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

Authors: Yuxiao Wang[1]*†, Ruei-Zeng Lin[1]†, Meghan Harris[1], Bianca Lavayen[1], Bruce McCreedy, Daniel Getts[1]*

Affiliations:

[1] Myeloid Therapeutics Inc., Cambridge, MA 02139, USA

* Corresponding authors. Email: daniel@myeloidtx.com (D.G.); ywang@myeloidtx.com (Y.W.);

† These authors contributed equally

Abstract:

To realize the potential of genome editing for broad therapeutic applications, new tools are needed for delivery of multi-kilobase (kb) payloads at desired genomic sites with high precision. Here we introduce the CRISPR-Enabled Autonomous Transposable Element (CREATE), an RNA-based genome editing system that merges CRISPR/Cas9 with the human L1 retrotransposon to insert gene-sized payloads without DNA donors or double strand breaks. CREATE enables L1-mediated reverse transcription and integration of an RNA-encoded payload gene specifically at two Cas9-induced nick sites. CREATE is delivered using mRNA components and achieves integration of >1 kb payload in mammalian cells, opening the door to mRNA mediated therapeutic genome editing in vivo.

One-Sentence Summary: An mRNA-based, large-scale genome editing technology integrating CRISPR/Cas9 and L1 retroelement.

Main Text:

The development of CRISPR-Cas9 systems enabled genomic editing through the site-specific introduction of double strand breaks (DSBs), which can be harnessed for the insertion of a DNA fragment via the homology-directed repair (HDR) pathway (*1*). This approach requires a DNA template for HDR (often delivered via viral vectors) and is prone to introduce large chromosomal deletions and translocations, all of which pose safety concerns and have slowed its clinical applications (*2–5*). Prime Editing (PE) was subsequently developed, fusing a Cas9 nickase with the reverse transcriptase from Moloney Murine Leukemia Virus (M-MLV). Avoiding DSB, this approach is limited to the insertion of up to 40 base pairs (bp), encoded within the prime editing guide RNA (pegRNA) (*6*). Attempts to overcome this size limitation, including Twin-Prime and template-jumping prime editors, have failed to extend the payload capacity beyond a few hundred basepairs (bps) (*7*, *8*). This is not sufficient for the treatment of most genetic disorders. Additional approaches have been developed that combine PE with a sequence-specific serine recombinase, enabling the integration of a DNA fragment from a donor plasmid (*8*, *9*). Despite their potential, the practical application of these systems is hindered by their significant complexity that requires the co-delivery of multiple components: the PE editor with pegRNA, the serine recombinase, and a DNA donor plasmid (*10*, *11*). Clinically, the delivery of a DNA template either through plasmids or viral vectors is cumbersome, has size limitations and is expensive. A system that can be delivered using mRNA, can insert large genomic cargos at desired genomic sites with no off-target insertions would pave the way for the next generation of gene editors (*10*).

By using the human LINE-1 elements, we have addressed this need by developing the CRISPR-Enabled Autonomous Transposable Element (CREATE) system. The LINE1 retrotransposon (referred to as L1) is a mobile genetic element that comprises 17% of the human genome (*12*). It naturally replicates via an RNA intermediate, capable of efficiently reverse transcribing large mRNA sequences into cDNA and integrating into human genome (*13*, *14*). This significant processivity inspired us to harness L1 and its replication cycle to achieve programmable, large-scale genome editing. L1 replication involves a bicistronic mRNA encoding two proteins, ORF1p and ORF2p. The ORF1p protein is an RNA-binding protein that interacts with L1 mRNA transcripts (*15*). ORF2p encodes an endonuclease (EN) and a reverse transcriptase (RT) domain (*16*, *17*). A distinctive feature of L1 is its 'cis preference,' where ORF1p and ORF2p associate with their own mRNA transcript to form a ribonuclear protein complex (RNP) (*18*), facilitating the reverse transcription and insertion of the L1 mRNA (*19*, *20*). Recent high-resolution cryo-EM structures of ORF2p complexed with DNA and RNA substrates have shed light on the process of target site recognition and initiation of reverse transcription (*21*, *22*). These structures revealed that the EN domain of ORF2p recognizes and cleaves the 5’TTTT/AA3’ consensus DNA sequence which then hybridizes with the poly-A tail of L1 mRNA. The rest of ORF2p forms a groove that tightly binds the RNA:DNA heteroduplex, allowing the RT domain to initiate the target-primed reverse transcription (TPRT) reaction for the synthesis of the first cDNA strand. The mechanism of second-strand cDNA synthesis is hypothesized to involve an upstream nick and strand exchange through microhomology to initiate TPRT. The ORF2p RT domain was shown to be highly processive, capable of reverse transcription of long RNA sequences. However, in its native form, the L1 EN domain lacks sequence specificity, precluding its application in targeted, programmable gene editing. A recent study attempting to combine Cas9 with R2 retroelement via a direct fusion approach was unable to achieve complete retrotransposition and integration (*23*).

In this study we report the development of the CREATE system that combines the precision and programmability of CRISPR/Cas9 with the large-scale genome re-writing capability of L1. To demonstrate targeted gene delivery in mammalian cells, we inserted a 1.1 kb gene comprising a promoter and green fluorescent protein (GFP) into several sites in the human genome guided by Cas9. Mechanistic studies revealed that the CREATE system is ultra-specific with no observed off-target events. Together our findings show that the CREATE system is a programmable and highly specific gene delivery platform capable of delivering a payload of any size with only RNA components, thus offering considerable prospects for diverse therapeutic applications.

Design of the CREATE editing system

The CREATE system encodes, on a single mRNA, the codon optimized L1 components ORF1p and ORF2p, separated by the native inter-ORF sequence which mediates bicistronic expression (Fig. 1A). To facilitate nuclear import, the ORF2p is fused with an N-terminal SV40 nuclear localization signal (NLS) and a C-terminal nucleoplasmin NLS. A key design element implemented to prevent nonspecific DNA cleavage and integration was the inactivation of the catalytic residue within the EN domain of ORF2p (mutation D205A). The payload cassette consisting of a promoter and the gene of interest is placed at the 3'UTR of the CREATE mRNA (Fig. 1A). To ensure that the payload cannot be expressed by direct translation, the payload is encoded in the anti-sense orientation (Fig. 1A). Critically, the payload is flanked by sequences designed to hybridize with single-guide RNA (sgRNA) target sites in the genome to prime reverse transcription. These sequences are referred to as primer binding site 1 (PBS1) and reverse complement primer binding site 2 (RC-PBS2).

The CREATE mRNA when delivered into the cells will be translated into ORF1p and $ORF2p^{ENdead}$, which then co-assemble with the mRNA to form CREATE RNP and enter cell nucleus (Fig. 1B, step 1). Natural L1 retrotransposition relies on nicking of the non-targeting DNA strand by the EN domain of ORF2p (*21*, *22*). To harness this mechanism, a $Cas9^{H840A}$ non-target strand nickase was employed to introduce a single strand nick guided by sgRNA1 (Fig 1B, step 2). The liberated DNA 3'-flap then hybridizes with the PBS1 sequence on CREATE mRNA, serving as the primer for the RT domain of ORF2p for first strand cDNA synthesis (Fig. 1B, step 3). The newly synthesized cDNA strand includes both the payload and the PBS2 sequences in the sense orientation (Fig. 1B, step 4). Subsequently, PBS2 hybridizes with the 3' flap released from the second nicked site generated by sgRNA2, invoking the L1 template jump mechanism to complete the second strand synthesis and integration (Fig. 1B, step 5-6). The outcome of the editing cycle is the successful integration of the payload cassette between the PBS sites, discarding the L1 components, resulting in safe non-replicative gene insertion.

CREATE-mediated large payload gene delivery in mammalian cells

To demonstrate that CREATE can be used to deliver functional genes, we designed a reporter payload consisting of an EF1α core promoter-driven GFP with an SV40 poly(A) signal (totaling 1.1 kb). The AAVS1 site, a well-characterized genomic safe harbor site, was selected to show programmable targeted insertion. We chose two sgRNAs that have been validated previously, targeting sites within AAVS1 that are 90 bp apart (*7*). Successful editing will result in replacement of the 90 bp DNA segment with the 1.1 kb payload cassette (Fig. 1A). Baldwin et al showed that ORF2p can efficiently initiate TPRT with 7-20 bp primer length (*22*). Therefore, we

designed PBS1 and RC-PBS2 in this range to anneal with the DNA flaps released from two sgRNA cut sites, and positioned them on either side of the payload.

Both Cas9 and ORF2p are large multidomain proteins, each exceeding 100 kD. We were concerned that direct Cas9-ORF2p fusions may compromise activity and the ability to enter cell nucleus. Since ORF2p could scan genomic DNA for entry sites, we hypothesized that it could independently identify sites nicked by Cas9, eliminating the need for a direct fusion. To test this concept, we developed a HEK293T cell line engineered to stably express a SpCas9 non-targeting strand nickase (293T-nCas9^{H840A}). Flow cytometry was used to quantify the percentage of cells expressing GFP as a measurement of successful payload integration. Transfection of cells with CREATE mRNA and dual sgRNAs resulted in GFP expression on Day 3 and Day 8 (Fig. 1C). No GFP expression was observed in 293T-nCas9^{H840A} cells transfected with non-targeting control sgRNAs or in wildtype HEK293T cells (Fig. 1C). These results confirmed sgRNA-guided integration and excluded the possibility of Cas9-independent, nonspecific insertion. Furthermore, we demonstrated that payload expression requires co-transfection of both sgRNAs, as delivery of either sgRNA alone did not result in GFP positive cells (Fig. 1C). This result confirmed that dual nicking events on opposite DNA strands are essential for complete TPRT and successful integration. Such requirement inherently enhances specificity, as a single independent sgRNA off-target binding would be insufficient to cause payload integration.

To further appreciate the temporal requirements for delivery of the CREATE components, sequential transfection of RNA for the individual CREATE elements was performed. The specific protocols included (1) Simultaneous transfection of CREATE mRNA and sgRNAs (One-shot Protocol); (2) Transfecting CREATE mRNA four hours prior to sgRNAs (Two-step, mRNA-first Protocol); (3) Transfecting sgRNAs four hours before CREATE mRNA (Two-step, sgRNA-first Protocol). In Protocols #2 and #3, the culture media was replaced immediately prior to the second transfection. Observations support superior performance of the two-step protocols when compared to the one-shot approach (Fig. 1D). Notably, the mRNA-first strategy yielded the highest CREATE editing efficiency, reaching ~1.2% GFP expression on day 3 and remaining stable on day 8. The improvement is likely due to the necessary time for the ORF1p and ORF2p mRNA to be translated, co-assemble with CREATE mRNA and subsequently enter the cell nucleus. Delivery of sgRNAs before the CREATE RNP complex forms might result in the single strand nicks being repaired by the cell before the CREATE editing process can occur. Further optimization of the time interval between mRNA and sgRNA transfections, along with adjustments in sgRNA concentrations, culminated in around 1.5% of the cells expressing the GFP payload (fig. S1).

Characterization of site-specific integration of payload

Fluorescence-activated cell sorting (FACS) was employed to enrich the CREATE edited cells to approximately 70% GFP positive, followed by passage of the cells for an additional two weeks (Fig. 2A). During this time GFP expression remained stable (>70%), indicative of successful and stable genomic integration of the payload. Specific integration into the AAVS1 site was determined using PCR on the post-sort GFP-positive cells. A band corresponding to the expected size associated with insertion at the AAVS1 site was observed along with unedited allele (Fig. 2B). Prime editing and related approaches have been observed to incur indel formation at the sgRNA target sites (*8, 11*). To further confirm specific integration at AAVS1 site and assess the editing accuracy at the payload-genome junctions, we performed amplicon sequencing on

CREATE-edited GFP-positive cells from two independent experiments. At the PBS2-AAVS1 junction, 95% of sequencing reads showed expected editing outcome without indels in both experiments (Fig. 2C and 2D). A study using a pair of pegRNAs to perform PE also observed varying degree of indel formation (sometimes close to 80%) when different pegRNAs were tested (*8*).

Investigation of CREATE editing mechanism

To further appreciate the critical elements for CREATE gene editing, a reductionist approach, using functional mutants was applied. We first examined whether CREATE editing requires non-target strand ($Cas9^{H840A}$) rather than target strand nickase ($Cas9^{D10A}$) activity. When we transfected CREATE mRNA and dual AAVS1-targeting sgRNAs into HEK293T cells stably expressing either $Cas9^{D10A}$ or $Cas9^{H840A}$ nickases, we observed GFP expression and integration exclusively in the presence of $Cas9^{H840A}$ nickase (Fig. 3A), confirming the natural L1 replication mechanism that initiates with non-target strand cleavage. In addition, the activity of $Cas9^{H840A}$ also destroys the PAM sequence after successful editing, mitigating concerns about continuous nicking of the edited genome. Next we introduced D702A mutation to inactivate the reverse transcriptase (RT) domain of L1 ORF2p ($CREATE^{RTdead}$). This modification entirely abolished the integration and expression of the GFP payload (Fig. 3A). This result ruled out the possibility of contaminating plasmid DNA-mediated integration and expression, which would occur independent of the RT activity of L1. Similarly, the removal of the entire EN domain (CREATE $mRNA^{\Delta EN}$) resulted in a marked decrease in GFP insertion efficiency, indicating that while the EN domain must be rendered inactive to avert non-specific editing, its complete removal is detrimental to the editing process. This is consistent with the recent reports showing that L1 ORF2p functions through an integrated genomic DNA binding and TPRT pathway. Complete removal of the EN domain may sabotage the required structural interactions between the DNA strand and the L1 retrotransposon disabling the DNA binding/scanning necessary for full function of ORF2p.

PBS length and sgRNA sites influence editing efficiency

To determine if PBS length impacts insertion efficiency, we designed CREATE mRNA with different PBS1 and RC-PBS2 hybridization sequence lengths (17 bp, 30 bp, and 50 bp) and measured their capacity to insert GFP payload at the AAVS1 sites targeted with the same dual sgRNAs. Consistent with a recent study investigating the poly-A length required for efficient TPRT by L1 ORF2 (*21*), we observed that 17-30 bp appear to be ideal for efficient editing, and longer length led to a reduction of GFP integration efficiency (Fig. 3B).

CREATE can be used for exogenous gene delivery and search-and-replacement, as the segment between sgRNA1 and sgRNA2 will be deleted and replaced by the payload (Fig. 1A). To explore this capability, we investigated the influence of the distance between the two sgRNA cut sites on editing efficiency. By fixing sgRNA2 and selecting different sgRNA1s with cut sites located further away, we designed several CREATE mRNA/sgRNAs to replace 90 bp, 481 bp and 976 bp DNA segment with the GFP expression cassette. Our results demonstrated that CREATE can achieve successful replacement in all three cases, although with lower efficiencies when the target sites were further apart (Fig. 3B). Notably this exceeded the capability of prime editing-based approach, which can only replace < 100bp sequences (*8*).

CREATE paired sgRNAs and matched PBS sequences impart high fidelity and precision

To demonstrate the importance of the dual PBS sites flanking the payload we designed CREATE mRNA with either PBS1 or a RC-PBS2 sequence. In the presence of AAVS1 targeting sgRNAs, the removal of either PBS1 or RC-PBS2 resulted in no GFP expression on day 3 or day 8 (Fig. 3C). Additionally, swapping the position of PBS1 and PBS2 while simultaneously encoding them in a reverse-complementary sequence ("RC-PBS1" and "PBS2) restored the GFP expression, suggesting correct insertion of the payload cassette in the opposite orientation. The requirement for dual PBS flanking sites and dual sgRNA imparts a level of specificity, which was confirmed further when we reviewed the top 5 predicted off-target sites for sgRNA1 and sgRNA2, by PCR. We observed no PCR product for any of the predicted off target sites (fig. S2). Together these data not only validated the proposed mechanism of CREATE but also showed the editing process has a number of inherent checkpoints that prevent off-target activity. These include 1) correct base-pairing of PBS1 with sgRNA1 cut sites; 2) correct base-pairing of reverse transcribed PBS2 with sgRNA2 cut site and 3) adjacent sgRNA1 and sgRNA2 nicking.

Targeted insertion of large DNA fragments into different genomic loci

To determine the ability of harnessing the CREATE system to edit and deliver genes specifically into sites in the genome beyond AAVS1, we engineered CREATE to deliver GFP to other genomic loci including *HEK3*, *PRNP*, and *IDS*. The *HEK3* locus, situated on chromosome 1 is frequently chosen to benchmark gene editing tools. The *PRNP* locus on chromosome 20 encodes the prion protein implicated in multiple neurodegenerative diseases, and *IDS* gene on chromosome X is associated with Hunter's syndrome, a lysosomal storage disease. For each targeted site, a pair of sgRNAs 70-90 bp apart were selected. CREATE mRNA, which carried the corresponding 30 bp PBS1 and RC-PBS2 flanking the GFP expression cassette, was deployed for each target site. Flow cytometric analysis on day 3 demonstrated all three sites exhibited comparable editing efficiency, as determined by GFP expression (Fig. 4A). We used FACS to enrich PRNP locus edited cells to >60% GFP positivity (Fig. 4B). To confirm specific integration, we amplified the expected junction region and performed TA-cloning of the PCR product. Sanger sequencing of the individual colonies from the TA-cloning revealed correct insertion at the expected junction at PRNP site, validated the programmable, targeted editing capability of CREATE (Fig. 4C).

CREATE as a fully RNA-based large-scale genome editing system in human liver cells

Finally, an important attribute of CREATE gene editing is the ability to deliver all components as RNA into mammalian cells (Fig. 5A). This was achieved in HEK293T and the immortalized liver cell line Huh7, whereby RNA transfection resulted in up to 1% GFP gene re-expression (Fig. 5B). Together these findings confirmed the ability to use RNA to deliver all necessary CREATE components resulting in stable integration of cargos not possible with other available approaches.

Discussion

Advances in genome editing have moved at a significant pace in the last decade fueled by editing tools taken from prokaryotes. However, these tools have not enabled the delivery of large gene sized insertions. The complexity of most genetic disorders, which commonly involve a spectrum of mutations, insertions, or deletions across extensive genomic region requires new tools to be developed. CREATE represents a major step forward in gene editing. The system not only significantly increases the cargo size that can be inserted into the genome without DNA donors, but its built-in regulatory mechanisms reduce the potential for off-target genotoxcity. PE requires encoding of desired edits within a compact pegRNA, severely limiting the payload size. In contrast, the payload for CREATE is encoded in the 3'UTR of CREATE mRNA, which can accommodate multi-kb sequences with no apparent size limit. Additionally, by using the highly processive RT domain from L1 ORF2p, CREATE can to complete cDNA synthesis much larger than what can be achieved with the M-MLV RT domain used in PE. We demonstrated this ability inserting a 1.1kb GFP cassette in multiple mammalian cell lines at ~2% integration efficiency, while the largest payload reported to be delivered via PE-derived approach is ~1kb with less than 0.003% integration (*24*). There remains significant opportunity to further optimize the CREATE gene delivery system. Optimization of RNA sequences, delivery methods, modulation of the cellular DNA repair and RNA processing pathways. In addition, as the programmable targeting (Cas9 nickase) and the component for RT and integration (ORF2p) do not need to be covalently linked, this may also facilitate the exploration of alternative programmable RNA-guided DNA endonucleases such as Cas12, TnpB and Fanzor to expand the capability of the CREATE system (*25–27*)

Reference

1. J. Y. Wang, J. A. Doudna, CRISPR technology: A decade of genome editing is only the beginning. *Science* **379**, eadd8643 (2023).

2. M. L. Leibowitz, S. Papathanasiou, P. A. Doerfler, L. J. Blaine, L. Sun, Y. Yao, C.-Z. Zhang, M. J. Weiss, D. Pellman, Chromothripsis as an on-target consequence of CRISPR–Cas9 genome editing. *Nat. Genet.* **53**, 895–905 (2021).

3. E. Brunet, M. Jasin, Chromosome Translocation. *Adv. Exp. Med. Biol.* **1044**, 15–25 (2018).

4. Y. Song, Z. Liu, Y. Zhang, M. Chen, T. Sui, L. Lai, Z. Li, Large-Fragment Deletions Induced by Cas9 Cleavage while Not in the BEs System. *Mol. Ther. - Nucleic Acids* **21**, 523–526 (2020).

5. M. Kosicki, K. Tomberg, A. Bradley, Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).

6. A. V. Anzalone, P. B. Randolph, J. R. Davis, A. A. Sousa, L. W. Koblan, J. M. Levy, P. J. Chen, C. Wilson, G. A. Newby, A. Raguram, D. R. Liu, Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).

7. C. Zheng, B. Liu, X. Dong, N. Gaston, E. J. Sontheimer, W. Xue, Template-jumping prime editing enables large insertion and exon rewriting in vivo. *Nat. Commun.* **14**, 3369 (2023).

8. A. V. Anzalone, X. D. Gao, C. J. Podracky, A. T. Nelson, L. W. Koblan, A. Raguram, J. M. Levy, J. A. M. Mercer, D. R. Liu, Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. *Nat. Biotechnol.* **40**, 731–740 (2022).

9. M. T. N. Yarnall, E. I. Ioannidi, C. Schmitt-Ulms, R. N. Krajeski, J. Lim, L. Villiger, W. Zhou, K. Jiang, S. K. Garushyants, N. Roberts, L. Zhang, C. A. Vakulskas, J. A. Walker, A. P. Kadina, A. E. Zepeda, K. Holden, H. Ma, J. Xie, G. Gao, L. Foquet, G. Bial, S. K. Donnelly, Y. Miyata, D. R. Radiloff, J. M. Henderson, A. Ujita, O. O. Abudayyeh, J. S. Gootenberg, Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases. *Nat. Biotechnol.* **41**, 500–512 (2023).

10. S. Tang, S. H. Sternberg, Genome editing with retroelements. *Science* **382**, 370–371 (2023).

11. P. J. Chen, D. R. Liu, Prime editing for precise and highly versatile genome manipulation. *Nat. Rev. Genet.* **24**, 161–177 (2023).

12. I. H. G. S. Consortium, W. I. for B. R. Research: Center for Genome, E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, T. S. Centre:, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, W. U. G. S. Center, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, U. D. J. G. Institute:, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.-F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, B. C. of M. H. G. S. Center:, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, R. G. S. Center:, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, G. and C. UMR-8030:, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. of G. A. Biotechnology: Institute of Molecular, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, G. S. Center:, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, B. G. I. G. Center:, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, M. S. C. Biology: The Institute for Systems, L. Hood, L. Rowen, A. Madan, S. Qin, S. G. T. Center:, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, U. of O. A. C. for G. Technology:, B. A. Roe, F. Chen, H. Pan, M. P. I. for M. Genetics:, J. Ramser, H. Lehrach, R. Reinhardt, C. S. H. L. Center: Lita Annenberg Hazen Genome, W. R. McCombie, M. de la Bastide, N. Dedhia, G. R. C. for Biotechnology:, H. Blöcker, K. Hornischer, G. Nordsiek, R.

Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H.-C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. R. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. A. Smit, E. Stupka, J. Szustakowki, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S.-P. Yang, R.-F. Yeh, S. management: N. H. G. R. I. Health: US National Institutes of, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, S. H. G. Center:, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, U. of W. G. Center:, M. V. Olson, R. Kaul, C. Raymond, D. of M. B. Medicine: Keio University School of, N. Shimizu, K. Kawasaki, S. Minoshima, U. of T. S. M. C. at Dallas:, G. A. Evans, M. Athanasiou, R. Schultz, O. of S. Energy: US Department of, A. Patrinos, T. W. Trust:, M. J. Morgan, Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

13. C. R. Beck, J. L. Garcia-Perez, R. M. Badge, J. V. Moran, LINE-1 Elements in Structural Variation and Disease. *Annu. Rev. Genom. Hum. Genet.* **12**, 187–215 (2011).

14. J. S. Han, Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. *Mob. DNA* **1**, 15 (2010).

15. E. Khazina, V. Truffault, R. Büttner, S. Schmidt, M. Coles, O. Weichenrieder, Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. *Nat. Struct. Mol. Biol.* **18**, 1006–1014 (2011).

16. S. L. Mathias, A. F. Scott, H. H. K. Jr., J. D. Boeke, A. Gabriel, Reverse Transcriptase Encoded by a Human Transposable Element. *Science* **254**, 1808–1810 (1991).

17. Q. Feng, J. V. Moran, H. H. Kazazian, J. D. Boeke, Human L1 Retrotransposon Encodes a Conserved Endonuclease Required for Retrotransposition. *Cell* **87**, 905–916 (1996).

18. A. J. Doucet, A. E. Hulme, E. Sahinovic, D. A. Kulpa, J. B. Moldovan, H. C. Kopera, J. N. Athanikar, M. Hasnaoui, A. Bucheton, J. V. Moran, N. Gilbert, Characterization of LINE-1 Ribonucleoprotein Particles. *PLoS Genet.* **6**, e1001150 (2010).

19. K. K. Kojima, Different integration site structures between L1 protein-mediated retrotransposition in cis and retrotransposition in trans. *Mob. DNA* **1**, 17 (2010).

20. D. A. Kulpa, J. V. Moran, Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat. Struct. Mol. Biol.* **13**, 655–660 (2006).

21. A. Thawani, A. J. F. Ariza, E. Nogales, K. Collins, Template and target-site recognition by human LINE-1 in retrotransposition. *Nature*, 1–8 (2023).

22. E. T. Baldwin, T. van Eeuwen, D. Hoyos, A. Zalevsky, E. P. Tchesnokov, R. Sánchez, B. D. Miller, L. H. D. Stefano, F. X. Ruiz, M. Hancock, E. Işik, C. Mendez-Dorantes, T. Walpole, C.

Nichols, P. Wan, K. Riento, R. Halls-Kass, M. Augustin, A. Lammens, A. Jestel, P. Upla, K. Xibinaku, S. Congreve, M. Hennink, K. B. Rogala, A. M. Schneider, J. E. Fairman, S. M. Christensen, B. Desrosiers, G. S. Bisacchi, O. L. Saunders, N. Hafeez, W. Miao, R. Kapeller, D. M. Zaller, A. Sali, O. Weichenrieder, K. H. Burns, M. Götte, M. P. Rout, E. Arnold, B. D. Greenbaum, D. L. Romero, J. LaCava, M. S. Taylor, Structures, functions and adaptations of the human LINE-1 ORF2 protein. *Nature*, 1–13 (2023).

23. M. E. Wilkinson, C. J. Frangieh, R. K. Macrae, F. Zhang, Structure of the R2 non-LTR retrotransposon initiating target-primed reverse transcription. *Science* **380**, 301–308 (2023).

24. J. Wang, Z. He, G. Wang, R. Zhang, J. Duan, P. Gao, X. Lei, H. Qiu, C. Zhang, Y. Zhang, H. Yin, Efficient targeted insertion of large DNA fragments without DNA donors. *Nat. Methods* **19**, 331–340 (2022).

25. B. Zetsche, J. S. Gootenberg, O. O. Abudayyeh, I. M. Slaymaker, K. S. Makarova, P. Essletzbichler, S. E. Volz, J. Joung, J. van der Oost, A. Regev, E. V. Koonin, F. Zhang, Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163**, 759–771 (2015).

26. M. Saito, P. Xu, G. Faure, S. Maguire, S. Kannan, H. Altae-Tran, S. Vo, A. Desimone, R. K. Macrae, F. Zhang, Fanzor is a eukaryotic programmable RNA-guided endonuclease. *Nature* **620**, 660–668 (2023).

27. T. Karvelis, G. Druteika, G. Bigelyte, K. Budre, R. Zedaveinyte, A. Silanskas, D. Kazlauskas, Č. Venclovas, V. Siksnys, Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature* **599**, 692–696 (2021).

28. K. Clement, H. Rees, M. C. Canver, J. M. Gehrke, R. Farouni, J. Y. Hsu, M. A. Cole, D. R. Liu, J. K. Joung, D. E. Bauer, L. Pinello, CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).

Acknowledgments: We thank R. Vale (HHMI Janelia) & S. Mukherjee (Columbia U) for critical reading of the manuscript; R. Hofmeister (MyeloidTx) for reviewing and editing of the manuscript; and the entire Myeloid Therapeutics team for support.

Funding: This study was funded entirely by Myeloid Therapeutics

Author contributions:

Conceptualization: YW, DG, RL, BM

Methodology: YW, DG, RL, MH

Experiments and Data Analysis: RL, MH, BL,YW

Writing: YW, RL, DG

Competing interests: Y.W., R.L., M.H., B.L. and D.G. are current employees of Myeloid Therapeutics. B.M. was a past employee of Myeloid Therapeutics. All authors hold equity interest in Myeloid Therapeutics.

Data and materials availability: Sequences including primers, sgRNAs are available in the supplementary materials.

Supplementary Materials

Materials and Methods

Figs. S1 to S2

Tables S1 to S4

Fig. 1. CREATE editing design and proof-of-concept in mammalian cells. (**A**) Schematics of the CREATE mRNA design. Payload cassette (EF1α core promoter driven GFP with SV40 polyA signal) is encoded in the antisense direction. Red diamond indicate stop codon after ORF1 and ORF2-NLS. (**B**) CREATE editing mechanism. (1) L1 components ORF1p and ORF2p are produced from CREATE mRNA and co-assemble into ribonuclear complex with CREATE mRNA. (2) Non-target strand nicking by $Cas9^{H840A}$ and sgRNAs. (3) and (4) the DNA 3'-flap hybridizes with PBS1 to prime 1st strand cDNA synthesis mediated by ORF2p RT domain. (5) 2nd strand synthesis is primed by PBS2 that hybridizes with the reverse transcribed RC-PBS2 sequence in the newly synthesized cDNA strand. (6) Targeted insertion of the payload between PBS sites into the genome without the sequences encoding L1 components. (**C**) CREATE editing in mammalian cells requires Cas9 nickase activity and dual sgRNAs. Statistical analysis was performed using two-way ANOVA with Dunnett's multiple comparisons test comparing against the first sample. **** indicates $p < 0.0001$. Statistical significance is labeled for samples with $p < 0.01$.(**D**) Optimization of transfection protocol improved editing efficiency. Protocol 1 (one shot), Protocol 2 (mRNA first) and Protocol 3 (sgRNA first). Data shown are representative of at least two independent experiments.

Fig. 2. Analysis of payload integration at AAVS1 site. (**A**) AAVS1 edited cells before and after FACS enrichment of GFP positive cells. (**B**) PCR primers were designed to amplify region that covers the AAVS1 locus marked by sgRNAs for integration. The unedited allele showed expected 570 bp band (red arrow) while the edited allele showed an expected increase in size due to payload integration (green arrow). (**C**) The percentage of indels at the integration junction. Each plot shows percentage sequencing reads with or without indels. Data from two independent experiments are shown. (**D**) Distribution of indels and substitutions at the junction sites. Black arrow indicates sgRNA cut sites at PBS1 and PBS2 junctions. In Exp 2, insertion of polyA sequences was observed and indicated on the graph. Data from (**C**) and (**D**) were analyzed and plotted using CRISPResso2 (*28*).

Fig. 3. Mechanistic analysis of CREATE editing. (**A**) Mutational analysis of the key enzymatic activities and protein domains involved in editing. (**B)** Influence of the length of PBS sequences in CREATE mRNA and the distance between sgRNA nick sites in the genome on editing efficiency. (**C**) CREATE editing requires two PBS sequences that matches the two sgRNA target sites. Statistical analysis was performed using two-way ANOVA with Dunnett's multiple comparisons test comparing against the first sample. **** $p < 0.0001$.** $p<0.01$. Statistical significance is labeled for samples with $p < 0.01$. Data shown are representative of at least two independent experiments.

Fig. 4. CREATE editing mediates targeted gene delivery to multiple genomic loci. (**A**) Editing of three additional genomic loci by designing unique sgRNAs and matching PBSs. (**B**) FACS enrichment of the PRNP edited cells confirmed stable integration. Statistical analysis was performed using two-way ANOVA with Dunnett's multiple comparisons test comparing against the first sample. **** $p < 0.0001$. *** $p<0.001$.** $p<0.01$. Statistical significance is labeled for samples with $p < 0.01$. Data shown are representative of at least two independent experiments. (**C**) Sanger sequencing of PCR amplicons covering the junction region confirmed correct integration of payload at the PRNP locus.

Fig. 5. CREATE as a fully RNA-based gene delivery system to multiple cell types (**A**) Co-delivery of CREATE editing components as mRNA/sgRNA. (**B**) RNA-based GFP payload insertion at AAVS1 locus in Huh7 and HEK293T cells. Statistical analysis was performed using two-way ANOVA with Dunnett's multiple comparisons test comparing against the first sample. **** $p < 0.0001$. *** $p<0.001$. Statistical significance is labeled for samples with $p <0.01$. Data shown are representative of at least two independent experiments.

Fig. 1

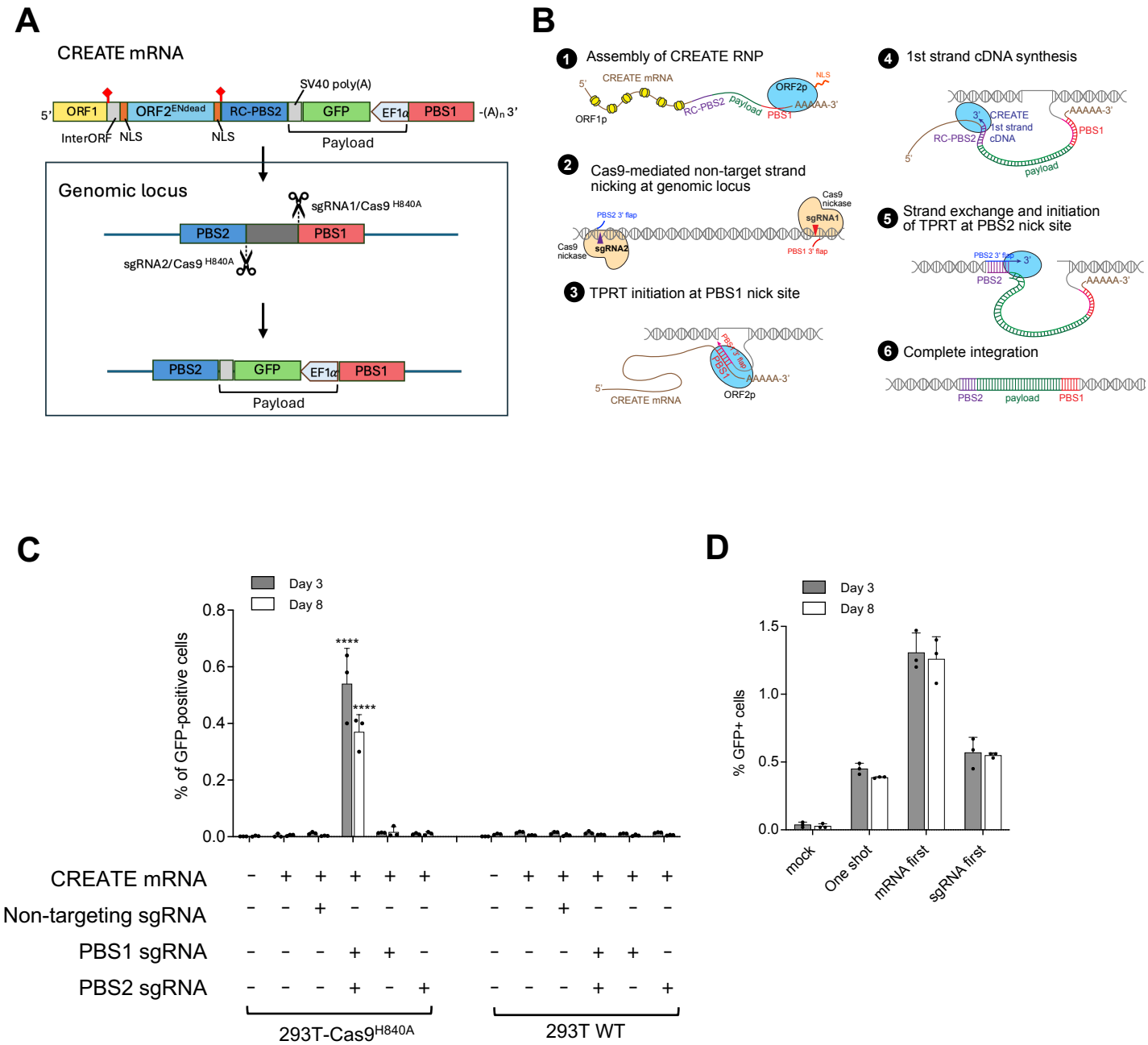

A
CREATE mRNA
SV40 poly(A)
5'
ORF1
ORF2ENdead
RC-PBS2
GFP
EF1α
PBS1
-(A)n 3'
InterORF
NLS
NLS
Payload
Genomic locus
sgRNA1/Cas9 H840A
PBS2
PBS1
sgRNA2/Cas9 H840A
PBS2
GFP
EF1α
PBS1
Payload
B
1 Assembly of CREATE RNP
CREATE mRNA
5'
ORF1p
RC-PBS2
payload
PBS1
ORF2p
NLS
AAAAA-3'
2 Cas9-mediated non-target strand nicking at genomic locus
PBS2 3' flap
Cas9 nickase
sgRNA2
Cas9 nickase
sgRNA1
PBS1 3' flap
3 TPRT initiation at PBS1 nick site
5'
CREATE mRNA
ORF2p
AAAAA-3'
4 1st strand cDNA synthesis
3'
CREATE 1st strand cDNA
RC-PBS2
AAAAA-3'
PBS1
5'
payload
5 Strand exchange and initiation of TPRT at PBS2 nick site
PBS2 3' flap
3'
PBS2
AAAAA-3'
6 Complete integration
PBS2
payload
PBS1
C
Day 3
Day 8
% of GFP-positive cells
0.8
0.6
0.4
0.2
0.0
****
****
CREATE mRNA
Non-targeting sgRNA
PBS1 sgRNA
PBS2 sgRNA
293T-Cas9H840A
293T WT
D
Day 3
Day 8
% GFP+ cells
1.5
1.0
0.5
0.0
mock
One shot
mRNA first
sgRNA first

Fig. 2

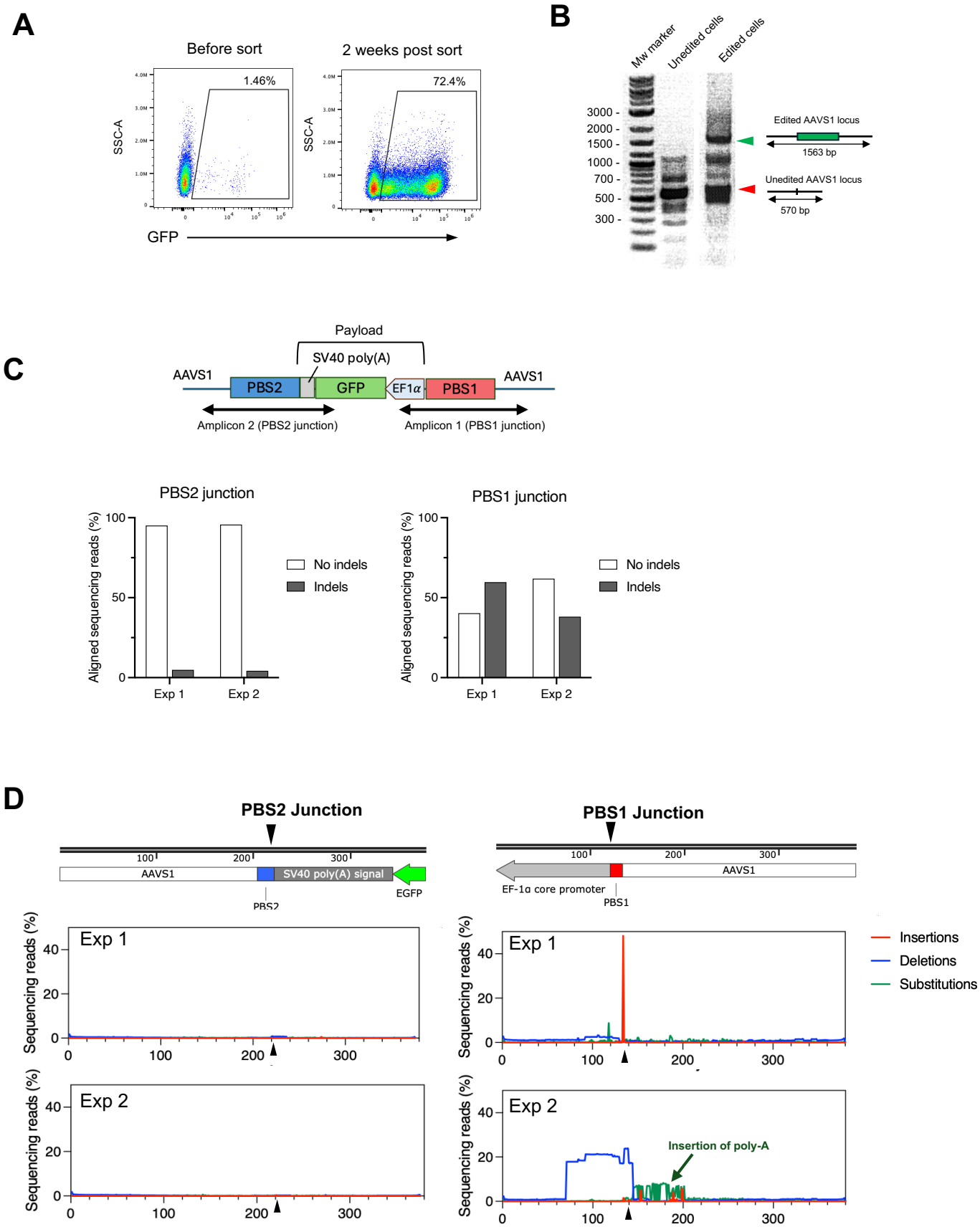
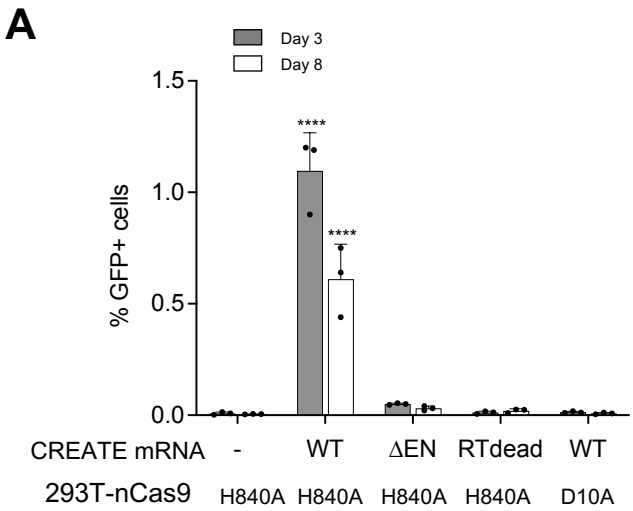
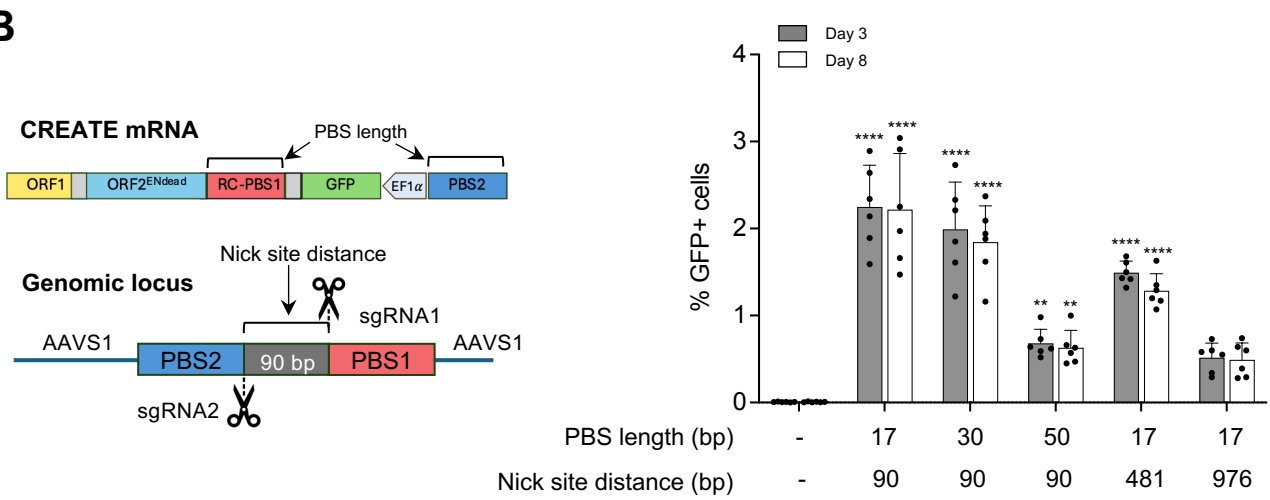

A
Before sort
2 weeks post sort
1.46%
72.4%
SSC-A
GFP
B
Mw marker
Unedited cells
Edited cells
3000 -
2000 -
1500 -
1000 -
700 -
500 -
300 -
Edited AAVS1 locus
1563 bp
Unedited AAVS1 locus
570 bp
C
Payload
SV40 poly(A)
AAVS1
PBS2
GFP
EF1α
PBS1
AAVS1
Amplicon 2 (PBS2 junction)
Amplicon 1 (PBS1 junction)
PBS2 junction
PBS1 junction
Aligned sequencing reads (%)
100
50
0
No indels
Indels
Exp 1
Exp 2
D
PBS2 Junction
PBS1 Junction
100
200
300
AAVS1
SV40 poly(A) signal
EGFP
PBS2
EF-1α core promoter
PBS1
Sequencing reads (%)
Exp 1
Exp 2
Insertions
Deletions
Substitutions
Insertion of poly-A

Fig. 3


A
Day 3
Day 8
% GFP+ cells
1.5
1.0
0.5
0.0
****
****
CREATE mRNA
-
WT
ΔEN
RTdead
WT
293T-nCas9
H840A
H840A
H840A
H840A
D10A


B
CREATE mRNA
PBS length
ORF1
ORF2ENdead
RC-PBS1
GFP
EF1α
PBS2
Nick site distance
Genomic locus
sgRNA1
AAVS1
PBS2
90 bp
PBS1
AAVS1
sgRNA2
Day 3
Day 8
% GFP+ cells
4
3
2
1
0
PBS length (bp)
-
17
30
50
17
17
Nick site distance (bp)
-
90
90
90
481
976

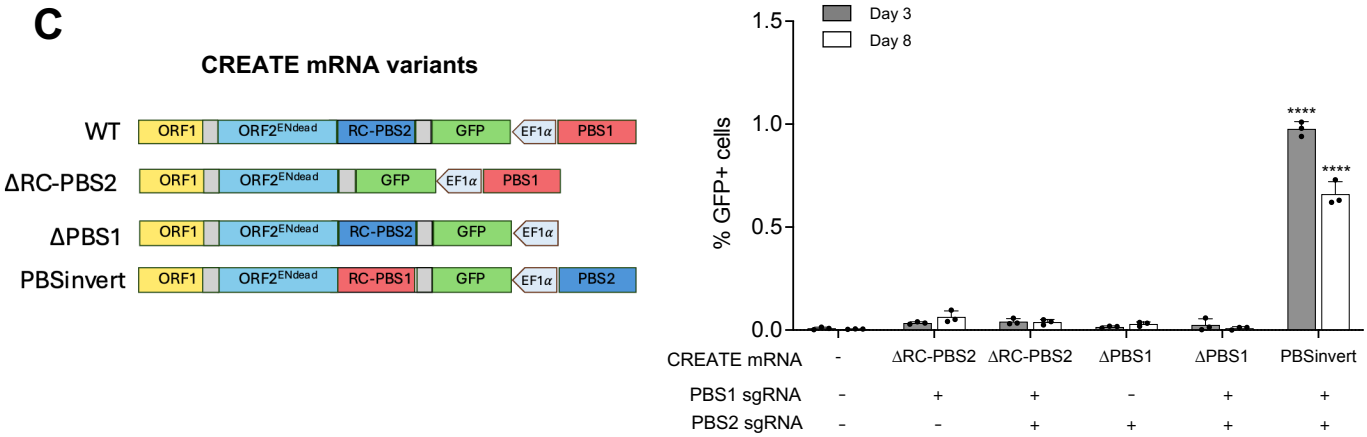

C
CREATE mRNA variants
WT
ΔRC-PBS2
ΔPBS1
PBSinvert
ORF1
ORF2ENdead
RC-PBS2
RC-PBS1
GFP
EF1α
PBS1
PBS2
Day 3
Day 8
% GFP+ cells
1.5
1.0
0.5
0.0
CREATE mRNA
-
ΔRC-PBS2
ΔRC-PBS2
ΔPBS1
ΔPBS1
PBSinvert
PBS1 sgRNA
PBS2 sgRNA

Fig. 4

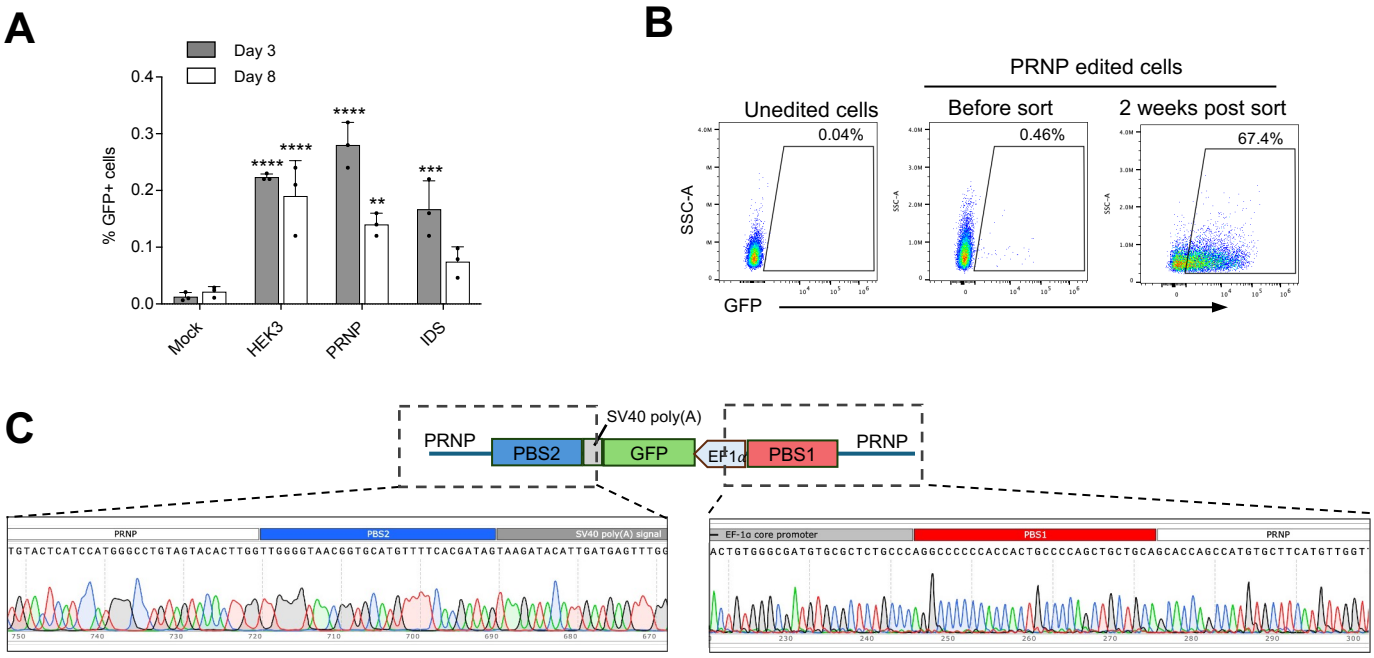
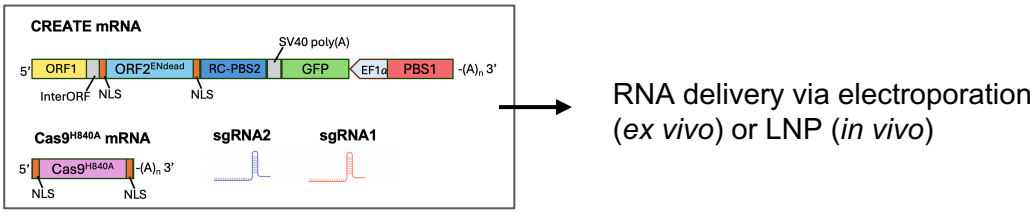
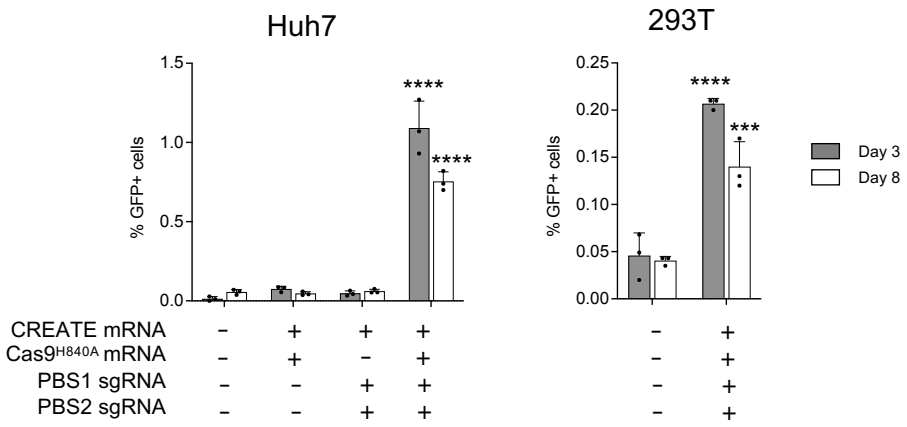

A
Day 3
Day 8
% GFP+ cells
0.4
0.3
0.2
0.1
0.0
****
****
****
**
***
Mock
HEK3
PRNP
IDS
B
PRNP edited cells
Unedited cells
Before sort
2 weeks post sort
0.04%
0.46%
67.4%
SSC-A
GFP
C
PRNP
PBS2
SV40 poly(A)
GFP
EF1α
PBS1
PRNP
PRNP
PBS2
SV40 poly(A) signal
EF-1α core promoter
PBS1
PRNP

Fig. 5

A


CREATE mRNA
SV40 poly(A)
5'
ORF1
ORF2ENdead
RC-PBS2
GFP
EF1α
PBS1
-(A)n 3'
InterORF
NLS
NLS
Cas9H840A mRNA
sgRNA2
sgRNA1
5'
Cas9H840A
-(A)n 3'
NLS
NLS
RNA delivery via electroporation
(ex vivo) or LNP (in vivo)

B


Huh7
293T
% GFP+ cells
1.5
1.0
0.5
0.0
0.25
0.20
0.15
0.10
0.05
0.00
****
****
****
***
Day 3
Day 8
CREATE mRNA
Cas9H840A mRNA
PBS1 sgRNA
PBS2 sgRNA