EMBO reports

CRISPR-Enabled Autonomous Transposable Element (CREATE) for RNA-based gene editing and delivery

Yuxiao Wang, Ruei-Zeng Lin, Meghan Harris, Bianca Lavayen, Neha Diwanji, Bruce McCreedy, Robert Hofmeister, and Daniel Getts

Corresponding authors: Yuxiao Wang (ywang@myeloidtx.com) , Daniel Getts (daniel@myeloidtx.com)

Review Timeline:

Submission Date:	12th Jul 24
Editorial Decision:	28th Aug 24
Revision Received:	28th Nov 24
Editorial Decision:	10th Dec 24
Revision Received:	14th Dec 24
Accepted:	19th Dec 24

Editor: Esther Schnapp

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. Depending on transfer agreements, referee reports obtained elsewhere may or may not be included in this compilation. Referee reports are anonymous unless the Referee chooses to sign their reports.)

Dear Dr. Wang,

Thank you for the submission of your manuscript to EMBO reports. We have now received the full set of referee reports that is pasted below.

As you will see, the referees acknowledge that the findings are potentially interesting. However, they also have several suggestions for how the study should be improved, and I think all points are good and should be addressed. Please let me know in case you disagree, and we can discuss this further, also in a video chat, if you like.

I would thus like to invite you to revise your manuscript with the understanding that the referee concerns must be fully addressed and their suggestions taken on board. Please address all referee concerns in a complete point-by-point response. Acceptance of the manuscript will depend on a positive outcome of a second round of review. It is EMBO reports policy to allow a single round of major revision only and acceptance or rejection of the manuscript will therefore depend on the completeness of your responses included in the next, final version of the manuscript.

We realize that it is difficult to revise to a specific deadline. In the interest of protecting the conceptual advance provided by the work, we recommend a revision within 3 months (28th Nov 2024). Please discuss the revision progress ahead of this time with the editor if you require more time to complete the revisions.

You can either publish the study as a short report or as a full article. For short reports, the revised manuscript should not exceed 27,000 characters (including spaces but excluding materials & methods and references) and 5 main plus 5 expanded view figures. The results and discussion sections must further be combined, which will help to shorten the manuscript text by eliminating some redundancy that is inevitable when discussing the same experiments twice. For a normal article there are no length limitations, but it should have more than 5 main figures and the results and discussion sections must be separate. In both cases, the entire materials and methods must be included in the main manuscript file.

IMPORTANT NOTE: we perform an initial quality control of all revised manuscripts before re-review. Your manuscript will FAIL this control and the handling will be DELAYED if the following APPLIES:
1) A data availability section providing access to data deposited in public databases is missing. If you have not deposited any data, please add a sentence to the data availability section that explains that.
2) Your manuscript contains statistics and error bars based on n=2. Please use scatter blots in these cases. No statistics should be calculated if n=2.

When submitting your revised manuscript, please carefully review the instructions that follow below. Failure to include requested items will delay the evaluation of your revision.

1) a .docx formatted version of the manuscript text (including legends for main figures, EV figures and tables). Please make sure that the changes are highlighted to be clearly visible.

2) individual production quality figure files as .eps, .tif, .jpg (one file per figure). See https://wol-prod-cdn.literatumonline.com/pb-assets/embo-site/EMBOPress_Figure_Guidelines_061115-1561436025777.pdf for more info on how to prepare your figures.

3) We replaced Supplementary Information with Expanded View (EV) Figures and Tables that are collapsible/expandable online. A maximum of 5 EV Figures can be typeset. EV Figures should be cited as 'Figure EV1, Figure EV2" etc... in the text and their respective legends should be included in the main text after the legends of regular figures.

- For the figures that you do NOT wish to display as Expanded View figures, they should be bundled together with their legends in a single PDF file called *Appendix*, which should start with a short Table of Content. Appendix figures should be referred to in the main text as: "Appendix Figure S1, Appendix Figure S2" etc. See detailed instructions regarding expanded view here:

- Additional Tables/Datasets should be labeled and referred to as Table EV1, Dataset EV1, etc. Legends have to be provided in a separate tab in case of .xls files. Alternatively, the legend can be supplied as a separate text file (README) and zipped together with the Table/Dataset file.

4) a .docx formatted letter INCLUDING the reviewers' reports and your detailed point-by-point responses to their comments. As part of the EMBO Press transparent editorial process, the point-by-point response is part of the Review Process File (RPF), which will be published alongside your paper.

5) a complete author checklist, which you can download from our author guidelines . Please insert information in the checklist that is also reflected in the manuscript. The completed author checklist will also be part of the RPF.

6) Please note that all corresponding authors are required to supply an ORCID ID for their name upon submission of a revised

manuscript (). Please find instructions on how to link your ORCID ID to your account in our manuscript tracking system in our Author guidelines

7) Before submitting your revision, primary datasets produced in this study need to be deposited in an appropriate public database (see https://www.embopress.org/page/journal/14693178/authorguide#datadeposition). Please remember to provide a reviewer password if the datasets are not yet public. The accession numbers and database should be listed in a formal "Data Availability" section placed after Materials & Method (see also https://www.embopress.org/page/journal/14693178/authorguide#datadeposition). Please note that the Data Availability Section is restricted to new primary data that are part of this study. * Note - All links should resolve to a page where the data can be accessed. *
If your study has not produced novel datasets, please mention this fact in the Data Availability Section.

8) At EMBO Press we ask authors to provide source data for the main manuscript figures. Our source data coordinator will contact you to discuss which figure panels we would need source data for and will also provide you with helpful tips on how to upload and organize the files.

9) Our journal also encourages inclusion of *data citations in the reference list* to directly cite datasets that were re-used and obtained from public databases. Data citations in the article text are distinct from normal bibliographical citations and should directly link to the database records from which the data can be accessed. In the main text, data citations are formatted as follows: "Data ref: Smith et al, 2001" or "Data ref: NCBI Sequence Read Archive PRJNA342805, 2017". In the Reference list, data citations must be labeled with "[DATASET]". A data reference must provide the database name, accession number/identifiers and a resolvable link to the landing page from which the data can be accessed at the end of the reference. Further instructions are available at https://www.embopress.org/page/journal/14693178/authorguide#referencesformat

10) Regarding data quantification (see Figure Legends: https://www.embopress.org/page/journal/14693178/authorguide#figureformat)

The following points must be specified in each figure legend:

- the name of the statistical test used to generate error bars and P values,

- the number (n) of independent experiments (please specify technical or biological replicates) underlying each data point,

- the nature of the bars and error bars (s.d., s.e.m.),

- If the data are obtained from n {less than or equal to} 2, use scatter blots showing the individual data points.

Discussion of statistical methodology can be reported in the materials and methods section, but figure legends should contain a basic description of n, P and the test applied.

- Please also include scale bars in all microscopy images.

11) The journal requires a statement specifying whether or not authors have competing interests (defined as all potential or actual interests that could be perceived to influence the presentation or interpretation of an article). In case of competing interests, this must be specified in your disclosure statement. Further information: https://www.embopress.org/competing-interests

12) All Materials and Methods need to be described in the main text using our 'Structured Methods' format, which is required for all research articles. According to this format, the Methods section includes a Reagents and Tools Table (listing key reagents, experimental models, software and relevant equipment and including their sources and relevant identifiers) followed by a Methods and Protocols section describing the methods using a step-by-step protocol format. The aim is to facilitate adoption of the methodologies across labs. More information on how to adhere to this format as well as a downloadable template (.docx) for the Reagents and Tools Table can be found in our author guidelines: https://www.embopress.org/page/journal/14693178/authorguide#structuredmethods.

An example of a Method paper with Structured Methods can be found here: https://www.embopress.org/doi/full/10.1038/s44320-024-00037-6#sec-4.

We would also welcome the submission of cover suggestions, or motifs to be used by our Graphics Illustrator in designing a cover.

As part of the EMBO publication's Transparent Editorial Process, EMBO reports publishes online a Review Process File (RPF) to accompany accepted manuscripts. This File will be published in conjunction with your paper and will include the referee

reports, your point-by-point response and all pertinent correspondence relating to the manuscript.

You are able to opt out of this by letting the editorial office know (emboreports@embo.org). If you do opt out, the Review Process File link will point to the following statement: "No Review Process File is available with this article, as the authors have chosen not to make the review process public in this case."

I look forward to seeing a revised form of your manuscript when it is ready. Please use this link to submit your revision: https://embor.msubmit.net/cgi-bin/main.plex

Best regards,
Esther

Esther Schnapp, PhD
Senior Editor
EMBO reports

Referee #1:

Wang and Lin et al. describe a new genome editing tool called CREATE that harnesses a repurposed L1 system plus a Cas9-nickase to enable all RNA mediated insertion of large payloads. A system that enables all RNA encoded large payload insertions remains one of the more exciting areas of therapeutic genome editing as the delivery of DNA donors needed for other gene insertion technologies such as PASTE, LSR recombinases and PASSIGE remains a challenge. As such CREATE should generate substantial interest in the genome editing community. CREATE seems to be highly specific for insertions at the targeted site but is not highly efficient with insertions in the low single digit range. The authors show CREATE works across 4 target sites and 3 cell types including primary T cells. The authors should be commended for doing a series of experiments with different mutations and truncations to the PBS, ORF1p, ORF2p or Cas9 nickase to demonstrate the proposed mechanism of CREATE insertions is supported. Overall CREATE is a very exciting technology and this should open the door for additional technology development or enable a variety of new research possibilities in biology or therapeutic approaches in medicine.

Minor points
- Please elaborate on any potential caveats based on the design of the hybridization probes used for NGS evaluation of CREATE specificity in Figure 1D. Will this approach only detect full length insertions or can this approach detect partial or abortive insertions of the cargo sequence?
- Please elaborate on the NGS amplicon sequencing in Figure 3. The authors comment on indel formation versus insertions at PBS1 and PBS2 but are all the insertions full length or do you see partial insertions where CREATE primes at PBS1 and initiates reverse transcription but the full cargo is not integrated to any reason?
- What locus was edited in Figure 4c?

Referee #2:

Review of EMBOR-2024-59975V1

CRISPR-Enabled Autonomous Transposable 1 Element (CREATE) for RNA-based gene editing and delivery

Authors: Yuxiao Wang1*†, Ruei-Zeng Lin1†, Meghan Harris1, Bianca Lavayen1, Neha Diwanji1, Bruce McCreedy, Robert Hofmeister1 and Daniel Getts1*

In this manuscript, the authors have devised an RNA-based gene delivery tool that utilizes an endonuclease mutant version (D205A) of the human Long INterspersed Element-1 (LINE-1 or L1) retrotransposon together with a Cas9 (H840A) nickase and two sgRNAs to deliver an GFP reporter gene into the AAVS1 site in HEK293T cells and several other human cell lines.

A previous study by Tao and colleagues (Tao et al., Nature Comm., 2022) showed that human L1 can mediate insertions into Cas9 mediated double strand breaks. They also showed that L1 can integrate into single strand breaks created by Cas9 nickase albeit very inefficiently. The authors of the current study appear to have increased the efficiency of L1 insertions into single strand nicks created by the Cas9 nickase, as well as increased specificity of L1 insertions into nicks using their CREATE system.

The CREATE RNA is essentially a modified L1 retrotransposition construct (Moran et al., Cell, 1996) that has been modified with a GFP reporter gene (Ostertag et al., NAR, 2000) in the 3' UTR (GFP lacks the intron used in retrotransposition assays) that is flanked by unique sequences called primer binding sites (PBS) that correspond to target sites within the AAVS1 locus that are 90 bp apart. The CREATE RNA is transfected with a nickase Cas9 mutant (H840A) that cleaves a single strand of genomic DNA

opposite the target guide RNA and two sgRNAs that target the genomic PBS site sequences within the AAVS1 locus. The innovation over the Tao et al., study is the use of the PBS sequences which create two nicks that are 90 bps apart in opposite strands of the DNA at the target AAVS2 locus and the inclusion of PBS sites in the CREATE RNA that flank the GFP reporter which are hypothesized to facilitate the integration of the GFP reporter gene between the genomic PBS sites in the AAVS1 locus.

Overall, the manuscript is logically presented, the methods are sufficiently detailed, and the data in general supports the conclusions. However, there are several issues that should be addressed, which are outlined below:

1) The authors of this study have utilized a Cas9 nickase (H840A) to create a pair of single strand DNA lesions on opposite strands of DNA at the AAVS1 locus. They also utilize an endonuclease mutant L1 ORF2 (D205A), which mediates insertion of the GFP reporter gene into the DNA lesions created by the Cas9 nickase. The authors conclude that the insertion is mediated via "natural" L1 TPRT. Although their data suggests that EN-deficient L1 ORF2 (D205A) is required to mediate the insertions after Cas9 nicking, it seems more likely that the insertion is mediated by the Endonuclease independent retrotransposition (ENi) pathway, which is an important aspect of L1 biology that is likely relevant to their study. Wild type or canonical TPRT involves the sequential nicking of the target DNA by the L1 EN followed by reverse transcription. In a 2002 study by Morrish and colleagues (Morrish et al., Nature Genetics, 2002), the authors characterized an endonuclease-independent L1 retrotransposition pathway (ENi) in which they showed that endonuclease deficient ORF2p (i.e., ORF2 D205A and H230A) L1 mutants could integrate into cells that contained DNA lesions. Later studies expanded on these findings demonstrating the ENi could occur at mammalian Telomeres (Morrish et al., Nature, 2007). Also relevant to the current study is a recent study (Tao et al., Nature Comm., 2022) that demonstrated that both wild type and EN mutant L1s can integrate into Cas9 lesions. It would be a welcome addition to the manuscript if the authors would reference these works and re-evaluate and discuss their findings in light of these studies.

2) The diagram in figure 1a is a bit too small and somewhat difficult to follow. It would be helpful if the authors could:
a. Enlarge the insertion model diagram
b. Indicate the location of the PBS target sites, PAM sites and cleavage sites of the Cas9 nickase
c. Label the 5' and 3' ends of all DNA/RNA strands including sgRNAs in all steps of diagram

3) The authors claim that the genomic sequence at the AAVS1 locus between the PBS sites is replaced (deleted) by integration of the GFP reporter gene. How do the authors know that the sequence between the PBS sites is deleted and how do they think the sequence might be deleted?

4) Related to the above question, the authors confirm the integration of the GFP gene into the AAVS1 locus on chromosome 19 using target hybridization and next gen sequencing. However, this data seems only to provide information regarding the general accuracy of the insertion into the AAVS1 locus on ch19 (Fig 1d). The authors need to fully characterize some of the insertions to confirm junction sites and verify whether the intervening sequence between the PBS sites has been deleted. Also, characterizing full insertions will help determine whether insertions occur via canonical TPRT or variations thereof (i.e., Endonuclease Independent L1 retrotransposition (ENi)).

5) In figure 3, the authors should provide some examples of indels and fully characterize at least some of the insertions.

6) LINE 89: "Natural L1 retrotransposition relies on nicking of the non-targeting DNA strand 90 by the EN domain of ORF2p". I am not sure what the authors mean by "natural" L1 retrotransposition. In canonical TPRT, the wild type L1 ORF2p EN cleaves a single strand of genomic DNA at a degenerate consensus sequence of 5'-TTTT/A-3' where the "/" indicates the EN cleavage site to expose a 3'-OH on the nicked strand of DNA. My understanding is that in the author's assay in which they utilize the ORF2p D205A mutant that lacks EN activity, the Cas9 nickase nicks a single strand of the DNA, thereby replacing the ORF2p EN activity.

7) LINEs 54-55: "transcription (TPRT) process mediated by the RT domain of ORF2p that converts the L1 mRNA into cDNA." For the TPRT mechanism, please cite: Luan et al, 1993 (PMID: 7679954)

8) LINE 99: Delete or modify the statement: "safe non-replicative gene insertion" as the data does not support this assertion.

9) Abstract LINE 15: "LINE1" should be written as "LINE-1". LINE-1 is an abbreviation for Long INterspersed Element-1.

10) LINE 182: "These results are with a recent study investigating the poly-A length required for efficient TPRT by L1 ORF2" Can the authors please explain how the results obtained through modulating the PBS lengths relate to poly-A tail length.

Referee #3:

* General comments:

Targeted integration of a large gene is definitely important in the genome editing fiend. To date, although various tools including primg editing combined with recombinase (PASSIGE or PASTE), CRISPR-associated transposon (CAST), Bridge RNA technology, retrotransposon system have been developed, it is still unsatisfactory to conduct targeted gene insertion with high efficiency in human cells. In the present study, Wang et al. demonstrat a new type of targeted integration strategy using human LINE1 (L1) retrotransposon element combined with CRISPR-Cas9 system, named CRISPR-Enabled Autonomous Transposable Element (CREATE). It is very smart to inactivate endonuclease domain of ORF2 (modORF2) to reduce genome-wide off-target integration. Although average integration efficincies were not high enough (approximately 1%) in human cells, it is valuable to open this method to the field. I would like to raise a few issues to strengthen this study.

* Specific comments:
1. As dipicted in Figure 1a, the necessity of "ORF1p" is vague. The original role of ORF1p is an RNA-binding protein that interacts with L1 mRNA transcripts. However, in the present study, the L1 mRNA is provided outside the cells by the authors. I assume that ORF1p might not be necessary for the CREATE activity. If successful without the ORF1p, it would be benefitial in terms of shorter length of L1 mRNA.

2. The CREATE requires double-nicking system by nCas9, instead of endonuclease of ORF2, which unfortunely results in unwanted indel mutations at the target site. Because it is not unvoidable, I should raise a fundamental question, "Is the CREATE a tool without causing a DNA double-stand break (DSB)?" Although the CREATE does not generate DSBs directly, double-nicking is typically accompanied with DSB.

3. In line 12, "double-stand breaks" -> "double-strand breaks"

Response to review of manuscript EMBOR-2024-59975V1

We sincerely thank all referees for their thorough review and insightful comments. Below is a point-by-point response. Blue colored text are the original comments.

Referee #1:

Wang and Lin et al. describe a new genome editing tool called CREATE that harnesses a repurposed L1 system plus a Cas9-nickase to enable all RNA mediated insertion of large payloads. A system that enables all RNA encoded large payload insertions remains one of the more exciting areas of therapeutic genome editing as the delivery of DNA donors needed for other gene insertion technologies such as PASTE, LSR recombinases and PASSIGE remains a challenge. As such CREATE should generate substantial interest in the genome editing community. CREATE seems to be highly specific for insertions at the targeted site but is not highly efficient with insertions in the low single digit range. The authors show CREATE works across 4 target sites and 3 cell types including primary T cells. The authors should be commended for doing a series of experiments with different mutations and truncations to the PBS, ORF1p, ORF2p or Cas9 nickase to demonstrate the proposed mechanism of CREATE insertions is supported. Overall CREATE is a very exciting technology and this should open the door for additional technology development or enable a variety of new research possibilities in biology or therapeutic approaches in medicine.

Minor points

- Please elaborate on any potential caveats based on the design of the hybridization probes used for NGS evaluation of CREATE specificity in Figure 1D. Will this approach only detect full length insertions or can this approach detect partial or abortive insertions of the cargo sequence?

Response: we thank the reviewer for this important question. The hybridization probes used for NGS were 120 nt probes tiled across the entire length of the payload gene. During library preparation, genomic DNA was enzymatically fragmented and captured by the probes, followed by Illumina short-read sequencing (150 bp x 2). This approach has limitations in distinguishing full-length vs partial insertions. While it can identify the presence of insertions at specific genomic loci, the short read lengths and variable probe capture efficiency prevent accurate quantification of full-length vs partial integrations.

To address this limitation, we have now performed Nanopore long-read sequencing of the amplicon covering the entire length of the integrated payload. This allows us to sequence full-length insertions and characterize partial integrations. We have included this new analysis in Fig. 4E and Appendix Fig. S3 and S4. Our results show that over 70% of the confirmed integration alignment showed full-length or near full-length transgene (< 20 bp total indel length).

- Please elaborate on the NGS amplicon sequencing in Figure 3. The authors comment on indel formation versus insertions at PBS1 and PBS2 but are all the insertions full length or do you see partial insertions where CREATE primes at PBS1 and initiates reverse transcription but the full cargo is not integrated to any reason?

Response: see above. We have performed Nanopore long read sequencing which addresses this question.

- What locus was edited in Figure 4c?

Response: we have clarified in Figure 4c that AAVS1 locus was targeted for integration of payload.

Referee #2:

In this manuscript, the authors have devised an RNA-based gene delivery tool that utilizes an endonuclease mutant version (D205A) of the human Long INterspersed Element-1 (LINE-1 or L1) retrotransposon together with a Cas9 (H840A) nickase and two sgRNAs to deliver an GFP reporter gene into the AAVS1 site in HEK293T cells and several other human cell lines.

A previous study by Tao and colleagues (Tao et al., Nature Comm., 2022) showed that human L1 can mediate insertions into Cas9 mediated double strand breaks. They also showed that L1 can integrate into single strand breaks created by Cas9 nickase albeit very inefficiently. The authors of the current study appear to have increased the efficiency of L1 insertions into single strand nicks created by the Cas9 nickase, as well as increased specificity of L1 insertions into nicks using their CREATE system.

The CREATE RNA is essentially a modified L1 retrotransposition construct (Moran et al., Cell, 1996) that has been modified with a GFP reporter gene (Ostertag et al., NAR, 2000) in the 3' UTR (GFP lacks the intron used in retrotransposition assays) that is flanked by unique sequences called primer binding sites (PBS) that correspond to target sites within the AAVS1 locus that are 90 bp

apart. The CREATE RNA is transfected with a nickase Cas9 mutant (H840A) that cleaves a single strand of genomic DNA opposite the target guide RNA and two sgRNAs that target the genomic PBS site sequences within the AAVS1 locus. The innovation over the Tao et al., study is the use of the PBS sequences which create two nicks that are 90 bps apart in opposite strands of the DNA at the target AAVS2 locus and the inclusion of PBS sites in the CREATE RNA that flank the GFP reporter which are hypothesized to facilitate the integration of the GFP reporter gene between the genomic PBS sites in the AAVS1 locus.

Overall, the manuscript is logically presented, the methods are sufficiently detailed, and the data in general supports the conclusions. However, there are several issues that should be addressed, which are outlined below:

1) The authors of this study have utilized a Cas9 nickase (H840A) to create a pair of single strand DNA lesions on opposite strands of DNA at the AAVS1 locus. They also utilize an endonuclease mutant L1 ORF2 (D205A), which mediates insertion of the GFP reporter gene into the DNA lesions created by the Cas9 nickase. The authors conclude that the insertion is mediated via "natural" L1 TPRT. Although their data suggests that EN-deficient L1 ORF2 (D205A) is required to mediate the insertions after Cas9 nicking, it seems more likely that the insertion is mediated by the Endonuclease independent retrotransposition (ENi) pathway, which is an important aspect of L1 biology that is likely relevant to their study. Wild type or canonical TPRT involves the sequential nicking of the target DNA by the L1 EN followed by reverse transcription. In a 2002 study by Morrish and colleagues (Morrish et al., Nature Genetics, 2002), the authors characterized an endonuclease-independent L1 retrotransposition pathway (ENi) in which they showed that endonuclease deficient ORF2p (i.e., ORF2 D205A and H230A) L1 mutants could integrate into cells that contained DNA lesions. Later studies expanded on these findings demonstrating the ENi could occur at mammalian Telomeres (Morrish et al., Nature, 2007). Also relevant to the current study is a recent study (Tao et al., Nature Comm., 2022) that demonstrated that both wild type and EN mutant L1s can integrate into Cas9 lesions. It would be a welcome addition to the manuscript if the authors would reference these works and re-evaluate and discuss their findings in light of these studies.

Response: we thank the reviewer for highlighting the work on ENi retrotransposition and its potential relevance to our study. We agree that our initial use of the term "natural" L1 TPRT was

imprecise and potentially misleading. We have revised our description to more accurately reflect the CREATE mechanism, and provided a detailed step-by-step diagram showing the proposed mechanism (Fig. 1B). We have added a new section in the discussion that contextualizes CREATE within the broader landscape of L1 biology, including ENi and integration at Cas9-induced breaks. We now cite and discuss the works by Morrish et al. (2002, 2007) and Tao et al. (2022).

2) The diagram in figure 1a is a bit too small and somewhat difficult to follow. It would be helpful if the authors could:

a. Enlarge the insertion model diagram

b. Indicate the location of the PBS target sites, PAM sites and cleavage sites of the Cas9 nickase

c. Label the 5' and 3' ends of all DNA/RNA strands including sgRNAs in all steps of diagram

Response: we have made the requested changes and updated Fig. 1B.

3) The authors claim that the genomic sequence at the AAVS1 locus between the PBS sites is replaced (deleted) by integration of the GFP reporter gene. How do the authors know that the sequence between the PBS sites is deleted and how do they think the sequence might be deleted?

Response: we have added a detailed integration mechanism depiction in Fig. 1B and also clarified in the main text. Briefly, ORF2p synthesizes new DNA strands primed from PBS1 and PBS2 that anneal together and contains the payload; excision of the original, unedited DNA strands and ligation of the nicks would result in the replacement of the 90 bp sequences between the nicks with the payload (Fig. 1B)

4) Related to the above question, the authors confirm the integration of the GFP gene into the AAVS1 locus on chromosome 19 using target hybridization and next gen sequencing. However, this data seems only to provide information regarding the general accuracy of the insertion into the AAVS1 locus on ch19 (Fig 1d). The authors need to fully characterize some of the insertions to confirm junction sites and verify whether the intervening sequence between the PBS sites has been deleted. Also, characterizing full insertions will help determine whether insertions occur via canonical TPRT or variations thereof (i.e., Endonuclease Independent L1 retrotransposition (ENi)).

Response: to characterize the full insertions and analyze the junction site structure, we have performed further NGS analysis and Nanopore long-read sequencing. The results have been added

as Fig. 4E, EV3, EV4 and Appendix Figure S1-S4. We also discussed the potential insertion mechanism in light of this analysis.

5) In figure 3, the authors should provide some examples of indels and fully characterize at least some of the insertions.

Response: we have characterized the dominant indel patterns observed from amplicon sequencing. The analysis is added to the manuscript (Fig. EV3, EV4 and Appendix Fig. S1-S2)

6) LINE 89: "Natural L1 retrotransposition relies on nicking of the non-targeting DNA strand 90 by the EN domain of ORF2p". I am not sure what the authors mean by "natural" L1 retrotransposition. In canonical TPRT, the wild type L1 ORF2p EN cleaves a single strand of genomic DNA at a degenerate consensus sequence of 5'-TTTT/A-3' where the "/" indicates the EN cleavage site to expose a 3'-OH on the nicked strand of DNA. My understanding is that in the author's assay in which they utilize the ORF2p D205A mutant that lacks EN activity, the Cas9 nickase nicks a single strand of the DNA, thereby replacing the ORF2p EN activity.

Response: we thank the reviewer for the clarification. Indeed as the reviewer described, the CREATE relies on Cas9 to nicks the non-target strand to release a 3' single-strand DNA flap, which then serve as the primer for TPRT by hybridizing with the PBS1 site on CREATE mRNA, reminiscent of the mechanism of canonical TPRT. We have modified the text to clarify this point and provided a detailed step-by-step diagram showing the proposed mechanism (Fig. 1B).

7) LINEs 54-55: "transcription (TPRT) process mediated by the RT domain of ORF2p that converts the L1 mRNA into cDNA." For the TPRT mechanism, please cite: Luan et al, 1993 (PMID: 7679954)

Response: this citation is added.

8) LINE 99: Delete or modify the statement: "safe non-replicative gene insertion" as the data does not support this assertion.

Response: we have modified the sentence to remove this statement. The sentence now reads "The outcome of the editing cycle is the successful integration of the payload cassette between the PBS sites without insertion of retrotransposition competent L1 sequences."

9) Abstract LINE 15: "LINE1" should be written as "LINE-1". LINE-1 is an abbreviation for Long INterspersed Element-1.

Response: we have made the requested changes

10) LINE 182: "These results are with a recent study investigating the poly-A length required for efficient TPRT by L1 ORF2" Can the authors please explain how the results obtained through modulating the PBS lengths relate to poly-A tail length.

Response: we agree that our original statement was confusing and speculative. We have removed this sentence and instead focus on describing our empirical findings regarding optimal PBS length for CREATE efficiency. Below is a brief discussion regarding our original thoughts:

Thawani et al. investigated the optimal poly(A) length of mRNA template for ORF2p to initiate TPRT, and found that a poly(A) length of 20-50 nucleotides was optimal. This range closely aligns with our observed optimal PBS length of 20-30 nucleotides for CREATE efficiency (Thawani *et al*, 2023) . We speculated that the similarity in optimal length between the poly(A) tail for efficient TPRT by ORF2p and the PBS for CREATE is not coincidental. In CREATE, the PBS sequence may functionally mimic the role of the poly(A) tail by hybridizing with target DNA sequence and engage ORF2p for TPRT initiation.

Referee #3:

* General comments:

Targeted integration of a large gene is definitely important in the genome editing fiend. To date, although various tools including primg editing combined with recombinase (PASSIGE or PASTE), CRISPR-associated transposon (CAST), Bridge RNA technology, retrotransposon system have been developed, it is still unsatisfactory to conduct targeted gene insertion with high efficiency in human cells. In the present study, Wang et al. demonstrat a new type of targeted integration strategy using human LINE1 (L1) retrotransposon element combined with CRISPR-Cas9 system, named CRISPR-Enabled Autonomous Transposable Element (CREATE). It is very smart to inactivate endonuclease domain of ORF2 (modORF2) to reduce genome-wide off-target integration. Although average integration efficincies were not high enough (approximately 1%) in human cells, it is valuable to open this method to the field. I would like to raise a few issues to strengthen this study.

Response: we thank the reviewer for the positive evaluation of the manuscript.

* Specific comments:

1. As dipicted in Figure 1a, the necessity of "ORF1p" is vague. The original role of ORF1p is an RNA-binding protein that interacts with L1 mRNA transcripts. However, in the present study, the L1 mRNA is provided outside the cells by the authors. I assume that ORF1p might not be necessary for the CREATE activity. If successful without the ORF1p, it would be benefitial in terms of shorter length of L1 mRNA.

Response: we thank the reviewer for the suggestion. We tested a CREATE mRNA construct with ORF1p removed in primary T cells. Surprisingly, this construct failed to produce successful integration of the payload. This new data is now included in Fig. EV5. We think that ORF1p may serve several essential functions in the CREATE system:

a) RNP formation: ORF1p likely facilitates the assembly of CREATE ribonucleoprotein complexes, which may be necessary for efficient nuclear import and/or interaction with the target DNA (Freeman *et al*, 2019) .

b) Protection from cellular RNA degradation: ORF1p binding may shield the CREATE mRNA from cellular RNA decay mechanisms (Monot *et al*, 2013) .

2. The CREATE requires double-nicking system by nCas9, instead of endonuclease of ORF2, which unfortunely results in unwanted indel mutations at the target site. Because it is not unvoidable, I should raise a fundamental question, "Is the CREATE a tool without causing a DNA double-stand break (DSB)?" Although the CREATE does not generate DSBs directly, double-nicking is typically accompanied with DSB.

Response: we thank the reviewer for raising this important question. In Ran et al, 2013 the author used $Cas9^{D10A}$ nickase with paired guide RNA to introduce DSB (Ran *et al*, 2013) . They systemically tested the off-set between the two sgRNAs and found that generally, an off-set of less than 100 bp are needed to observe DSB induced indels, and an off-set of > 150 bp will not cause spontaneous DSB formation. We propose that systemic testing of the off-set distance between the nicking sgRNA can identify the optimal design that maximize editing efficiency while minimizing DSB formation. We have added a discussion regarding DSB formation in the Discussion section.

3. In line 12, "double-stand breaks" -> "double-strand breaks"

Response: we corrected the mistake.

References:

Freeman BT, Sokolowski M, Roy-Engel AM, Smither ME & Belancio VP (2019) Identification of charged amino acids required for nuclear localization of human L1 ORF1 protein. *Mobile DNA* 10: 20

Monot C, Kuciak M, Viollet S, Mir AA, Gabus C, Darlix J-L & Cristofari G (2013) The Specificity and Flexibility of L1 Reverse Transcription Priming at Imperfect T-Tracts. *PLoS Genet* 9: e1003499

Ran FA, Hsu PD, Lin C-Y, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y, *et al* (2013) Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* 154: 1380–1389

Thawani A, Ariza AJF, Nogales E & Collins K (2023) Template and target site recognition by human LINE-1 in retrotransposition. *Nature*: 1–3

Dear Dr. Wang,

Thank you for the submission of your revised manuscript. We have now received the enclosed reports from the referees and I am happy to say that all support its publication now. Only a few editorial requests will need to be addressed before we can proceed with the official acceptance of your manuscript:

- Your ms has 5 main figures but separate results and discussion sections. Please either add one more main figure or combine the results and discussion sections to publish your ms as a short report. The max character count for short reports is 30.000 including spaces but excluding Methods and References. Usually, combining results and discussions helps to eliminate some redundancy in the text so it becomes shorter.

- Please add up to 5 keywords to the ms file.

- Please correct the conflict of interest subheading to "Disclosure Statement and Competing Interests"

- Bruce McCreedy needs an affiliation number/sign superscript next to his name, please add.

- The author credits need to be removed from the ms file. All credits need to be entered during online ms submission.

- In the author checklist, please answer all questions on statistics and send us a new, completed list.

- Please also acknowledge all funders mentioned in the ms in our online submission system in the separate entries: Luffing Future LLC; Myeloid Therapeutics are currently missing.

- In the Appendix file, please add page numbers on the title page for all Appendix items.

- Please remove the Reagents & Tools table from the ms file and upload it as a separate file using the template from our GTA: https://www.embopress.org/page/journal/14693178/authorguide#manuscriptpreparation
Primer sequences can all be added to this table.

- Please upload all Source Data as one folder per figure.

- Materials and Methods should be called "Methods".

- Please provide the specific URL for the PRJNA1189019 dataset in the data availability section.

- Please provide the exact p values in the legends of figures 2B, 3A-C, 4A, 5B, C; EV5 B, as reasonable.

- Please write the abstract in present tense as per journal policy (demonstrated should be demonstrate).

EMBO press papers are accompanied online by A) a short (1-2 sentences) summary of the findings and their significance, B) 2-3 bullet points highlighting key results and C) a synopsis image that is exactly 550 pixels wide and 200-600 pixels high (the height is variable). The synopsis image should provide a sketch of the major findings, like a graphical abstract. Please note that text needs to be readable at the final size. Please send us this information along with the final manuscript.

I look forward to seeing a final version of your manuscript as soon as possible.

Best regards,
Esther

Esther Schnapp, PhD
Senior Editor
EMBO reports

Referee #1:

the authors have addressed all my concerns

Referee #2:

The authors have sufficiently addressed my comments.

Referee #3:

The authors have mostly answered the issues I raised in the earlier review. I recommend the publication of this revised version.

The authors addressed the remaining editorial issues.

Dr. Yuxiao Wang
Myeloid Therapeutics
300 Technology Square Suite 203
Cambridge, Massachusetts 02139
United States

Dear Dr. Wang,

I am very pleased to accept your manuscript for publication in the next available issue of EMBO reports. Thank you for your contribution to our journal.

Your manuscript will be processed for publication by EMBO Press. It will be copy edited and you will receive page proofs prior to publication. Please note that you will be contacted by Springer Nature Author Services to complete licensing and payment information.

You may qualify for financial assistance for your publication charges - either via a Springer Nature fully open access agreement or an EMBO initiative. Check your eligibility: https://www.embopress.org/page/journal/14693178/authorguide#chargesguide

Should you be planning a Press Release on your article, please get in contact with embo_production@springernature.com as early as possible in order to coordinate publication and release dates.

If you have any questions, please do not hesitate to contact the Editorial Office. Thank you for your contribution to EMBO Reports.

Yours sincerely,

Esther Schnapp, PhD
Senior Editor
EMBO reports

>>> Please note that it is EMBO Reports policy for the transcript of the editorial process (containing referee reports and your response letter) to be published as an online supplement to each paper. If you do NOT want this, you will need to inform the Editorial Office via email immediately. More information is available here: https://www.embopress.org/transparent-process#Review_Process

EMBO Press Author Checklist

Corresponding Author Name: Yuxiao Wang
Journal Submitted to: EMBO Reports
Manuscript Number: EMBOR-2024-59975

USEFUL LINKS FOR COMPLETING THIS FORM
The EMBO Journal - Author Guidelines
EMBO Reports - Author Guidelines
Molecular Systems Biology - Author Guidelines
EMBO Molecular Medicine - Author Guidelines

Reporting Checklist for Life Science Articles (updated January
This checklist is adapted from Materials Design Analysis Reporting (MDAR) Checklist for Authors. MDAR establishes a minimum set of requirements in transparent reporting in the life sciences (see Statement of Task: 10.31222/osf.io/9sm4x). Please follow the journal's guidelines in preparing your

Please note that a copy of this checklist will be published alongside your article.

Abridged guidelines for figures

1. Data

The data shown in figures should satisfy the following conditions:

- the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
- ideally, figure panels should include only measurements that are directly comparable to each other and obtained with the same assay.
- plots include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical
- if n<5, the individual data points from each experiment should be plotted. Any statistical test employed should be justified.
- Source Data should be included to report the data underlying figures according to the guidelines set out in the authorship guidelines on Data

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

- a specification of the experimental system investigated (eg cell line, species name).
- the assay(s) and method(s) used to carry out the reported observations and measurements.
- an explicit mention of the biological and chemical entity(ies) that are being measured.
- an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.
- the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- a statement of how many times the experiment shown was independently replicated in the laboratory.
- definitions of statistical methods and measures:
  - common tests, such as t-test (please specify whether paired vs. unpaired), simple χ^2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - exact statistical test results, e.g., P values = x but not P values < x;
  - definition of 'center values' as median or average;
  - definition of error bars as s.d. or s.e.m.

Please complete ALL of the questions below.
Select "Not Applicable" only when the requested information is not relevant for your study.

Materials

Newly Created Materials	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
New materials and reagents need to be available; do any restrictions apply?	Not Applicable	

Antibodies	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
For **antibodies** provide the following information: - Commercial antibodies: RRID (if possible) or supplier name, catalogue number and or/clone number - Non-commercial: RRID or citation	Not Applicable	

DNA and RNA sequences	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Short novel DNA or RNA including primers, probes: provide the sequences.	Yes	Appendix

Cell materials	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Cell lines: Provide species information, strain. Provide accession number in repository **OR** supplier name, catalog number, clone number, and/**OR** RRID.	Yes	Reagents and Tools Table
Primary cultures: Provide species, strain, sex of origin, genetic modification status.	Not Applicable	
Report if the cell lines were recently **authenticated** (e.g., by STR profiling) and tested for mycoplasma contamination.	Not Applicable	

Experimental animals	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Laboratory animals or Model organisms: Provide species, strain, sex, age, genetic modification status. Provide accession number in repository **OR** supplier name, catalog number, clone number, **OR** RRID.	Not Applicable	
Animal observed in or captured from the field: Provide species, sex, and age where possible**.**	Not Applicable	
Please detail **housing and husbandry conditions**.	Not Applicable	

Plants and microbes	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Plants: provide species and strain, ecotype and cultivar where relevant, unique accession number if available, and source (including location for collected wild specimens).	Not Applicable	
Microbes: provide species and strain, unique accession number if available, and source.	Not Applicable	

Human research participants	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
If collected and within the bounds of privacy constraints report on age, sex and gender or ethnicity for all study participants.	Not Applicable	

Core facilities	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
If your work benefited from core facilities, was their service mentioned in the acknowledgments section?	Not Applicable	

Design

Study protocol	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
If study protocol has been **pre-registered, provide DOI in the manuscript**. For clinical trials, provide the trial registration number **OR** cite DOI.	Not Applicable	
Report the **clinical trial registration number** (at ClinicalTrials.gov or equivalent), where applicable.	Not Applicable	

Laboratory protocol	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Provide DOI OR other citation details if **external detailed step-by-step protocols** are available.	Not Applicable	

Experimental study design and statistics	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Include a statement about **sample size** estimate even if no statistical methods were used.	Yes	Figure legends
Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. **randomization procedure**)? If yes, have they been described?	Yes	Materials and Methods
Include a statement about **blinding** even if no blinding was done.	Yes	Materials and Methods
Describe **inclusion/exclusion criteria** if samples or animals were excluded from the analysis. Were the criteria pre-established? If sample or data points were omitted from analysis, report if this was due to attrition or intentional exclusion and provide justification.	Yes	Materials and Methods
For every figure, are **statistical tests** justified as appropriate? Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it. Is there an estimate of variation within each group of data? Is the variance similar between the groups that are being statistically compared?	Yes	Materials and Methods

Sample definition and in-laboratory replication	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
In the figure legends: state number of times the experiment was **replicated** in laboratory.	Yes	Figure legends
In the figure legends: define whether data describe **technical or biological replicates**.	Yes	Figure legends

Ethics

Ethics	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Studies involving **human participants**: State details of **authority granting ethics approval** (IRB or equivalent committee(s), provide reference number for approval.	Not Applicable	
Studies involving **human participants**: Include a statement confirming that **informed consent** was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.	Not Applicable	
Studies involving **human participants:** For publication of **patient photos**, include a statement confirming that consent to publish was obtained.	Not Applicable	
Studies involving experimental **animals**: State details of **authority granting ethics approva**l (IRB or equivalent committee(s), provide reference number for approval. Include a statement of compliance with ethical regulations.	Not Applicable	
Studies involving **specimen and field samples:** State if relevant **permits** obtained, provide details of authority approving study; if none were required, explain why.	Not Applicable	

Dual Use Research of Concern (DURC)	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Could your study fall under dual use research restrictions? Please check biosecurity documents and list of **select agents and toxins** (CDC): https://www.selectagents.gov/sat/list.htm	Not Applicable	
If you used a select agent, is the security level of the lab appropriate and reported in the manuscript?	Not Applicable	
If a study is subject to dual use research of concern regulations, is the name of the **authority granting approval and reference number** for the regulatory approval provided in the manuscript?	Not Applicable	

Reporting

The MDAR framework recommends adoption of discipline-specific guidelines, established and endorsed through community initiatives. Journals have their own policy about requiring specific guidelines and recommendations to complement MDAR.

Adherence to community standards	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
State if relevant guidelines or checklists (e.g., **ICMJE, MIBBI, ARRIVE, PRISMA**) have been followed or provided.	Not Applicable	
For **tumor marker prognostic studies,** we recommend that you follow the **REMARK** reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	Not Applicable	
For **phase II and III randomized controlled trials**, please refer to the **CONSORT** flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	Not Applicable	

Data Availability

Data availability	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Have **primary datasets** been deposited according to the journal's guidelines (see 'Data Deposition' section) and the respective accession numbers provided in the Data Availability Section?	Yes	Data availability section
Were **human clinical and genomic datasets** deposited in a public access-controlled repository in accordance to ethical obligations to the patients and to the applicable consent agreement?	Not Applicable	
Are **computational models** that are central and integral to a study available without restrictions in a machine-readable form? Were the relevant accession numbers or links provided?	Not Applicable	
If publicly available data were reused, provide the respective **data citations in the reference list**.	Not Applicable	