on human lifespan, a fascinating but still controversial subject.

Specifically, they tested two predictions. These predictions suggest that the genomes of an older man should harbor more TEs compared to those of an older woman, due to the toxic effect of the Y chromosome and less efficient epigenetic regulation with age. In addition, they also tested whether TE expression varies with the number of Y chromosome using RNA resequencing data from 25 blood samples from individuals with different karyotypic compositions (46,XX, 46,XY, 47,XXY and 47,XYY).

**However, the two reviewers noted several points which it seems important to address. One important point concerns the methodology of part 1, in which you present the analysis of a selected subset of GTEx data. Although reviewer 1 appreciates your honesty about your hindsight on the results of this part, the justification of the choice of data and filters used is missing. Reviewer 2 also suggested to look at the TEs that are expected to be more likely involved in somatic mutations.**

As I believe this is a relevant topic that will be of interest to the community, I encourage the authors to respond carefully to the various points highlighted by the reviewers, and to resubmit their article.

### Major modifications for recommender and reviewers

First, the authors gratefully thank Anne for having accepted to be the recommender of their work and the reviewers for their pertinent and helpful comments.

A thorough re-analysis of the RNAseq data (in the gonosome aneuploidy dataset generated in this study) enabled us to identify an outlier among the 47,XXY individuals: the Txxy21a individual has non homogenous XXY karyotype (80% of mitoses are XX and 20% XXY). Therefore, the Txxy21a individual was excluded. This does not alter our conclusions, but has led us to update some p-values and some figures, tables and supplementary data listed below:

- **Data S1**, **Data S2**, **Data S3**: update.
- **Figure 1-A, Fig S2, Fig S9**: update of figures (one less individual to show on the figures) and p-values, when applicable.
- **Fig S3, Fig S4**: update because the lists of differentially expressed genes have changed a little.
- **Figure 2, Fig S11, Fig S12**: update because the lists of differentially expressed transposable elements (TE) have changed a little.
- **Table S1**: removal of Txxy21a subject.
- **Table S2**: values update.
- **Tables S3, S4, S5**: updates to lists of differentially expressed genes or TEs that were slightly different.
- **Tables S9, S10, S11, S12**: modification of the overlaps (indicated by * next to the TE family names in the table) between the GTEx datasets and the gonosome aneuploidy dataset

because the lists of TE differentially expressed obtained using the latter dataset were slightly different.

In response to reviewer comments, we have made significant changes to the manuscript. First, we have restructured the manuscript by separating the results and discussion sections for clarity (line number and figure/table/data number have changed compared to the previous manuscript). Additionally, we have expanded the introduction and discussion sections, adding new references and improved results and materials and methods sections. Furthermore, certain figures and tables previously included in the supplementary material have been moved to the main text to better illustrate findings (**Figure 4**) or are new (**Figure 3, Fig S5 to S8, Fig S10, Fig S13, Fig S14, Fig S19, Fig S21, Table S13, Data S4, Data S5**). We have also incorporated new analyses based on reviewers' feedback which are now detailed in the Materials and Methods section. Notably, the GTEx analysis has been rerun with a larger sample size, resulting in updates to the Materials and Methods and Results sections, figures and tables. Colleagues who have contributed to these revisions have been added to the list of authors (Miriam Merenciano, Daniel Siqueira de Oliveira, Alessandro Brandulas-Cammarata).

## Reviews

# Reviewer 1

In this manuscript, Teoli and colleagues analyze expression data from human individuals to test the hypothesis of the Toxic Y, which could be involved in explaining the shorter lifespan of males compared to females in our species and many others. They used two independent datasets: 1. the blood RNA-seq data from the Genotype-Tissue Expression (GTEx) project and 2. a blood RNA-seq dataset from several individuals with different karyotypes generated by the authors. Analyses of dataset 1 provided some support for the prediction that TEs reactivate in old individuals, particularly in males. However, the structure of the data prevented the author from drawing firm conclusions and the results are quite noisy. The analyses of the second dataset provide better support for the second prediction, which suggests that TE expression is associated with the number of Y chromosomes.

Overall, I think the paper is concise, clear and well written. The analyses seem to have been carried out carefully and I really appreciate that the methods used are described in detail. I also appreciate that the authors do not overstate their findings and are honest about the limitations of their study, particularly the small sample size. The results are very interesting and should be of interest to a wide range of biologists.

**1) My only major concern is with the analyses of dataset 1. In short, the authors greatly reduce the size of the genotype-tissue expression dataset by applying very strict filtering. This strong filtering is not properly justified and results in a very small dataset to analyse. As RNA-seq data are generally very noisy, this small dataset prevents the authors from drawing conclusions. I'm not sure that this strong filtering is necessary, and it should be better justified. For example, in "To reduce data heterogeneity and get closer to the individuals we sampled in parallel to constitute our dataset", I do not understand why you need a dataset "close" to the individuals you sampled: you do not analyse the two datasets together, but separately. Reducing the heterogeneity of the data only makes sense if that heterogeneity is biasing your results, which is not necessarily the case. Removing half of the dataset by eliminating non-white or Latino/Hispanic individuals seems to me to be too strong a filter and is not justified. It greatly reduces your statistical power. The same rationale applies to the other filters, such as excluding people with viruses, cancer or dementia. I understand how these disorders might affect gene or TE expression, but I'm not sure how this should affect the putative toxic Y effect. I think you should try to re-analyse this dataset with very minimal or no filtering, possibly including these inter-sample differences (ethnicity, disorders, etc.) as covariates in your statistical model when comparing TE expression between sexes and age groups. Alternatively, you should properly justify why these filters are necessary (e.g. strong interactions between ethnicity and age class in TE expression).**

Authors' reply: the authors sincerely thank the reviewer for the invaluable feedback. Thanks to a collaboration with Marc Robinson-Rechavi lab, we have repeated the TE expression analysis with an extended GTEx dataset of 318 libraries whose features are detailed here :

https://www.bgee.org/search/raw-data?pageType=raw_data_annots&data_type=RNA_SEQ&data=6306a7ecbbd28a9f97bf00883598e03eee560c64

Therefore, we have updated Materials and Methods, Results, Discussion concerning GTEx data. New figures and tables were also added (**Fig S21**, **Table S13**). Results obtained from these new data do not change our conclusions previously obtained using a filtered GTEx dataset.

**2) Other comments:**

Authors' reply: the authors thank the reviewer for his/her corrections. See details below.

**-Abstract:**
**"Lifespan differences between sexes is a puzzling question". For me, a difference is not a question.**
**"a toxic genomic impact in this trait" What is the trait ? Male lifespan ?**

Authors' reply: Thank you for your suggestions. The abstract has been revised.

**-L59: Should be "Z-linked", no ?**

Authors' reply: Correct. Please find the correction line **67**.

**-L65-68: The use of "is caused by" and 'is still controversial" seems a bit contradictory to me.**

Authors' reply: We have removed these sentences and we have added more information on the previous studies testing the Y toxic hypothesis in *D. melanogaster*. Please see lines **92 to 100**).

**-L69-80: I think you should quickly define 46,XX, 46,XY, 47XXY, 47,XYY. It may not be clear to all readers what this means.**

Authors' reply: We have added more information to clarify the different karyotypes used in this work in the introduction section. Please see lines **103-104**: "Furthermore, men with 47,XYY and 47,XXY abnormal karyotypes (with an extra Y or X chromosome, respectively)"

And lines **112 to 114**: "46,XX females (normal female karyotype), 46,XY males (normal male karyotype), as well as males with abnormal karyotypes, such as 47,XXY and 47,XYY."

**-L104: "female-biased protein-coding genes". A bit hard to read and understand at the beginning. Maybe it could be reformulated as, e.g. "genes overexpressed in females"?**

Authors' reply: We have replaced this sentence. Please find the update lines **147 to 149**: "As expected, we found that while most of the upregulated genes (13/19, 68.42%) were Y-linked genes, most of the downregulated ones (14/18, 77.78%) were X-linked genes."

**-L104: "most" → How many ?**

Authors' reply: We have removed this sentence. Furthermore, we have studied the overlap between differentially upregulated genes in XXY compared to XY found in our study and those found in Zhang et al. 2020 study, and between differentially upregulated genes in XXY compared to XY found in our study and the candidate XCI escapees reported in Wainer-Katsir et al. 2019 study (see **Supplementary text**).

**-L106: "It is indeed well-known that the extra-X chromosomes in 47,XXY or other karyotypes (e.g triple X) are inactivated" Not clear. The previous sentences refer to XX and XY samples, not XXY. The connection between this sentence and the previous ones should be explained. Does the sentence "We also found that most (13/20, 65%) male-biased protein-coding genes are Y-linked genes, and that most female-biased protein-coding genes are known X chromosome inactivation escapees" discuss only results comparing XX and XY or also other genotypes ?**

Authors' reply: We have removed this paragraph from the previous manuscript. A new paragraph concerning XCI escapees was reworded and added in the **Supplementary text**.

**Fig 1 legend: I think there is an error. I think it should be : " […] and less than 0.01 (**) for 47,XYY samples compared to ##46,XX samples## (P = 0.0095) and near 0.05 for 47,XYY samples compared to 46,XY samples (P = 0.067) using Wilcoxon test"**

Authors' reply: We thank the reviewer for pointing this out. This sentence no longer appears in the **figure 1** legend but in the paragraph lines **170 to 190** in the results section. Note that p-values are slightly different from the previous manuscript because we have excluded the Txxy21a individual as explained above.

**Fig S9: What do represent the blue and grey arrows ?**

Authors' reply: We have added this precision in figure legend when necessary (**Fig S3**).

**Fig S11: What is the x axis ?**

Authors' reply: This figure concerning gene ontology and pathways enrichment was replaced by new figures generated using the R package clusterProfiler (**Fig S5 to S8**). Nevertheless, concerning the former figure, we found this answer in the scientific paper describing g:profiler2 (R package we used for the Gene ontology): "The locations on the x-axis are always fixed and ordered in a way that the terms from the same GO subtree are located closer to each other. This helps to highlight different enriched GO sub-branches as they form peaks in the Manhattan plot..." from Kolberg et al. 2020 (PMID: 33564394).

**L120-175: I think it is important to write and discuss that an important observation is that XXY and XYY individuals have more TE expression than XX and XY individuals, which suggests that the number of sex chromosomes is a major determinant of TE expression (more than the number of Y chromosomes). The difference in TE expression between XXY and XYY is small and not significant. So the claim that you show an over-expression of TE is related to the number of Y chromosomes in the karyotype seems a bit too strong to me.**

Authors' reply:

Schematically and if we considered only p-values, the "add" of one Y chromosome increased significantly or near significantly the TE expression (47,XXY compared to 46,XX: p-value = 0.029, 47,XYY compared to 46,XY: p-value = 0.067) whereas the "add" of one X chromosome did not (47,XXY compared to 46,XY: p-value = 0.345) (see lines **180 to 185** and **Figure 1-A**).

Then, if we considered the trend shown on the **Figure 1-A**, "Moreover, we saw that the addition of a sex chromosome, either X or Y, tended to increase the overall amounts of TE transcripts probably due to an increase in genomic material and thus in TE load. However, the addition of a Y chromosome seemed to increase the expression of TEs even more than the addition of a X chromosome (Fig. 1A). These results suggested that the presence of the Y chromosome might be associated with an increase in TE transcripts, as postulated in the toxic Y hypothesis, and could contribute to a global deregulation of TEs." (this paragraph was added in the Results section, lines **185 to 190**).

We also added these sentences in the Discussion section:

"Additionally, it seemed that the presence of an additional sex chromosome, either X or Y, tended to result in an elevated overall abundance of TE transcripts, likely due to increased genomic material and, consequently, TE load. Notably, consistent with the toxic Y hypothesis, the addition of a Y chromosome appeared to amplify TE expression even more than the addition of an X chromosome." (lines **379 to 382**).

Furthermore, p-values we found not significant herein for some pairwise karyotype comparisons may become significant in a larger dataset. Then we suggest that conducting a study using a larger sample size would be of great interest (lines **431-432**).

**L155-156: ”Strong change” & “Clear trend”. Considering that these changes and trends are not statistically significant, I think you should tone down this sentence a bit.**

Authors' reply: We agree with the reviewer. The discussion section was reworded (notably, see lines **375 to 384**).

# Reviewer 2

Review of Teoli et al.

In this paper the authors reanalyze human blood RNA-seq data from the GTEx project and produce a new expression dataset for individuals carrying 47, XXY and 47, XYY karyotypes. They use these data to test for a toxic Y effect in humans, with the expectation that older men and individuals carrying more Y copies should also display increased Y-linked TE activity.

I identified several potential issues that I detail more below. **In particular, I do not think that the present results strongly support a scenario where the reactivation of Y-linked TEs may lead to increased somatic transposition, which seems to be at the core of the current version of this work. A possible line of explanation closer to the observations might be an effect of the Y chromosome on the integrity of genome-wide heterochromatin (Brown et al., 2020).**

Authors' reply: We thank the reviewer for the comment. We agree that our data does not support a scenario where the activation of Y-linked TEs may lead to increased somatic transposition. The focus of this work was to associate the presence of the Y chromosome to an increased TE expression. This increased expression can thus be translated to an increased transposition rate, but we have not checked that in this manuscript (we refer only to the TE expression and not to TE reinsertion or somatic mutation rate along the article). We have provided some clarifications lines **439-440** ("However, the generation of new somatic mutations with age and the impact of these new TE insertions in the aging process is still to be fully determined in humans (Pabis et al. 2024).").

Indeed, several works supporting the toxic Y effect described age-dependent TE expression in males, but not an increased rate in TE transposition (Schneider et al., 2023, Tsai et al., 2024). Therefore, we discussed other mechanisms than transposition that may explain the deleterious effect of TEs (see lines **434 to 455**).

**1) Lack of references to the extant literature on transposable elements activity in humans**

Authors' reply: We have added literature on transposable elements activity in humans in the introduction section (please see lines **76 to 80**). We also made a new figure focused on known active TE groups in humans according to Kojima et al. 2018 (**Fig S10**) and described the results lines **200 to 209**.

**The authors hypothesize that the Y chromosome may host transposable elements that are reactivated when host's regulation weakens during aging, generating somatic mutations as new TE copies insert elsewhere in the genome. However, the vast majority of elements they identify as being differentially expressed in their data is made of endogenous retroviruses (see table 4 in (Kojima, 2018)). This is inconsistent with the fact that in humans, LINEs and Alu are the only elements that seem able to effectively transpose. From my knowledge, the most recently active endovirus in humans may be HERV-K (Subramanian et al., 2011), which still contains intact ORFs, but I am not aware of any evidence for any active transposition (Maksakova et al., 2008). Most examples of transcripts derived from LTR TEs in humans do not correspond to active transposition, but to domestication by the host. It is also surprising and concerning that non-LTR RTs are only rarely detected given that they are more likely to actually transpose.**

Authors' reply: We agree with the reviewer's comment. As mentioned before, the focus of the manuscript was to associate the presence of the Y chromosome to a general increase TE

expression. Whether this can be associated with an increased transposition rate and, thus with an increased somatic mutation rate, is out of the scope of the paper. We added some clarifications throughout the manuscript (introduction and discussion sections). However, we agree that some of the differentially expressed TE subfamilies are not necessarily active TE subfamilies in humans. Nevertheless, we think that these findings did not exclude the Y toxic hypothesis since TE insertions can modify in many different ways the expression and structure of genes not only by their transposition (Casacuberta and Gonzalez, 2013, Schneider et al., 2023). We have detailed that in the discussion section. Please see lines **441 to 455**.

Besides that, we made new graphs focused on the expression of TE groups cited by Kojima et al. (2018) which are likely actively transposing in humans: L1 (and L1HS), AluY, AluS, SVA, and HERVK elements. These new graphs show TE expression levels according to karyotype for each of these TE groups (**Fig S10**). The results are described lines **200 to 209**.

**L149-150: The authors filter out exonic insertions and insertions found in lncRNA. What about intronic TEs? Disruptions in the splicing process (including intron retention) seem to increase with age (Bhadra et al., 2020). There may also be differences in splicing between males and females caused by the Y chromosome, at least in Drosophila (Wang et al., 2018). This might also contribute to an excess of TE-derived sequences.**

Authors' reply: We thank the reviewer for raising this point. We are not entirely convinced that the toxic effect of the Y chromosome is exclusively linked to autonomous TE expression. Given that the analysis with the TEspex software, which filtered out exonic insertions and insertions found in lncRNA, significantly diminished the statistical power by reducing the number of reads aligned on TE sequences, and considering the substantial bioinformatic resources required to suppress all intronic TEs, we have opted to remove it. We agree with the reviewer that then we might have TE transcripts coming from intron retention or pervasive intragenic transcription together with transcripts coming from autonomous TE activation. However, it has been shown that TE expression and not only TE transposition might have an important effect in aging (Pabis et al. 2024, Schneider et al. 2023, Tsai et al. 2024).

**2) Possible methodological issues**

**• Why do the authors only focus on blood, at least for the GTEx dataset? I understand the need to compare the results from GTEx with the data the authors obtained, but why not take advantage of the whole GTEx experiment? This is particularly surprising given that some other studies have derived valuable insights when investigating transcripts derived from endogenous retroviruses from this same dataset (She et al., 2022).**

Authors' reply: We agree with the reviewers' advice. We focused only on blood for the GTEx datasets to obtain a similar dataset as the gonosome aneuploidy dataset we have generated to check the association of the Y chromosome with TE expression in the different karyotypes. Blood samples are among the least invasive samples that we can collect from living humans. However, we have now extended the GTEx dataset including all blood RNAseq data available from Marc Robinson-Rechavi's team (318 libraries). The features of these 318 libraries are detailed here:

https://www.bgee.org/search/raw-data?pageType=raw_data_annots&data_type=RNA_SEQ&data=6306a7ecbbd28a9f97bf00883598e03eee560c64

We have updated Materials and Methods, Results, Discussion, Figures and Tables concerning GTEx data to include this larger GTEx dataset.

Otherwise, we agree that it would be of great interest to study the relation between TE expression and age or sex in all body tissues available in the GTEx dataset. However, another team has already done this work (Bogu et al. 2019, BioRxiv preprint) using a different methodological approach than ours. Moreover, this would require a substantial amount of work, which is not compatible with the end of our student's PhD. We will consider your request for future work.

**• The authors seem to assume that the presence of TE transcripts is associated with TE insertion which generate somatic mutations, but transcription of TE-derived sequences is not enough to prove that an element is still active and transposing.**

Authors' reply: We agree with the reviewer's comment. As discussed in a previous point, we have clarified in the introduction and discussion sections that our work does not assume that the presence of TE transcripts is associated with an increase of somatic mutations.

**• The authors never align short reads to the reference genome, but instead use a reference transcriptome (fo ex. Using the kallipso method) or consensus sequences of transposable elements. It is therefore difficult to test directly for a toxic Y effect since the position information of TEs is lost, and alternate transcripts cannot be exhaustively identified. It may be worth investigating whether methods that realign reads on the reference genome (using, e.g., STAR) can also identify some interesting TEs (Schwarz et al., 2022). Reference (She et al., 2022) may be an interesting starting point given its similarity with the present work.**

Authors' reply: Undoubtedly, a limitation of our study is the lack of access to the genomes of individuals under investigation, with only their transcriptome available. This is the reason why we have worked with a reference transcriptome, which is more conservative. With respect to the quantification of TE expression, we have opted to use TEcount software from TEtools. With this tool we map all the reads to TE sequence insertions, that are then regrouped by TE subfamily. We do not need to use consensus of TE sequences. While we acknowledge availability of other bioinformatic tools, we opted for kallisto due to its rapid execution and TEtools because it does allow us to use the maximum of reads that map to each TE insertion. Following acceptance of this manuscript, we intend to make the data available from the European Genome-phenome Archive (EGA), enabling other research teams to reanalyze it using alternative tools.

**• The authors do not attempt to identify families that may be in excess (or absent) on the Y chromosome. This could be a way to start testing more directly for a Y effect.**

Authors' reply: We thank the reviewer's suggestion. It would be of great interest to know the particular differentially expressed TE copies. In order to achieve that, access to individuals' genomes would be needed. Moreover, we do not expect TE copies inserted in the Y chromosome to be more expressed than other copies inserted in other chromosomes. The fact that the Y chromosome is rich in TE insertions increases the probability to find more active TEs in males compared to females. However, there is no evidence indicating a specific increase in the expression of insertions on the Y chromosome compared to other chromosomes. To answer to the reviewer's suggestion, we made three new analyzes although the conclusions were limited by the use of a reference transcriptome rather than individual's genome:

- We first assessed whether the observed proportion of TE copies for a particular differentially expressed TE subfamily inserted in the Y chromosome matched the expected proportion on the Y chromosome if TE copies of this TE subfamily were equally distributed across all chromosomes. Please see lines **690 to 703** in the Materials and Methods part, **Data S4** and **Fig S14**, and legend of some other figures (**Figures 2 and 4, Fig S12, Fig S17, Fig S20**).
- Second, we show a positive correlation between the mean of normalized counts for each TE subfamily across all individuals and the number of copies located in the Y chromosome among all TE copies for a specific subfamily. See **Fig S13**.
- Finally, we showed that, regardless of whether we examined Y-enriched TE subfamilies, Y-depleted TE subfamilies, TE subfamilies that they are neither Y-enriched no Y-depleted, or solely TE subfamilies with no copies in the Y chromosome, we consistently observed the highest TE expression in 47,XYY individuals and the lowest in 46,XX individuals. See **Fig S14**.

**• I would recommend the authors focus only on TEs for which complete copies can be identified. Fossils and old TEs are unlikely to transpose and insert. I would also suggest focusing on LINES and Alu first.**

Authors' reply: We have generated new figures showing TE expression in the different karyotypes grouping the insertions in TE orders. We found the same trend as before for LTR retrotransposons and for non-LTR retrotransposons (SINE, LINE, SVA), but not for DNA elements (see **Figure 1**). We have also generated a figure focused on the Alu group and we have found the same pattern as for SINEs. Furthermore, as indicated above, we made new graphs focused on TE groups known to be able to effectively transpose according to Kojima et al. 2018. These new graphs display TE expression according to karyotype for each of these TE groups individually (see **Fig S10**).

**• A quick examination of some of the elements shown on Figure 1A (such as LTR22B2 or LTR19A, see below) shows that they are likely solo LTR sequences (see below, taken from Repbase), which are unlikely to transpose and generate somatic mutations. The fact that these sequences are highly repetitive makes me wonder whether the methods used can accurately assign reads to consensi.**

Authors' reply: The reviewer is correct and indeed, when sequences are highly repetitive there may be some difficulties to map the reads against TE consensus. Say this, we believe this should not significantly affect our analysis. Our research does not focus on studying TE new insertion or somatic mutations, but instead on the global expression of the TE subfamilies. For that we have used TEcount, which used all the insertion sequences belonging to a subfamily to make the mapping, and then the results are analyzed by subfamily. We believe this minimize the loss of reads that could escape. Moreover, we acknowledge that tools designed for studying TEs, despite being tailored for repeat sequences, may have inherent limitations.

############## Consensus sequences of LTR-RTs mentioned above ###############

>LTR22B2 ERV2 Homo sapiens
tgttggggttcaatcaggctggtgggaaaaatattaaagatagttatagtaatagtcaaaaactctcttg
gaaggccgtgagagtttgcatagcttcggtaattgctgtggctgaaggcagccagggtctctttgcagga
gccagaaagattagggtgcaagtacaaaggaatgtgggaagtttatcttactaacctgtttacttatatg
ggcttaagactaacctttgtcctaccgcgggtactttactgcctcctactgggagcgggmgggggtcggc
agaagtttattacccgcaaatggtgtttgctttaggcctcggaacctggcctttaatctttaccctctag
tggtgtttactcacaacttttgttaattagtcttactgaataaatgcgagtctcactagctgatcagggc

cgagtcgcaactgtttacagaactcagcttggagcctgtaagcggctcggaccctcagctggactggcag
agcagaatatctgtgtgtcagtgtacgtttattcatccgtcgccgaatcaggggtctgcaaggaacagac
cccccgcagctagtgcccccgcgaaaggagcgctgcctca

>LTR19A ERV3 Homo sapiens
tgacagagcaggagcatcgccatcttggacaagcactgccattttaaagttccccttgatcaaaaaccgc
ctaaatccaacccaaagggcatcagcctaatggctaakgtcagcatgaccataaaccacaaatgacatct
ccgaccagaaacattccaaccctaagataaaccccctcccyraccagagacatgccagccccgagataacc
tcccctccggccagagagatgtcagccccaasataacctcccccttcaaccagagacattccaaccccaca
ataaacttctcccccacacagaaacattccaagcctgtgataaagctctctcaccctaaaaccccttaaat
actcttagtctgtaagagagagtgctcctgactgaaatcggccagaagcccctctcaggtttattctcca
aaataaacctgtctttgactgttgagccgcttttcrtgtttctttcctctttctttaactcttaca

##########################################################################

**• It may be good to check whether genes found differentially expressed harbour TE sequences from the families that are identified as also being differentially expressed.**

Authors' reply: We thank the reviewer for raising this point. We made an overlap between differentially expressed genes and differentially expressed TE subfamilies. Please see lines **739 to 749** in the Materials and Methods section and lines **278 to 296** in the Results section. We have also included in the manuscript an enrichment analysis of TEs in upstream regions of differentially expressed genes (lines **717 to 737** in the Materials and Methods section and lines **265 to 276** in the Results section).

**3) Miscellaneous issues with the presentation**

**-The authors do not present nor discuss much the results obtained from GTEx, with all figures and tables given in the Supplementary Material. This biases the narrative by putting the emphasis on the data the authors collected. Most figures are in the Supplementary Material.**

Authors' reply: We thank the reviewer's advice. We have now added more information on the analysis of the GTEx data in the results section (lines **298 to 332**), including 2 figures (**Figures 3 and 4**).

**-Figure 1B is not very convincing, especially given the rather small sample sizes. It would be interesting to also display the age of individuals for each karyotype instead of showing that information only in Table S7.**

Authors' reply: We thank the reviewer for pointing this out. We have added the age of each individual to **Figure 1**. We also did that for some of the new graphs generated (when applicable) (**Fig S10, Fig S14**).

Authors gratefully thanks the reviewer for all these pertinent references listed below. Some of them were now included in the references of the manuscript.

---

References

Bhadra, M., Howell, P., Dutta, S., Heintz, C., & Mair, W. B. (2020). Alternative splicing in aging and longevity. Human Genetics, 139(3), 357–369. https://doi.org/10.1007/s00439-019-02094-6

Brown, E. J., Nguyen, A. H., & Bachtrog, D. (2020). The Drosophila Y Chromosome Affects Heterochromatin Integrity Genome-Wide. Molecular Biology and Evolution, 37(10), 2808–2824. https://doi.org/10.1093/molbev/msaa082

Kojima, K. K. (2018). Human transposable elements in Repbase: Genomic footprints from fish to humans. Mobile DNA, 9(1), 2. https://doi.org/10.1186/s13100-017-0107-y

Maksakova, I. A., Mager, D. L., & Reiss, D. (2008). Endogenous retroviruses: Keeping active endogenous retrovirallike elements in check: the epigenetic perspective. Cellular and Molecular Life Sciences, 65(21), 3329–3347. https://doi.org/10.1007/s00018-008-8494-3

Schwarz, R., Koch, P., Wilbrandt, J., & Hoffmann, S. (2022). Locus-specific expression analysis of transposable elements. Briefings in Bioinformatics, 23(1), bbab417. https://doi.org/10.1093/bib/bbab417

She, J., Du, M., Xu, Z., Jin, Y., Li, Y., Zhang, D., Tao, C., Chen, J., Wang, J., & Yang, E. (2022). The landscape of hervRNAs transcribed from human endogenous retroviruses across human body sites. Genome Biology, 23(1), 231. https://doi.org/10.1186/s13059-022-02804-w

Subramanian, R. P., Wildschutte, J. H., Russo, C., & Coffin, J. M. (2011). Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. Retrovirology, 8(1), 90. https://doi.org/10.1186/1742-4690-8-90

Wang, M., Branco, A. T., & Lemos, B. (2018). The Y Chromosome Modulates Splicing and Sex-Biased Intron Retention Rates in Drosophila. Genetics, 208(3), 1057–1067. https://doi.org/10.1534/genetics.117.300637