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).

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

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.

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.

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).

- 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.

- 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.

- 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.

- 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.

- 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.

############## 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.

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. 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.

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 retroviral-like 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