

Discovery and Engineering of Retrons for Precise Genome Editing

Jesse D. Buffington^{1,*}, Hung-Che Kuo^{1,*}, Kuang Hu^{1,*}, You-Chiun Chang^{1,*}, Kamyab Javanmardi¹, Brittney Voigt², Yi-Ru Li¹, Mary E. Little¹, Sravan K. Devanathan¹, Blerta Xhemalçe¹, Ryan S Gray², and Ilya J. Finkelstein^{1,3,*}

¹*Department of Molecular Biosciences, University of Texas at Austin, Austin, TX, 78712, USA*

²*Departments of Nutrition and Pediatrics, Dell Pediatric Research Institute, The University of Texas at Austin, Dell Medical School, Austin, TX 78712, USA.*

³*Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, TX, 78712, USA*

*equal contribution

Abstract

Retrons are promising gene editing tools because they can produce multi-copy single-stranded DNA in cells via self-primed reverse transcription. However, their potential for inserting genetic cargos in eukaryotes remains largely unexplored. Here we report the discovery and engineering of highly efficient retron-based gene editors for mammalian cells and vertebrates. Through bioinformatic analysis of metagenomic data and functional screening, we identified novel retron reverse transcriptases (RTs) that are highly active in mammalian cells. Rational design further improved the editing efficiency to levels comparable with conventional single-stranded oligodeoxynucleotide donors but from a genetically encoded cassette. Small molecule inhibitors of non-homologous end joining factors and Cas9-DNA repair protein fusions further increase homology-directed repair. Retron editors also exhibited robust activity with Cas12a nuclease and Cas9 nickase, expanding the genomic target scope and bypassing the need for a DNA double-stranded break. Using a rationally engineered retron editor, we incorporate a split GFP epitope tag for live cell imaging. Finally, we develop an all-RNA delivery strategy to enable DNA-free gene editing in cells and vertebrate embryos. This work establishes retron editors as a versatile and efficient tool for precise gene editing, offering new opportunities for biotechnology and biomedical research.

Keywords: CRISPR, retron, gene editing, reverse transcriptase, templated repair

*correspondence: ilya@finkelsteinlab.org

Introduction

Precise genome editing is a cornerstone of biomedicine and gene therapy. However, installing specific edits via templated homology-directed repair (HDR) is limited by the challenge of delivering and integrating an exogenous template DNA into the genome [1–3]. The most efficient delivery methods are viral vectors and synthetic DNA donors [4–7]. Viral vectors can induce genotoxicity via insertional mutagenesis, are depleted during the gene editing window, and are not suitable for multiplexed applications [5, 7, 8]. Synthetic DNA templates must be transfected or electroporated into cells, do not target the nucleus, and are incompatible with RNA-based delivery in cells and organisms [4, 6]. Therefore, recent gene editing technologies have begun to harness reverse transcriptases (RTs) that can continuously generate the template DNA near its edit site. [9–35]. Among these approaches, retrons-RTs are especially promising tools because they are self-priming and can generate high copies of long single-stranded DNAs (ssDNAs) *in vivo*.

Retrons are bacterial anti-phage defense systems that consist of a self-priming reverse transcriptase (retron-RT), a cognate non-coding (ncRNA) that primes & templates reverse transcription, and an accessory protein that participates in anti-viral immunity [36–44]. The ncRNA consists of two main regions: the msr (msDNA-specific region) and the msd (multi-copy ssDNA-coding region) (Fig. 1A). The msr is located at the 5' end of the ncRNA and forms a specific structure that is recognized by the RT [45, 46]. This region typically contains one to three stable stem-loops with 7–10 base pair stems and 3–10 nucleotide loops. The msr also includes a highly conserved guanosine residue at the 5' end, which serves as the branching point for initiating reverse transcription (Fig. 1A). The msd is positioned downstream of the msr and can be divided into two parts: a dispensable region that can be replaced with a desired sequence (i.e., the donor DNA for genome editing) and a conserved region that is essential for the proper folding and function of the ncRNA. The RT primes from the msr and uses the msd as a template [45, 46]. The resulting msDNA remains covalently linked to the ncRNA through a 2',5'-phosphodiester bond formed between the branching guanosine residue in the msr and the 5' end of the msDNA [45, 46]. The host RNase H degrades the RNA-DNA hybrid to expose the ssDNA (Fig. 1A, right) [39]. By replacing the dispensable msd region with a user-programmable sequence, retrons-RTs can generate templates for homology-directed DNA repair in cells.

Recent studies have demonstrated the potential of retrons coupled with CRISPR-Cas9 to enhance precise genome editing in bacteria, yeast, plants, and mammalian cells [17, 19, 24, 30–32, 34] (Fig. 1B). Early studies fused the Cas9 single guide RNA (sgRNA) to a retron ncRNA with a modified donor msd [32]. This increases the local concentration of the donor template at a double-stranded DNA break (DSB), biasing repair toward templated HDR (Fig. 1C). Alternative designs fused the retron RT to Cas9 directly with the same ultimate goal [30, 34]. Despite these advances, the potential of retrons for precise genome editing has yet to be fully realized. To date, only a handful of retron RTs have been benchmarked in mammalian cells. Further engineering of the retron editor system, along with the ncRNA, and delivery methods can further improve editing efficiency. Developing a flexible framework for retron editor optimization will greatly expand the utility of this tool across all domains of life.

Here, we report the discovery and engineering of a highly efficient retron gene editor. We bioinformatically identified >500 high-confidence retrons from metagenomic sources. Using a functional reporter system, we screened 98 variants in mammalian cells and identified 17 RTs that were more active than the previously established Eco1-RT. Further rational design achieved editing efficiencies that are comparable to conventional single-stranded oligodeoxynucleotide (ssODN) donors. Steering DNA repair outcomes towards HDR via small molecule inhibitors and Cas9-DNA repair protein fusions boosted targeted DNA in-

sertion. Retron editors also function with Cas12a, significantly broadening their genomic target range. The nickase Cas9(D10A) also supports retron editing, and this activity can be improved with DNA repair protein fusions. We apply retron editors for installing in-frame epitopes for live cell imaging in U2OS cells. Finally, we demonstrate all RNA-based retron editing in cell lines and vertebrates. Broadly, this work establishes retron editors as a powerful tool for templated cargo insertion, opening new gene editing modalities.

Results

A metagenomic survey identifies retron-RTs active in mammalian cells

We hypothesized that a metagenomic survey of retron-RTs would uncover variants that improve homology-directed repair in heterologous hosts. Towards this goal, we developed a pipeline to phenotypically assess retron-RTs using fluorescent proteins in HEK293T cells (**Fig. 1D**). The reporter expresses RFP and GFP that are separated by a ribosomal skipping T2A sequence [47]. RFP has a 9 base pair (bp) deletion ($\Delta 9$) adjacent to a Y64L mutation [48]. These mutations ensure that RFP($\Delta 9$) is dark until the wild type (WT) sequence is restored via templated HDR following a Cas9-generated double-stranded break (DSB). GFP serves as a transfection control and also reports on Cas9-generated insertions and/or deletions (indels) that shift the open reading frame out of frame. As expected, transfecting Cas9 and a single-stranded oligodeoxynucleotide (ssODN) donor restored RFP fluorescence (**Fig. 1E, middle**). The well-characterized Eco1-RT also repaired RFP, although HDR activity was substantially lower than the ssODN (**Fig. 1E, left**). The dispensable msd region included 29 nt of homology flanking a 9 nt insertion that paired with the target strand and also reverted the Y64L mutation. Having established this assay, we proceeded to test additional RTs from diverse microbes.

Retron-RTs are ubiquitous throughout bacteria, but only a handful have been tested experimentally [33]. Therefore, we developed a bioinformatics pipeline to identify new retron-RTs from metagenomic sources (**Fig. S1A**). We first annotated all RTs in the NCBI database of non-redundant bacterial and archaeal genomes, as well as 2M partially assembled bacterial genomes from the human microbiome [49, 50]. We reasoned that human microbiome-derived RTs will also be active under physiological conditions. After identifying likely retron-RTs, we searched for the msr-msd non-coding RNA (ncRNA) and accessory proteins. This search identified >500 high-confidence, non-redundant retrons with well-annotated msr-msds (**Fig. S1B**). We classified these systems into a phylogenetic tree and sorted them into 11 clades following a prior bioinformatic survey (**Fig. 1F**) [51]. The highest-confidence systems across multiple clades were prioritized for experimental characterization in mammalian cells.

Next, we screened 98 retron-RTs using flow cytometry and the RFP($\Delta 9$) reporter (**Fig. 1G**). As expected, transfecting the Cas9-sgRNA plasmid with a ssODN partially restored the RFP signal (**Fig. 1G, inset**). Thirty-one RTs (32% of all tested systems) restored RFP fluorescence (defined as >1% RFP+ cells). Eco1-RT restored RFP+ signal in 5% of the cells. Ten RTs had >2-fold higher repair activity than Eco1-RT (**Fig. 1G**). Mva1-RT, derived from *Myxococcus vastator*, had an editing efficiency that was 6-fold higher than Eco1-RT in this transient repair assay (**Fig. 1G, inset**). The best-performing RTs all belonged to clade 9, indicating that these enzymes are especially active in mammalian cells, and/or that our bioinformatics workflow was most accurate in predicting the msr-msd sequences from this clade. The top systems also showed mostly RFP+ cells via confocal microscopy (**Figs. 1E, S2A**). To test these RTs in a genomic reporter, we also integrated the RFP cassette into the AAVS1 locus of HEK293T cells. While Eco1-RT was minimally able to restore RFP signal, the top six RTs were all active in the genetically integrated assay (**Fig. S2B**).

Escherichia fergusonii (Efe1)-RT was the most active in the genomic reporter, restoring RFP+ signal ~10-fold better than Eco1-RT.

Retron-RTs co-evolve with a cognate msr-msd, but their ability to reverse transcribe from the msr-msd of other retrons is unknown. Therefore, we explored the feasibility of using two or more orthogonal RTs for multiplexed retron editing (**Fig. 1H**). We tested the RFP repair activity of six active novel RTs, along with Eco1-RT, when co-expressed with the msr-msd from other systems. Mva1-RT, the most broadly cross-reactive RT, only shared 33-36% amino acid sequence identity and 49-59% msr-msd nucleotide identity with all other systems (**Fig. S3A**). In contrast, Efe1-RT shared 44-46% amino acid sequence identity but remained exclusive to its cognate msr-msd. The remaining RTs could use a broad range of msr-msds, including those from Eco1-RT. *Vibrio rotiferianus* (Vro1)- and *Vibrio aphrogenes* (Vap1)-RTs showed comparable activity with *Proteus* sp. (Psp1) msr-msd and their native msr-msds. Notably, all RTs shared a similar msr-msd secondary structure, including a palindromic repeat and an extended msd hairpin (**Fig. S3B**). We conclude that retron-RTs are more flexible in msr-msd utilization than previously appreciated and that caution should be taken when combining multiple systems in a single cell line. Further, Efe1-RT is the most active enzyme in the genomic reporter assay and strikes an excellent balance between high activity and specificity for gene editing applications.

Next, we tested the top-performing RTs for their ability to integrate a 10 nt cargo into native genomic loci (**Fig. 2**). Two rounds of PCR were used to generate libraries for next-generation DNA sequencing (NGS) (**Fig. 2A**). The first PCR reaction was primed outside the homology arms to avoid amplifying the reverse transcribed ssDNA. The second PCR reaction barcoded the amplicons for short-read NGS. Cas9 and a ssODN were used as a positive control, and to benchmark the RT fidelity. We distinguished retron editing efficiency from Cas9-generated indels by scoring the percentage of modified reads that had the intended insert relative to all modified reads. Editing efficiencies ranged from 8-30%. Efe1-RT showed the highest editing activity at EMX1 and CFTR, consistent with the genetically integrated RFP(Δ 9) assay (**Fig. 2B**). Thus we selected Efe1-RT for all subsequent experiments.

Deep sequencing of the insert at the EMX1 locus showed that >99% of Efe1-RT driven insertion events contained the intended 10 nt cargo (**Fig. 2C, D**). The most frequent imperfect insertions included indels and single nucleotide substitutions that are the signature of alternative end joining pathways [52, 53]. To determine whether RT fidelity is contributing to this error, we conducted a similar analysis for the Cas9+ssODN experiment at EMX1 (**Fig. 2D**). Efe1-RT error rates were $\sim 10^{-3} - 10^{-4}$ (errors/nt), consistent with the substitution rates measured for high fidelity RTs (**Fig. 2D**) *in vitro* [54, 55]. In contrast, ssODN substitution rates were 10-fold greater at the same locus. These results indicate that RT-based editing fidelity exceeds that of conventional ssODNs and that repair fidelity may be limited by cell-intrinsic repair pathways (also see below).

We next tested retron editing across five native loci and as a function of the homology arm length. In all cases, inserting 10 nt was most efficient with 50 nt homology arms (7-28% insertion rates across five loci) (**Fig. 2E**). Editing efficiency was strongly correlated with Cas9 cleavage activity at each locus. Insertion efficiency generally matched, and in the case of EMX1 and HBB, exceeded the edit rates with Cas9 + ssODN. We conclude that short homology arms support the highest insertion rates in our system.

Rational Engineering of Retron Editors Increases Insertion Activity

To further improve insertion activity, we optimized ncRNA expression, nuclear localization signals (NLSs), and the nuclease-RT linker in an Efe1-RT-based retron editor (**Fig. 3**). The integrated RFP(Δ 9) reporter

was used for rapid iterative screens. Splitting the sgRNA and msr-msd into two transcripts increased editing efficiency (**Fig. 3A,B**). In the "split" design, sgRNA expression was driven by a U6 promoter, and the msr-msd was transcribed via the H1 promoter (**Fig. 3A**). These results suggest that fusing the sgRNA and the msr-msd may impact Cas9 and/or RT activity, possibly by misfolding the structural elements of each ncRNA, or by imposing additional steric constraints. All subsequent experiments used the split design.

Nuclear import can be a rate-limiting factor in mammalian gene editing [56–61]. Therefore, we tested 25 N- and C-terminal NLS combinations that were previously reported to improve Cas9-based gene editing (**Fig. 3C & Table 5**). A combination of N- and C-terminal bipartite SV40 (BP-SV40) NLSs showed the highest RFP repair activity. Adding cMyc-SV40 to the N-terminus and two or more SV40s to the C-terminus increased Cas9 cleavage 1.6-fold relative to an N-terminal SV40 and C-terminal NLP signal, as measured by a reduction in GFP+ cells. However, this did not increase templated DNA insertion (**Fig. 3C**). Retron-RTs interact extensively with their ncRNA and msDNA via their C-terminal domains [41, 62]. An extended C-terminal NLS may impair this interaction, reducing overall repair activity, but not Cas9-catalyzed DNA cleavage. We conclude that nuclear import is likely not the limiting factor for templated insertion.

The linker between the nuclease and the RT can also impact editing outcomes. We tested 15 linkers with various physical properties, including a ribosomal skipping T2A sequence between Cas9 and Efe1-RT (**Fig. 3D and Table 6**). Separating the two enzymes via a T2A sequence reduced RFP+ cells by 20% relative to the reference linker, (SGGS)₂-XTEN-(SGGS)₂ [63]. Next, we tested a panel of flexible (e.g., (GGS)_N) and rigid (e.g., (KL(A/E)AA)_n) linkers. Retron editors accommodated a broad range of linker designs (**Table 6**). However, multimerizing the RT on Cas9 via SpyTags or SunTags reduced RFP+ signal by 85–90% [64]. As Cas9-Sun/SpyTag-Cas9 fusions are reportedly active in mammalian cells, we speculate that multimerization disrupts RT activity [64, 65]. We conclude that retron editors can accommodate a broad range of linker geometries, and even split enzyme designs.

To expand the retron editor target range, we tested Efe1-RT with WT AsCas12a and AsCas12a Ultra at five genomic loci (**Fig. 3E-G**) [66]. Cas12a and Efe1-RT were fused via an (SGGS)₂-XTEN-(SGGS)₂ linker, and the crRNA and msr-msd were expressed from independent promoters. In all cases, AsCas12a Ultra-RT fusions had higher insertion activity than WT AsCas12a (**Fig. 3E**). This higher activity is due to the increased cleavage by AsCas12a Ultra relative to WT enzyme (**Fig. S4**). Deep sequencing of the BRD8 locus confirmed that 99% of the inserts had the insert sequence (**Fig. 3F**). The most frequent error was an indel outside of the immediate insert site, followed by mismatches within the insert. Base substitution rates across five genomic loci closely matched the pattern that was observed at EMX1 and were lower than those with Cas9 + ssODN donor (**Figs. 3G, 2D**). In sum, retron editors can be assembled from a broad range of ncRNA, NLS, & linker configurations, and can be paired with Cas9 and Cas12a nucleases to expand their target range.

Channeling repair pathway choice increases insertion efficiency.

DNA repair via non-homologous end joining (NHEJ) limits templated DNA insertion in mammalian cells [53, 67]. To further increase retron editing efficiency, we sought to channel repair away from NHEJ via two approaches (**Fig. 4A**). First, we focused on small molecule inhibitors that inhibit NHEJ or enhance HDR. AZD7648 and M3814 inhibit the DNA-dependent protein-kinase catalytic subunit (DNA-PKcs) to improve templated repair of Cas9 breaks [68–71]. TAK-931 is a CDC7-selective inhibitor that arrests cells in S phase, thereby increasing the HDR time window [72]. We first established the optimal working concentrations for each inhibitor (**Fig. S5A**). All three inhibitors improved insertion activity, with the strongest improvements

with AZD7648 at all loci (**Fig. 4B, left**). In contrast, M3814 showed strong improvements at all loci except HBB. Next, we tested whether retrons editors can insert larger cargos with 50 nt homology arms, and how this is modulated by inhibitors or DNA repair proteins (**Fig. 4C**). AZD7648 stimulated insertion of 25 and 50 nt inserts by 8.8 and 6.0-fold respectively at EMX1 (**Fig. 4D**). TAK-931 showed more modest 2.4 and 1.8 fold editing increases for 25 and 50 nt cargos respectively. A comparison of Cas9-generated indels and retron-driven insertions confirmed that AZD7648 doesn't increase Cas9 cleavage, but increases the utilization of a template ssDNA for genomic repair. Further, AZD7648 reduced the mutational signature at the target site, further highlighting the utility of repair pathway modulation in retron editing applications.

Cas9 fusion proteins can improve templated DNA insertion locally without perturbing repair pathways globally [53]. We focused on three fusions that have been previously characterized across multiple loci and cell types (**Figs. 4A, B, right**) [73–75]. Fusing Cas9 with the HDR-promoting CtIP and a dominant negative RNF168 (dnRNF168) increased retron editing efficiency 1.8-2.5 fold across five loci. A dominant-negative mutant of 53BP1 (DN1S) had variable effects across the five loci, with no improvements at EMX1. (**Fig. 4B, right**) [74]. Fusion of Cas9 to the N-terminal region of human Geminin limits Cas9 expression to the S/G2/M phases of the cell cycle when HDR is most active [76]. However, Cas9-hGem1/100 fusions either decreased or had no effect on retron editing activity (**Fig. 4B, right**). This result may reflect the mechanistic differences between ssODN and double-stranded donor DNA repair pathways (see **Discussion**). Combining Cas9-CtIP-dnRNF168 fusions with AZD7648 and TAK-931 did not improve editing any further at EMX1 (**Fig. S5B**). This finding is consistent with the hypothesis that Cas9-CtIP-dnRNF168 and AZD7648 both down-regulate NHEJ, albeit via different mechanisms. In contrast, retron-mediated insertion with Cas9-DN1S and Cas9-hGem1/110 fusions was higher with AZD7648 and TAK-931, but never exceeded AZD7648 or Cas9-CtIP-dnRNF168 alone. We conclude that selective inhibition of DNA-PKcs significantly improves retron-mediated gene editing.

Next, we tested retron editing activity with the nickase Cas9(D10A), and its fusions with DNA repair proteins (**Fig. S5C-E**). On its own, Cas9(D10A) paired with Efe1-RT caused an insertion in 1% of all reads at EMX1. However, fusing Cas9(D10A) to CtIP-dnRNF168 and a hyperactive HR recombinase hRad51(K133R) increased insertion rates >20-fold [77] (**Fig. S5D**). We also reasoned that fusing Cas9(D10A) to a helicase could improve HDR by peeling back the target and non-target DNA strands for the retron-synthesized DNA. We tested this hypothesis with Rep-X, an engineered helicase with exceptional processivity [78]. Cas9(D10A)-Rep-X helicase fusions increased insertion efficiency >20-fold, suggesting that retron editors can be used for DSB-free editing. Additional studies will be required to further optimize Cas9(D10A)-retron-RT editors and their combinations with DNA repair factors (see **Discussion**).

Retron editing in cell lines and vertebrates

As a proof of principle, we next used retrons editors to insert a split GFP for live cell imaging of endogenously expressed proteins in U2OS cells. GFP1-10, comprised of the first 10 GFP β -strands, is expressed from an integrated & inducible promoter (**Fig. S6**). The 11th β -strand is fused to the protein of interest via a short linker. GFP1-10 is not fluorescent until it is complemented by GFP11 because chromophore maturation requires a critical GFP11-encoded glutamic acid [79]. Thus, fusing GFP11 to a target protein allows visualization of sub-cellular localization in live cells [80–82].

We targeted two proteins for endogenous GFP11 tagging via retrons editors. The dispensable msd region was replaced with a 197 nt DNA template that included 70 nt homology arms, a nine nt linker encoding Gly-Gly-Gly, and the 48 nt GFP11 epitope. To maximize the integration efficiency, cells were

transfected with Cas9-CtIP-dnRNF168 (see **Channeling repair pathway choice increases insertion efficiency**). 72-hours post-transfection, GFP1-10 was induced for 48 hours and GFP+ cells were collected using fluorescence-activated cell sorting (FACS). Confocal imaging of the insertion region confirmed the genomic edit (**Fig. S6C**). These results demonstrate that Efe1-RT can synthesize 200 nt ssDNAs. More broadly, this approach can be readily used for installing epitopes, disease-specific alleles, and other large insertions across the entire proteome from a genetically encoded cassette.

To expand retror editors beyond plasmid-based expression, we tested delivery in an all-RNA format (**Fig. 5**). In these experiments, Cas9 and Efe1-RT mRNAs were *in vitro* transcribed, 5'-capped, and 3'-polyA-tailed. The msr-msd was *in vitro* transcribed, and the sgRNAs were chemically synthesized (**Fig. 5A**). The four RNA cocktail were transfected into HEK293T cells and editing efficiency was scored by deep sequencing. Insertion of a 10 nt cargo was 9.9, 11.4, and 9.3% for the *EMX1*, *CFTR*, and *AAVS1* loci, respectively (**Fig. 5B**). This lower editing efficiency is likely due to the reduced Cas9 and RT expression from mRNA relative to plasmid-based delivery. Optimizing the UTRs, chemical modifications, and mRNA to ncRNA ratio may further boost editing at therapeutically relevant loci. We conclude that retror editors are compatible with direct RNA delivery.

We next tested the repair of a pathogenic mutation via mRNA-based retror editing in zebrafish. We designed a sgRNA to target a pathogenic mutation in the Kinesin Family Member 6 *kif6*^{ut20} gene. Mutations in *kif6*^{ut20} cause scoliosis in zebrafish and are linked to neurological defects in humans [83]. The msd was designed to correct two base mutations that reverted a pathogenic Pro→Thr substitution. In addition, we introduced a silent T→C mutation that abolished a Bsal cleavage site for downstream restriction enzyme analysis (**Fig. 5E-F**). Embryos were injected with the sgRNA, msr-msd, and a fused Cas9-RT or split Cas9 and Efe1-RT mRNAs. Genomic DNA was harvested 24 hours post-injection, and submitted to NGS, restriction enzyme digestion, and Sanger sequencing (**Fig. 5D-F**). As expected, untreated controls did not show any editing. Injecting Cas9 and Efe1-RT mRNAs, along with the two ncRNAs, induced edits up to 10% of all reads from crude genomic preparations. The fused Cas9-RT mRNA showed reduced editing as compared to the split mRNAs (**Fig. 5E**). This may be due to the stability of the fusion mRNA in zebrafish. To further confirm retror editing, we sub-cloned the genomic DNA and subjected it to restriction enzyme and Sanger sequencing analysis (**Fig. 5F-G**). Bsal treatment of the edited, but not WT and *kif6*^{ut20} embryos resulted in a single band, indicating the expected cleavage pattern. Sanger sequencing of the insert site sub-cloned from an edited embryo also showed the expected edits. Collectively, these results indicate that retror editors are active in cell lines and zebrafish embryos when delivered as an all-RNA package.

Discussion

Here, we engineer retror editors for precise genome engineering in mammalian cells and zebrafish embryos. Our metagenomic survey yielded 17 RTs that outperform Eco1-RT in mammalian cells. Efe1-RT, the lead candidate, is highly active, specific for its cognate RNA, and is capable of generating at least ~200 nt ssDNAs *in vivo* (**Fig. 1**). The most active RTs in our survey were all derived from clade 9. A recent bacterial functional screen also concluded that RTs in this clade generate high ssDNA levels in *E. coli* [17]. Additional structure-function studies will be required to dissect the mechanistic basis for this higher activity. Finally, we demonstrated retror editor delivery in an all-RNA format (**Fig. 5**). We anticipate that further optimizations, such as increasing mRNA and ncRNA stability will further improve editing efficiency (see below).

We iteratively improved on prior designs by optimizing the NLS, nuclease-RT linkers, and nuclease combinations (**Fig. 3**). The linker and NLS are key design components for both base and prime editors

[84, 85]. In contrast, we show that retron editors can accommodate diverse NLS and linker combinations. Importantly, retron editors are compatible with Cas9, Cas12a, and even the nickase Cas9(D10A) (**Fig. 3, S5**). Cas12a-based retron editors further expand the potential target range and create opportunities for multiplexed retron editing due to Cas12a's ability to process its crRNA [84]. Based on our results, we anticipate that retron-RTs will be compatible with other established, i.e. transcription activator-like effectors (TALEs), and emerging RNA/DNA-guided nucleases. Importantly, further development of nickase-based retron editors promises to avoid the induction of double-stranded DNA breaks (**Fig. S5**).

Retron editors are uniquely capable of synthesizing high copy numbers of the repair template at the excision site [48, 86, 87]. Boosting ssDNA synthesis via RT and ncRNA engineering will further improve the processivity, fidelity, and ultimately, edit length and efficiency. For example, rational engineering of a retron RT-based prime editor boosted editing efficiency >8 -fold [11]. The design of the msr-msd and repair template relies on empirical testing of different msr-msd designs and offers an additional area for improvement. This includes modifying the insertion and truncation site of the native msd, homology arm length, target/non-target strand selection, and overall RNA structure. RNA circularization, structured RNA pseudoknots, and chemical modifications also increase ncRNA stability in mammalian cells [88]. Further optimization in these directions may enhance the overall efficiency of retron editing. In general principles and predictive algorithms for msr-msd and repair templates will further improve retron editors.

The repair of DSBs with a single-stranded DNA donor proceeds via single-strand template repair (SSTR) [89–91]. Our results are also consistent with an SSTR-based repair mechanism for retron editors. First, NHEJ inhibitors increase retron editing and reduce NHEJ-associated indels for both Cas9 and Cas12a across all tested target sites (**Fig. 4**). Second, fusing Cas9 with a dominant-negative allele of 53BP1, or with the DNA resection promoting CtIP, also increases retron editor efficiency. Third, we observed that 36–50 nt homology arms are optimal for retron editors (**Fig. 2**). Similarly, SSTR is maximized with 30–60 nt homology arms [92, 93]. SSTR is a RAD52-dependent process in yeast and human cells, suggesting that nuclease-RAD52 and/or RT-RAD52 fusions may boost retron editing [75, 94]. SSTR competes with two error-prone DSB repair pathways: classical NHEJ and polymerase theta-mediated end joining (TMEJ) [52, 95]. Dual inhibition of NHEJ and TMEJ may further synergize with RAD52 fusions [47, 69]. Rational design of asymmetric templates, cleavage-blocking mutations, and dual Cas9 nickases are additional avenues for maximizing editing efficiency [93, 96]. Mechanistic studies of retron editor-mediated repair will further improve editing outcomes across all domains of life.

In conclusion, retron editors are emerging as a highly promising gene editing tool. Their unique ability to accurately insert or replace sizeable DNA segments opens up possibilities for correcting complex genetic mutations that were previously challenging to address. Compatibility with an all-RNA formulation opens new avenues for therapeutic delivery into cells and organisms. Additionally, retron editors are poised to broaden the scope of high-throughput functional screens, allowing for the characterization of complex genetic variants with single-base resolution. Integration of retron editors into existing screening pipelines holds great promise for advancing our understanding of gene function and regulation, ultimately paving the way for novel therapeutic interventions and biotechnological applications.

Materials and Methods

Oligonucleotides and Plasmids

Retron-RT and ncRNAs gene blocks were ordered from IDT or Twist Biosciences and cloned into a GFP dropout entry vector via Golden Gate assembly. mRNAs were purchased from Cisterna Biologics. Retron editor expression plasmids were assembled by combining the sgRNA, retrone msr-msd, the RT, and SpCas9 or AsCas12a in a *ccdB* dropout mammalian expression vector using Golden Gate assembly.

Tissue Culture

HEK293T cells were generously provided by Professor Xiaolu A. Cambronne. HEK293T and U2OS were cultured in Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum (Gibco) and 1% Penicillin-Streptomycin (Gibco). All cell lines were cultured and maintained at 37 °C and 5% CO₂. U2OS Flp-In TREx - GFP1-10 were generated by stably integrating a GFP1-10 construct at the single FRT locus through dual transfection of pCDNA5/FRT/TO/Intron-eGFP1-10 and pOG44 plasmids in a 1:10 ratio and subsequent selection with hygromycin at 200 µg/mL and blasticidin at 15 µg/mL.

Metagenomic Discovery

To find new retrons, we searched the NCBI genome and metagenomic contigs for the retrone-RT gene using hidden Markov models (HMMER, [97, 98]) and a database containing all experimentally validated retrone-RT with a permissive e-value threshold of 10^{-4} [33, 49, 50]. We next established a pipeline to detect structured RNAs near retrone RT sequences. Initially, each experimentally validated msr-msd transcript and its close counterparts served as seeds for CMfinder 0.4.1, an RNA motif predictor that leverages both folding energy and sequence covariation [99]. Covariate models were then crafted with Infernal suite's cmbuild and used to search for analogous structures around the start of the RT open reading frame (ORF). We manually inspected the msr-msd regions of all retrone candidates, especially those that didn't return any hits via the automated pipeline. RNAfold was used to inspect structured regions, and to compare them to known msr-msd transcripts [100]. We then used MAFFT-Q-INS-i for multiple alignments, focusing on identifying conserved sequences in related genomes. Subsequently, we pruned the alignments at the a1 and a2 areas, cycling back to earlier pipeline steps to acquire more sequences from both ends. R-scape was used to pinpoint covarying base pairs in the suggested consensus structures, mitigating the influence of phylogenetic correlations and base composition biases not attributed to conserved RNA structure. Accessory retrone genes adjacent to the RT were annotated using the Pfam database as a query and HMMER. High-confidence retrone systems were manually inspected to confirm the expected protein domains, catalytic residues, and ncRNA structure. To remove redundant sequences, all putative hits were clustered with CD-HIT [101], setting a sequence identity threshold at 90% and an alignment overlap of 80%. The clustered datasets were then transformed into phylogenetic trees, and candidates for validation were selected based on their positioning in the tree.

To organize the retrone RTs into phylogenetic groups, we employed MAFFT software and used progressive methods for multiple sequence alignments (MSAs) [102]. An MSA was constructed from the RT0–7 domain of 1,912 sequences, sourced from a dataset of 9,141 entries previously categorized as retrone/retron-like RTs and an additional 16 RTs from experimentally verified retrons [33]. Phylogenetic trees were gen-

erated using FastTree, applying the WAG evolutionary model, combined with a discrete gamma model featuring 20 rate categories. Specifically, the RT tree was crafted using IQ-TREE v1.6.12, incorporating 1000 ultra-fast bootstraps (UFBoot) and the SH-like approximate likelihood ratio test (SH-aLRT) with 1000 iterations [103]. The best-fit model, identified by ModelFinder as the LG+F+R10 due to its minimal Bayesian Information Criterion (BIC) value among 546 protein models, was used. The RT Clades' internal nodes exhibited UFBoot and SH-aLRT support values exceeding 85% [101].

Plasmid-based fluorescent reporter assays

1.2×10^5 HEK293T cells were seeded in a 24-well plate 18-24 hours before transfection. 0.35 μ g of the retrorviral editor plasmid and 0.35 μ g of the fluorescent reporter plasmid were co-transfected using Lipofectamine 2000 (Invitrogen). Cells were trypsinized and collected for flow analysis 72 hours after transfection. Flow analysis was conducted on a Novocyte flow cytometer (ACEA Biosciences). Cells were gated to exclude dead cells and doublet and 10,000 cells were analyzed in all samples. Cells were then gated by FITC-A (x-axis) and PE-Texas Red-A (y-axis). The editing efficiency was reported as the percentage of cells in the quadrant of FITC-A and PE-Texas Red-A.

Genomic reporter assays

The fluorescent reporter was cloned in a plasmid designed for Bxb1 recombinase-driven landing pad system [104]. This plasmid was transfected into landing pad HEK293T cells followed by doxycycline induction and AP1903 selection to generate stably integrated reporter cells. 1.2×10^5 of HEK293T reporter cells were seeded in a 24-well plate 18-24 hours before transfection. 1 μ g of the retrorviral editor plasmid was transfected using Lipofectamine 2000. 48-72 hours after transfection, cells were treated with doxycycline to induce the expression of the fluorescent reporter. Cells were trypsinized and collected for flow analysis and genomic DNA extraction (Qiagen DNeasy Blood and Tissue kit).

Confocal imaging

HEK293T cells were seeded and transfected as described for the plasmid-based fluorescent reporter assay (see above). 48 hours post-transfection, cells were seeded into 15 mm glass-bottom cell culture dishes (NEST) and incubated for an additional 24 hours. Cells were then imaged with a Nikon Ti2 Spinning Disk Confocal Microscope at 20x magnification. 8858 x 8858-pixel images were acquired and processed using ImageJ.

U2OS FlpIn TReX GFP1-10 cells were transfected with 1 μ g retrorviral editor plasmids that target either the N- or C-terminus of the protein of interest to insert a GFP₁₁ fragment. 48-72 hours after transfection, cells were expanded into 10-cm plates. Cells with high GFP intensities were then sorted via a cell sorter (Sony MA900). For confocal imaging, cells were incubated with doxycycline for 48-72 hours before imaging to induce the expression of the GFP1-10 construct. Image acquisition was performed with live cells under spinning-disk confocal microscopy (Olympus).

Next generation DNA sequencing

Genomic samples were subjected to two rounds of PCR for NGS library preparation 2. The first round of PCR was performed using the KOD One PCR master mix (TOYOBO). Primers were designed ~600 base

pairs away from the cut site on each side to avoid amplifying the RT-generated ssDNA. PCR products were gel purified and barcoded via a second round of PCR with Illumina P5/P7 adapters using Q5 HotStart High Fidelity master mix (NEB). PCR amplicons were sequenced on an Illumina Novaseq. Reads were demultiplexed using NovaSeq Reporter (Illumina). Alignment of amplicon sequences to a reference sequence was performed using CRISPResso2 [105]. The aligned sequences were first checked for the expected cargo insertion. Next, the homology arms and insertion sequences were further checked for any mismatches or deletions. We considered a “perfect edit” as any sequence that had the expected insertion only, with no additional mismatches and deletions. If these events also had a mismatch or deletion, we counted it as an “imperfect” edit. Any insertions, deletions, or mismatches that did not have the intended cargo were surmised to be from nuclease-only activity.

In vitro transcription of the msr-msd

The msr-msd was PCR amplified from a plasmid or gene block with a T7 promoter. The ncRNA was generated using the HiScribe T7 High Yield RNA Synthesis Kit (NEB) according to the manufacturer’s protocol. RNA products were purified using the RNeasy Mini Kit (Qiagen).

Zebrafish Maintenance and Gene Editing

All zebrafish experiments were performed according to University of Texas at Austin IACUC standards. Embryos were raised at 28.5 °C in fish water (0.15% Instant Ocean in reverse osmosis water). Wildtype and *kif6*^{ut20(P293T)} mutants were in-crossed, and one-cell stage embryos were injected with 1 nL of the injection mix using a microinjector-pump system (World Precision Instruments Nanoliter Injector and PV 820 Pneumatic PicoPump). Injected embryos and un-injected sibling controls were incubated at 28.5 °C in fish water until 24 hr post fertilization, at which point, surviving embryos were euthanized in excess Tricaine (0.4% MS-222). Genomic DNA was extracted from individual embryos using the HotSHOT Method [106]. Briefly, embryos were transferred into 50 mM NaOH and heated to 95 °C for 20 min. The samples were neutralized with $\frac{1}{4}$ volume of 1 M Tris-HCl, pH 8, before downstream analysis. For NGS sequencing, genomic DNA was PCR amplified to extend the amplicon with Illumina adapters using the KOD One PCR master mix (TOYOBO). PCR amplicons were directly sequenced on an Illumina NovaSeq sequencer.

Data Availability

Raw FASTQ files and processed data tables are available at GenBank. Source data are provided with this paper.

Declarations

Author Contributions K.J., H.-C.K., K.H. and I.J.F. conceived the project. K.H. carried out the bioinformatic discovery and NGS data analysis. J.B., H.-C.K., Y.-C.C. and K.J. performed all cell line experiments and analyzed the data. B.V. and R.S.G. performed the zebrafish experiments. Y.-R.L. S.K.D., and B.X. established the U2OS cell lines and performed some experiments. M.E.L. assisted with cloning. J.B., H.-C.K.,

K.H., Y.-C.C., and I.J.F. prepared the figures and wrote the manuscript with input from all co-authors. I.J.F. secured the funding and supervised the project.

Acknowledgments We thank Daphne Sahaya for cloning assistance, and all members of the Finkelstein lab for discussions. Professor Xiaolu A. Cambronne provided access to a flow cytometer and the analysis software. Confocal imaging was conducted at the Center for Biomedical Research Support with the assistance of Paul Olyphint. This work was supported by a Sponsored Research Agreement from Retronix (to I.J.F.), the Welch Foundation grant F-1808 (to I.J.F.), and the College of Natural Sciences Catalyst grant for seed funding.

Competing interests The authors have filed patent applications related to retrorv editors. I.J.F. is a scientific consultant for Retronix. The other authors declare no competing interests.

References

1. Fichter, K. M., Setayesh, T. & Malik, P. Strategies for precise gene edits in mammalian cells. *Molecular Therapy. Nucleic Acids* **32**, 536–552 (2023).
2. Jacinto, F. V., Link, W. & Ferreira, B. I. CRISPR/Cas9-mediated genome editing: From basic research to translational medicine. en. *Journal of Cellular and Molecular Medicine* **24**. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/jcmm.14916>, 3766–3778 (2020).
3. Jang, H.-K., Song, B., Hwang, G.-H. & Bae, S. Current trends in gene recovery mediated by the CRISPR-Cas system. en. *Experimental & Molecular Medicine* **52**. Publisher: Nature Publishing Group, 1016–1027 (2020).
4. Li, L., Hu, S. & Chen, X. Non-viral delivery systems for CRISPR/Cas9-based genome editing: Challenges and opportunities. *Biomaterials* **171**, 207–218 (2018).
5. Luther, D., Lee, Y., Nagaraj, H., Scaletti, F. & Rotello, V. Delivery approaches for CRISPR/Cas9 therapeutics in vivo: advances and challenges. *Expert Opinion on Drug Delivery* **15**. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/17425247.2018.1517746>, 905–913 (2018).
6. Wang, H.-X. et al. CRISPR/Cas9-Based Genome Editing for Disease Modeling and Therapy: Challenges and Opportunities for Nonviral Delivery. *Chemical Reviews* **117**. Publisher: American Chemical Society, 9874–9906 (2017).
7. Yip, B. H. Recent Advances in CRISPR/Cas9 Delivery Strategies. en. *Biomolecules* **10**. Number: 6 Publisher: Multidisciplinary Digital Publishing Institute, 839 (2020).
8. David, R. M. & Doherty, A. T. Viral Vectors: The Road to Reducing Genotoxicity. *Toxicological Sciences* **155**, 315–325 (2017).
9. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. eng. *Nature* **576**, 149–157 (2019).
10. Chen, P. J. et al. Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. English. *Cell* **184**. Publisher: Elsevier, 5635–5652.e29 (2021).
11. Doman, J. L. et al. Phage-assisted evolution and protein engineering yield compact, efficient prime editors. en. *Cell* **186**, 3983–4002.e26 (2023).
12. Tang, S. & Sternberg, S. H. Genome editing with retroelements. *Science (New York, N.Y.)* **382**, 370–371 (2023).
13. Bhattarai-Kline, S. et al. Recording gene expression order in DNA by CRISPR addition of retror barcodes. en. *Nature* **608**. Publisher: Nature Publishing Group, 217–225 (2022).
14. Fishman, C. B. et al. *Continuous Multiplexed Phage Genome Editing Using Recombitrons* en. Pages: 2023.03.24.534024 Section: New Results. 2023.

15. Hwang, J., Ye, D.-Y., Jung, G. Y. & Jang, S. Mobile genetic element-based gene editing and genome engineering: Recent advances and applications. *Biotechnology Advances* **72**, 108343 (2024).
16. Kaur, N. & Pati, P. K. Retron Library Recombineering: Next Powerful Tool for Genome Editing after CRISPR/Cas. *ACS Synthetic Biology* **13**. Publisher: American Chemical Society, 1019–1025 (2024).
17. Khan, A. G. *et al.* *An experimental census of retrons for DNA production and genome editing* en. Pages: 2024.01.25.577267 Section: New Results. 2024.
18. Lee, G. & Kim, J. *Engineered retrons generate genome-independent protein-binding DNA for cellular control* en. 2023.
19. Lim, H. *et al.* Multiplex Generation, Tracking, and Functional Screening of Substitution Mutants Using a CRISPR/Retron System. *ACS Synthetic Biology* **9**. Publisher: American Chemical Society, 1003–1009 (2020).
20. Lin, Q. *et al.* Prime genome editing in rice and wheat. en. *Nature Biotechnology* **38**. Publisher: Nature Publishing Group, 582–585 (2020).
21. Liu, J. *et al.* Generation of DNAzyme in Bacterial Cells by a Bacterial Retron System. *ACS Synthetic Biology* **13**. Publisher: American Chemical Society, 300–309 (2024).
22. Liu, W. *et al.* Retron-mediated multiplex genome editing and continuous evolution in *Escherichia coli*. *Nucleic Acids Research* **51**, 8293–8307 (2023).
23. Ramirez-Chamorro, L., Boulanger, P. & Rossier, O. Strategies for Bacteriophage T5 Mutagenesis: Expanding the Toolbox for Phage Genome Engineering. English. *Frontiers in Microbiology* **12**. Publisher: Frontiers (2021).
24. Roy, K. R. *et al.* *Dissecting quantitative trait nucleotides by saturation genome editing* en. Pages: 2024.02.02.577784 Section: New Results. 2024.
25. Schubert, M. G. *et al.* High-throughput functional variant screens via in vivo production of single-stranded DNA. *Proceedings of the National Academy of Sciences* **118**. Publisher: Proceedings of the National Academy of Sciences, e2018181118 (2021).
26. Simon, A. J., Morrow, B. R. & Ellington, A. D. Retroelement-Based Genome Editing and Evolution. *ACS Synthetic Biology* **7**. Publisher: American Chemical Society, 2600–2611 (2018).
27. Ellington, A. J. & Reisch, C. R. Efficient and iterative retron-mediated in vivo recombineering in *Escherichia coli*. *Synthetic Biology* **7**, ysac007 (2022).
28. Farzadfar, F. & Lu, T. K. Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. en. *Science* **346**, 1256272 (2014).
29. González-Delgado, A., Lopez, S. C., Rojas-Montero, M., Fishman, C. B. & Shipman, S. L. *Simultaneous multi-site editing of individual genomes using retron arrays* en. 2023.

30. Kong, X. *et al.* Precise genome editing without exogenous donor DNA via retron editing system in human cells. en. *Protein & Cell* **12**, 899–902 (2021).
31. Lopez, S. C., Crawford, K. D., Lear, S. K., Bhattarai-Kline, S. & Shipman, S. L. Precise genome editing across kingdoms of life using retron-derived DNA. en. *Nature Chemical Biology* **18**, 199–206 (2022).
32. Sharon, E. *et al.* Functional Genetic Variants Revealed by Massively Parallel Precise Genome Editing. en. *Cell* **175**, 544–557.e16 (2018).
33. Simon, A. J., Ellington, A. D. & Finkelstein, I. J. Retrons and their applications in genome engineering. en. *Nucleic Acids Research* **47**, 11007–11019 (2019).
34. Zhao, B., Chen, S.-A. A., Lee, J. & Fraser, H. B. Bacterial Retrons Enable Precise Gene Editing in Human Cells. eng. *The CRISPR journal* **5**, 31–39 (2022).
35. Jiang, W. *et al.* High-efficiency retron-mediated single-stranded DNA production in plants. *Synthetic Biology* **7**, ysac025 (2022).
36. Bobonis, J. *et al.* Bacterial retrons encode phage-defending tripartite toxin–antitoxin systems. en. *Nature* **609**, 144–150 (2022).
37. Gao, L. *et al.* Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* **369**. Publisher: American Association for the Advancement of Science, 1077–1084 (2020).
38. Millman, A. *et al.* Bacterial Retrons Function In Anti-Phage Defense. en. *Cell* **183**, 1551–1561.e12 (2020).
39. Palka, C., Fishman, C. B., Bhattarai-Kline, S., Myers, S. A. & Shipman, S. L. Retron reverse transcriptase termination and phage defense are dependent on host RNase H1. en. *Nucleic Acids Research* **50**, 3490–3504 (2022).
40. Rychlik, I., Sebkova, A., Gregorova, D. & Karpiskova, R. Low-Molecular-Weight Plasmid of *Salmonella enterica* Serovar Enteritidis Codes for Retron Reverse Transcriptase and Influences Phage Resistance. *Journal of Bacteriology* **183**, 2852–2858 (2001).
41. Wang, Y. *et al.* Cryo-EM structures of *Escherichia coli* Ec86 retrons reveal architecture and defence mechanism. en. *Nature Microbiology* **7**. Publisher: Nature Publishing Group, 1480–1489 (2022).
42. Carabias, A. *et al.* Retron-Eco1 assembles NAD+-hydrolyzing filaments that provide immunity against bacteriophages. English. *Molecular Cell* **0**. Publisher: Elsevier (2024).
43. Wang, Y. *et al.* Defense mechanism of a bacterial retron supramolecular assembly en. Pages: 2023.08.16.553469 Section: New Results. 2023.
44. Azam, A. H. *et al.* Viruses encode tRNA and anti-retron to evade bacterial immunity en. Pages: 2023.03.15.532788 Section: New Results. 2023.

45. Hsu, M. Y., Eagle, S. G., Inouye, M. & Inouye, S. Cell-free synthesis of the branched RNA-linked msDNA from retron-Ec67 of *Escherichia coli*. eng. *The Journal of Biological Chemistry* **267**, 13823–13829 (1992).
46. Shimamoto, T., Inouye, M. & Inouye, S. The formation of the 2',5'-phosphodiester linkage in the cDNA priming reaction by bacterial reverse transcriptase in a cell-free system. eng. *The Journal of Biological Chemistry* **270**, 581–588 (1995).
47. Schimmel, J. *et al.* Modulating mutational outcomes and improving precise gene editing at CRISPR-Cas9-induced breaks by chemical inhibition of end-joining pathways. eng. *Cell Reports* **42**, 112019 (2023).
48. Savic, N. *et al.* Covalent linkage of the DNA repair template to the CRISPR-Cas9 nuclease enhances homology-directed repair. *eLife* **7** (ed de Massy, B.) Publisher: eLife Sciences Publications, Ltd, e33761 (2018).
49. Almeida, A. *et al.* A unified catalog of 204,938 reference genomes from the human gut microbiome. en. *Nature Biotechnology* **39**. Publisher: Nature Publishing Group, 105–114 (2021).
50. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research* **35**, D61–D65 (2007).
51. Mestre, M. R., González-Delgado, A., Gutiérrez-Rus, L. I., Martínez-Abarca, F. & Toro, N. Systematic prediction of genes functionally associated with bacterial retrons and classification of the encoded tripartite systems. en. *Nucleic Acids Research* **48**, 12632–12647 (2020).
52. Hussmann, J. A. *et al.* Mapping the genetic landscape of DNA double-strand break repair. eng. *Cell* **184**, 5653–5669.e25 (2021).
53. Nambiar, T. S., Baudrier, L., Billon, P. & Ciccia, A. CRISPR-based genome editing through the lens of DNA repair. en. *Molecular Cell* **82**, 348–388 (2022).
54. Potapov, V. *et al.* Base modifications affecting RNA polymerase and reverse transcriptase fidelity. *Nucleic Acids Research* **46**, 5753–5763 (2018).
55. Yasukawa, K. *et al.* Next-generation sequencing-based analysis of reverse transcriptase fidelity. *Biochemical and Biophysical Research Communications* **492**, 147–153 (2017).
56. Goeckel, M. E. *et al.* Modulating CRISPR gene drive activity through nucleocytoplasmic localization of Cas9 in *S. cerevisiae*. en. *Fungal Biology and Biotechnology* **6**, 2 (2019).
57. Liu, P. *et al.* Improved prime editors enable pathogenic allele correction and cancer modelling in adult mice. en. *Nature Communications* **12**, 2121 (2021).
58. Maggio, I. *et al.* Integrating gene delivery and gene-editing technologies by adenoviral vector transfer of optimized CRISPR-Cas9 components. en. *Gene Therapy* **27**, 209–225 (2020).

59. Suzuki, K. *et al.* In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. en. *Nature* **540**, 144–149 (2016).
60. Wu, Y. *et al.* Highly efficient therapeutic gene editing of human hematopoietic stem cells. en. *Nature Medicine* **25**, 776–783 (2019).
61. Luk, K. *et al.* Optimization of Nuclear Localization Signal Composition Improves CRISPR-Cas12a Editing Rates in Human Primary Cells. en. *GEN Biotechnology* **1**, 271–284 (2022).
62. Inouye, S., Hsu, M.-Y., Xu, A. & Inouye, M. Highly Specific Recognition of Primer RNA Structures for 2'-OH Priming Reaction by Bacterial Reverse Transcriptases*. *Journal of Biological Chemistry* **274**, 31236–31244 (1999).
63. Tan, J., Zhang, F., Karcher, D. & Bock, R. Engineering of high-precision base editors for site-specific single nucleotide replacement. *Nature Communications* **10**, 439 (2019).
64. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. English. *Cell* **159**. Publisher: Elsevier, 635–646 (2014).
65. Hinrichsen, M. *et al.* A new method for post-translationally labeling proteins in live cells for fluorescence imaging and tracking. *Protein Engineering, Design and Selection* **30**, 771–780 (2017).
66. Zhang, L. *et al.* AsCas12a ultra nuclease facilitates the rapid generation of therapeutic cell medicines. en. *Nature Communications* **12**. Publisher: Nature Publishing Group, 3908 (2021).
67. Carusillo, A. & Mussolini, C. DNA Damage: From Threat to Treatment. eng. *Cells* **9**, 1665 (2020).
68. Fok, J. H. L. *et al.* AZD7648 is a potent and selective DNA-PK inhibitor that enhances radiation, chemotherapy and olaparib activity. en. *Nature Communications* **10**. Publisher: Nature Publishing Group, 5065 (2019).
69. Wimberger, S. *et al.* Simultaneous inhibition of DNA-PK and Pol \square improves integration efficiency and precision of genome editing. *Nature Communications* **14**, 4761 (2023).
70. Fu, Y.-W. *et al.* Dynamics and competition of CRISPR–Cas9 ribonucleoproteins and AAV donor-mediated NHEJ, MMEJ and HDR editing. en. *Nucleic Acids Research* **49**. Publisher: Oxford University Press, 969 (2021).
71. Riesenber, S. *et al.* Simultaneous precise editing of multiple genes in human cells. *Nucleic Acids Research* **47**, e116 (2019).
72. Iwai, K. *et al.* Molecular mechanism and potential target indication of TAK-931, a novel CDC7-selective inhibitor. *Science Advances* **5**, eaav3660 (2019).
73. Reint, G. *et al.* Rapid genome editing by CRISPR-Cas9-POLD3 fusion. en. *eLife* **10**, e75415 (2021).

74. Jayavaradhan, R. *et al.* CRISPR-Cas9 fusion to dominant-negative 53BP1 enhances HDR and inhibits NHEJ specifically at Cas9 target sites. en. *Nature Communications* **10**, 2866 (2019).
75. Carusillo, A. *et al.* A novel Cas9 fusion protein promotes targeted genome editing with reduced mutational burden in primary human cells. en. *Nucleic Acids Research* **51**, 4660–4673 (2023).
76. Gutschner, T., Haemmerle, M., Genovese, G., Draetta, G. F. & Chin, L. Post-translational Regulation of Cas9 during G1 Enhances Homology-Directed Repair. en. *Cell Reports* **14**, 1555–1566 (2016).
77. Rees, H. A., Yeh, W.-H. & Liu, D. R. Development of hRad51–Cas9 nickase fusions that mediate HDR without double-stranded breaks. en. *Nature Communications* **10**, 2212 (2019).
78. Arslan, S., Khafizov, R., Thomas, C. D., Chemla, Y. R. & Ha, T. Engineering of a superhelicase through conformational control. en. *Science* **348**, 344–347 (2015).
79. Barondeau, D. P., Putnam, C. D., Kassmann, C. J., Tainer, J. A. & Getzoff, E. D. Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proceedings of the National Academy of Sciences* **100**. Publisher: Proceedings of the National Academy of Sciences, 12111–12116 (2003).
80. Kamiyama, D. *et al.* Versatile protein tagging in cells with split fluorescent protein. en. *Nature Communications* **7**. Publisher: Nature Publishing Group, 11046 (2016).
81. Pinaud, F. & Dahan, M. Targeting and imaging single biomolecules in living cells by complementation-activated light microscopy with split-fluorescent proteins. eng. *Proceedings of the National Academy of Sciences of the United States of America* **108**, E201–210 (2011).
82. Cabantous, S., Terwilliger, T. C. & Waldo, G. S. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. en. *Nature Biotechnology* **23**. Publisher: Nature Publishing Group, 102–107 (2005).
83. Konjikusic, M. J. *et al.* Mutations in Kinesin family member 6 reveal specific role in ependymal cell ciliogenesis and human neurological development. en. *PLOS Genetics* **14** (ed Pa- zour, G. J.) e1007817 (2018).
84. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nature reviews. Genetics* **19**, 770–788 (2018).
85. Zhao, Z., Shang, P., Mohanraju, P. & Geijsen, N. Prime editing: advances and therapeutic applications. English. *Trends in Biotechnology* **41**. Publisher: Elsevier, 1000–1012 (2023).
86. Carlson-Stevermer, J. *et al.* Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing. en. *Nature Communications* **8**. Publisher: Nature Publishing Group, 1711 (2017).

87. Aird, E. J., Lovendahl, K. N., St. Martin, A., Reuben S. Harris & Gordon, W. R. Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. en. *Communications Biology* **1**. Publisher: Nature Publishing Group, 1–6 (2018).
88. Chen, P. J. & Liu, D. R. Prime editing for precise and highly versatile genome manipulation. en. *Nature Reviews Genetics* **24**. Publisher: Nature Publishing Group, 161–177 (2023).
89. Gallagher, D. N. *et al.* A Rad51-independent pathway promotes single-strand template repair in gene editing. eng. *PLoS genetics* **16**, e1008689 (2020).
90. Gallagher, D. N. & Haber, J. E. Single-strand template repair: key insights to increase the efficiency of gene editing. en. *Current Genetics* **67**, 747–753 (2021).
91. Richardson, C. D. *et al.* CRISPR–Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. en. *Nature Genetics* **50**. Publisher: Nature Publishing Group, 1132–1139 (2018).
92. Okamoto, S., Amaishi, Y., Maki, I., Enoki, T. & Mineno, J. Highly efficient genome editing for single-base substitutions using optimized ssODNs with Cas9-RNPs. en. *Scientific Reports* **9**. Publisher: Nature Publishing Group, 4811 (2019).
93. Schubert, M. S. *et al.* Optimized design parameters for CRISPR Cas9 and Cas12a homology-directed repair. en. *Scientific Reports* **11**. Publisher: Nature Publishing Group, 19482 (2021).
94. Paulsen, B. S. *et al.* Ectopic expression of RAD52 and dn53BP1 improves homology-directed repair during CRISPR–Cas9 genome editing. en. *Nature Biomedical Engineering* **1**. Publisher: Nature Publishing Group, 878–888 (2017).
95. Wyatt, D. W. *et al.* Essential Roles for Polymerase θ-Mediated End Joining in the Repair of Chromosome Breaks. eng. *Molecular Cell* **63**, 662–673 (2016).
96. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. en. *Nature Biotechnology* **34**. Publisher: Nature Publishing Group, 339–344 (2016).
97. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. en. *Nucleic Acids Research* **39**, W29–W37 (2011).
98. Eddy, S. R. Accelerated Profile HMM Searches. en. *PLoS Computational Biology* **7** (ed Pearson, W. R.) e1002195 (2011).
99. Yao, Z., Weinberg, Z. & Ruzzo, W. L. CMfinder—a covariance model based RNA motif finding algorithm. en. *Bioinformatics* **22**, 445–452 (2006).
100. Lorenz, R. *et al.* ViennaRNA Package 2.0. en. *Algorithms for Molecular Biology* **6**, 26 (2011).
101. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. en. *Bioinformatics* **28**, 3150–3152 (2012).

102. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. en. *Molecular Biology and Evolution* **30**, 772–780 (2013).
103. Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. en. *Molecular Biology and Evolution* **32**, 268–274 (2015).
104. Matreyek, K. A., Stephany, J. J., Chiasson, M. A., Hasle, N. & Fowler, D. M. An improved platform for functional assessment of large protein libraries in mammalian cells. *Nucleic Acids Research* **48**, e1 (2020).
105. Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. eng. *Nature Biotechnology* **37**, 224–226 (2019).
106. Meeker, N. D., Hutchinson, S. A., Ho, L. & Trede, N. S. Method for Isolation of PCR-Ready Genomic DNA from Zebrafish Tissues. EN. *BioTechniques*. Publisher: Taylor & Francis (2007).

Figures

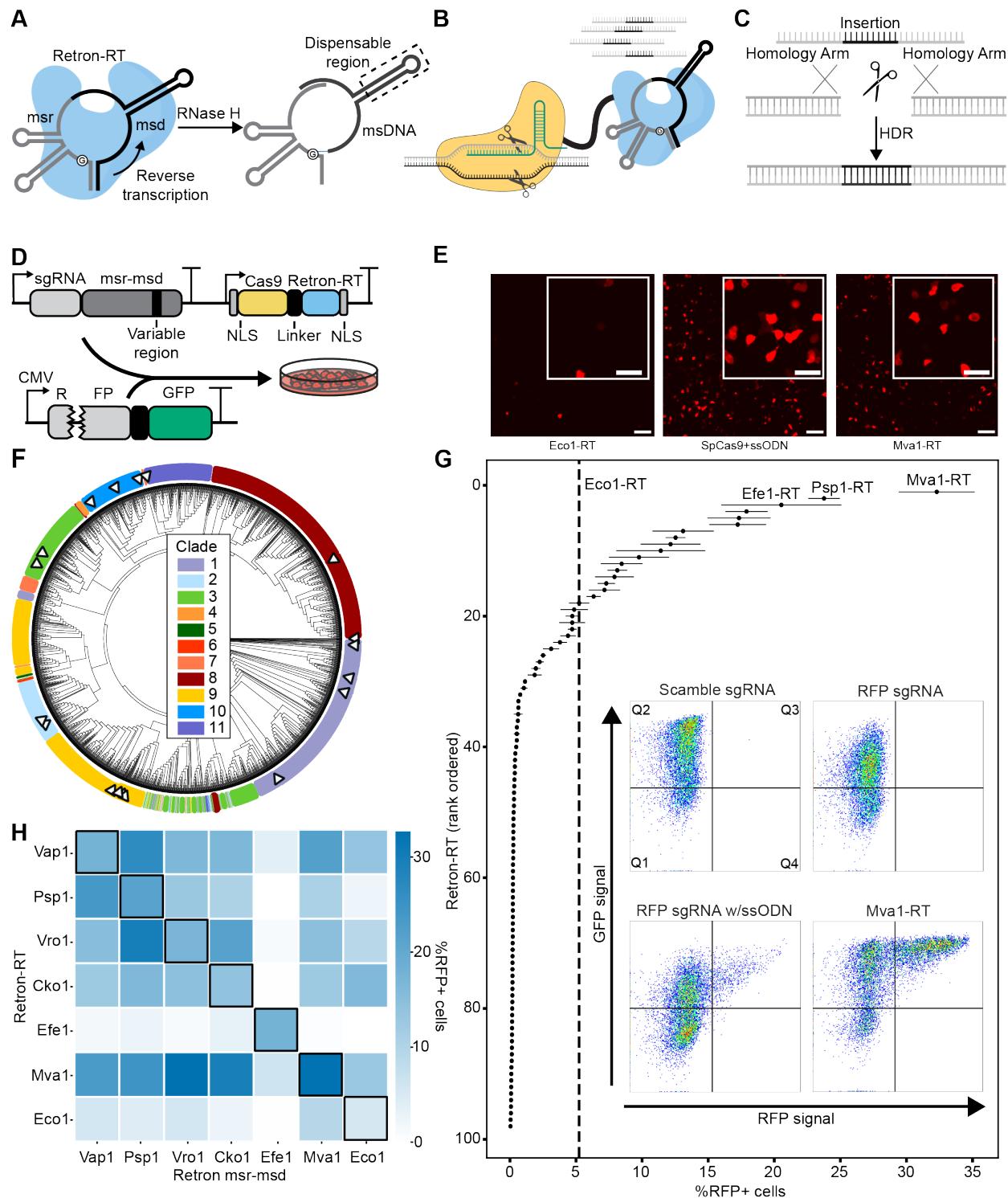


Figure 1: A metagenomic survey reveals highly active RTs in mammalian cells. A) Retron-RTs self-prime from a non-coding RNA, termed the *msr-msd*. *msr*: gray; *msd* stem: black; variable region: black, in dotted box. The arrow indicates the direction of reverse transcription. B) Schematic of a retron editor. The RT is linked to Cas9 (shown) or another nuclease. C) Reverse transcription of the variable region of the *msd* generates a ssDNA template for homology-directed repair of the cleavage site. D) A plasmid-encoded fluorescent reporter assay. The RFP has a 9bp deletion proximal to a Y64L mutation to completely turn off RFP fluorescence. The reporter is co-transfected with a plasmid that encodes the retron editor, along with an *msd* that repairs the broken RFP. RFP+ cells are imaged via confocal microscopy and quantified via flow cytometry. E) Confocal microscopy images of cells transfected with Cas9-Eco1-RT (left), Cas9 + ssODN (middle), and Cas9-Mva1-RT (right). F) Phylogenetic classification of novel retron systems discovered from metagenomic sources. These systems are classified into clades, as described in [51]. G) Rank ordered list of RFP repair efficiency with 98 metagenomically discovered retron-RTs using flow cytometry. Dashed line: RFP+ repair with Eco1-RT. Inset: flow cytometry data for Cas9 with a scrambled sgRNA (top left); RFP-targeting sgRNA (top right); RFP sgRNA and a ssODN repair template (bottom left); RFP sgRNA and Mva1-RT (bottom right). Error bars: mean of three biological replicates. The three most active RTs are labeled. H) Gene editing activity of the six most active retron-RTs, along with Eco1-RT with a cognate (diagonal) or non-cognate *msr-msd*. Flow cytometry was used to score activity with the transient RFP reporter. Heat map: mean of three biological replicates.

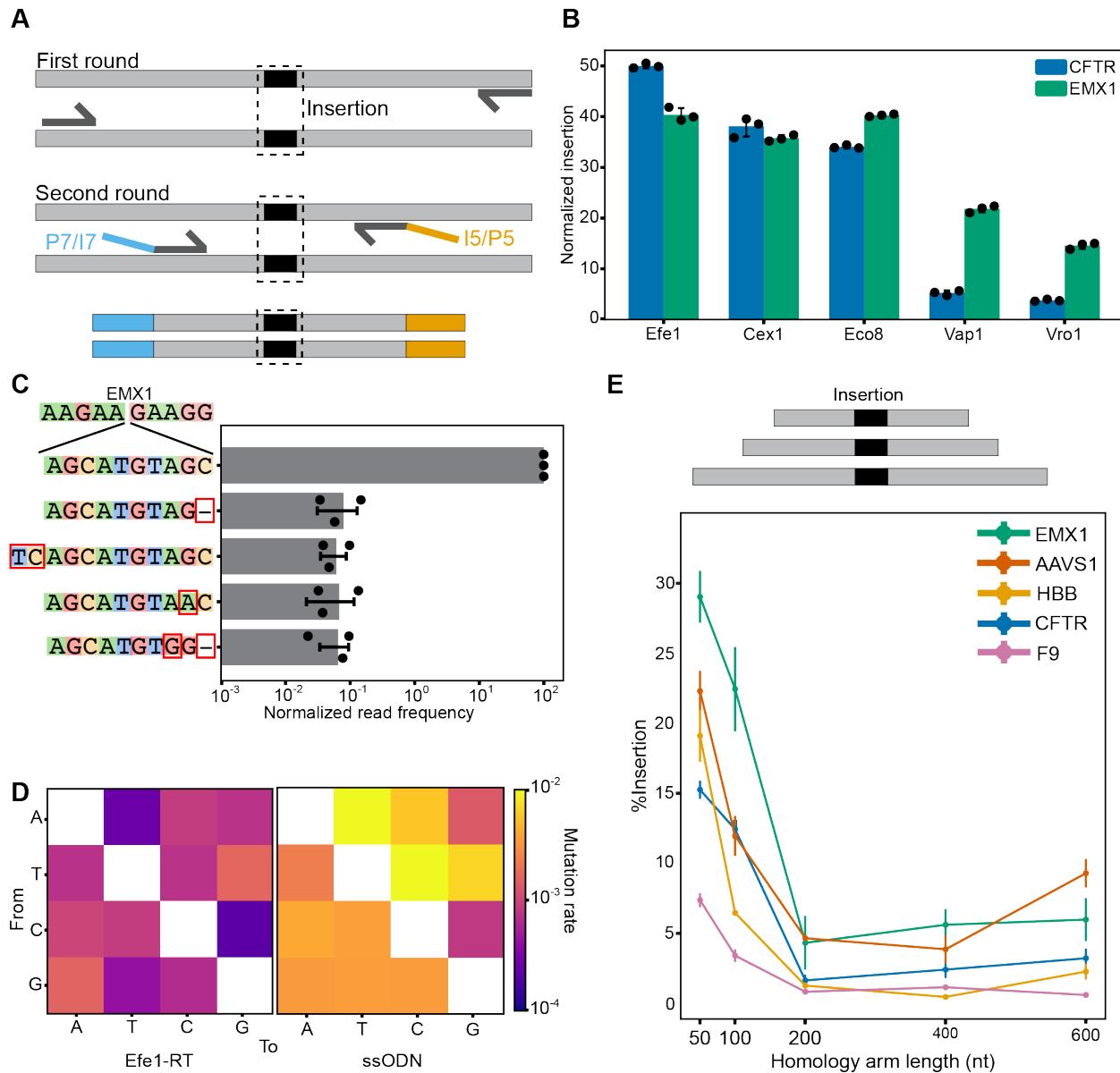


Figure 2: Efe1-RT catalyzes precise genomic insertions across multiple loci. A) Schematic of the NGS library preparation strategy. Genomic DNA is first amplified with primers that are outside the homology arms to avoid amplifying the retron-synthesized msDNA. After gel extraction, a second round of PCR amplifies and barcodes the insertion site for deep sequencing. Blue, orange: universal Illumina P5/I5 and P7/I7 adapters and indices. B) Normalized insertion efficiency for the top 5 retron-RTs at the CFTR and EMX1 loci. Error bars: mean of three biological replicates (dots). C) The relative insertion frequency of a 10 nt cargo at the EMX1 locus, along with the four most frequent misincorporated sequences. The most common errors are a deletion or insertion at the periphery of the homology arms. Error bars: mean of three biological replicates (dots). D) Efe1-RT substitution errors (left) are less frequent than ssODN insertion (right) at the EMX1. Substitution rates are computed from the insert in the EMX1 locus. Heat map: three biological replicates. E) Schematic (top) and results (bottom) of the insertion efficiency with Efe1-RT as a function of the homology arm length. Templatized insertion is most active with 50 nt homology arms at five genomic loci. Error bars: mean of three biological replicates (dots).

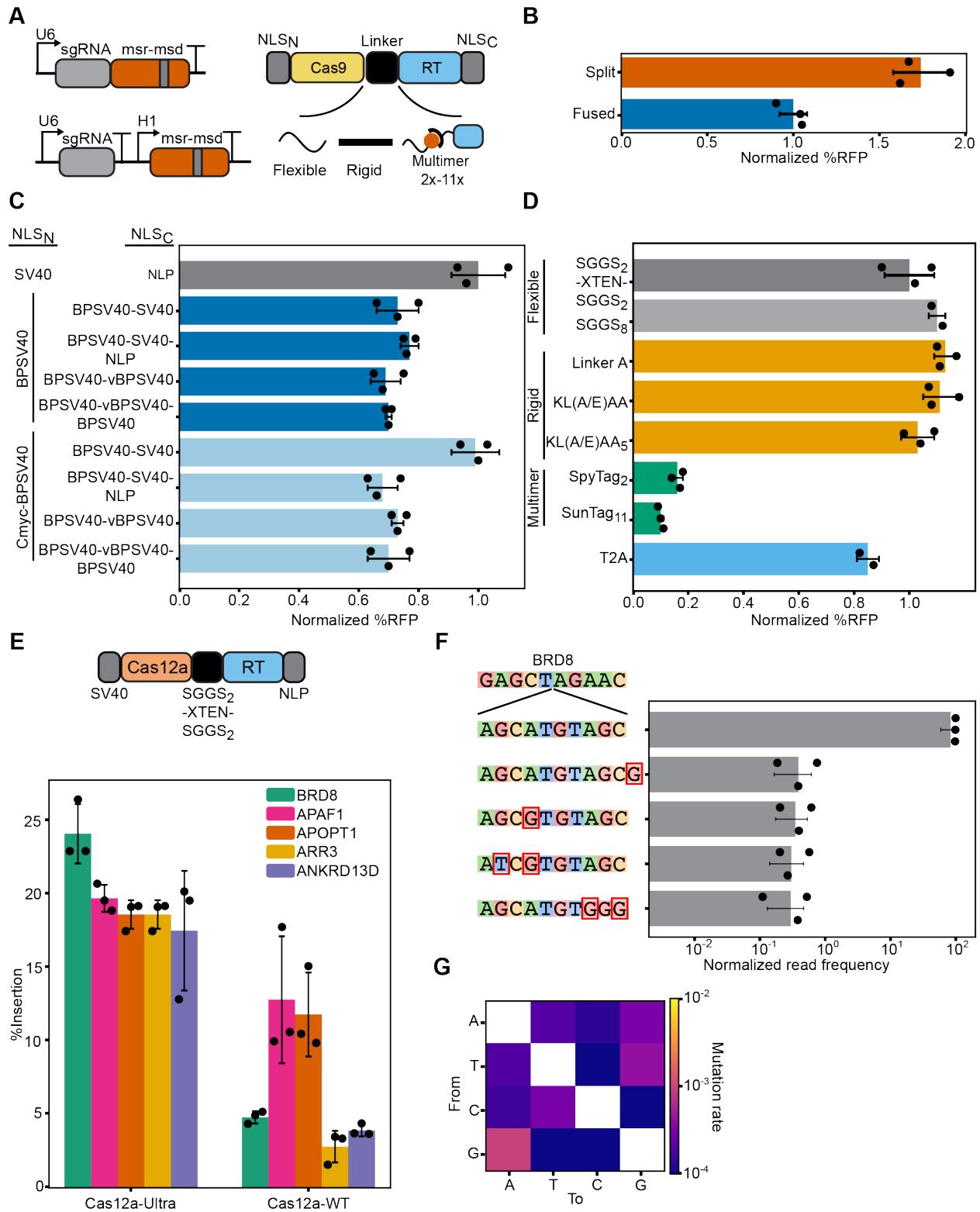


Figure 3: Rational Engineering of an Efe1-RT-based retron editor. A) We tested the effect of splitting the sgRNA and msr-msd (left), the identity of the nuclear localization sequences (NLSs), and the linker between the Cas9 and Efe1-RT (bottom). B) Expressing the sgRNA and msr-msd increased gene editing by 60% relative to a fused sgRNA-msr-msd design. C) Optimization of the N- and C-terminal nuclear localization sequences (NLSs). Gray: reference design that was used for normalization. Error bars: mean of three biological replicates (dots). D) Optimization of the linker peptide between the Cas9 and Efe1-RT. Retron editors tolerate a broad range of flexible (light gray) and rigid (orange) linkers. Splitting the two enzymes via a ribosomal skipping peptide (T2A, light blue) also retains most activity. However, multimerization domains abrogated activity (green). Dark gray: reference design that was used for normalization. Error bars: mean of two or three biological replicates (dots). E) Schematic (top) and editing activity of a Cas12a-based retron editor at five genomic loci. Error bars: mean of three biological replicates (dots). F) The relative insertion frequency of a 10 nt cargo at the BRD8 locus, along with the four most frequent misincorporated sequences. In contrast to Cas9-based retron editors, Cas12a retron editors generate substitution errors in the insert. Error bars: mean of three biological replicates (dots). G) Insert substitution frequencies across five loci. Heat map: mean of three biological replicates.

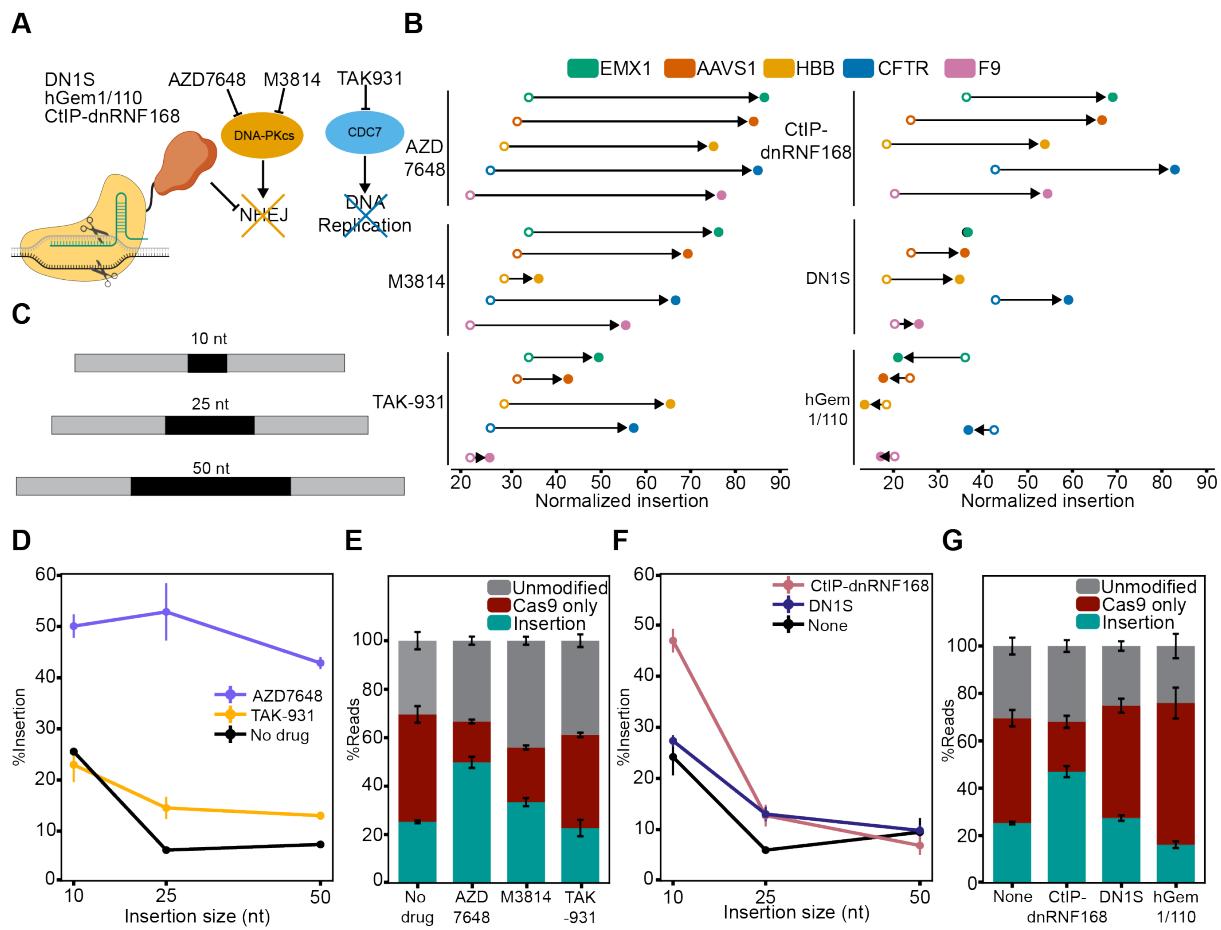


Figure 4: Inhibiting non-homologous end joining boosts templated insertion. A) Two strategies that boost templated insertion. Left: Cas9 is fused to proteins that alter DNA repair pathway choice. Right: small molecule inhibition of DNA-dependent protein-kinase catalytic subunit (DNA-PKcs) or CDC7. B) The effect of inhibitors (left) and Cas9 fusions (right) on the relative rate of templated insertion at five loci. AZD7648 (left, top) and Cas9-CtlP-dnRNF168 (right, top) both have the strongest effect at all tested loci. Open circles: editing with no inhibitor or DNA repair protein. Closed circle: editing with the indicated inhibitor or DNA repair protein fused to Cas9. All circles indicate the mean across three biological replicates. Arrow: change in editing efficiency with the indicated inhibitor. Fusions increase the relative rates of templated repair across all loci. Closed and open circles: mean of three biological replicates. C) Schematic of experiments with 50 nt homology arms and increasing insert lengths at the EMX1 locus. D) AZD7648 increases the insertion efficiency across all cargo sizes tested in this study. Error bars: mean of three biological replicates. E) AZD7648 outperforms TAK-931 and M3814 in boosting 10 nt insertion efficiency at EMX1 without increasing mutational signature or Cas9-generated indels. Error bars: mean of three biological replicates. F) Insertion efficiency decreases for all Cas9-repair protein fusions at EMX1. Error bars: mean of three biological replicates. G) Cas9 fused to CtlP-dnRNF168 increased insertion efficiency of 10 nt cargo at EMX1 without increasing mutational signature or Cas9-generated indels compared to no DNA repair fusion, DN1S, and hGem1/110. Error bars: mean of three biological replicates.

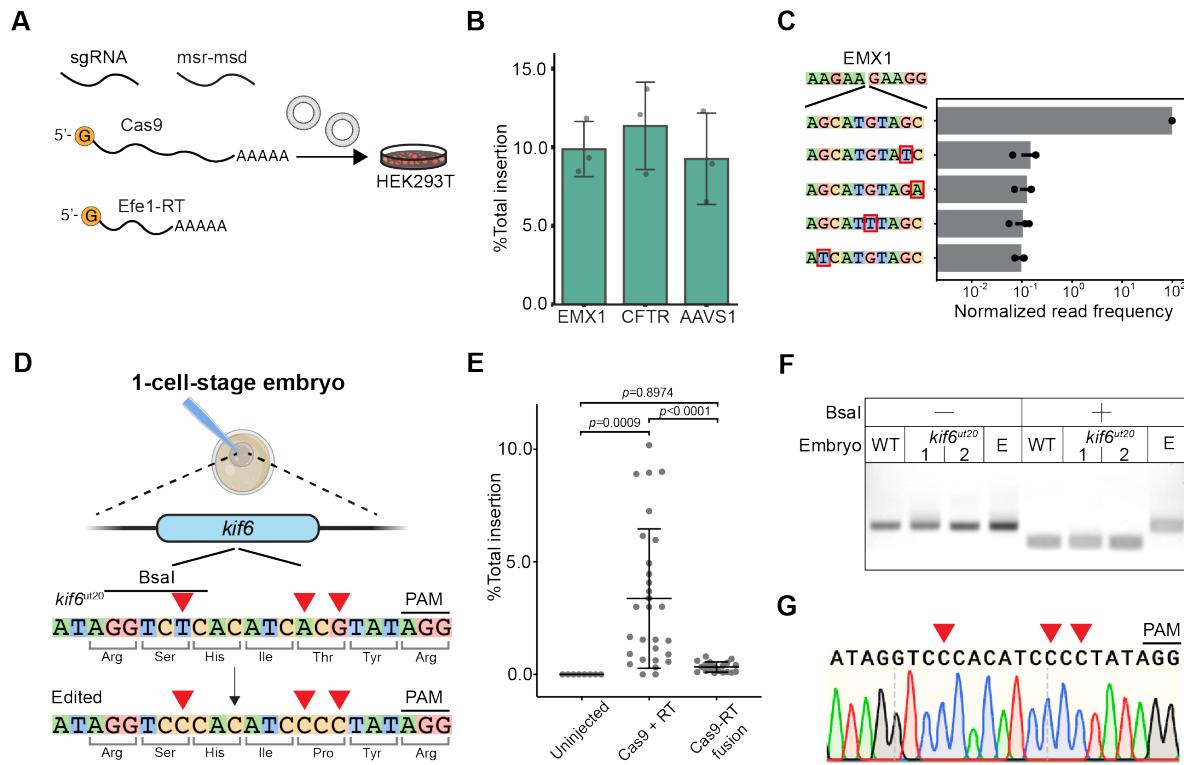


Figure 5: Retron editor delivery via an all-RNA package. A) Illustration of RNA-based retron editing. Cas9 and Efe1-RT are delivered as capped and poly-A-tailed mRNAs. The sgRNA and msr-msd are mixed with the mRNAs in a 10:1 molar ratio before transfection into HEK293T cells. B) Insertion efficiency of a 10 nt insert at the indicated loci with RNA-based delivery. Error bars: mean of three biological replicates (dots). C) The relative insertion frequency of a 10 nt cargo at the EMX1 locus, along with the four most frequent misincorporated sequences. Error bars: mean of three biological replicates (dots). D) Schematic of the three substitutions that are introduced in *kif6* by mRNA injection. The first mutation is silent but abolishes a Bsal cut site. E) Deep sequencing of embryos 24 hr post RNA injection for the indicated conditions. Each point is a single embryo. p-values are determined by a One-Way ANOVA test. F) Bsal restriction enzyme digests of the edit site sub-cloned from WT, *kif6*^Δ, and edited embryos. The Bsal cut site is abolished in edited, but not WT or *kif6*^Δ embryos. G) Sanger sequencing of the edited embryo from (E) confirms precise editing at the three expected sites (red triangles).