Reviewers' Comments:

Reviewer #1: The authors revisit the question of LINE-1 activity in human tissues by analyzing multiple existing datasets and some newly generated datasets on tumor tissues. The major purported advance is the development of a new computational method to assess more accurately LINE-1 activity, which the authors' claim improves their ability to detect true LINE-1 transcription compared to previous publications. Of note, Cell has published at least two papers in the past 5 years exploring this question -(Evrony GD et al Cell 2012 and Upton KR et al Cell 2015) with somewhat different conclusions about the degree of activity of retrotransposons in the human brain. The current paper purports that LINE-1 activity is minimal in the brain relative to other tissues, potentially provoking further debate on this issue.
The primary messages and conclusions of this paper are:
1) That LINE-1 is more active than previously appreciated in somatic tissues from adult humans
2) LINE activity is correlated with Insertions and Deletions in human cancers and thus may represent an important driver of malignancy
Claim 1 is demonstrated using a large dataset of published RNAseq experiments taken from cell lines and various human tissues. The novel bioinformatics method that the authors have developed allows them to filter out 'pervasive transcription' of different LINE subfamilies, which has not been previously considered as a confounding factor for defining LINE activity. By accounting for the abundance of different LINE subfamilies in the human genome relative to transcripts detected, it was possible to define a strong overall correlation, which could imply that much of the signal detected is pervasive transcription rather than active LINE-1 transcription. This is further demonstrated by analyzing nuclear vs cytoplasmic LINE transcript abundance and the expression of poly-adenylated transcripts relative to 5' signal in cell lines. Finally the authors use deviation from the expected genomic/transcriptomic correlations indicative of pervasive transcription as a means of detecting active transcription in different tissues.:

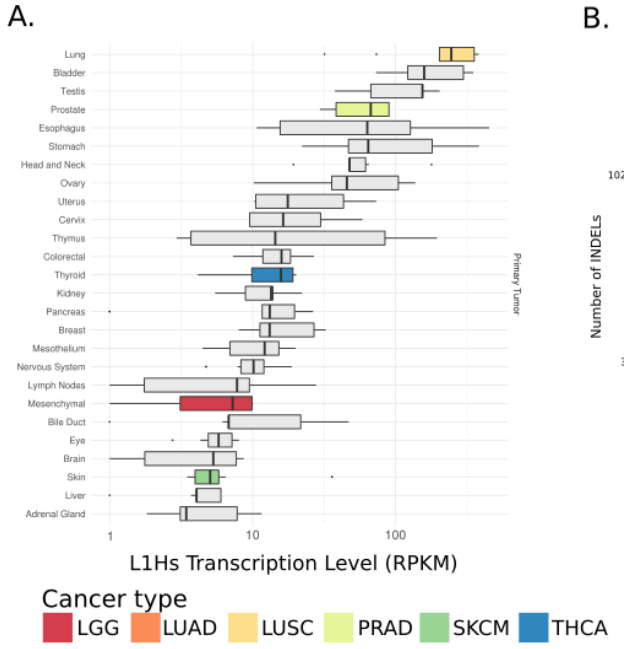
Claim 1 - Major Points
Measuring L1 Transcripts in Human Tissues
The approach taken in Figure 1b to address whether differences exist between tissues is confusing and difficult to draw simple conclusions from. Even when tissues showed deviation from a perfect correlation the correlations were still quite strong ($p>0.9$) suggesting that the signal is not likely to be very significant. Indeed the author's claim later that signal is very low in the GTEx datasets analyzed (highest levels 47 RPKM, 34% of samples with detectable levels).
That's actually play in our favor. Since most of the methods don't take pervasive transcription into account, all this noise data is considered when using RNA-seq data. We should make that clear in the text.

When figure 2 is introduced it becomes apparent that the majority of signal detected is from the LIH subfamily and that signal from the others represents primarily pervasive transcripts rather than active transcription (L1PA2 has low levels). Given this observation it could be of interest to re-plot figures 1A and 1B to show how L1H levels vary between tissues. Presumably the majority of the deviation from the genomic/transcriptomic correlation is driven by elevated L1H transcript abundance but it would be useful to know if any of this data supports a significant decrease in transcript levels relative to genomic abundance, which would possibly represent an important source of 'noise' in the data.
This is a valid suggestion. We could highlight the expression of L1Hs in the plots a showing that L1Hs is actually an outlier, with fewer (still thousands) instances but a higher than expected amount of reads from RNA-seq.

This is even more difficult to interpret relative to figure 3, in which actual L1H transcript abundance is now shown for each tissue. In some cases there appears to be distinctions in the apparent variability observed from the correlation plot in Figure 1B and the transcript abundance plot in Figure 3 - whole blood for example shows a high degree of variability and deviation from a perfect correlation in Figure 1B, but appears to have very low levels of activity in Figure 3 (RPKM<10 in all cases). It is also curious why the most active sites of LIH activity in Figure 3 are not shown in Figure 1 since the dataset is the same - would the tibial nerve not be expected to show the lowest correlation for L1H in Figure 1B if it were plotted?
That's actually a very good point, I should double check that. But the expectation is not really correct. The correlation is calculated for tens of L1 subfamilies, having a single outlier (L1Hs) wouldn't drive a huge change in the correlation, however, as we mention in the text, it was this change in correlation that made me pursue the problem further and create TeXP. It is also an important panel, because it shows that most algorithms today use this data to calculate L1 subfamily expression

Correlations with Age/Weight/Turnover
It is unclear whether the tests done for age and weight for each tissue in Figure 3 have been corrected for multiple testing. Also the correlations presented in Table S5 are not terribly striking. Given that one of the significant correlations is found for Brain (Cerebellar Hemisphere) where virtually no signal is detected in Figure 3 it is hard for me to believe that the other correlations are biologically significant without further validation.
How is the Spleen included as a tissue with 'low cell turnover rates (p35)'? Nucleated cells in the spleen are predominantly lymphocytes and blood lineages, which exhibit some of the highest turnover rates of all cells.
I also think that Liver and Pancreas are hard to define as tissues with low cell turnover rates in humans - provide a reference to support this please. Heart and skeletal muscle (which are among the top most active tissues in this dataset) are on the other hand well known to exhibit very low turnover rates.
For the final statement about tissues with high turnover where nerve and prostate are listed there is also no references to support this claim. I was not aware that prostate and nerve cells were highly proliferative.
I would have to check where the spleen samples where extracted from, I think it doesn't take the blood lineages into account. If this was true, spleen should be very close to blood cells in the GTEx tsne/PCA plot. (03-11, Checked, that's not the case),.

Claim 2 relies on analysis of another published set of data from human cancers to specifically explore the possible role of LINE activity in **driving** malignancies. This is an active area of research with many lines of existing evidence to support a role for LINE-1 in certain cancers. In this manuscript the authors claim to identify a novel mechanism by which LINE-1 transcription causes mutations through activation of the error prone DSB repair pathway due to LINE endonuclease activity.
Claim 2 - Major Point.
Cancer Cells Show Elevated LINE Activity
Here many samples were screened (2504) but no indication for multiple testing corrections is made - please elaborate on whether the appropriate statistics were performed. That some tumors were lower and some were higher might argue that noise (technical for different sample sets or biological for different cancers) could be a confounding factor.
These were the datasets available at the time, we tried to focus on tumors that had a good range of expression of their healthy pairs. Lung was an obvious choice because Tubio et al suggests that Lung is actually the tissue with highest levels of transduction. I created a new figure for the K99 grant which shows that we cover most of the L1 expression range with these tumor sites.



Why were these specific cancer types selected? Colon cancer appears to be associated with LINE activity but was not included in this study - is there a reason for this?
We can add colon cancer as one of the analyzed tumors

For Figure 4A no numerical values are presented on the axis - it is hard to appreciate how strong the signal is. At face value the correlation here appears to be almost entirely driven by the two lung cancers, which could be a spurious correlation rather than causative. The fact that Thyroid carcinoma has the highest INDELS but is among the lowest for LINE activity seems to point to this direction. And the absence of a scale for the quantity of LINE transcripts in 4A makes it hard to know if it is detected at all in this tissue given that melanoma is stated to have 2.38x less activity than healthy skin which had roughly 10-15 RPKM signal on average.
We can include the axis values for the figure 4A, we are sorry that this escaped our revision. We can further filter the number of indels to indels overlapping TT|AAA and show that the correlation is even better. Although we know that L1 endonuclease recognizes other motifs.

Minimal (no?) discussion of why the Thyroid cancer has the most striking enrichment of INDEL near the putative LINE motif is presented in text that might argue against this.
Easy to address

Conclusions
The strengths of the current study are that by tweaking the computational approach to measure transcript abundance the authors claim to be able to more accurately assess levels of LINE activity across human tissues.

Substantial revisions should be made to address a number of key concerns with the claim that LINE activity is elevated in healthy human tissues and arguments regarding relationships between cellular division rates and age/weight need to be toned down barring additional biological evidence.
This is really easy to address.

Absent demonstration that LINE activity can result in INDELs it is hard to see this as a complete manuscript ready for Cell. There have already been substantial discussions about the potential role of LINE family members in the development of cancer. For the purposes of a complete manuscript for publication in Cell I would argue that the authors should perform well-controlled experiments that can definitively demonstrate a causal link between LINE-1 activity and the types of INDELs described in Figures 4 and 5. This could be done in cell lines using targeting sequencing approaches as a read out similar to those used to generate the datasets that the authors have analyzed.
XXXXXXXXX This is hard…

Additional Minor Points
The conclusions that LINE-1 is especially active in epithelial cells is not really supported by the analysis - the human tissue reference used is RNAseq from bulk tissues composed of hundreds of possible cell types and the low levels of detection do not allow for a direct inference into the source of activity in a given tissue type.

Reviewer #3: The authors developed methodology to identify the transcriptional activity of likely active LINE-1 elements by deconvolving LINE-1 associated expression signals from a variety of different samples (normal and tumor samples) taking into account pervasive transcription of presumably non-active elements. The authors then went on applying their methodology on various datasets, and partially are able confirm as well as in part are able to challenge prior assumptions about the activity of LINE elements in various somatic tissues. This methodology allows the authors to present a more comprehensive overview of active LINE-1 transcription across cell types then done in any prior study, to my knowledge. This work is interesting and potentially very impactful.
The authors present an association between LINE-1 transcriptional activity and cell turnover, as well as a

seeming link between LINE-1 transcriptional activity and genomic instability in cancer, which in principle is extremely interesting given the extent of LINE-1 activity in tumors. The line subfamily L1H is shown to be highly active in tumors by the authors. The authors propose a frequent mechanism of mutation formation in cancer whereby target sites of LINE elements accumulate small insertions and deletions at appreciable rates. I found this manuscript to be well-written and the take-home message exciting. The TeXP computational methodology to deconvolve signatures of LINE expression appears interesting and potentially useful for the field. There are, nonetheless a number of remaining questions that the authors will need to address:

1. Error bars and P values were missing for some display items including Figure 4 and Supplemented Figures 2, 3, 4, 5 and 6. These need to be included to be able to judge whether trends presented are significant (most appear to be significant, but it would be good to be reassured through p-values).
Easy to address

2. I am not sure Figure 4A showing general correlation between LINE-1 transcription and Indels is as meaningful as the authors argue in their manuscript, since there could be other reasons for correlation between Indel load and expression (non-causative association, i.e. highly mutated tumors could express LINE-1 elements more highly). This issue could be addressed by the authors through demonstration of an association between gene expression and indel load near the TTT|AA motif. Indeed, the authors would in my view be able to present compelling evidence for their theory (and their model shown in in Fig. 5) if this correlation (indels near TTT|AA with line expression) would exceed the correlation currently shown in Figure 4A (representing all classes of Indels).
Easy to address. And this is actually a good suggestion. But, L1Hs is known to create DSB in other motifs too

3. The authors assume that 5' expression signals mostly derive from autonomous LINE transcription whereas 3' transcription would be derived from a combination of autonomous and pervasive transcription. The authors should present more details on the theory behind this presumption.
Easy to explain… We kind of do it already, but we can make it more clear in the main text

4. The authors should provide a sense of how the mutational rates of Indels near the TTT|AA motif vary across tumor types. Do relative rates (TTT|AA-associated Indels versus all somatic Indels) peak in those tumors that on average show highest LINE-1 expression?
EASY!

Minor:

Second sentence of the last results paragraph: I did not understand what was meant by 'is an extremely costly?' (I assumed the authors meant to state that LINE-1 insertion detection in exomes is very challenging, a statement to which I would agree)