In the proposed manuscript, Rebollo et al explore the use of the long-read Oxford Nanopore (ONT) sequencing technology to describe qualitatively and quantitatively the transcriptional landscape of transposable elements in the *Drosophila* gonads. The authors generate, sequence and analyze long-read cDNA libraries (one replicate from each tissue type), as well as compare the obtained results to previously published short-read datasets.

The manuscript is timely, as we can observe a growing interest in the use of ONT technology for transcriptome analysis, especially in the field of DNA repeats. As such, this work will certainly be useful for many researchers, even more so, that it includes clear explanations of all wet-lab and data analysis approaches. The manuscript exposes clearly the strengths, as well the limitations of the techniques used. The manuscript is also quite unique in its detailed comparison between short- and long-read transcriptomic datasets, which shows how these two technologies can complement each other on different levels.

The manuscript could benefit from the following improvements or clarifications:

Major comments:

Regarding "unique" and "unique best" mapping reads vs multi-mappers:

1- According to the text and the Table S1, the "unique best" mapping reads represent 91% and 99% of the sequenced libraries. This however, relates to total reads, only small percentage of which maps to TEs. Thus, effectively, this percentage is true for genes. What are the fractions of "unique" and "unique best" reads for TE-mapping reads only? Supposedly, these could be different than for single copy genes. How does this compare with short-read libraries? To which extend long-reads reduce or overcome the multimapping issue? This would be interesting to show clearly and it is also important for the downstream analysis.

If there is a significant fraction of multi-mapping reads among all TE-mapping reads, are these reads taken under account in transcript abundance quantifications? If present, these multi-mappers should contribute to the family-level transcript estimates (Fig 2), and should be taken under account when quantifying copy-specific expression (Fig 3 and 4). If, for example, for a given TE family, there are significantly more multi-mapping reads than unique best aligners, any conclusions on how all the copies contribute the total transcript levels would be impossible to make. The analysis of expressed vs non-expressed TE copies would still hold true, but only for genomic loci that present enough sequence variation to produce transcripts with unique best alignments.

The authors should clarify this by providing the statistics of multi-mapped reads for TEs, performing any additional analysis if necessary and adjusting their conclusions if required.

2 -The above point brings up the question of sequence variation between genomic copies of different expressed TE families, which, in the current version of the manuscript, is not much discussed. Expression of evolutionary younger TEs, with lower sequence divergence, would obviously be more difficult to quantify. This would be particularly relevant for the search of full-length TE transcripts (Fig 4), which would carry less informative sequence variation. The authors could include sequence variation of genomic copies (as they do for sequence variants of transcripts in Fig 5) in their analysis or, minimally, they should comment on the limitations that could be related to the potential lack of such

variation. Again, this issue will be less relevant if no (or very few) TE-specific multi-mappers are in fact found in the libraries.

Regarding mapping reads to features:

3– A substantial number of TEs is located in intronic sequences. Taking under account how the authors assign reads to features, would intronic TEs in expressed genes be taken under account or omitted? This is not clear. Theoretically, in such cases, both features (the gene and the intronic TE) could be fully covered. In other words, are TEs belonging to the "intron" category if Fig 1E found only (or primarily) in non-expressed genes?

Table S1 and line 215:
Please add read length statistics (median, N50) separately for both samples. Read length will influence some of the downstream analysis, thus it would be important to indicate it.

Additional minor comments:

Methods:
Please specify how much total RNA was used as input for the TeloPrime cDNA amplification.

Lines 238-240:
The use of percent ranges (e.g. 37-48%) is misleading, when in fact only two samples are analyzed. Replacing with the two obtained values only, would be more accurate.

Fig 1B and D:
Transcript coverage in Fig 1B should be plotted as a function of transcript length, similarly as done for the figure panel 1D.

Related to the above point (lines 229-230 and Fig 1D), the text of the results sections should not omit the fact that good correlation is achieved only for short transcripts. Although this detailed explanation is coming later in the text, it would be easier for the reader, if the point of underrepresentation of long transcripts was clarified up front.

Fig 1C:
Please correct, testis > testes

Line 246-247 and Fig 1F:
The authors to some extend contrast TE transcripts with gene transcripts by stating: "on the other hand, gene transcripts may reach 5kb". Although this is true based on the data, it should be taken under account that overall genes are much more highly expressed than TEs, increasing the chance of detection of underrepresented long transcripts. If the authors wish to make such comparison, they should include transcript abundance as a contributing factor.

Overall, due to lack of replicates and important difference on coverage, any conclusions regarding comparison between samples should be made very carefully. The authors do acknowledge this and mostly remain careful in their conclusions (e.x. line 293-294). However, throughout the paragraph (lines

279-294) the authors should avoid direct comparisons of read numbers between female and male libraries. Without any kind of normalization statements such as “but both families with higher transcripts in males” are not very meaningful. Also, regarding the observation that the global proportion of reads mapping to TEs is significantly higher in testes than in ovaries, it was not clear from the text, if this was also true for the previously published data (Larat et al 2017). If so, this would strengthen the obtained result, as it does in Fig 2C for the expression at the family level.

Finally, are TEs with male-specific transcripts (HETA, TAHRE) enriched on the Y chromosome?

Fig 3A:
Please unify: “copy number” = “# of genomic copies”

Line 360:
Please remove the abbreviation “v.r.t.”

Line 356 and the results section below:
In light of the demonstration that the technical approach taken is strongly underestimating very long transcripts (Fig 1), the section title should be toned down to “are rarely detected” rather than “rarely transcribed”. Also, the authors should remind the reader of this technical limitation here and tone down their conclusion as to whether the detected transcripts are fully reflecting the transcripts presents in the tissues investigated.

Fig 5:
Please enlarge fonts for TE family names