nature portfolio

Peer Review File

LINE-1 insertion intermediates recombine with one another or with DNA breaks to form genome rearrangements

Corresponding Author: Dr Carlos Mendez-Dorantes

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
In this manuscript, Mendes-Dorantes et al. adapted a well-established retrotransposition reporter system to detect and study specific forms of genomic rearrangements associated with LINE-1 retrotransposition in cultured human cells. Using this novel assay, the authors were able to detect events where insertion resolution involved distant DNA breaks or a second concomitant insertion, resulting in translocation. The design of the reporter system is original, clever, and thoroughly characterized and controlled. Overall, the experimental data are strong and the claims are well supported. However, I have concerns related to the biological insights or conclusions that can be deduced from this approach.

Major points:
1. The system requires precise in-frame fusion between the split GFP exons flanking the predicted junction. As a consequence, this assay is essentially biased toward identifying homologous recombination (HR) events, rather than less perfect junctions that may involve other repair pathways. While this limitation is duly acknowledged by the authors in the discussion, it remains a significant constraint, particularly given that a recent preprint from some of the co-authors suggests that HR is not a primary driver of L1-mediated genomic rearrangements. I suggest that the authors include a second reporter where the GFP split is engineered within the intron and lacks homologous sequences. I recognize that splicing of the reporter during retrotransposition would no longer be possible, preventing the direct exclusion of plasmid-plasmid recombination. However, this can be controlled using an RT-defective construct. With such a reporter, imprecise junctions between the L1 and a break (or between two L1s) would lead to an intron-containing GFP that could be spliced and expressed.
2. The detection of rearrangements by ONT whole-genome sequencing is exciting. It is remarkable that the authors could recover rare events and characterize them through whole genome sequencing. This aspect of the work deserves a more detailed description and discussion. What is the coverage and read length? Is the number of reads containing GFP consistent with the frequency of events observed in the GFP+ population? Are there other types of recombination events (e.g, with endogenous L1 loci)? Would it be possible to detect and sequence a broader spectrum of rearrangements using a conventional L1-GFP retrotransposition assay coupled with ONT whole-genome sequencing (maybe at a higher depth)? This would allow for a better estimate of the relative contribution of HR-driven rearrangements compared to other types of rearrangements ? (assuming the frequency reported in the manuscript of 5 retrotransposition-mediated rearrangements per 75 canonical L1 insertions is accurate, I would expect this to be feasible).
3. The involvement of BRCA1 in the described rearrangements is clear, yet not entirely unexpected given the strict requirement for homology inherent to the reporter system (as detailed in point #1). This methodological bias raises questions regarding: 1) whether other rearrangements that do not rely on perfect homology also require BRCA1; and 2) the extent to which this phenotype is specific to the cell line used. It is worth noting that U2OS cells appear to support endonuclease-independent retrotransposition (according to the results shown in Fig. 1), a process that necessitates specific repair pathway alterations, such as deficiencies in non-homologous end-joining and p53 defects (see Morrish et al. Nat Genet 2002 and Nature 2007). Thus, testing the role of BRCA1 using a more unbiased approach (such as those suggested in the above points), and a broader set of cell lines, would be instrumental in clarifying whether BRCA1 is a universal requirement for L1-associated rearrangements or if it is restricted to the specific HR-driven events selected by the current version of the reporter, or reflects specific alterations of the host cellular pathways.

Minor points:
4. Are the 293 cells referenced actually 293 or 293T? (they may tolerate L1 retrotransposition very differently).

5. The schematics illustrating the chromosome junctions are not aways easy to interpret at a glance.
6. Could you clarify whether the different Flip-In cell lines with various homology arms are populations (selected for Flp recombination but non-clonal) or single clones for each construct (which may be subject to clonal effects, and would need replicate clones)?
7. The percentage of 30% of retrotransposon-derived sequences in the human genome is likely to solely include non-LTR retrotransposons

Reviewer #2

(Remarks to the Author)
This manuscript presented a well designed investigation of L1 mediated recombination using complementary GFP based reporters paired with long read sequencing. The study showed that L1 cDNA intermediates recombined with double strand breaks and with one another to generate genome rearrangements, with outcomes strongly dependent on sequence homology and influenced by BRCA1 associated repair pathways. The experimental design was clear, controls and quantitation were appropriate, and the analyses were rigorous, yielding a coherent mechanistic picture of how L1 activity can contribute to structural variation in human cells. The comments that follow outline several suggestions and minor clarifications that would further strengthen the manuscript.

Major comments

1. The current assays rely on the homology of GFP a single highly active L1. Endogenous L1 to L1 recombination spans a range of homology lengths, GC content, and sequence divergence across alleles and subfamilies. To strengthen generality, I suggest adding an experiment with variable recombination targets such as using different active L1s, or at least focused discussion that justifies recombination rate variation. Clarifying how these variables influence pathway choice and product structure will help translate reporter based rates to endogenous L1 to L1 recombination.

2. The ONT data summarized insertion classes and length distributions in GFP positive cells, but the manuscript did not link insertion length to recombination frequency or product structure. Because L1 insertions are frequently 5 prime truncated, a brief exploratory analysis using existing calls would be informative, for example binning insertions by length or 5 prime truncation class and comparing the incidence of recombination associated signatures.

3. The manuscript did not comment on 3 prime transduction. I suggest adding a brief statement on whether any events were detected in the long read data. If so, what is their role on recombination. If none or very few were observed, say so explicitly and note the detection limits so readers can calibrate expectations.

4. Converting the recombination readout into an estimated rate per cell per induced break or per canonical retrotransposition event would aid interpretation. A simple calculation based on cutting efficiency and GFP normalization would help readers judge how often these events might occur in relevant biological settings that include cancers with L1 activity.

5. Given the reported ratio of canonical insertions to recombination events, an estimated number of L1 mediated rearrangements per genome for a representative L1 active cancer cell line would help establish whether the phenomenon is rare enough to dismiss or potentially consequential in tumors.

Minor comments

1. Somewhere in the text or figure legend a global definition of usage of uppercase and lowercase letters in “GFP” should be given, such as the meaning of 5’GFp and 3’gFP.

2. Move critical details now in Methods into the relevant figure legends. For example, the normalization of GFP frequencies to a co-transfected GFP plasmid, the n for each condition, and the gating strategy. You already report n and stats in captions; adding the normalization detail where the frequencies first appear would make the results more self-contained.

Reviewer #3

(Remarks to the Author)
In their manuscript “L1 insertion intermediates recombine with one another or with DNA breaks to form genome rearrangements”, Mendez-Dorantes and colleagues use modified L1 reporter constructs to examine the potential spectrum of L1-mediated chromosomal rearrangements and gain mechanistic insight regarding their formation. The work presented complements recent pre-prints from their own research group and the Tubio lab which also present evidence for L1-mediated rearrangements; the former using traditional L1 reporters, and the later using long-read whole genome sequencing of human tumours. The present study uses an ingenious split GFP reporter system to detect structural rearrangements arising from the recombination of “in-progress” L1 integrations arising by target-primed reverse transcription (TPRT) with a stably integrated GFP sequence in a known chromosomal location, or by the recombination of two synchronous TPRT events. By design, this impressive experimental system identifies homology-driven structural rearrangements, in this case mediated by the “split” GFP sequence; presumably, similar endogenous L1-mediated rearrangements would be mediated by homology in the native L1 sequence. Overall, the results are intriguing and complement parallel emerging insights in this area, building on a decades-spanning body of work demonstrating the large-scale genomic impact of L1 retrotransposition.

However, there are a few areas where I think this experimental system could be exploited more thoughtfully or pushed further to provide more solid evidence and additional mechanistic insight on how these rearrangements likely come about.

1. Cas9-mediated targeted ONT sequencing of L1 insertions arising from recombination between the fixed Chr11 “FP” sequence and the reporter-delivered “GFP” sequence is an excellent strategy to characterise structural rearrangements resulting in GFP-positive cells. However, I do not understand why the guide RNA directing adaptor ligation was targeted to the GFP sequence. This design results in ONT long-reads encompassing the 3’ flank (chr11) -or- the 5’ flank (chromosome targeted by transfected “GF” L1 reporter). The entire rearrangement-spanning structure is never contained within a single read, and the evidence for L1-mediated rearrangement is limited to sequencing confirmation of the reconstituted GFP. Targeting the chr11 genomic sequence perhaps a few hundred bp downstream of the fixed Chr11 “FP” integrant would result in a large number of identical ONT reads extending away from the recombination site but also would generate ONT reads beginning downstream of the Chr11 recombination site, encompassing the entire reconstituted GFP cassette, and extending into the genomic sequence comprising the other participant in the recombination event. A major advantage of characterising L1 insertions via long-read sequencing is the ability to span entire L1 integrants including their structural hallmarks and associated genomic perturbations in a single read representative of a single DNA molecule. The above refinement of the authors’ targeted ONT sequence strategy would take full advantage of this experimental platform.

2. The dependence of GFP expression on the introduction of a Cas9-mediated DNA break at the “FP” chromosome 11 integrant demonstrates nicely that both homology and a DNA break are required for recombination-driven L1-mediated chromosomal rearrangements. This finding is reinforced in the WGS analysis which identified rearrangements putatively arising from the recombination of two “in progress” TPRT events. To further strengthen the conclusion that translocations arise from synchronous TPRT events, rather than sequentially, it would be interesting to stagger the transfection of the “GF” and “FP reporters, allowing retrotransposition events from the first reporter to become fully integrated before the second reporter is introduced.

3. The use of robust cultured cell lines presents a key opportunity to generate clonal cell lines from GFP positive cells putatively harbouring L1-mediated chromosomal rearrangements. While some rearrangements clearly are unstable and will be lost during clonal expansion, I expect some could persist. The split reporter assay could be redesigned using a selectable (e.g. neomycin) rather than screenable marker, to enforce retention of the L1-mediated translocation. The generation of clonal cell lines would allow more robust validation of chromosomal translocations and other L1-mediated structural rearrangements.

In revision, I request that the authors generate clonal cell lines, identify L1-mediated translocations, and formally demonstrate these rearrangements using (ideally) a karyotyping/chromosome painting approach.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This revised manuscript represents a beautiful piece of work. The authors have now addressed all my initial questions and concerns, and I fully support its publication in Nature Communications.

Reviewer #2

(Remarks to the Author)
My concerns have been satisfactorily addressed, and I endorse the publication of this work.

Reviewer #3

(Remarks to the Author)
The authors have more than sufficiently addressed my comments. The revised manuscript work advances significantly our understanding the role of LINE-1 activity in cancer genome instability, and I look forward to seeing it published.

Point-by-point responses to the reviewers' comments for NCOMMS-25-74535.

We thank the reviewers for their suggestions to improve the manuscript. The reviewers' comments are shown in *italics/blue*.

Reviewer #1 (Remarks to the Author):
In this manuscript, Mendes-Dorantes et al. adapted a well-established retrotransposition reporter system to detect and study specific forms of genomic rearrangements associated with LINE-1 retrotransposition in cultured human cells. Using this novel assay, the authors were able to detect events where insertion resolution involved distant DNA breaks or a second concomitant insertion, resulting in translocation. The design of the reporter system is original, clever, and thoroughly characterized and controlled. Overall, the experimental data are strong and the claims are well supported. However, I have concerns related to the biological insights or conclusions that can be deduced from this approach. While this limitation is duly acknowledged by the authors in the discussion, it remains a significant constraint, particularly given that a recent preprint from some of the co-authors suggests that HR is not a primary driver of L1-mediated genomic rearrangements.

We thank the reviewer for acknowledging the novelty and rigorous validation of our experimental system to study chromosomal rearrangements mediated by L1 insertion intermediates. We completely agree with the reviewer's point that studies of cancer genomes suggest that HR-mediated recombination events are uncommon as compared to end-joining. Modeling the latter has proven more difficult; as described below, we have attempted to build reporters of imperfectly end-joined recombination products though these are plagued by significant background signals and do not appear to be usable. We will continue to innovate in this area. We assert that the reporters we do report in this manuscript will help the field, perhaps precisely by providing insights into why HR-mediated recombinations are not more commonplace. For the current publication, we address the reviewer's concerns related to the biological insights to our findings with new experiments, new analyses and an extended discussion about the implications of our work on L1 retrotransposition in cancer genomes, as detailed in response to point 1., immediately below, and also in response to point 3, through the lens of a role for BRCA1.

Major points:

1. The system requires precise in-frame fusion between the split GFP exons flanking the predicted junction. As a consequence, this assay is essentially biased toward identifying homologous recombination (HR) events, rather than less perfect junctions that may involve other repair pathways. While this limitation is duly acknowledged by the authors in the discussion, it remains a significant constraint, particularly given that a recent preprint from some of the co-authors suggests that HR is not a primary driver of L1-mediated genomic rearrangements. I suggest that the authors include a second reporter where the GFP split is engineered within the intron and lacks homologous sequences. I recognize that splicing of the reporter during retrotransposition would no longer be possible, preventing the direct exclusion of plasmid-plasmid recombination. However, this can be controlled using an RT-defective construct. With such a reporter, imprecise junctions between the L1 and a break (or between two L1s) would lead to an intron-containing GFP that could be spliced and expressed.

We acknowledge the limitation of our split GFP recombination assays: these reporters do not model *ALL* the possible genome rearrangements that could be mediated by the chromosomal breaks induced by L1 retrotransposition but are limited to a subset of possible outcomes. In a separate manuscript[1] from our group, we have comprehensively characterized the genomic alterations caused by L1 retrotransposition, and building reporters to study each type in depth will have value going forward. We suggest the focus of our study here is to address how L1 cDNAs use extensive homologous sequences to recombine with each other or with other DNA breaks to form genome rearrangements, and we appreciate that the reviewer agrees that we have been rigorous in this effort. The reviewer proposed a strategy to study the recombination of L1 cDNA intermediates through mechanisms independent of homologous sequences such as imperfect end joining resolution. While this is desirable, unfortunately, the reviewer also predicts a significant problem with this strategy, as we describe in point 4., below. To address this concern, we have performed the following experiments and analyses in our revised manuscript:

1) We re-analyzed our ONT WGS data and now report chromosomal rearrangements caused by L1 retrotransposition that did not reconstitute GFP and that did not exploit extensive homology:

By sequencing L1 expression models and cancer genomes, we and others have shown that L1 cDNA insertion intermediates commonly recombine with each other using limited homology or no homology to generate diverse genome rearrangements[1,2]. Consistent with these results, in our revised manuscript, we now report L1 retrotransposition-mediated chromosomal rearrangements that failed to reconstitute GFP expression and that did not exploit extensive homologous sequences in our ONT WGS data from GFP+ cells derived from U20S cells and HEK293 cells (SFig 11, SFig 7). We propose that these rearrangements were mediated by the joining of L1 cDNA products generated by induced L1 expression to other distant DNA breaks of unknown etiology. We report chromosomal rearrangements containing microhomology (n=2) or 8-9 bp non-templated insertions (n=2) at the recombination junctions between the L1 cDNA products and the distant chromosomal breaks, suggesting that limited or no microhomology was exploited in the formation of these L1-mediated rearrangements. While these outcomes do not generate GFP and thus are not indicated by our reporter system, their inclusion in the manuscript provide an opportunity to highlight end join events for our readers.

2) We also now report chromosomal rearrangements caused by L1 retrotransposition that did not reconstitute GFP but likely used extensive homology to recombine with endogenous full-length copies of L1.

Our unbiased analysis for detecting L1-mediated rearrangements independent of GFP reconstitution revealed two chromosomal rearrangements that did not restore GFP but were mediated by the recombination of L1 cDNA intermediates and endogenous full-length copies of L1 (L1PA3) (SFig 11). These examples implicate the involvement of extensive homologous sequences in the formation of these rearrangements like in our GFP recombination events. These findings are significant in part because the prevalence of L1 sequences throughout the genome means there are many possible substrates for these types of recombinations.

3) We have added new experimental data showing that sequence divergence (mismatches) between L1 cDNA products limit their recombination, which we propose is a likely mechanism for suppressing these rearrangements in cancer genomes.

Although 17% of our genome is composed of L1 sequences, only a few hundred L1 copies (L1Hs) are full-length and retrotransposition-competent, which are the endogenous L1 elements that we are modeling in our study. These retrotransposition-competent L1 elements have some single nucleotide polymorphisms (SNPs)[2], which could generate mismatches during the annealing of two L1 cDNA products preventing their recombination via DNA heteroduplex rejection[3]. Additionally, all potentially active L1 (L1Hs and L1PA2) have nucleotide variants in their 3' UTR that distinguish them from evolutionarily older subfamilies of L1; this reduces homology with other high copy number L1s that may happen to be located proximal to dsDNA break. We have now added a new Main Figure Panel 7 to our manuscript showing that mismatches between L1 cDNA products significantly limits their recombination to form genome rearrangements in manner dependent on the mismatch repair factor MSH2 (See Figure 7 in the manuscript).

Briefly, to examine the hypothesis that imperfectly matched sequences are less likely to recombine with one another, we generated another variant of the L1-3' gFP reporter plasmid such that L1 cDNA intermediates would be produced containing three mismatches within the span of the 150 bp of overlapping homology sequence to L1 cDNA products generated by the L1-5'GFp reporter plasmid (Figure 7A). We found that 3 mismatches between the 150bp of overlapping sequence of the L1 cDNA products significantly reduces the frequency of GFP recombination events (Figure 7B). We then sought to test whether the mismatch repair factor MSH2 suppressed the recombination of L1 cDNA products containing mismatches. For this, we depleted MSH2 from cells using a pool of siRNA, which we found did not drastically influence the frequency of canonical L1 retrotransposition (SFig 12D) or GFP recombination of L1 cDNA products with 150 bp of identical homology (Fig 7C). In contrast, we found that depletion of MSH2 caused a significant increase in frequency of recombination between L1 cDNA products containing mismatches/ sequence divergence. Together, these findings show that mismatches between L1 cDNA products limit their recombination to form genome rearrangements in a manner dependent on the mismatch repair factor MSH2 likely via its role in heteroduplex rejection.

These results highlight L1 sequence divergence as a factor that prevents their recombination and maintains genome integrity.

4) As requested by the reviewer, we constructed a split GFP reporter to model recombination of L1 cDNA intermediates through mechanisms independent of homologous sequences such as imperfect end joining resolution. However, as was predicted by the reviewer, the suggested experimental set-up to engineer a split GFP within an intron to capture events mediated by imperfect end joining resolution resulted in a high background of false-positive GFP signal likely resulting from plasmid-plasmid recombination.

*See **Figure i**.* As suggested by the reviewer, we developed a split GFP reporter assay designed to generate L1 insertion intermediates that could recombine via imperfect end joining to reconstitute an *intron-containing* GFP reporter gene (**Figure i**). We designed and generated two L1 plasmids (L1-5'Gfp-INT plasmid and L1-RON-3'gFP plasmid) predicted to produce L1 cDNA intermediates containing a 5'GFP fragment fused to a 5'Intron fragment and a 3'Intron fragment fused to a 3'GFP fragment, respectively. The two parts of the reporter shared no overlapping homologous sequences, and their recombination / end-joining within the span of the intron would reconstitute a functional intron-containing GFP cassette. We introduced this split GFP reporter assay in U20S cells and found a robust induction of GFP+ cells via flow cytometry. However, we failed to see a reduction GFP+ cells upon mutation of the reverse transcriptase domain of ORF2p in both L1 plasmids, suggesting that the induction of GFP+ cells is independent of L1 retrotransposition. We found that induction of GFP+ cells required the transfection of both L1-5'Gfp-INT plasmid and L1-RON-3'gFP plasmid, consistent with a model wherein induction of GFP+ cells resulted from plasmid-plasmid recombination or *trans*-splicing, either of which would not monitor the intended events: recombination of L1 cDNA intermediates via imperfect end joining. Accordingly, these results highlight the importance of splicing of the antisense intron in L1 retrotransposition reporters to selectively model L1 retrotransposition biology. We decided to present these results of this assay to the reviewer here (**Figure i**). Since this strategy did not successfully model chromosomal rearrangements or LINE-1 biology as we intended, and because the manuscript is already at a target length for *Nature Communications*, we propose not to include these data in the manuscript, though wanted to share them here for our reviewers.

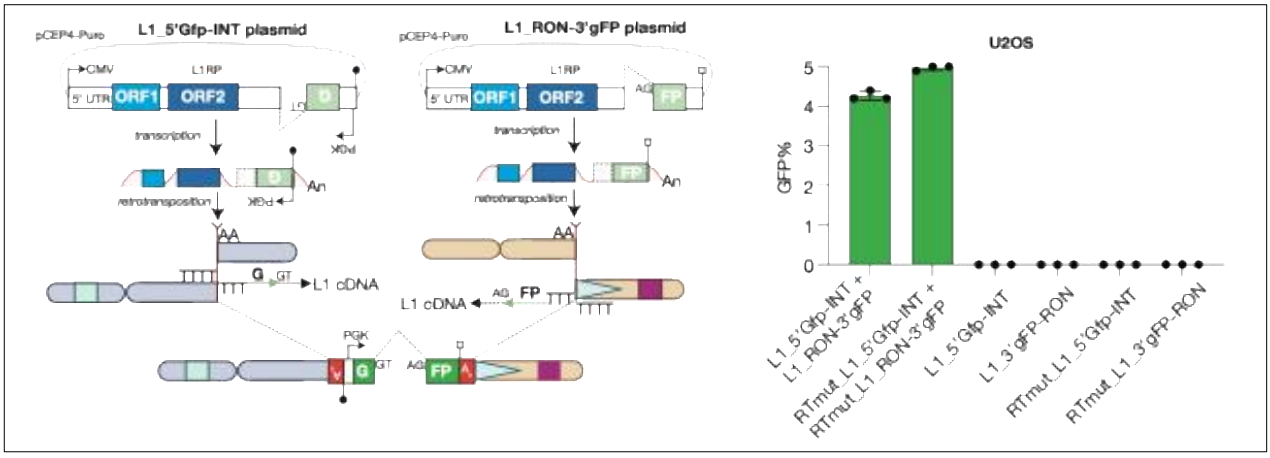

Figure i. A split GFP reporter assay appears to reconstitute GFP expression in cells via plasmid-plasmid recombination but not via L1 retrotransposition. *Left.* Shown is a schematic of a split GFP reporter assay (L1-5'Gfp-INT plasmid and L1-RON-3'gFP plasmid) designed to generate L1 cDNA insertion intermediates that could recombine via imperfect end joining to reconstitute an intron-containing GFP cassette (**Figure i**). *Right.* Shown are the percentage of GFP+ cells induced in U20S cells after being transfected with WT or RT mutant L1 plasmids containing either the 5'Gfp-INT reporter or the RON-3'gFP reporter. We both the WT and RT mutant L1 plasmids robustly induce GFP+ cells, which suggest that the induction of GFP+ cells is independent of L1 retrotransposition and instead is dependent on plasmid-plasmid recombination. Transfection were done in triplicates, and GFP+ cells were analyzed via flow cytometry five days after transfection.

5) Lastly, we examined the impact of varying the length of overlapping homology (ranging from 150bp to 6bp between L1 cDNA products) on their frequency for rearrangement to reconstitute GFP expression (See **Figure 4B**). We found GFP recombination events were detectable in cells when L1 cDNA products share homology as short as 25 bp. The frequency of these GFP recombination events was low, 0.25%, and (as expected) still

dependent on L1 ORF2p reverse transcriptase activity. We believe our GFP recombination reporter assay reaches its lower limit of detection for indicating L1 cDNA recombination when sequences share homology shorter than 25 bp. Indeed, we failed to see an induction of GFP recombination events when L1 cDNA products share 12 bp or 6 bp of homology. Thus, we could not model recombination events with more limited microhomology in the current experiment setting (i.e., could not reproducibly measure RT-dependent GFP+ events at lower frequencies).

In summary, we recognize that our GFP reporters do not model all possible rearrangements generated by L1 retrotransposition. However, this system permits us to study rearrangements that can clearly occur in cells with deleterious consequences for genome stability, and to delineate roles of key molecular contributors.

2. The detection of rearrangements by ONT whole-genome sequencing is exciting. It is remarkable that the authors could recover rare events and characterize them through whole genome sequencing. This aspect of the work deserves a more detailed description and discussion. What is the coverage and read length? Is the number of reads containing GFP consistent with the frequency of events observed in the GFP+ population? Are there other types of recombination events (e.g, with endogenous L1 loci)? Would it be possible to detect and sequence a broader spectrum of rearrangements using a conventional L1-GFP retrotransposition assay coupled with ONT whole-genome sequencing (maybe at a higher depth)? This would allow for a better estimate of the relative contribution of HR-driven rearrangements compared to other types of rearrangements ? (assuming the frequency reported in the manuscript of 5 retrotransposition-mediated rearrangements per 75 canonical L1 insertions is accurate, I would expect this to be feasible).

The reviewer here is requesting us to 1) clarify the details about the ONT sequencing runs, 2) report any rearrangement with endogenous L1 copies, and 3) characterize other types of chromosomal rearrangements generated by L1 retrotransposition (similar to the Reviewer 1st Major Point).

1) Statistics of our ONT sequencing runs:

We now provide statistics for each of the ONT sequencing runs we completed for this study in Table 4. This table lists the total number of reads generated, the average read length per run (5kb-10kb), the percent of reads mapping to the human genome per run, and percent of reads aligning to our predicted recombination sequence per run (less than 1% for every run). Long reads aligning to our reporter sequence represent chromosomal rearrangements (if they span a complete recombination) or canonical retrotransposition insertions (if they partially map to a spliced GFP). Subsequent alignment to the T2T human genome reference allowed us to characterize re-construct the sequence features of each rearrangement and insertions. **We provide readers a comprehensive list of all rearrangements and insertions providing a supporting sequencing read for each, including the Read Label ID and the respective sequence.** Notably, we recovered a higher number of rearrangements and canonical insertions in U20S cells than in HEK293 cells, which is consistent with a higher GFP+ frequency in U20S versus HEK293 induced with our reporter assays.

2) We now report in our revised manuscript chromosomal rearrangements caused by L1 retrotransposition that did not reconstitute GFP but likely used extensive homology to recombine with endogenous full-length copies of L1.

Our unbiased analysis of our ONT sequencing data of GFP+ cells for the detection of L1-mediated rearrangements independent of GFP reconstitution revealed two chromosomal rearrangements that did not restore GFP but were mediated by the recombination of L1 cDNA intermediates and endogenous full-length copies of L1 (L1PA3) (SFig 11). These examples implicate intervals of extensive homology in the formation of these rearrangements similar to our GFP recombination events.

3) Characterize other types of chromosomal rearrangements generated by L1 retrotransposition (Similar to the Reviewer 1st Major Point).

As highlighted above, we now report in our revised manuscript additional rearrangements mediated by L1 retrotransposition that did not reconstitute GFP expression (SFig 11, SFig 7). We propose that these rearrangements were mediated by the joining of L1 cDNA products generated by our L1 expression plasmids

to distant DNA breaks of unknown etiology. We report chromosomal rearrangements containing limited microhomology (n=2) or 8-9 bp non-templated insertions (n=2) at the recombination junctions between the L1 cDNA products and the distant chromosomal breaks, suggesting that limited or no microhomology was used in the formation of these L1-mediated rearrangements. Also, we report additional rearrangements mediated by L1 retrotransposition in HEK293 GFP+ cells that we profiled using ONT WGS (SFig 7, Table 2), including five that reconstituted GFP expression, and one that failed to reconstitute GFP expression, as well as 17 canonical L1 insertions. In the ONT WGS of U2OS GFP+ cells induced by the recombination of L1 cDNA products with each other, we report a total of 5 chromosomal rearrangements that reconstituted GFP expressions and 6 chromosomal rearrangements that failed to reconstitute GFP expression, as well as 75 canonical L1 insertions. Finally, we have validated chromosomal rearrangements mediated by L1 retrotransposition using PCR validation of some of the rearrangements we report in this manuscript (See Fig 2E, Fig 5C), as well as via metaphase spread analysis of GFP+ cells showing an additional fragment of chromosome 11 confirming the involvement of the chr11 locus in rearrangements mediated by L1 cDNA in U2OS FRT cells (SFig 6).

Lastly, in a separate manuscript, we comprehensively profiled genomic alterations caused by L1 retrotransposition in single cell clones established after exposing cells to L1 expression using human RPE-1 hTERT cells[1], which are stable nearly diploid cells. In that work, we found on average 10 canonical insertions and two chromosomal rearrangements mediated by L1 retrotransposition per clone. Hence, we direct readers to that other manuscript and emphasize a different scope in the current study. We are gratified that the reviewer agrees that we have rigorously and thoroughly validated that our GFP recombination reporter systems produce chromosomal rearrangements and indicate with GFP a subset of those outcomes, which is the focus of the current manuscript.

3. The involvement of BRCA1 in the described rearrangements is clear, yet not entirely unexpected given the strict requirement for homology inherent to the reporter system (as detailed in point #1). This methodological bias raises questions regarding: 1) whether other rearrangements that do not rely on perfect homology also require BRCA1; and 2) the extent to which this phenotype is specific to the cell line used. It is worth noting that U2OS cells appear to support endonuclease-independent retrotransposition (according to the results shown in Fig. 1), a process that necessitates specific repair pathway alterations, such as deficiencies in non-homologous end-joining and p53 defects (see Morrish et al. Nat Genet 2002 and Nature 2007). Thus, testing the role of BRCA1 using a more unbiased approach (such as those suggested in the above points), and a broader set of cell lines, would be instrumental in clarifying whether BRCA1 is a universal requirement for L1-associated rearrangements or if it is restricted to the specific HR-driven events selected by the current version of the reporter, or reflects specific alterations of the host cellular pathways.

We have addressed the reviewer's concern by showing that the requirement of BRCA1 for L1-mediated rearrangements using homologous sequences is not specific to U20S cells, which are cells widely used in the DNA repair field to probe BRCA1 functions[4]. Indeed, we have replicated the key finding regarding BRCA1 in HEK293 cells and RPE-1 cells:

- BRCA1 is required for homologous recombination between L1 cDNA products and chromosomal breaks that share homologous sequences in HEK293 cells (SFig 12B). Here, we found that depletion of BRCA1 (using siRNA) reduced the frequency of these recombination events.

- BRCA1 is required for the recombination between two L1 cDNA products that share homologous sequences in RPE-1 cells (SFig 12C). Here, we found that the recombination frequency between L1 cDNA products that share homologous sequences was reduced in RPE-1 hTERT p53-/- BRCA1-/- versus RPE-1 hTERT p53-/- BRCA1+/+ cells (SFig 12C).

Together, these results show that BRCA1 is a key factor involved in the formation of rearrangements mediated by the recombination of L1 cDNA intermediates with one another or with chromosomal breaks.

Based on these experiments, we agree with the reviewer that we cannot rule out the possibility that BRCA1 is a suppressor for other types of L1-mediated rearrangements not modeled by our GFP recombination reporter assays. Indeed, in a separate manuscript, we report that L1 retrotransposition contributes to the formation of diverse large-scale genome rearrangements and elicits chromosomal instability[1]. Since BRCA1 is a

suppressor of canonical L1 retrotransposition and BRCA1 is required for long-term growth of cells with L1 expression[5], we have now edited our Discussion to highlight that future work is required to address the impact of BRCA1 on the integrity of the genome of cells with ongoing L1 retrotransposition. We expect this will be an important direction of future research, although is beyond the scope what can be reasonably covered in the current manuscript which focuses on L1 cDNA products recombining using homology.

Minor points:

4. Are the 293 cells referenced actually 293 or 293T? (they may tolerate L1 retrotransposition very differently).

The cells we used in our study are HEK293 cells; these were modified to contain the Flp-In T-Rex system and are originally derived from Invitrogen (R71007). We have edited the **Methods** under Cell Lines to specify the Invitrogen catalog number of the HEK293 T-Rex cells for the readers. Indeed, all experiments in HEK293 cells were performed in these cells.

5. The schematics illustrating the chromosome junctions are not aways easy to interpret at a glance.

We thank the reviewer for this comment and have addressed this by making the schematics representing chromosomal rearrangements in our figures larger and less crowded, adding additional labels showing what each schematic represents, and adding clarifying descriptions in the figure legends.

6. Could you clarify whether the different Flip-In cell lines with various homology arms are populations (selected for Flp recombination but non-clonal) or single clones for each construct (which may be subject to clonal effects, and would need replicate clones)?

The U20S Flp-In cell lines included in our original submission of the manuscript are single clones for each construct. Nonetheless, we have now included additional experiments replicating many of the key findings from our study in other cell lines, including HEK293FRT and RPE-1 hTERT cells. Together, these results suggest that the results derived from the engineered U20S clonal cell lines are not clonal effects. Here is a list of these experiments we are now including in our manuscript.

- Sequencing of genome rearrangements induced by the recombination of L1 cDNA products and chromosomal breaks in HEK293 cells, showing these events are not specific to U2OS cells (SFig 7, Table 2). For this, we performed ONT whole-genome sequencing of GFP+ cells derived from HEK293 3'gFP (150bp Hom) cells transfected with L1-5' GFp plasmid and the Cas9/sg150bp expressing plasmid.

- Requirement of long homologous sequences between L1 cDNA products and chromosomal breaks for their recombination in HEK293 cells (SFig 9), showing that this key result is not a clonal effect of our U20S single clones. For this, we established an additional set of single clones in HEK293FRT cells containing each of the three 3' gFP sequences at the FRT locus, where we performed our split GFP reporter assay to assess how length of homology between L1 cDNA products and chromosomal breaks on affects their recombination frequency. Similar to U20S single clones (Fig 3), we found that long homologous sequences between L1 cDNA products and chromosomal breaks favor their recombination in HEK293 cells (SFig 9).

- Recombination between L1 cDNA products requires long homologous sequences in parental HEK293 cells in addition to parental U20S cells (Fig 4D).

- BRCA1 is required for recombination between L1 cDNA products and chromosomal breaks that share homologous sequences in HEK293 cells in addition to U2OS cells (SFig 12B).

- BRCA1 is required for recombination between L1 cDNA products that share homologous sequences in RPE-1 cells (SFig 12C).

In summary, we have replicated key findings from our study in additional cell lines.

7. The percentage of 30% of retrotransposon-derived sequences in the human genome is likely to solely include non-LTR retrotransposons.

Thank you to the reviewer for noting this statement in our introduction. We have now edited our first two sentences of our introduction to be inclusive of all retrotransposons to make a distinction between LTR and non-LTR retrotransposons in the human genome.

"Nearly half of our genome is composed of repetitive DNA elements derived from the activity of retrotransposons [6,7]. Retrotransposons are mobile DNA sequences that copy and paste in the genome via RNA intermediates and include long terminal repeat (LTR) retrotransposons (e.g. endogenous retroviruses) and non-LTR retrotransposons (e.g., LINE-1 (L1), Alu and SVA)."

Reviewer #2 (Remarks to the Author):

This manuscript presented a well-designed investigation of L1 mediated recombination using complementary GFP based reporters paired with long read sequencing. The study showed that L1 cDNA intermediates recombined with double strand breaks and with one another to generate genome rearrangements, with outcomes strongly dependent on sequence homology and influenced by BRCA1 associated repair pathways. The experimental design was clear, controls and quantitation were appropriate, and the analyses were rigorous, yielding a coherent mechanistic picture of how L1 activity can contribute to structural variation in human cells. The comments that follow outline several suggestions and minor clarifications that would further strengthen the manuscript.

We thank the reviewer for acknowledging the strengths of our experimental system to investigate chromosomal rearrangements mediated by L1 insertion intermediates using GFP reporters paired with long-read sequencing. As described in our point-by-point responses to follow, we have now addressed the suggestions and requests for clarifications of the reviewer with the addition of new experiments, new analyses and changes to the text.

Major comments

1. The current assays rely on the homology of GFP a single highly active L1. Endogenous L1 to L1 recombination spans a range of homology lengths, GC content, and sequence divergence across alleles and subfamilies. To strengthen generality, I suggest adding an experiment with variable recombination targets such as using different active L1s, or at least focused discussion that justifies recombination rate variation. Clarifying how these variables influence pathway choice and product structure will help translate reporter-based rates to endogenous L1 to L1 recombination.

We thank the reviewer for asking us to consider the impact of diverse L1 retrotransposon sequences on the frequency on L1-to-L1 recombination for generating genome rearrangements. Retrotransposition-competent L1 elements have single nucleotide polymorphisms (SNPs)[2], both between one another and (more so) between their sequences and evolutionarily older sequences which are 'fixed' in the genome. These SNPs then represent potential mismatches during the annealing of two L1 cDNA products or the annealing of a cDNA intermediate of transposition and another genomic L1, limiting their recombination via DNA heteroduplex rejection. We have now added a new Figure Panel 7 to our manuscript showing that mismatches between L1 cDNA products significantly limits their recombination to produce genome rearrangements in manner dependent on the mismatch repair factor MSH2 (See Figure 7 in the manuscript or Figure i in the response to Reviewer 1).

As described above, to examine this hypothesis, we generated another variant of the L1-3' gFP reporter plasmid such that L1 cDNA intermediates would be produced containing three mismatches within the span of the 150 bp of overlapping sequence to L1 cDNA products generated by the L1-5'GFp reporter plasmid (Figure 7A). We found that 3 mismatches between the 150bp of overlapping sequence of the L1 cDNA products significantly reduces the frequency of GFP recombination events (Figure 7B). We then sought to test whether the mismatch repair factor MSH2 suppressed the recombination of L1 cDNA products containing mismatches. For this, we depleted MSH2 from cells using a pool of siRNA, which we found did not drastically influence the frequency of canonical L1 retrotransposition (SFig 12D) or GFP recombination of L1 cDNA products with 150

bp of identical homology (Fig 7C). In contrast, we found that depletion of MSH2 caused a significant increase in frequency of recombination between L1 cDNA products containing mismatches/ sequence divergence. Together, our work shows that mismatches between L1 cDNA products limit their recombination to form genome rearrangements in a manner dependent on the mismatch repair factor MSH2 likely via its role in heteroduplex rejection.

These results highlight the impact of sequence divergence between L1 copies to prevent their recombination. We conclude that stringent requirements for homology serve to limit recombination between high copy number repeats and maintain genome integrity.

2. The ONT data summarized insertion classes and length distributions in GFP positive cells, but the manuscript did not link insertion length to recombination frequency or product structure. Because L1 insertions are frequently 5 prime truncated, a brief exploratory analysis using existing calls would be informative, for example binning insertions by length or 5 prime truncation class and comparing the incidence of recombination associated signatures.

The design of our L1 split GFP reporters requires that L1 cDNA insertion intermediates be processed/ 5' truncated to reconstitute the GFP cassette in the recombination events. Accordingly, both of our split GFP reporter systems (i.e., the system modeling recombination of two L1 cDNA products with each other and the system modeling recombination between one L1 cDNA and a chromosomal break) selectively capture 5' truncated insertions at the GFP recombination junctions. Namely, the L1-5'GFp and the L1-3'gFP reporter plasmids generates 5' truncated insertions that are 864 bp or 825 bp long respectively in the recombination junctions. Other studies have highlighted that L1-induced chromosomal rearrangements can produce 5' truncated reverse transcribed products at breakpoint junctions. We have added the following two sentences in the Results section to clarify that 5' truncated insertions of consistent lengths are captured in the recombination events.

-"*As expected, the L1 insertion cDNAs recombined with chromosomal breaks in these GFP recombination events would all be 5' truncated insertions (864 bp) without additional L1 sequence from the L1-5'GFp reporter construct."*

-*"Indeed, we anticipate that two L1 5' truncated insertions (864 bp for the L1-5'GFp reporter and 825 bp for the L1-3'gFP reporter) would be captured at the GFP recombination junction."*

In addition, we comprehensively characterized all the genomic alterations caused by L1 retrotransposition in our ONT WGS of GFP+ cells including those that failed to reconstitute the GFP cassette. Through this analysis, we identified five additional chromosomal rearrangements mediated by the recombination of L1 cDNA products and other chromosomal breaks of unknown etiology, which we have now added as an additional supplementary figure (SFig 11) and discuss in the Results section. Briefly, all these events represent 5' truncated L1 insertions of varying lengths longer than ~800bp and ranging from 1491 bp to 4807 bp that recombined with distant chromosomal breaks to cause genome rearrangements.

Hence, we now report that 5' truncated L1 insertions of varying lengths could be involved in genome rearrangements, which is consistent with previous studies[1,2,8-10].

3. The manuscript did not comment on 3 prime transduction. I suggest adding a brief statement on whether any events were detected in the long read data. If so, what is their role on recombination. If none or very few were observed, say so explicitly and note the detection limits so readers can calibrate expectations.

We did not detect 5' transductions or 3' transductions derived from our L1 transgene expression systems in either canonical L1 insertions or chromosomal rearrangements. Indeed, we believe that these types of sequence additions are not well modeled by our L1 transgenes since transcription of these L1 constructs is expected to consistently start at the engineered CMV promoter and not extend downstream of the polyA signal. Accordingly, we have added the following sentence in the discussion of the L1 insertions detected from our ONT WGS analysis of GFP+ cells:

"Notably, we did not detect 5' transductions[11] or 3' transductions[12-14]; insertions containing additional sequences upstream or downstream of L1 may not be modeled by our transgene system since transcription starts at the CMV promoter and does not extend downstream of the polyA signal."

4. Converting the recombination readout into an estimated rate per cell per induced break or per canonical retrotransposition event would aid interpretation. A simple calculation based on cutting efficiency and GFP normalization would help readers judge how often these events might occur in relevant biological settings that include cancers with L1 activity.

We have now added a new Supplementary Figure Panel (SFig S3D) showing the frequency of GFP recombination between L1 cDNA products and chromosomal breaks (1.9%) relative to canonical L1 retrotransposition (42.2%, hence normalized frequency is 4.5%), relative to Cas9 cutting efficiency (Indel mutagenesis) (41.4%, hence normalized frequency is 4.6%), and relative to transfection efficiency (74%, hence normalized frequency is 2.6%). The normalized GFP frequency does not drastically change based on these distinct approaches of normalizing the GFP recombination events, and we maintain the approach of normalizing GFP recombination events to transfection efficiency in most analyses. In addition, a recent study[2] of cancer genomes from Jose Tubio's group during revision of this manuscript estimates that one per 40 insertions gives rise to a genome rearrangement (2.5%), which is near the frequency of our GFP reporter assay.

5. Given the reported ratio of canonical insertions to recombination events, an estimated number of L1 mediated rearrangements per genome for a representative L1 active cancer cell line would help establish whether the phenomenon is rare enough to dismiss or potentially consequential in tumors.

A recent study of cancer genomes from Jose Tubio's group and published in *Science* during the revision of this manuscript shows that chromosomal rearrangements mediated by the recombination of two independent L1 insertion intermediates occur and can arise early in tumorigenesis[2]. These findings indicate that L1 retrotransposition activity and associated genomic consequences can act as early drivers of cancer genome evolution. We suggest that L1-mediated rearrangements may not be as frequent as L1 insertions or other forms of mutation (i.e. in context of the overall tumor mutational burden, TMB) but may still be consequential in tumor development. We now clearly report that frequency of L1 retrotransposition-mediated rearrangements versus canonical insertions in our discussion and consider the potential consequences of these events in cancers. We reference the key results of from Tubio's group showing these rearrangements are cancer drivers. We suggest that our split GFP reporters provide direct evidence of interactions between *synchronous* or *concurrent* intermediates of retrotransposition as well as causal evidence that L1 insertion intermediates mediate genome rearrangements, which is a finding also recognized in a Perspective article (DOI: 10.1126/science.aeg466) accompanying Jose Tubio's group study[2] in *Science*.

Minor comments

1. Somewhere in the text or figure legend a global definition of usage of uppercase and lowercase letters in "GFP" should be given, such as the meaning of 5'GFp and 3'gFP.

We thank the reviewer for this suggestion. We have now edited the following two sentences in the text to clarify our usage of uppercase and lowercase letters

- To establish a GFP reporter for L1 retrotransposition-mediated rearrangements, we first modified the GFP-AI reporter in the 3' UTR of L1 such that retrotransposition would generate insertions containing an incomplete GFP sequence (***5' GFP fragment***, or hereinafter referred to as ***5' GFp*** where the lowercase p represents the deleted sequence) (Fig 1D).

-Using the FRT/Flp recombinase system in U2OS cells[15], we generated a cell line containing a chromosomal copy of the remaining ***3' GFP sequence*** (or hereinafter referred to as ***3' gFP*** where the lowercase represents the deleted sequence) (Fig 1D, SFig 2).

2. Move critical details now in Methods into the relevant figure legends. For example, the normalization of GFP

frequencies to a co-transfected GFP plasmid, the n for each condition, and the gating strategy. You already report n and stats in captions; adding the normalization detail where the frequencies first appear would make the results more self-contained.

We thank the reviewer for this suggestion. We have made two edits to the figure legends.
We added the following description from the **Methods** in **Figure 1C** legend to explain our GFP normalization strategy for all GFP-based reporter assays in our study:

"GFP normalization for all reporter assays: The percentage of GFP+ cells induced by the reporter assays were normalized relative to the percentage of GFP+ cells from cells transfected with the GFP plasmid."

We also added the following sentence from the **Methods** in the captions of **Figure 1E** and **Figure 4C** to describe our gating strategy for flow cytometry analysis of GFP+ cells: *"Flow cytometry analysis and GFP normalization:*

Singlets were gated on side-scatter versus forward scatter and GFP+ cells were gated on GFP versus autofluorescence (PE). Cells with autofluorescence were detected on a diagonal line, whereas cells showing increased green fluorescence (GFP+) were gated above the autofluorescence diagonal line. The percentage of GFP+ cells from the reporter assays were normalized relative to the percentage of GFP+ cells from cells transfected with the GFP plasmid."

Reviewer #3 (Remarks to the Author):

In their manuscript "L1 insertion intermediates recombine with one another or with DNA breaks to form genome rearrangements", Mendez-Dorantes and colleagues use modified L1 reporter constructs to examine the potential spectrum of L1-mediated chromosomal rearrangements and gain mechanistic insight regarding their formation. The work presented complements recent pre-prints from their own research group and the Tubio lab which also present evidence for L1-mediated rearrangements; the former using traditional L1 reporters, and the later using long-read whole genome sequencing of human tumours. The present study uses an ingenious split GFP reporter system to detect structural rearrangements arising from the recombination of "in-progress" L1 integrations arising by target-primed reverse transcription (TPRT) with a stably integrated GFP sequence in a known chromosomal location, or by the recombination of two synchronous TPRT events. By design, this impressive experimental system identifies homology-driven structural rearrangements, in this case mediated by the "split" GFP sequence; presumably, similar endogenous L1-mediated rearrangements would be mediated by homology in the native L1 sequence. Overall, the results are intriguing and complement parallel emerging insights in this area, building on a decades-spanning body of work demonstrating the large-scale genomic impact of L1 retrotransposition. However, there are a few areas where I think this experimental system could be exploited more thoughtfully or pushed further to provide more solid evidence and additional mechanistic insight on how these rearrangements likely come about.

We thank the reviewer for acknowledging the novelty and utility of our experimental system to study chromosomal rearrangements mediated by L1 insertion intermediates. As described below, we have engaged with the reviewer's suggestions and addressed their concerns with additional experiments, new analyses, and extended Discussion on the implications of our work in cancer genomes and the limitations.

1. Cas9-mediated targeted ONT sequencing of L1 insertions arising from recombination between the fixed Chr11 "FP" sequence and the reporter-delivered "GFP" sequence is an excellent strategy to characterise structural rearrangements resulting in GFP-positive cells. However, I do not understand why the guide RNA directing adaptor ligation was targeted to the GFP sequence. This design results in ONT long-reads encompassing the 3' flank (chr11) -or- the 5' flank (chromosome targeted by transfected "GF" L1 reporter). The entire rearrangement-spanning structure is never contained within a single read, and the evidence for L1-mediated rearrangement is limited to sequencing confirmation of the reconstituted GFP. Targeting the chr11 genomic sequence perhaps a few hundred bp downstream of the fixed Chr11 "FP" integrant would result in a large number of identical ONT reads extending away from the recombination site but also would generate ONT reads beginning downstream of the Chr11 recombination site, encompassing the entire reconstituted GFP cassette, and extending into the genomic sequence comprising the other participant in the recombination

event. A major advantage of characterising L1 insertions via long-read sequencing is the ability to span entire L1 integrants including their structural hallmarks and associated genomic perturbations in a single read representative of a single DNA molecule. The above refinement of the authors' targeted ONT sequence strategy would take full advantage of this experimental platform.

We agree with the reviewer for recognizing that the Cas9-mediated targeted ONT sequencing approach is an excellent strategy to characterize the structural rearrangements resulting from our split GFP recombination reporter between L1 cDNA intermediates and chromosomal breaks.

We would like to first clarify our strategy for how we applied Cas9-mediated targeted ONT sequencing to the GFP sequence in GFP+ U2OS cells. We specifically targeted ONT sequencing at the 3' end of reconstituted GFP within the 3'GFP sequence (3' gFP sequence) that integrated stably in the FTR locus (chromosome 11 (chr11:7784) in U2OS prior to GFP reconstitution. This location is ~350bp away from the Cas9-mediated chromosomal break induced inside cells where the recombination junction occurred. Hence, the targeted ONT long-reads mapping to the 5' end of GFP will 1) span the GFP rearrangement junction between L1 cDNA products containing the 5'GFp sequence and the chromosomal breaks induced inside cells containing the 3'gFP sequence, 2) encompass the overlapping homologous sequences, 3) include sequences from both recombining substrates, and 4) extend to distinct chromosomal locations wherever the retrotransposition reaction had initiated (SFig 5A). Moreover, all the ONT long-reads mapping to the 3' end of the GFP map to chromosome 11 (chr11:7784) where the FRT is located (Fig 2). Hence, our data in Figure 2 validates that reconstitution of GFP results from the formation of genome rearrangements via recombination of L1 cDNA products and chromosomal breaks.

We thank the reviewer for the suggestion to perform Cas9-targeted ONT sequencing at the chr11:7784 versus at the 3' end of the GFP sequence at the FRT locus. This would be an inefficient approach since there is a complete LacZ gene (~3.5 kb) integrated between the 3' GFp sequence and the chr11:7784, which places our desired target sequences far from the targeted ONT sequencing. However, we have validated the structure of the rearrangements of interest using PCR analysis spanning the rearrangement junctions (including the complete GFP sequence), e.g., validating a translocation rearrangement and an inversion rearrangement (shown in Fig 2E), and also by metaphase FISH showing an additional fragment of Chr11 in GFP+ cells compared to the parental U2OS reporter cells confirming the involving of Chr11 in genome rearrangements (SFig 6). In addition, we validated GFP recombination events in another cell line by performing ONT whole-genome sequencing analysis of sorted GFP+ cells induced in 150bpHom-3' gFP HEK293 cells transfected with the plasmids expressing WT L1-5'GFp and Cas9/sg150bp (SFig 7A-B, Table 2). Through this analysis, we validated that that GFP recombination events represent chromosomal rearrangements between L1 cDNA intermediates and a Cas9-mediated DSB in a separate cell line (SFig 7D-E). Altogether, spanning ONT reads and several orthogonal pieces of data support our models of L1 retrotransposition-mediated chromosomal rearrangements capable of being read-out with this reporter.

2. The dependence of GFP expression on the introduction of a Cas9-mediated DNA break at the "FP" chromosome 11 integrant demonstrates nicely that both homology and a DNA break are required for recombination-driven L1-mediated chromosomal rearrangements. This finding is reinforced in the WGS analysis which identified rearrangements putatively arising from the recombination of two "in progress" TPRT events. To further strengthen the conclusion that translocations arise from synchronous TPRT events, rather than sequentially, it would be interesting to stagger the transfection of the "GF" and "FP reporters, allowing retrotransposition events from the first reporter to become fully integrated before the second reporter is introduced.

We thank the reviewer for this suggestion. In our initial efforts to characterize this experimental system we tested something similar, and in our original submission, we showed that sequential transfection of the L1-5'GFp reporter plasmid and the Cas9/150bp expressing plasmid into 150bp-Hom-3'gFP U20S cells failed to induce GFP+ cells (SFig 4). In contrast, we showed that co-transfection both plasmids the L1-5'GFp reporter plasmid and the Cas9/150bp expressing plasmid into 150bp-Hom-3'gFP U20S cells robustly induced GFP+ cells (SFig 4). These results support our conclusions that both 1) ongoing retrotransposition (cDNA production) and 2) active chromosomal breakage are concurrently required to elicit GFP recombination events. This result raised the question for the reviewer of whether interactions between two L1 cDNAs similarly require them to be

in intermediate states contemporaneously (as opposed to one fully integrated). In other words, is *co*-transfection of our second split GFP reporter plasmids (L1-5'GFp and our L1-3'gFP (150bp Hom)) needed to model recombination between L1 cDNA products and induce GFP+ cells. To address this, we tested whether we could induce GFP+ cells when we sequentially transfected parental U2OS cells with our split GFP reporter plasmids (L1-5'GFp plasmid or L1-3'gFP (150 Hom) plasmid) modeling the recombination of L1 cDNA products (See **SFig 10B**). For this, we transfected U20S cells first with either L1-5'GFp plasmid or L1-3'gFP (150 Hom) plasmid, expanded transfected cells for a week to allow plasmid dilution, and then re-transfected each of these previously transfected cells with the either plasmid (**SFig 10B**). From this experiment, we failed to detect an induction of GFP+ cells. In contrast, we found that only simultaneous co-transfection of the split GFP reporter plasmids (L1-5'GFp plasmid and L1-3'gFP (150 Hom) plasmid) induce GFP+ cells. This experiment provides compelling evidence that two "in progress" TPRT events are required to elicit GFP recombination events and supports inferences of these types of interactions based on analyses of cancer genomes[2].

3. The use of robust cultured cell lines presents a key opportunity to generate clonal cell lines from GFP positive cells putatively harbouring L1-mediated chromosomal rearrangements. While some rearrangements clearly are unstable and will be lost during clonal expansion, I expect some could persist. The split reporter assay could be redesigned using a selectable (e.g. neomycin) rather than screenable marker, to enforce retention of the L1-mediated translocation. The generation of clonal cell lines would allow more robust validation of chromosomal translocations and other L1-mediated structural rearrangements. In revision, I request that the authors generate clonal cell lines, identify L1-mediated translocations, and formally demonstrate these rearrangements using (ideally) a karyotyping/chromosome painting approach.

As highlighted above, we have validated chromosomal rearrangements mediated by L1 retrotransposition and reported by our GFP recombination assays using PCR validation of the rearrangements (See Fig 2E, Fig 5C), as well as via metaphase spread analysis of pools of GFP+ cells showing an additional fragment of chromosome 11 confirming involvement of chr11 in rearrangements between L1 cDNA and a dsDNA break in chr11(chr11:7784) in U2OS FRT cells (See SFig 6). In addition, we validated GFP recombination events in another cell line (HEK293) by performing ONT whole-genome sequencing analysis of sorted GFP+ cells induced in 150bpHom-3' gFP cells transfected with the plasmids expressing WT L1-5'GFp and Cas9/sg150bp (SFig 7A-B, Table 2). Through this analysis, we validated that GFP reconstitutions represent chromosomal rearrangements between L1 cDNA intermediates and a Cas9-mediated DSB (SFig 7D-E). Moreover, we performed an unbiased analysis of ONT WGS sequencing data for rearrangements mediated by L1 retrotransposition but that failed to reconstitute GFP expression (now included in our revised manuscript (SFig 11, SFig 7). It appears that these rearrangements are mediated by joining of L1 cDNA products to distant DNA breaks of unknown etiology. We report chromosomal rearrangements containing microhomology (n=2) or 8-9 bp non-templated insertions (n=2) at the junctions between L1 cDNA products and distant break sites, suggesting that limited or no microhomology was exploited in the formation of these rearrangements. We also report two chromosomal rearrangements that did not restore GFP but were mediated by the recombination of L1 cDNA intermediates and endogenous full-length copies of L1 (L1PA3) (SFig 11). These examples implicate extensive homologies are involved in the formation of these recombination events. All together, we count 5 rearrangements that reconstituted GFP expression using extensive homology, one rearrangement that failed to reconstitute GFP expression but was mediated by L1 retrotransposition, and 17 canonical L1 insertions in our HEK293 GFP+ cells by profiling cells using ONT WGS (SFig 7, Table 2). In the ONT WGS of U2OS GFP+ cells where GFP depends on recombination of two L1 cDNA products with each other, we report a total of 5 chromosomal rearrangements that reconstituted GFP expressions and 6 chromosomal rearrangements that failed to reconstitute GFP expression, as well as 75 canonical L1 insertions. In summary, we provide comprehensive validation of chromosomal rearrangements mediated by L1 retrotransposition detected with our split GFP reporter assays in the revised manuscript.

Lastly, in a separate manuscript, we profiled all types of genomic alterations caused by L1 retrotransposition by sequencing RPE-1 hTERT immortalized single cell clones after L1 expression[1]. RPE-1 are stable nearly diploid cells. In this work, we found on average 10 canonical insertions and two chromosomal rearrangements mediated by L1 retrotransposition per clone. We add citations to this other work in response to the reviewer's request to sequence rearrangements independent of our GFP recombination reporters, and focus on these reporter systems of the current manuscript.

1 Mendez-Dorantes, C. *et al.* Chromosomal rearrangements and instability caused by the LINE-1 retrotransposon. *bioRxiv*, 2024.2012.2014.628481 (2024). https://doi.org/10.1101/2024.12.14.628481

2 Zumalave, S. *et al.* Concurrent L1 retrotransposition events promote reciprocal translocations in human tumorigenesis. *Science (New York, N.Y.)* **392**, eaee4513 (2026). https://doi.org/10.1126/science.aee4513

3 Sugawara, N., Goldfarb, T., Studamire, B., Alani, E. & Haber, J. E. Heteroduplex rejection during single-strand annealing requires Sgs1 helicase and mismatch repair proteins Msh2 and Msh6 but not Pms1. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 9315-9320 (2004). https://doi.org/10.1073/pnas.0305749101

4 Anantha, R. W. *et al.* Functional and mutational landscapes of BRCA1 for homology-directed repair and therapy resistance. *eLife* **6** (2017). https://doi.org/10.7554/eLife.21350

5 Ardeljan, D. *et al.* Cell fitness screens reveal a conflict between LINE-1 retrotransposition and DNA replication. *Nature structural & molecular biology* **27**, 168-178 (2020). https://doi.org/10.1038/s41594-020-0372-1

6 Mendez-Dorantes, C. & Burns, K. H. LINE-1 retrotransposition and its deregulation in cancers: implications for therapeutic opportunities. *Genes & development* **37**, 948-967 (2023). https://doi.org/10.1101/gad.351051.123

7 Hoyt, S. J. *et al.* From telomere to telomere: The transcriptional and epigenetic state of human repeat elements. *Science (New York, N.Y.)* **376**, eabk3112 (2022). https://doi.org/10.1126/science.abk3112

8 Rodriguez-Martin, B. *et al.* Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nature genetics* **52**, 306-319 (2020). https://doi.org/10.1038/s41588-019-0562-0

9 Symer, D. E. *et al.* Human l1 retrotransposition is associated with genetic instability in vivo. *Cell* **110**, 327-338 (2002). https://doi.org/10.1016/s0092-8674(02)00839-5

10 Gilbert, N., Lutz-Prigge, S. & Moran, J. V. Genomic deletions created upon LINE-1 retrotransposition. *Cell* **110**, 315-325 (2002). https://doi.org/10.1016/s0092-8674(02)00828-0

11 Beck, C. R., Garcia-Perez, J. L., Badge, R. M. & Moran, J. V. LINE-1 elements in structural variation and disease. *Annual review of genomics and human genetics* **12**, 187-215 (2011). https://doi.org/10.1146/annurev-genom-082509-141802

12 Tubio, J. M. C. *et al.* Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science (New York, N.Y.)* **345**, 1251343 (2014). https://doi.org/10.1126/science.1251343

13 Moran, J. V., DeBerardinis, R. J. & Kazazian, H. H., Jr. Exon shuffling by L1 retrotransposition. *Science (New York, N.Y.)* **283**, 1530-1534 (1999). https://doi.org/10.1126/science.283.5407.1530

14 Goodier, J. L., Ostertag, E. M. & Kazazian, H. H., Jr. Transduction of 3'-flanking sequences is common in L1 retrotransposition. *Human molecular genetics* **9**, 653-657 (2000). https://doi.org/10.1093/hmg/9.4.653

15 Kelso, A. A., Lopezcolorado, F. W., Bhargava, R. & Stark, J. M. Distinct roles of RAD52 and POLQ in chromosomal break repair and replication stress response. *PLoS genetics* **15**, e1008319 (2019). https://doi.org/10.1371/journal.pgen.1008319