



Published in final edited form as:

Nat Chem Biol. 2022 February ; 18(2): 199–206. doi:10.1038/s41589-021-00927-y.

Precise genome editing across kingdoms of life using retron-derived DNA

Santiago C. Lopez^{1,3}, Kate D. Crawford^{1,3}, Sierra K. Lear^{1,3}, Santi Bhattarai-Kline¹, Seth L. Shipman^{1,2,*}

¹Gladstone Institute of Data Science and Biotechnology, San Francisco, CA, USA

²Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA, USA

³Graduate Program in Bioengineering, University of California, San Francisco and Berkeley, CA, USA

Abstract

Exogenous DNA can be a template to precisely edit a cell's genome. However, the delivery of *in vitro*-produced DNA to target cells can be inefficient, and low abundance of template DNA may underlie the low rate of precise editing. One potential tool to produce template DNA inside cells is a retron, a bacterial retroelement involved in phage defense. However, little effort has been directed at optimizing retrons to produce designed sequences. Here, we identify modifications to the retron non-coding RNA that result in more abundant reverse transcribed DNA. By testing architectures of the retron operon that enable efficient reverse transcription, we find that gains in DNA production are portable from prokaryotic to eukaryotic cells and result in more efficient genome editing. Finally, we show that retron RT-DNA can be used to precisely edit cultured human cells. These experiments provide a general framework to produce DNA using retrons for genome modification.

Graphical Abstract

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: <https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms>

*Correspondence to: seth.shipman@gladstone.ucsf.edu.

AUTHOR CONTRIBUTIONS STATEMENT

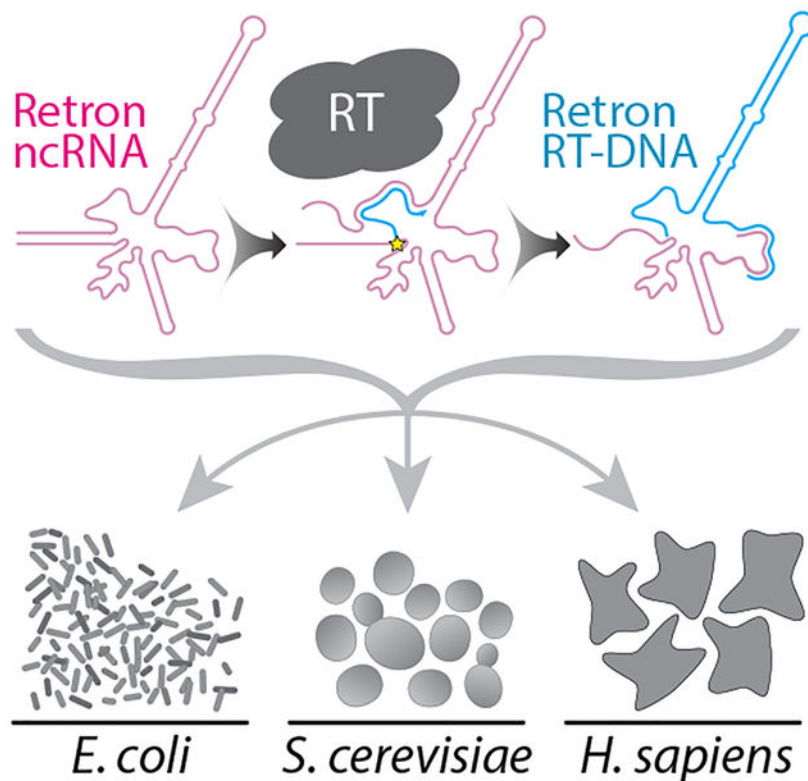
S.L.S., S.C.L. and K.D.C. conceived the study. S.C.L., K.D.C., S.K.L., S.B.K. and S.L.S. designed the work, performed experiments, analyzed the data, and wrote the manuscript.

COMPETING INTERESTS STATEMENT

S.L.S., S.C.L., and K.D.C. are inventors on patent applications related to the technologies described in this work.

CODE AVAILABILITY

Custom code to process or analyze data from this study is available on Github here: https://github.com/Shipman-Lab/retron_architectures.



INTRODUCTION

Exogenous DNA, which does not match the genome of the cell where it is harbored, is a fundamental tool of modern cell and molecular biology. This DNA can serve as a template to modify a cell's genome, subtly alter existing genes, or even insert wholly new genetic material that adds function or marks a cellular event such as lineage. Exogenous DNA for these uses is typically synthesized or assembled in a tube, then physically delivered to the cells that will be altered. However, it remains an incredible challenge to deliver exogenous DNA to cells in universally high abundance and without substantial variation between recipients¹. These technical challenges likely contribute to low rates of precise editing as well as unintended editing that occurs in the absence of template DNA²⁻⁴. Effort has been made to bias cells toward template-based editing by manipulating the proteins involved in DNA repair or tethering DNA templates to other editing materials to increase their local concentration⁵. However, a simpler approach may be to eliminate DNA delivery problems by producing the DNA inside the cell.

In recent years, it has been shown that retroelements can be used to produce DNA for genome editing within cells by reverse transcription⁶⁻⁹. This reverse transcribed DNA (RT-DNA) is produced in cells from plasmids, transgenes, or viruses, benefiting from transcriptional amplification to create high cellular concentrations that overcome inefficiencies in genome editing. One retroelement class that has been useful in this regard are bacterial retrons^{6,8,9}, which are elements involved in phage defense¹⁰⁻¹³. Retrons are attractive as tools for biotechnology due to their compact size, tightly defined sites of RT

initiation and termination, lack of known host factor requirements, and lack of transposable elements. Indeed, retron-generated RT-DNA has demonstrated utility in bacterial^{6,9} and eukaryotic⁸ genome editing.

Despite the potential of the retron as component of molecular biotechnology, it has so far been modified only as little as is necessary to produce an editing template. Given that the advantage of the retroelement approach is the increased cellular abundance of RT-DNA, we asked whether we could identify retron modifications that would yield even more abundant RT-DNA and increase in editing efficiency. Further, most work with retron has been carried out in bacteria, with only one functional demonstration of RT-DNA production in yeast⁸, and only a brief description of reverse transcription in mammalian cells (NIH3T3 mouse cells)¹⁴. Therefore, we wanted to engineer a more flexible architecture for retron expression across kingdoms of life, to serve as a universal framework for RT-DNA production.

Here, we used variant libraries in *E. coli* to show that extension of complementarity in the a1/a2 region of the retron non-coding RNA (ncRNA) increases production of RT-DNA. This effect generalized across different retrons and kingdoms, from bacteria to yeast. Moreover, retron DNA production across kingdoms was possible using a universal architecture. We found that increasing the abundance of RT-DNA in the context of genome engineering increased the rate of editing in both prokaryotic and eukaryotic cells, simultaneously showing that the template abundance is limiting for these editing applications and demonstrating a simple means of increasing genome editing efficiency. Finally, we show that the retron RT-DNA can be used as a template for editing human cells to enable further gains in both future research and therapeutic ventures.

RESULTS

Modifications to the retron ncRNA affect RT-DNA production

A typical retron operon consists of a reverse transcriptase (RT), a non-coding RNA (ncRNA) that is both the primer and template for the reverse transcriptase, and one or more accessory proteins¹⁵ (Fig 1a). The RT partially reverse transcribes the ncRNA to produce a single-stranded RT-DNA with a characteristic hairpin structure, which varies in length from 48–163 bases¹⁶. The ncRNA can be sub-divided into a region that is reverse transcribed (msd) and a region that remains RNA in the final molecule (msr), which are partially overlapping^{17–20}.

One of the first described retrons was found in *E. coli*, Eco1 (previously ec86)²⁰. In BL21 cells, this retron is both present and active, producing RT-DNA that can be detected at the population level, which is eliminated by removing the retron operon from the genome (Fig 1b). In the absence of this native operon, the ncRNA and RT can be expressed from a plasmid lacking the accessory protein, which is a minimal system for RT-DNA production. We quantified this RT-DNA using qPCR. Specifically, we compared amplification from primers that anneal to the msd region, which can use both the RT-DNA and plasmid as a template, to amplification from primers that only amplify the plasmid (Fig 1c, d). In *E. coli* lacking an endogenous retron, overexpression of the ncRNA and RT from a plasmid yielded an ~8–10 fold enrichment of the RT-DNA/plasmid region over the plasmid alone, which is evidence of robust reverse transcription (Fig 1d).

Given that the retron utility in biotechnology relies on increasing the RT-DNA abundance in cells above what can be achieved with delivery of a synthetic template, we set out to identify aspects of the ncRNA that could be modified to produce more abundant RT-DNA. To do this, we synthesized variants of the Eco1 ncRNA and cloned them into vector for expression, with the RT expressed from a separate vector. Our initial library contained variants that extended or reduced the length of the hairpin stem of the RT-DNA. This variant cloning took place in single-pot, Golden Gate reactions and the resulting libraries were purified and then cloned into an expression strain for analysis of RT-DNA production (Fig 1e). Cells harboring these library vector sets were grown overnight and then diluted and ncRNA expression was induced during growth for 5 hours.

We quantified the relative abundance of each variant plasmid in the expression strain by multiplexed Illumina sequencing before and after expression. After expression, we additionally purified RT-DNA from pools of cells harboring different retron variants by isolating cellular nucleic acids, treating that population with an RNase mixture (A/T1), and then isolating single-stranded DNA from double-stranded DNA using a commercial column-based kit. We then sequenced the RT-DNAs, comparing their relative abundance to that of their plasmid of origin to quantify the influence of different ncRNA parameters on RT-DNA production. To sequence the RT-DNA variants in this library, we used a custom sequencing pipeline to prep each RT-DNA without biasing toward any variant. This involved tailing purified RT-DNA with a string of polynucleotides using a template-independent polymerase (TdT), and then generating a complementary strand via an adapter-containing, inverse anchored primer. Finally, we ligated a second adapter to this double-stranded DNA and proceed to indexing and multiplexed sequencing (Extended Data Fig 1a, b).

In this first library, we modified the msd stem length from 0–31 base pairs, and found that stem length can have a large impact on RT-DNA production (Fig 2a). The RT tolerated modifications of the msd stem length that deviate by a small amount from the wild-type (wt) length of 25 base pairs. However, variants with stem lengths <12 and >30 produced less than half as much RT-DNA compared to the wt. Therefore, we used stem length of between 12 and 30 base pairs going forward.

In a second library, we investigated the effect of increasing the loop length at the top of what becomes the RT-DNA stem (Fig 2b). To do this, we created five random sequences of 70 bases each. We then synthesized variant ncRNAs incorporating 5–70 of these bases into the msd top loop. Thus, we tested five versions of each loop length, each with different base content, and then averaged each variant's RT-DNA production at every loop length. We did not include the wt loop in this library, so we normalized RT-DNA production to the 5 base loops, which are closest in size to the wt length of 4 bases. We found a substantial decline in RT-DNA production as loop length increases from 5 to ~14 bases, but almost no continued decline beyond that point, other than a single point at 28 bases which inexplicably produced more RT-DNA than its neighboring loops. While we were limited by our synthesis and sequencing parameters to 70 bases, our conclusion is that loops shorter than 14 bases are ideal for RT-DNA production; however, loops that extend beyond 14 bases do not additionally reduce RT-DNA production.

The other parameter we investigated was the length of a1/a2 complementarity, a region of the ncRNA structure where the 5' and 3' ends of the ncRNA fold back upon themselves, and which we hypothesized plays a role in initiating reverse transcription (Fig 2c). Because this region of the ncRNA is not reverse transcribed, we could not sequence the variants in the RT-DNA population directly. Instead, we introduced a nine-base barcode in an extended loop of the msd that we could sequence as a proxy for the modification (Fig 2d). We amplified these barcodes directly from the purified RT-DNA for sequencing (Fig 2e) or prepped the RT-DNA using the TdT extension method described above (Extended Data Fig 1c). In both cases, we found a similar effect: reducing the length of complementarity in this region below seven base pairs substantially impaired RT-DNA production, consistent with a critical role in reverse transcription (Fig 2f). However, extending the a1/a2 length resulted in increased production of RT-DNA relative to the wt length. Importantly, this is the first modification to a retron ncRNA that has been shown to increase RT-DNA production.

RT-DNA production in eukaryotic cells

We next wondered whether the increased production by the extended a1/a2 region would be a portable modification, both other retrons and to eukaryotic systems. To facilitate expression of Eco1 in eukaryotic cells, we inverted the operon from its native arrangement²¹. In the endogenous arrangement, the ncRNA is in the 5' UTR of the RT transcript, requiring internal ribosome entry for the RT from an RBS that is contained in or near the a2 region of the ncRNA. In eukaryotic cells, this arrangement puts the entire ncRNA between the 5' mRNA cap and the initiation codon for the RT. This increased distance between the cap and initiation codon, as well as the ncRNA structure and out-of-frame ATG codons, is expected to negatively affect RT translation^{21,22}. Moreover, altering the a1/a2 region in the native arrangement could have unintended effects on the RT translation. In the inverted architecture, the RT is driven by a pol II promoter directly with its initiation codon near the 5' end of the transcript and the ncRNA in the 3' UTR, where variations are unlikely to influence RT translation (Fig 3a).

We first tested this arrangement for Eco1 in *S. cerevisiae*, by placing the RT – ncRNA cassette under the expression of Galactose inducible promoter, on a single-copy plasmid. We detected RT-DNA production using a qPCR assay analogous to that described for *E. coli* above, comparing amplification from primers that could use the plasmid or RT-DNA as a template to amplification from primers that could anneal only to the plasmid. Here, we found that increasing the length of the Eco1 a1/a2 region from 12 to 27 base pairs resulted in more abundant RT-DNA production (Fig 3b, Extended Data Fig 2a). We then extended this analysis to another retron, Eco2¹⁹. We found a similar effect: though the wt ncRNA produced detectable RT-DNA, a version extending the a1/a2 region from 13 to 29 base pairs produced significantly more RT-DNA (Fig 3c, Extended Data Fig 2a). In each case, we compared induced to uninduced cells, which likely under-reports the total RT-DNA abundance if there is any transcriptional 'leak' from the plasmid in the absence of inducers. Indeed, we detected RT-DNA production in the uninduced condition relative to a control expressing a catalytically dead RT, indicating some transcriptional 'leak' (Extended Data Fig 2b).

We then moved from yeast to cultured human cells, HEK293T. Using a similar gene architecture to yeast, but with a genome-integrating cassette (Fig 3d), we found that Eco1 does not produce significant abundance of RT-DNA in human cells that we could detect by qPCR, regardless of a1/a2 length (Fig 3e), from a tightly regulated promoter (Extended Data Fig 2c). In contrast, Eco2 produces detectable RT-DNA, with both a wt and extended a1/a2 region (Fig 3f). In human cells, however, the introduction of an extended a1/a2 region diminishes, rather than enhances, production of RT-DNA. Nevertheless, this demonstrates RT-DNA production by a retron in human cells.

Improvements extend to applications in genome editing

In prokaryotes, retron-derived RT-DNA can be used as a template for recombineering^{6,9}. The retron ncRNA is modified to include a long loop in the msd that contains homology to a genomic locus along with one or more nucleotide modifications (Fig 4a). When RT-DNA from this modified ncRNA is produced along with a single stranded annealing protein (SSAP; e.g., lambda Red β), the RT-DNA is incorporated into the lagging strand during genome replication, thereby editing the genome of half of the cell progeny. This process is typically carried out in modified bacterial strains with numerous nucleases and repair proteins knocked out, because editing occurs at a low rate in wt cells⁹. Therefore, we asked whether increasing RT-DNA abundance using retrons with extended a1/a2 regions could increase the rate of editing in relatively unmodified strains.

We produced RT-DNA to edit a single nucleotide in the *rpoB* gene. We designed the retron using the same flexible architecture that we used for both the yeast and mammalian expression, with the ncRNA in the 3' UTR of the RT. We used a 12 base stem for the msd, which retains near-wt RT-DNA production. We constructed two versions of the editing retron, one with the wt 12-base a1/a2 region and another with an extended 22-base a1/a2 length. Using qPCR and PAGE analysis, we confirmed that the extended a1/a2 version produced more abundant RT-DNA (Fig 4b, c). Finally, we expressed each version of the ncRNA along with CspRecT, a high-efficiency SSAP²³, and mutL E32K, a dominant-negative mutL that eliminates mismatch repair at sites of single-base mismatch^{24,25} in BL21-AI cells that were unmodified, other than the removal of the endogenous Eco1 retron operon. Both ncRNAs resulted in appreciable editing after a single 16h overnight expression, but the extended version was significantly more effective (Fig 4d). To test whether the effect of the a1/a2 extension was locus-specific or generalized across genomic sites, we tested an additional three loci²⁶ for precise editing. We found that the engineered retron mediated editing at each additional loci, and that the efficiency of editing was improved by the a1/a2 extension at all three additional sites (Extended Data Fig 3). This shows that the abundance of the RT-DNA template for recombineering is a limiting factor for editing, and that modified ncRNA can be used to introduce edits at a higher rate.

Retron-derived RT-DNA can also be used to edit eukaryotic cells⁸. Specifically, in yeast, the ncRNA is modified to contain homology to a genomic locus and to add one or more nucleotide modifications in the loop of the msd, similar to the prokaryotic template. However, in this version, the ncRNA is on a transcript that also includes an SpCas9 gRNA and scaffold. When these components are expressed along with RT and SpCas9, the genomic

site is cut and repaired precisely using the RT-DNA as a template (Fig 4e). We tested our modified ncRNAs, using an architecture that was otherwise unchanged from a previously described version⁸. The ncRNA/gRNA transcript was expressed from a Galactose inducible promoter on a single-copy plasmid, flanked by ribozymes. Along with the plasmid-encoded ncRNA/gRNA, we expressed either Eco1 RT, Cas9, both the RT and Cas9, or neither, from Galactose inducible cassettes integrated into the genome. The ncRNA/gRNA was designed to target and edit the *ADE2* locus, resulting in both a two-nucleotide modification and a cellular phenotype, pink colonies.

Using the ncRNA with a 12-base a1/a2 length, we found that the expression of both the RT and Cas9 was necessary for editing based on pink colony counts, with only a small amount of background editing when we expressed Cas9 alone (Fig 4f, g). This is consistent with the reverse transcription of the ncRNA being required, rather than having the edit arise from the plasmid as donor. To test the effect of extending the a1/a2 region on genome editing efficiency, we designed two versions of the a1/a2 extended forms, both of which had a length of 27 base pairs, but differed in their a1/a2 sequence. We found that both versions outperformed the standard 12 base form for precise genome editing (Fig 4f, g). Consistent with our results in *E. coli*, this indicates that RT-DNA production is a limiting factor for precise genome editing, and that extended a1/a2 length is a generalizable modification that enhances retron-based genome engineering. We further confirmed these phenotypic results by sequencing the *ADE2* locus from batch cultures of cells (Fig 4h). Precise modifications of the site, resulting from edits that use the RT-DNA as a template, follow the same pattern as the phenotypic results, showing editing that depends on both the Cas9 nuclease and RT, and is increased by extension of the a1/a2 region.

We also found that the rates of precise editing determined by sequencing from batch cultures were consistently lower than those estimated from counting colonies. This is likely due to additional editing that continues to occur on the plate before counting, and our method of counting colonies as pink even if they were only partially pink. Another source of pink colonies could be any imprecise edits to the site that result in a non-functional *ADE2* gene. Indeed, we observed some *ADE2* loci that matched neither the wt nor precisely edited sequence. These occurred at a low rate (~1–3%) in all conditions, which was slightly elevated by Cas9 expression, but unaffected by RT expression/RT-DNA production (Extended Data Fig 4a). This, as well as the pattern of insertions, deletions, transitions, and transversions is consistent with a combination of sequencing errors and Cas9-produced indels (Extended Data Fig 4b, c).

As in the bacterial experiments, we tested whether the extended a1/a2 modification was a generalizable improvement by targeting additional loci across the genome. To this end, we generated wt and extended a1/a2 retrons to edit four additional loci²⁷ in yeast (TRP2, FAA1, CAN1, and LYP1). We found that for three out of four additional loci, the extended a1/a2 retrons yielded higher rates of precise editing, whereas one site showed lower, but still substantial rates of editing with the extended version (Extended Data Fig 5). Overall, across the nine sites tested in bacteria and yeast, the a1/a2 extension improved editing rates at eight sites.

Precise editing by retrons extends to human cells

Finally, we set out to test whether retron-produced RT-DNA could be used to for precise editing of human cells, as a step toward future therapeutic applications, as well as research applications seeking to unravel the mechanisms of genetic disease. Porting the editing machinery to cultured human cells required some additional modifications. In yeast, we produced both Cas9 and the retron RT from separate promoters. In human cells, expressing both of these proteins from a single promoter would greatly simplify the system and increase its portability. To identify an optimal single promoter architecture, we tested six arrangements in yeast: four fusion proteins using two different linker sequences with both orientations of Cas9 and Eco1 RT; and two versions where Cas9 and Eco1 RT were separated by a P2A²⁸ sequence in both possible orientations. These constructs were co-expressed with the best performing *ADE2* editing ncRNA/gRNA construct described above (extended v1, a1/a2 length 27 bp). We found that expression of these constructs resulted in a range of precise editing rates, with the Cas9-P2A-RT version yielding editing rates comparable to our previous versions based on two promoters (Fig 5a).

We then created two HEK293T cell lines that each harbored one of two integrating cassettes: Cas9 alone or Cas9-P2A-Eco1 RT (Fig 5b). We initially tested precise genome editing using a pol II driven ncRNA/gRNA flanked by ribozymes, as we had in yeast. However, we found no evidence of either precise editing or indels, consistent with previous reports of inefficient ribozyme-mediated gRNA release in human cells²⁹. Therefore, we changed the expression of our retron ncRNA/gRNA to be driven by a pol III H1 promoter, carried on a transiently transfected plasmid (Fig 5b). Six genomic loci (HEK3, RNF2, EMX1, FANCF, HEK4⁷, and AAVS1³⁰) were selected for editing, and an ncRNA/gRNA plasmid aiming to target and edit the site was generated.

The repair template was designed to introduce two distinct mutations, separated by at least 2 bp: the first introduced a single nucleotide change near the cut-site; the second recoded the PAM nucleotides (NGG → NHH, H: non-G nucleotide). The reasoning for this was two-fold: first, the multiple changes should both eliminate Cas9 cutting of the ncRNA/RT plasmid and re-cutting of the precisely recoded site; and second, these multiple, separated changes make it much less likely to mistakenly assign a Cas9-induced indel as a precise edit. As a technical aside, we would recommend against using single-base modifications to benchmark Cas9-induced precise editing applications as they are a common outcome of imprecise repair and can easily lead to inaccurate estimates of editing rate. We induced expression of the protein(s) for 24 hours, then transfected the ncRNA/gRNA plasmids, and subsequently harvested cells 3 days after transfection. Using targeted Illumina sequencing, we found precise editing of each site in the presence of the RT, well above the background rate of editing in the absence of the RT (Fig 5c–h). We believe that the small percentage of precise edits in the absence of the RT likely represents use of the plasmid as a repair template, and the gain in the editing rate in the presence of the RT indicates edits using RT-DNA as the template. Interestingly, we see that the rates of imprecise edits (indels) decline in the presence of the RT by roughly the same magnitude as the precise edits themselves, suggesting that the RT-DNA is being used to precisely edit sites that would have otherwise been edited imprecisely (Extended Data Figure 6).

DISCUSSION

The bacterial retron is a molecular component that can be exploited to produce designer DNA sequences *in vivo*. Our results yield a generalizable framework for retron RT-DNA production. Specifically, we show that a minimal stem length must be maintained in the *msd* to yield abundant RT-DNA and that the *msd* loop length affects RT-DNA production. We also show that there is a minimum length for the *a1/a2* complementary region. Perhaps most importantly, we demonstrate that the *a1/a2* region can be extended beyond its wt length to produce more abundant RT-DNA, and that increasing template abundance in both bacteria and yeast increases editing efficiency.

Importantly, these modifications are portable, both across retrons and across species. The extended *a1/a2* region produces more RT-DNA using Eco1 in bacteria and both Eco1 and Eco2 in yeast. Oddly, the extended *a1/a2* region did not increase RT-DNA production in cultured human cells. Further work will be necessary to optimize RT-DNA production in human cells specifically. Nonetheless, we provide a clear demonstration of retron-produced RT-DNA in human cells.

Retrons have been used to produce DNA templates for genome engineering^{6,8,9}, driven by the rationale that an intracellularly produced template eliminates the issues related to exogenous template delivery and availability. However, there have been no investigations of whether the RT-DNA templates are abundant enough to saturate the editing, or if even more template would lead to higher rates of editing. Our results establish that editing template abundance is limiting for genome editing in both bacteria and yeast, because extension of the *a1/a2* region, which increases the abundance of the RT-DNA, also increases editing efficiency.

Additionally, the inverted arrangement of the retron operon, with the ncRNA in the 3' UTR of the RT transcript, was found to produce RT-DNA in bacteria, yeast, and mammalian cells. This is the first time that a single, unifying retron architecture has been shown to be compatible with all of these host systems, simplifying comparisons and portability across kingdoms.

We also show, consistent with contemporaneous studies³¹, that the retron RT-DNA can be used as a template to precisely edit human cells. Further, our repair template design allows us to confidently call the precise editing rates. Importantly, we have also applied the same analysis to the Cas9 only conditions and reported the precise editing rates therein, and recommend that this approach be applied in future work. We believe that this will allow for estimations of the proportion of precise editing attributable to nuclease-only activity, and ultimately aid us in obtaining more realistic estimates of the precise editing rates attributable to the genome engineering tool of interest.

One major difference between the two eukaryotic systems (yeast/human) is the ratio of precise to imprecise editing. Yeast RT-DNA-based editing occurs at a ratio of ~74:1 precise edits to imprecise edits, while human editing inverted at a ratio of ~1:15 precise edits to imprecise edits. Whether this is a result of differences in repair pathways, or the substantial difference in the abundance of retron-produced RT-DNA between yeast and human cells that

we report here, it represents a clear direction for future research and technological advances in this area. In summary, this work represents an important advance in the versatile use of retron *in vivo* DNA synthesis and RT-DNA for genome editing across kingdoms.

METHODS

All biological replicates were collected from distinct samples and not the same sample measured repeatedly. Full statistics can be found in Supplementary Table 4.

Constructs and Strains

For bacterial expression, a plasmid encoding the Eco1 ncRNA and RT in that order from a T7 promoter (pSLS.436) was constructed by amplifying the retron elements from the BL21-AI genome and using Gibson assembly for integration into a backbone based on pRSFDUET1. The Eco1 RT was cloned separately into the erythromycin-inducible vector pJKR-O-mphR³² to generate pSLS.402. Eco1 ncRNA variants were cloned behind a T7/lac promoter in a vector based on pRSFDUET-1 with BsaI sites removed to facilitate Golden Gate cloning (pSLS.601), described further below. Eco1 RTs along with recombineering ncRNAs driven by T7/lac promoters (pSLS.491 and pSLS.492) were synthesized by Twist in pET-21(+).

Bacterial experiments were carried out in BL21-AI cells, or a derivative of BL21-AI cells. These cells harbor a T7 polymerase driven by a ParaB, arabinose-inducible, promoter. A knockout strain for the Eco1 operon (bSLS.114) was constructed from BL21-AI cells, using a strategy based on Datsenko and Wanner (2000)³³ to replace the retron operon with an FRT-flanked chloramphenicol resistance cassette. The replacement cassette was amplified from pKD3, adding homology arms to the Eco1 locus. This amplicon was electroporated into BL21-AI cells expressing lambda Red genes from pKD46 and clones were isolated by selection on 10µg/ml strength chloramphenicol plates. After genotyping to confirm locus-specific insertion, the chloramphenicol cassette was excised by transient expression of FLP recombinase to leave only a FRT scar.

For yeast expression, four sets of plasmids were generated. The first set of plasmids, designed to express the protein components for yeast genome editing, were based off of pZS.157⁸, a HIS3 yeast integrating plasmid for Galactose inducible Eco1RT and Cas9 expression (Gal1-10 promoter). A first set of variants of pZS.157, designed to compare the effect of wt vs. extended a1/a2 region lengths on genome editing were generated by PCR, and express either an empty cassette (pSCL.004), only Cas9 (pSCL.005), only the Eco1RT (pSCL.006), or both (pZS.157). A second set of variants was generated to test single-promoter expression of Cas9-Eco1RT variants. We designed 6 of such plasmids: Eco1RT-linker1-Cas9 (pSCL.71); Cas9-linker1-Eco1RT (pSCL.72); Eco1RT-linker2-Cas9 (pSCL.94); Cas9-linker2-Eco1RT (pSCL.95); Eco1RT-P2A-Cas9 (pSCL.102), and Cas9-P2A-Eco1RT (pSCL.103). The intervening sequences used were: Linker1 (GGTSSGGSGTAGSSGATSGG); Linker2 (SGGSSGGSSGSETPGTSESATPESGGSSGGSS)⁷; and P2A (ATNFSLLKQAGDVEENPGP)²⁸.

The second set of plasmids built for the genome editing experiments are based off of pZS.165⁸, a URA3+ centromere plasmid for Galactose (Gal7) inducible expression of a modified Eco1 retron ncRNA, which consists of an Eco1 msr -ADE2-targeting gRNA chimera, flanked by HH-HDV ribozymes. An initial variant of pZS.165 was generated by cloning an IDT-synthesized gBlock consisting of an Eco1 ncRNA (a1/a2 length: 12bp), which, when reverse transcribed, encodes a 200bp Ade2 repair template to introduce a stop codon (P272X) into the ADE2 gene (pSCL.002). Two additional plasmids were generated to extend the a1/a2 region of the Eco1 ncRNA to 27bp, with variations in the a1/a2 sequence (pSCL.039 and pSCL.040).

The third set of plasmids was built to assess the generalizability of the extended a1/a2 modification. The plasmids carrying wt length a1/a2 retrons are based off of pSCL.002, where the Ade2-targeting gRNA and Ade2-editing msd were replaced with analogous sequences to target and insert the following mutations: Can1 G444X (pSCL.106); Lyp1 E27X (pSCL.108); Trp2 E64X (pSCL.110); and Faa1 P233X (pSCL.112). The plasmids carrying extended length a1/a2 retrons are based off of pSCL.039, and were generated similarly to the wt length a1/a2 retron-encoding plasmids: Can1 G444X (pSCL.107); Lyp1 E27X (pSCL.109); Trp2 E64X (pSCL.111); and Faa1 P233X (pSCL.113).

The last set of plasmids, designed to compare the levels of RT-DNA production by the different retron systems, were derived from pSCL.002. IDT-synthesized gBlocks encoding a mammalian codon-optimized Eco1RT and ncRNA (wt), a dead Eco1RT and ncRNA (wt), and a human codon-optimized Eco2RT and ncRNA (wt), were cloned into pSCL.002 by Gibson Assembly, generating pSCL.027, pSCL.031 and pSCL.017, respectively. pSCL.027 was used to generate pSCL.028 by PCR, which carries a mammalian codon optimized Eco1RT and ncRNA (extended a1/a2: 27bp). Similarly, pSCL.017 was used to generate pSCL.034 by PCR, which carries a mammalian codon optimized Eco2RT and ncRNA (extended a1/a2: 29bp).

All yeast strains were created by LiAc/SS carrier DNA/PEG transformation³⁴ of BY4742³⁵. Strains for evaluating the genome editing efficiency of various retron ncRNAs were created by BY4742 integration of plasmids pZS.157, pSCL.004, pSCL.005 or pSCL.006 using KpnI-linearized plasmids for homologous recombination into the HIS3 locus. Transformants were isolated on SC-HIS plates. To evaluate effect of the length of the Eco1 ncRNA a1/a2 region on genome editing efficiency, these parental strains were transformed with episomal plasmids carrying the different retron ncRNA cassettes (pSCL.002, pSCL.039, or pSCL.040), and double transformants were isolated on SC-HIS-URA plates. The result was a set of control strains which were lacking one or both components of the genome editing machinery (i.e. Eco1RT, Cas9), and 3 strains which had all components necessary for retron-mediated genome editing but differed in the length of the Eco1 ncRNA a1/a2 region (12bp vs. 27bp).

Strains designed to assess the generalizability of the extended a1/a2 modification were created by transformation of a HIS3:pZS.157 yeast strain with plasmids carrying either wt or extended a1/a2 retrons for the editing of the four additional loci. Transformants were isolated on SC-HIS-URA plates. Strains to test single-promoter expression of Cas9-

Eco1RT variants were created by BY4742 integration of plasmids pSCL.71, pSCL.72, pSCL.94, pSCL.95, pSCL.102 or pSCL.103 using KpnI-linearized plasmids for homologous recombination into the HIS3 locus. Transformants were isolated on SC-HIS plates. These strains were then transformed with pSCL.39, and transformants isolated on SC-HIS-URA plates.

Strains designed to compare the levels of RT-DNA production by the different retron constructs were created by transformation of plasmids pSCL.027, pSCL.037 and pSCL.028 for Eco1 (wt, wt dead RT, and extended a1/a2, respectively) into BY4742; pSCL.017 and pSCL.031 for Eco2 (wt, and extended a1/a2, respectively) into BY4742. Transformants were isolated by plating on SC-URA agar plates. Expression of proteins and ncRNAs from all yeast strains was performed in liquid SC-Ura 2% Galactose media for 24h, unless specified.

For mammalian retron expression and quantification of RT-DNA production, synthesized gBlocks encoding human codon optimized Eco1 and Eco2 were cloned into a PiggyBac-integrating plasmid for doxycycline-inducible human protein expression (TetOn-3G promoter). Eco1 variants are wt retron-Eco1 RT and ncRNA (pKDC.018, with a1/a2 length: 12bp), extended a1/a2 length ncRNA (pKDC.019, with a1/a2 length: 27 bp), and a dead -Eco1 RT control (pKDC.020, with a1/a2 length: 27 bp). Eco2 variants were wt retron-Eco2 RT and ncRNA (pKDC.015, with a1/a2 length: 13bp), extended a1/a2 length ncRNA (pKDC.031, with a1/a2 length: 29 bp).

Stable mammalian cell lines for assessing RT-DNA production by wt and extended a1/a2 regions were created using the Lipofectamine 3000 transfection protocol (Invitrogen) and a PiggyBac transposase system. T25s of 50–70% confluent HEK293T cells were transfected using 8.3 ug of retron expression plasmids (pKDC.015, pKDC.018, pKDC.019, pKDC.020, or pKDC.031) and 4.2 ug PiggyBac transposase plasmid (pCMV-hyPBBase). Stable cell lines were selected with puromycin.

For assessment of retron-mediated precise genome editing in mammalian cells, two sets of plasmids were generated. The first set of plasmids, carrying either the SpCas9 gene or the SpCas9-P2A-Eco1RT construct, was built by restriction cloning of the respective genes, PCR amplified off of the aforementioned yeast vectors, into a PiggyBac-integrating plasmid for doxycycline-inducible human protein expression (TetOn-3G promoter). The second set of plasmids carries the ncRNA/gRNA targeting one of six loci in the human genome: HEK3 (pSCL.175); RNF2 (pSCL.176); EMX1 (pSCL.177); FANCF (pSCL.178); HEK4 (pSCL.179); and AAVS1 (pSCL.180). These were generated by restriction cloning of the ncRNA/gRNA cassette, built by primer assembly³⁶, into an H1 expression plasmid (FHUGW).

The ncRNA/gRNA cassette was designed as follows. The msd contains a repair template-encoding, 120bp sequence in its loop. The plasmid-encoded repair template is slightly asymmetric (49 bp of genome site homology upstream of Cas9 cut site; 71 bp of genome site homology downstream of cut site), and is complementary to the target strand – in practice, this means that after reverse transcription, the repair template RT-DNA is complementary to the non-target strand, as recommended in previous studies³⁷. The repair

template carries two distinct mutations: the first introduces a 1bp SNP at the Cas9 cut site; the second, designed to be at least 2bp away from the first mutation, recodes the Cas9 PAM (NGG → NHH, where H is any nucleotide beside G). The gRNA is 20bp.

Stable mammalian cell lines for assessing retron-mediated precise genome editing were created using the Lipofectamine 3000 transfection protocol (Invitrogen) and a PiggyBac transposase system. T25s of 50–70% confluent HEK293T cells were transfected using 8.3 ug of protein expression plasmids (pSCL.139 and pSCL.140) and 4.2 ug PiggyBac transposase plasmid (pCMV-hyPBBase). Stable cell lines were selected with puromycin.

Plasmids and strains are listed in Supplementary Tables 1 and 2. Primers used to generate and verify strains are listed in Supplementary Table 3. All plasmids will be made available on Addgene at the time of peer-reviewed publication.

qPCR

qPCR analysis of RT-DNA was carried out by comparing amplification from samples using two sets of primers. One set could only use the plasmid as a template because they bound outside the msd region (outside) and the other set could use either the plasmid or RT-DNA as a template because they bound inside the msd region (inside). Results were analyzed by first taking the difference in cycle threshold (CT) between the inside and outside primer sets for each biological replicate. Next, each biological replicate CT was subtracted from the average CT of the control condition (e.g. uninduced). Fold change was calculated as 2^{-CT} for each biological replicate. This fold change represents the difference in abundance of the inside versus outside template, where the presence of RT-DNA leads to fold change values >1.

For the initial analysis of EcoI RT-DNA when overexpressed in *E. coli*, the qPCR analysis used just three primers, two of which bound inside the msd and one which bound outside. The inside PCR was generated using both inside primers, while the outside PCR used one inside and one outside primer. For all other experiments, four primers were used. Two bound inside the msd and two bound outside the msd in the RT. qPCR primers are all listed in Supplementary Table 3.

For bacterial experiments, constructs were expressed in liquid culture, shaking at 37C for 6–16 hours after which a volume of 25ul of culture was harvested, mixed with 25ul H₂O, and incubated at 95C for 5 minutes. A volume of 0.3ul of this boiled culture was used as a template in 30ul reactions using a KAPA SYBR FAST qPCR mix.

For yeast experiments, single colonies were inoculated into SC-URA 2% Glucose and grown shaking overnight at 30C. To express the constructs, the overnight cultures were spun down, washed and resuspended in 1mL of water and passaged at a 1:30 dilution into SC-URA 2% Galactose, grown shaking for 24h at 30C. 250ul aliquots of the uninduced and induced cultures were collected for qPCR analysis. For qPCR sample preparation, the aliquots were spun down, resuspended in 50ul of water, and incubated at 100C for 15 minutes. The samples were then briefly spun down, placed on ice to cool, and 50ul of the supernatant was treated with Proteinase K by combining it with 29ul of water, 9ul of CutSmart buffer and 2ul

of Proteinase K (NEB), followed by incubation at 56C for 30 minutes. The Proteinase K was inactivated by incubation at 95C for 10 minutes, followed by a 1.5-minute centrifugation at maximum speed (~21,000 g). The supernatant was collected and used as a template for qPCR reactions, consisting of 2.5ul of template in 10ul KAPA SYBR FAST qPCR reactions.

For mammalian experiments, retron expression in stable HEK293T cell lines was induced using 1 ug/mL doxycycline for 24h at 37C in 6-well plates. 1ml aliquots of induced and uninduced cell lines were collected for qPCR analysis. qPCR sample preparation and reaction mix followed the yeast experimental protocol.

RT-DNA purification and PAGE analysis

To analyze RT-DNA on a PAGE gel after expression in *E. coli*, 2ml of culture were pelleted and nucleotides were prepared using a Qiagen mini prep protocol, substituting Epoch mini spin columns and buffers MX2 and MX3 for Qiagen components. Purified DNA was then treated with additional RNaseA/T1 mix (NEB) for 30 minutes at 37C and then single stranded DNA was isolated from the prep using an ssDNA/RNA Clean & Concentrator kit from Zymo Research. The purified RT-DNA was then analyzed on 10% Novex TBE-Urea Gels (Invitrogen), with a 1X TBE running buffer that was heated >80C before loading. Gels were stained with Sybr Gold (Thermo Fisher) and imaged on a Gel Doc imager (Bio-Rad).

To analyze RT-DNA on a PAGE gel after expression in *S. cerevisiae*, 5ml of overnight culture in SC-URA 2% Galactose was pelleted and RT-DNA was isolated by RNase A/T1 treatment of the aqueous (RNA) phase after TRIzol extraction (Invitrogen), following the manufacturer's recommendations with few modifications, as noted here. Cell pellets were resuspended in 500ul of RNA lysis buffer (100 mM EDTA pH8, 50 mM Tris-HCl pH8, 2% SDS) and incubated for 20 minutes at 85C, prior to the addition of the TRIzol reagent. The aqueous phase was chloroform-extracted twice. Following isopropanol precipitation, the RNA + RT-DNA pellet was resuspended in 265ul of TE and treated with 5ul of RNase A/T1 + 30ul NEB2 buffer. The mixture was incubated for 25 minutes at 37C, after which the RT-DNA was re-precipitated by addition of equal volumes of isopropanol. The resulting RT-DNA was analyzed on Novex 10% TBE-Urea gels as described above.

Variant library cloning

Eco1 ncRNA variant parts were synthesized by Agilent. Variant parts were flanked by BsaI type IIS cut sites and specific primers that allowed amplification of the sublibraries from a larger synthesis run. Random nucleotides were appended to the 3' end of synthesized parts so that all sequences were the same length (150 bases). The vector to accept these parts (pSLS.601) was amplified with primers that also added BsaI sites, so that the ncRNA variant amplicons and amplified vector backbone could be combined into a Golden Gate reaction using BsaI-HFv2 and T4 ligase to generate a pool of variant plasmids at high efficiency when electroporated into a cloning strain. Variant libraries were minipreped from the cloning strain and electroporated into the expression strain. Primers for library construction are listed in Supplementary Table 3. Variant parts are listed in Supplementary Data Set 1.

Variant library expression and analysis

Eco1 ncRNA variant libraries were grown overnight and then diluted 1:500 for expression. A sample of the culture pre-expression was taken to quantify the variant plasmid library, mixed 1:1 with H₂O and incubated at 95C for 5 minutes and then frozen at -20C. Constructs were expressed (arabinose and IPTG for the ncRNA, erythromycin for the RT) as the cells grew shaking at 37C for 5 hours, after which time two samples were collected. One was collected to quantify the variant plasmid library. That sample was mixed 1:1 with H₂O and incubated at 95C for 5 minutes and then frozen at -20C, identically to the pre-expression sample. The other sample was collected to sequence the RT-DNA. That sample was prepared as described above for RT-DNA purification.

The two variant plasmid library samples (boiled cultures) taken before and after expression were amplified by PCR using primers flanking the ncRNA region that also contained adapters for Illumina sequencing preparation. The purified RT-DNA was prepared for sequencing by first treating with DBR1 (OriGene) to remove the branched RNA, then extending the 3' end with a single nucleotide, dCTP, in a reaction with terminal deoxynucleotidyl transferase (TdT). This reaction was carried out in the absence of cobalt for 120 seconds at room temperature with the aim of adding only 5–10 cytosines before inactivating the TdT at 70C. A second complementary strand was then created from that extended product using Klenow Fragment (3'→5' exo-) with a primer containing an Illumina adapter sequence, six guanines, and a non-guanine (H) anchor. Finally, Illumina adapters were ligated on at the 3' end of the complementary strand using T4 ligase. In one variation, the loop of the RT-DNA for the a1/a2 library was amplified using Illumina adapter-containing primers in the RT-DNA, but outside the variable region from the purified RT-DNA directly. All products were indexed and sequenced on an Illumina MiSeq. Primers used for sequencing are listed in Supplementary Table 3.

Python software was custom written to extract variant counts from each plasmid and RT-DNA sample. In each case, these counts were then converted to a percentage of each library, or relative abundance (e.g., raw count for a variant over total counts for all variants). The relative abundance of a given variant in the RT-DNA sample was then divided by the relative abundance of that same variant in the plasmid library, using the average of the pre- and post-induction values, to control for differences in the abundance of each variant plasmid in the expression strain. Finally, these corrected abundance values were normalized to the average corrected abundance of the wt variant (set to 100%) or the loop length of 5 (set to 100%).

Recombineering expression and analysis

In experiments using the retron ncRNA to edit bacterial genomes, the retron cassette was co-expressed with CspRecT and mutL E32K from the plasmid pORTMAGE-Ec1²³ for 16 hours, shaking at 37C. After expression, a volume of 25ul of culture was harvested, mixed with 25ul H₂O, and incubated at 95C for 5 minutes. A volume of 0.3ul of this boiled culture was used as a template in 30ul reactions with primers flanking the edit site, which additionally contained adapters for Illumina sequencing preparation. These amplicons were

indexed and sequenced on an Illumina MiSeq instrument and processed with custom Python software to quantify the percent precisely edited genomes.

Yeast editing expression and analysis

For yeast genome editing experiments, single colonies from strains containing variants of the Eco1 ncRNA-gRNA cassette (wt or extended a1/a2 length for wt vs. extended a1/a2 region experiments; extended a1/a2 length v1 to test single-promoter expression of Cas9-Eco1RT variants) and editing machinery (-/+ Cas9, -/+ Eco1RT for wt vs. extended a1/a2 region experiments; Eco1RT-linker1-Cas9, Cas9-linker1-Eco1RT, Eco1RT-linker2-Cas9, Cas9-linker2-Eco1RT, Eco1RT-P2A-Cas9, Cas9-P2A-Eco1RT to test single-promoter expression of Cas9-Eco1RT variants) were grown in SC-HIS-URA 2% Raffinose for 24 hours, shaking at 30C. Cultures were passaged twice into SC-URA 2% Galactose (1:30 dilutions) for 24 hours, for a total of 48 hours of editing. At each timepoint (after 24h Raffinose, 24h Galactose, 48h Galactose), an aliquot of the cultures was harvested, diluted and plated on SC-URA low-ADE plates. Plates were incubated at 30C for 2–3 days, until visible and countable pink (ADE2 KO) and white (ADE2 WT) colonies grew. Editing efficiency was calculated in two ways. The first was by calculating the ratio of pink to total colonies on each plate for each timepoint. This counting was performed by an experimenter blind to the condition. The second was by deep sequencing of the target ADE2 locus. For this, we harvested cells from 250ul aliquots of the culture for each timepoint in PCR strips, and performed a genomic prep as follows. The pellets were resuspended in 120ul lysis buffer (see above), heated at 100C for 15 minutes and cooled on ice. 60ul of protein precipitation buffer (7.5M Ammonium Acetate) was added and the samples were gently inverted and placed at -20C for 10 minutes. The samples were then centrifuged at maximum speed for 2 minutes, and the supernatant was collected in new Eppendorf tubes. Nucleic acids were precipitated by adding equal parts ice-cold isopropanol, and incubating the samples at -20C for 10 minutes, followed by pelleting by centrifugation at maximum speed for 2 minutes. The pellets were washed twice with 200ul ice-cold 70% ethanol, and dissolved in 40ul of water. 0.5ul of the gDNA was used as template in 10ul reactions with primers flanking the edit site in ADE2, which additionally contained adapters for Illumina sequencing preparation (see Supplemental Table 4 for oligo sequences). Importantly, the primers do not bind to the ncRNA/gRNA plasmids. These amplicons were indexed and sequenced on an Illumina MiSeq instrument and processed with custom Python software to quantify the percent of P272X edits, caused by Cas9 cleavage of the target site on the ADE2 locus and repair using the Eco1 ncRNA-derived RT-DNA template.

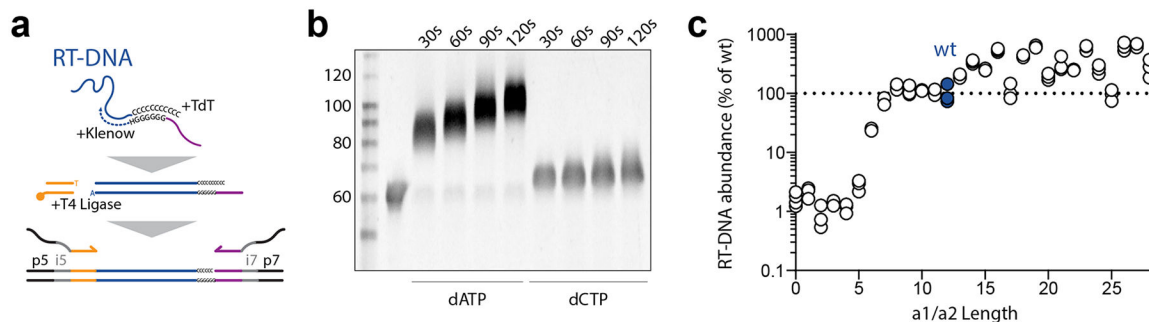
The editing experiments at additional loci were carried out as described above, with the difference that editing was quantified by amplifying 0.5ul of the gDNA with loci-specific primers, adapters for Illumina sequencing preparation. These primers are listed in Supplemental Table 3. Custom Python software was used to quantify the percent of precise edits, caused by Cas9 cleavage of the target site on the ADE2 locus and repair using the Eco1 ncRNA-derived RT-DNA template.

Human editing expression and analysis

For human genome editing experiments, Cas9 or Cas9-P2A-Eco1RT expression in stable HEK293T cell lines was induced using 1 ug/mL doxycycline for 24h at 37C in T12.5s. Then, cultures were transiently transfected with a plasmid constitutively expressing ncRNA/gRNA at a concentration of 5 ug plasmid per T12.5 using Lipofectamine 3000 (see site:plasmid list described above and in Supplemental Table 1). Cultures were passaged and doxycycline refreshed the following day for 48 more hours. 3 days post-transfection, cells were harvested for sequencing analysis.

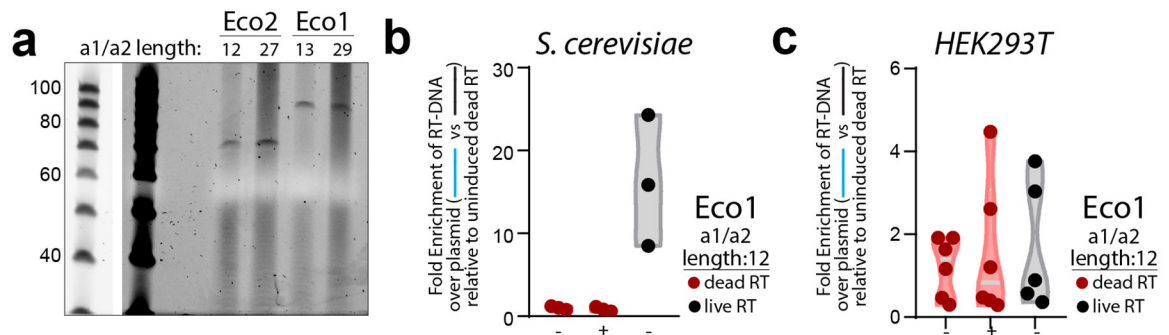
To prep samples for sequencing, cell pellets were processed and gDNA extracted using a QIAamp DNA mini kit, according to the manufacturer's instructions. DNA was eluted in 200uL of ultra-pure, nuclease free water. Then, 0.5ul of the gDNA was used as template in 12.5ul PCR reactions with primer pairs to amplify the locus of interest, which additionally contained adapters for Illumina sequencing preparation (see Supplemental Table 4 for oligo sequences). Importantly, the primers do not bind to the ncRNA/gRNA plasmids. The amplicons were purified using a QIAquick PCR purification kit according to the manufacturer's instructions, and the amplicons eluted in 12uL of ultra-pure, nuclease free water. Lastly, the amplicons were indexed and sequenced on an Illumina MiSeq instrument and processed with custom Python software to quantify the percent of on target precise and imprecise genomic edits.

Extended Data



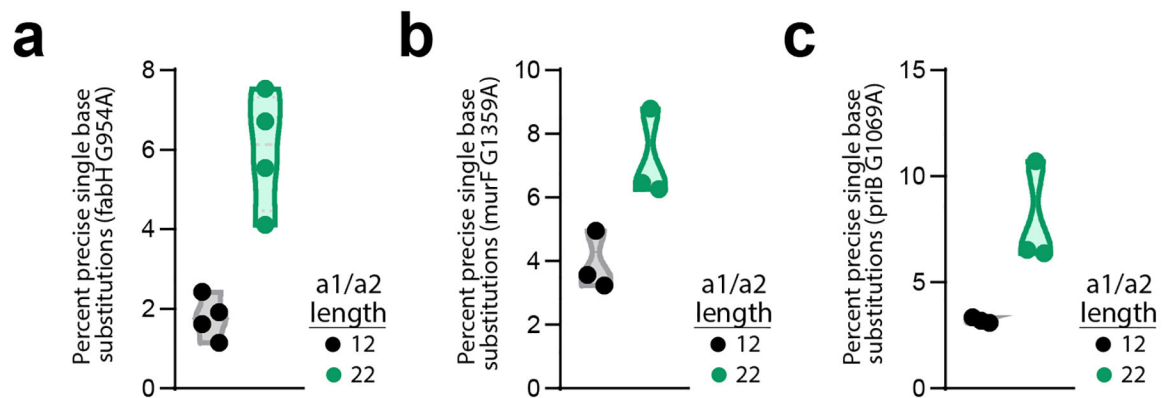
Extended Data Fig. 1. Extended Data Figure 1. Related to Figure 2.

a. Schematic of the sequencing prep pipeline for RT-DNA. **b.** Representative image of a PAGE analysis showing the addition of nucleotides to the 3' end of a single-stranded DNA, controlled by reaction time. The experiment was repeated twice with similar results. **c.** Alternate analysis of the RT-DNA for the a1/a2 length library, using a TdT-based sequencing preparation.



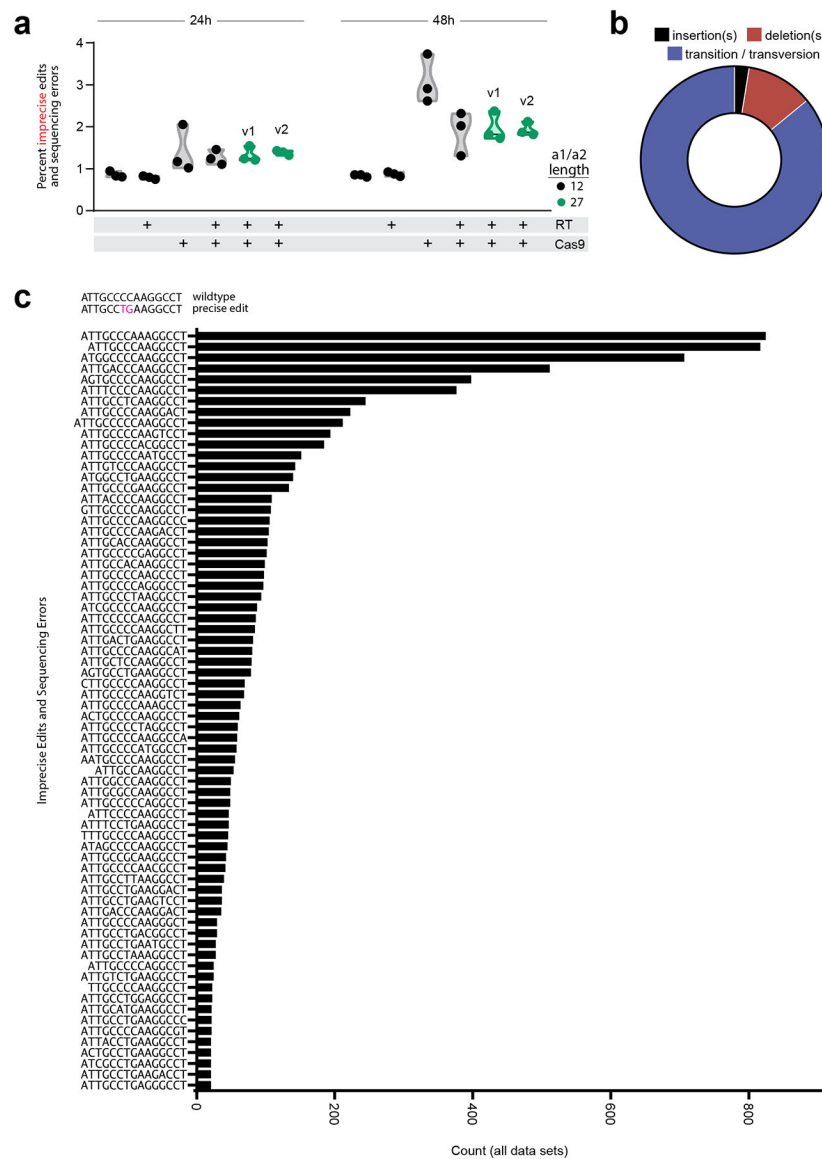
Extended Data Fig. 2. Extended Data Figure 2. Related to Figure 3.

a. Representative image of a PAGE analysis of Eco1 and Eco2 RT-DNA isolated from yeast. The ladder is shown at a different exposure to the left of the gel image. The experiment was repeated twice with similar results. **b.** Enrichment of the Eco1 RT-DNA/plasmid template when uninduced compared to a dead RT construct. Closed circles show each of three biological replicates, with red for the dead RT version and black for the live RT. **c.** Identical analysis as in b, but for Eco1 in HEK293T cells.



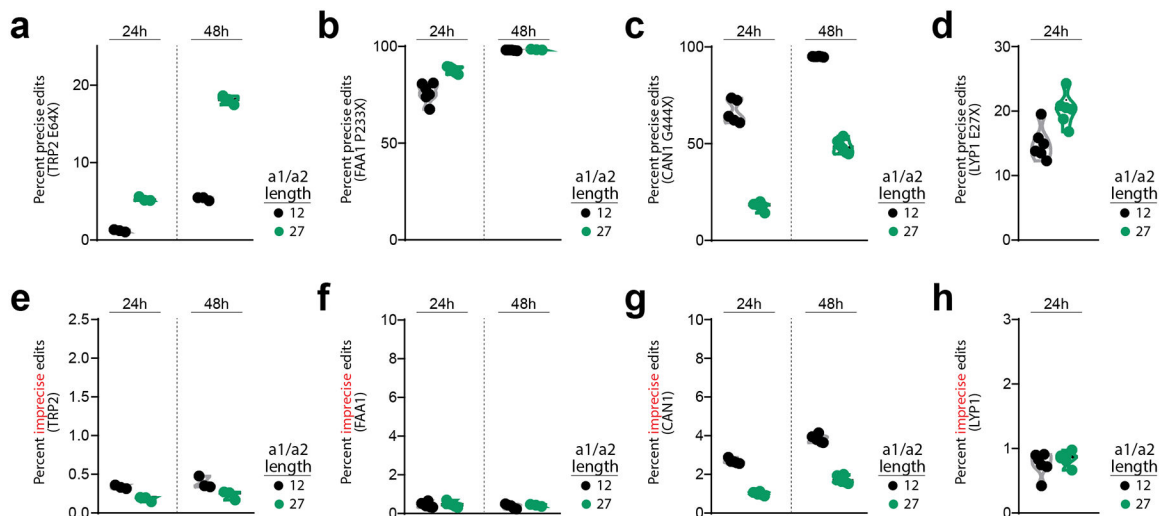
Extended Data Fig. 3. Extended Data Figure 3. Related to Figure 4a–d.

a-c. Percent of cells precisely edited, quantified by multiplexed sequencing, for the wt (black) and extended (green) recombineering constructs for three additional loci in *E. coli*.



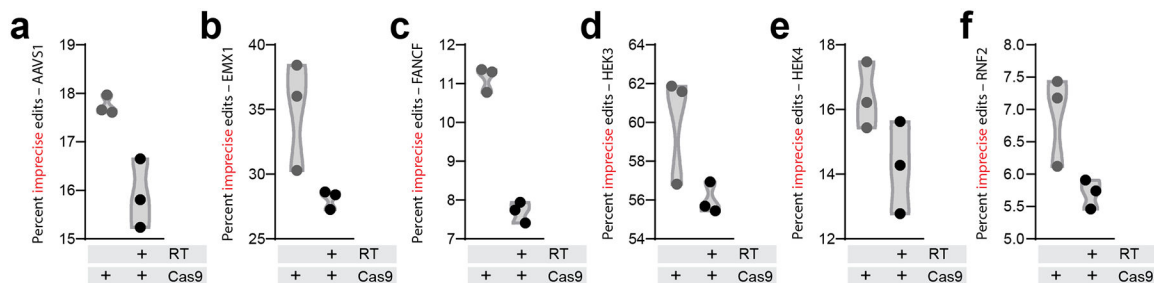
Extended Data Fig. 4. Extended Data Figure 4. Related to Figure 4h.

a. Percent of ADE2 loci with imprecise edits or sequencing errors at 24 and 48 hours. Closed circles show each of three biological replicates, with black for the wt a1/a2 length and green for the extended a1/a2 (two extended versions, v1 and v2). Induction conditions are shown below the graph for the RT and Cas9. **b.** Breakdown of the data in a. by type of edit/error. **c.** Imprecise edits and sequencing errors found in all data sets, ranked by frequency. Above the graph are the wt ADE2 locus and intended precise edit. On the Y axis are the imprecise edits and sequencing errors found. X axis represents count of each sequence in all data sets.



Extended Data Fig. 5. Extended Data Figure 5. Related to Figure 4e–h.

a–d. Percent of cells precisely edited, quantified by multiplexed sequencing, for the wt (black) and extended (green) recombineering constructs for four additional loci in *S. cerevisiae* at 24 and 48 hours. Cultures edited at the LYP1 E27X site were not viable beyond 24 hours. **e–h.** Percent of imprecise edits or sequencing errors for the loci in a–d.



Extended Data Fig. 6. Extended Data Figure 6. Related to Figure 5.

a–f. Percent of cells imprecisely edited (indels), quantified by multiplexed sequencing, in the presence of the ncRNA/gRNA plasmid and either Cas9 alone or Cas9 and EcoRI RT (as indicated below). Individual circles represent each of three biological replicates.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

Work was supported by funding from the Simons Foundation Autism Research Initiative (SFARI) Bridge to Independence Award Program, the Pew Biomedical Scholars Program, and a UCSF Program for Breakthrough Biomedical Research New Frontiers Research Award. S.L.S. acknowledges additional funding support from NIH/NIGMS (1DP2GM140917-01) and the L.K. Whittier Foundation. S.C.L. was supported by a Berkeley Fellowship for Graduate Study. K.D.C. was supported by an NSF Graduate Research Fellowship (2019247827), and a UCSF Discovery Fellowship. S.K.L. was supported by an NSF Graduate Research Fellowship (2034836). We thank Kathryn Claiborn for editorial assistance.

DATA AVAILABILITY

All data supporting the findings of this study are available within the article and its supplementary information. Sequencing data associated with this study is available through the NCBI BioProject database, with accession number PRJNA770365. All data required to re-create this manuscript's figures can be found in Source Data File 1. All uncropped gels can be found in Source Data File 2.

References:

1. Luo D & Saltzman WM Synthetic DNA delivery systems. *Nat Biotechnol* 18, 33–37, doi:10.1038/71889 (2000). [PubMed: 10625387]
2. Lin S, Staahl BT, Alla RK & Doudna JA Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* 3, e04766, doi:10.7554/eLife.04766 (2014). [PubMed: 25497837]
3. Paquet D et al. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* 533, 125–129, doi:10.1038/nature17664 (2016). [PubMed: 27120160]
4. Kosicki M, Tomberg K & Bradley A Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol* 36, 765–771, doi:10.1038/nbt.4192 (2018). [PubMed: 30010673]
5. Devkota S The road less traveled: strategies to enhance the frequency of homology-directed repair (HDR) for increased efficiency of CRISPR/Cas-mediated transgenesis. *BMB reports* 51, 437–443 (2018). [PubMed: 30103848]
6. Farzadfard F & Lu TK Synthetic biology. Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science* 346, 1256272, doi:10.1126/science.1256272 (2014). [PubMed: 25395541]
7. Anzalone AV et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157, doi:10.1038/s41586-019-1711-4 (2019). [PubMed: 31634902]
8. Sharon E et al. Functional Genetic Variants Revealed by Massively Parallel Precise Genome Editing. *Cell* 175, 544–557.e516, doi:10.1016/j.cell.2018.08.057 (2018). [PubMed: 30245013]
9. Schubert MG et al. High throughput functional variant screens via in-vivo production of single-stranded DNA. *bioRxiv*, 2020.2003.2005.975441, doi:10.1101/2020.03.05.975441 (2020).
10. Millman A et al. Bacterial Retrons Function In Anti-Phage Defense. *Cell* 183, 1551–1561.e1512, doi:10.1016/j.cell.2020.09.065 (2020). [PubMed: 33157039]
11. Bobonis J et al. Bacterial retrons encode tripartite toxin/antitoxin systems. *bioRxiv*, 2020.2006.2022.160168, doi:10.1101/2020.06.22.160168 (2020).
12. Bobonis J et al. Phage proteins block and trigger retron toxin/antitoxin systems. *bioRxiv*, 2020.2006.2022.160242, doi:10.1101/2020.06.22.160242 (2020).
13. Gao L et al. Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* 369, 1077–1084, doi:10.1126/science.aba0372 (2020). [PubMed: 32855333]
14. Mirochnitchenko O, Inouye S & Inouye M Production of single-stranded DNA in mammalian cells by means of a bacterial retron. *J Biol Chem* 269, 2380–2383 (1994). [PubMed: 7507924]
15. Lampson BC, Inouye M & Inouye S Retrons, msDNA, and the bacterial genome. *Cytogenetic and genome research* 110, 491–499, doi:10.1159/000084982 (2005). [PubMed: 16093702]
16. Simon AJ, Ellington AD & Finkelstein IJ Retrons and their applications in genome engineering. *Nucleic acids research* 47, 11007–11019, doi:10.1093/nar/gkz865 (2019). [PubMed: 31598685]
17. Inouye S, Hsu MY, Eagle S & Inouye M Reverse transcriptase associated with the biosynthesis of the branched RNA-linked msDNA in *Myxococcus xanthus*. *Cell* 56, 709–717, doi:10.1016/0092-8674(89)90593-x (1989). [PubMed: 2465092]
18. Lampson BC, Inouye M & Inouye S Reverse transcriptase with concomitant ribonuclease H activity in the cell-free synthesis of branched RNA-linked msDNA of *Myxococcus xanthus*. *Cell* 56, 701–707, doi:10.1016/0092-8674(89)90592-8 (1989). [PubMed: 2465091]

19. Lampson BC et al. Reverse transcriptase in a clinical strain of *Escherichia coli*: production of branched RNA-linked msDNA. *Science* 243, 1033–1038 (1989). [PubMed: 2466332]
20. Lim D & Maas WK Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in *E. coli* B. *Cell* 56, 891–904 (1989). [PubMed: 2466573]
21. Miyata S, Ohshima A, Inouye S & Inouye M In vivo production of a stable single-stranded cDNA in *Saccharomyces cerevisiae* by means of a bacterial retron. *Proceedings of the National Academy of Sciences of the United States of America* 89, 5735–5739, doi:10.1073/pnas.89.13.5735 (1992). [PubMed: 1378616]
22. Chappell SA, Edelman GM & Mauro VP Ribosomal tethering and clustering as mechanisms for translation initiation. *Proceedings of the National Academy of Sciences of the United States of America* 103, 18077–18082, doi:10.1073/pnas.0608212103 (2006). [PubMed: 17110442]
23. Wannier TM et al. Improved bacterial recombineering by parallelized protein discovery. *Proceedings of the National Academy of Sciences of the United States of America* 117, 13689–13698, doi:10.1073/pnas.2001588117 (2020). [PubMed: 32467157]
24. Aronshtam A & Marinus MG Dominant negative mutator mutations in the *mutL* gene of *Escherichia coli*. *Nucleic acids research* 24, 2498–2504, doi:10.1093/nar/24.13.2498 (1996). [PubMed: 8692687]
25. Nyerges Á et al. A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *Proceedings of the National Academy of Sciences of the United States of America* 113, 2502–2507, doi:10.1073/pnas.1520040113 (2016). [PubMed: 26884157]
26. Wang HH et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460, 894–898, doi:10.1038/nature08187 (2009). [PubMed: 19633652]
27. Zhang Y et al. A gRNA-tRNA array for CRISPR-Cas9 based rapid multiplexed genome editing in *Saccharomyces cerevisiae*. *Nature communications* 10, 1053, doi:10.1038/s41467-019-09005-3 (2019).
28. Liu JJ et al. CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* 566, 218–223, doi:10.1038/s41586-019-0908-x (2019). [PubMed: 30718774]
29. Knapp D et al. Decoupling tRNA promoter and processing activities enables specific Pol-II Cas9 guide RNA expression. *Nature communications* 10, 1490, doi:10.1038/s41467-019-09148-3 (2019).
30. Mali P et al. RNA-guided human genome engineering via Cas9. *Science* 339, 823–826, doi:10.1126/science.1232033 (2013). [PubMed: 23287722]
31. Kong X et al. Precise genome editing without exogenous donor DNA via retron editing system in human cells. *Protein & cell*, doi:10.1007/s13238-021-00862-7 (2021).
32. Rogers JK et al. Synthetic biosensors for precise gene control and real-time monitoring of metabolites. *Nucleic acids research* 43, 7648–7660, doi:10.1093/nar/gkv616 (2015). [PubMed: 26152303]
33. Datsenko KA & Wanner BL One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America* 97, 6640–6645, doi:10.1073/pnas.120163297 (2000). [PubMed: 10829079]
34. Gietz RD & Schiestl RH High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature protocols* 2, 31–34, doi:10.1038/nprot.2007.13 (2007). [PubMed: 17401334]
35. Brachmann CB et al. Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast (Chichester, England)* 14, 115–132, doi:10.1002/(sici)1097-0061(19980130)14:2<115::Aid-yea204>3.0.Co;2-2 (1998).
36. Tian S & Das R Primerize-2D: automated primer design for RNA multidimensional chemical mapping. *Bioinformatics (Oxford, England)* 33, 1405–1406, doi:10.1093/bioinformatics/btw814 (2017).
37. Richardson CD, Ray GJ, DeWitt MA, Curie GL & Corn JE Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol* 34, 339–344, doi:10.1038/nbt.3481 (2016). [PubMed: 26789497]

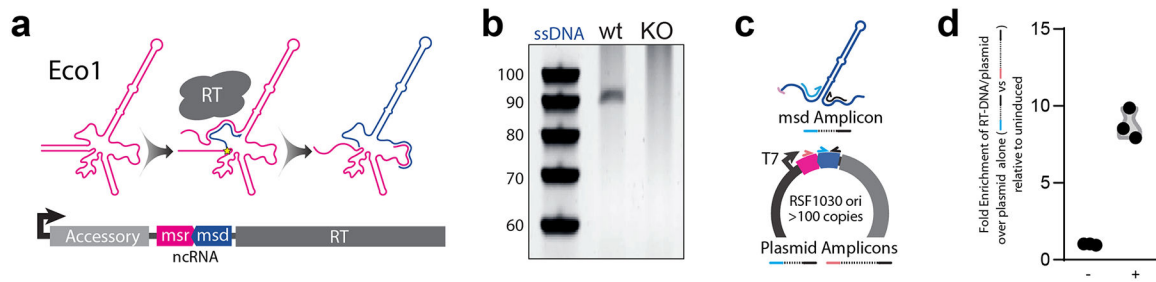


Figure 1. Bacterial retrons enable RT-DNA production.

a. Top: conversion of the ncRNA (pink) to RT-DNA (blue); bottom: schematic of the Eco1 retron operon. **b.** Representative image from N>3 PAGE analyses of endogenous RT-DNA produced from Eco1 in BL21-AI wild-type cells (wt) and a knockout of the retron operon (KO). **c.** qPCR analysis schematic for the RT-DNA. Blue/black primer pair will amplify using both the RT-DNA and the *msd* portion of the plasmid as a template. Red/black primer pair will only amplify using the plasmid as a template. **d.** Enrichment of the RT-DNA/plasmid template over the plasmid alone, relative to the uninduced condition, measured by qPCR. Induced vs. uninduced: $p = 0.0002$, unpaired t-test; N = 3 biological replicates. Circles represent each of three biological replicates.

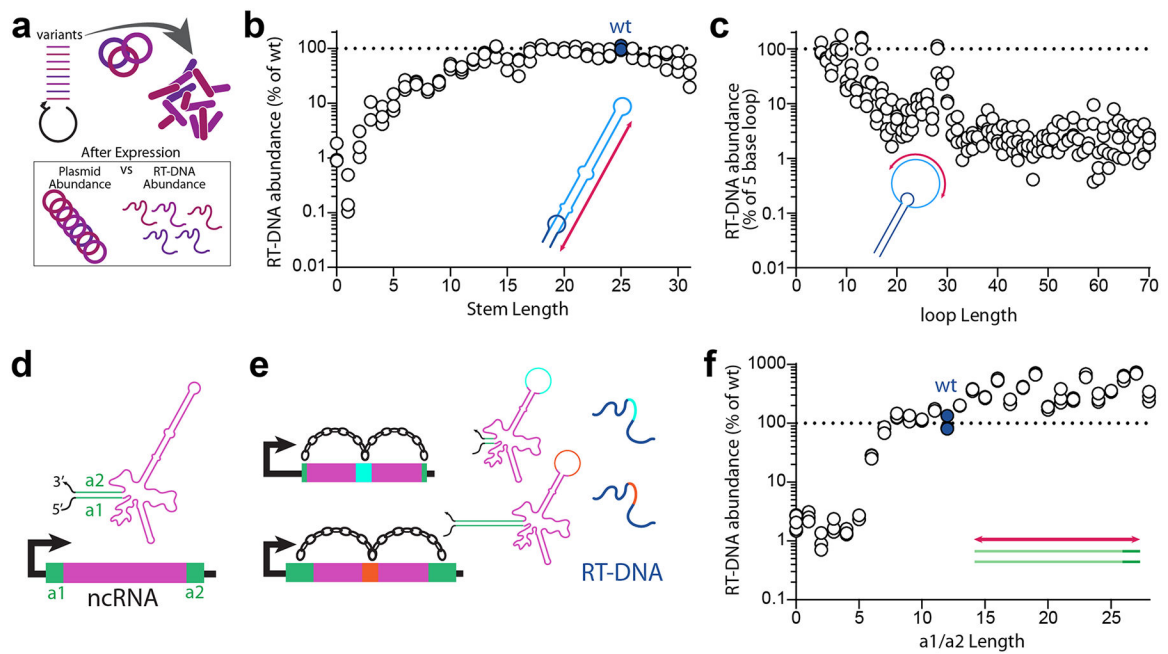


Figure 2. Modifications to retron ncRNA affect RT-DNA production.

a. Schematic of the variant library construction and analysis. **b.** Relative RT-DNA abundance of each stem length variant as a percent of wt. Circles represent each of three biological replicates. Wt length is shown in blue along with a dashed line at 100%. Effect of stem length: $p < 0.0001$, one-way ANOVA; $N = 3$ biological replicates. **c.** Relative RT-DNA abundance of each loop length variant as a percent of the value of 5 base loops. Circles represent each of three biological replicates, each of which is the average of five loops at that length with differing base content. A dashed line is shown at 100%. Effect of loop length: $p < 0.0001$, one-way ANOVA; $N = 3$ biological replicates. **d.** Schematic illustrating the a1 and a2 regions of the retron ncRNA. **e.** Variants of the a1/a2 region are linked to a barcode in the msd loop for sequencing. **f.** Relative RT-DNA abundance of each a1/a2 length variant as a percent of wt. Circles represent each of three biological replicates. Wt length is shown in blue along with a dashed line at 100%. Effect of a1/a2 length: $p < 0.0001$, one-way ANOVA; $N = 3$ biological replicates.

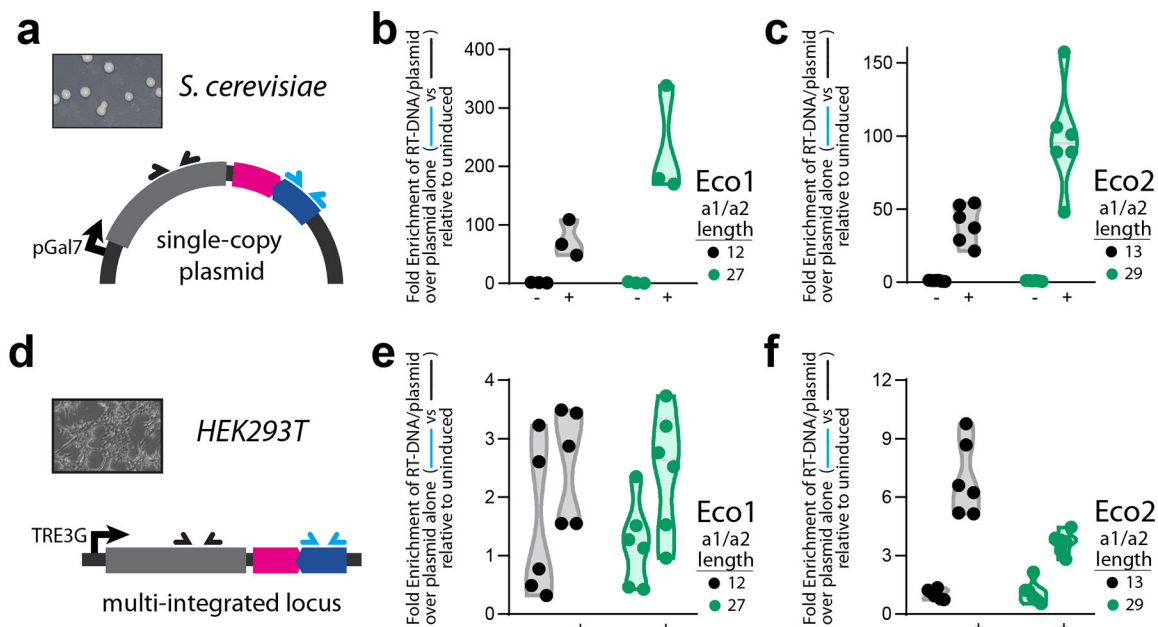


Figure 3. RT-DNA production in eukaryotic cells.

a. Schematic of the retron cassette for expression in yeast, with qPCR primers indicated.

b. Enrichment of the Eco1 RT-DNA/plasmid template over the plasmid alone by qPCR in yeast, with each construct shown relative to uninduced. Circles show each of three biological replicates, with black for the wt a1/a2 length and green for the extended a1/a2. One-way ANOVA, Sidak's multiple comparisons test (corrected): a1/a2 length 12, induced versus uninduced: $p = 0.2898$; a1/a2 length 27, induced versus uninduced: $p = 0.0015$; a1/a2 length 12 versus 27, induced: $p = 0.0155$; $N = 3$ biological replicates. **c.** qPCR of Eco2 in yeast, otherwise identical to **b.** One-way ANOVA, Sidak's multiple comparisons test (corrected): a1/a2 length 13, induced versus uninduced: $p = 0.006$; a1/a2 length 29, induced versus uninduced: $p < 0.0001$; a1/a2 length 13 versus 29, induced: $p < 0.0001$; $N = 6$ biological replicates. **d.** Schematic of the retron for expression in mammalian cells, with qPCR primers indicated. **e.** qPCR of Eco1 in HEK293T cells, otherwise identical to **b.** One-way ANOVA, Sidak's multiple comparisons test (corrected): a1/a2 length 12, induced versus uninduced: $p = 0.2897$; a1/a2 length 27, induced versus uninduced: $p = 0.1358$; a1/a2 length 12 versus 27, induced: $p = 0.9957$; $N = 5$ biological replicates. **f.** qPCR of Eco2 in HEK293T cells, otherwise identical to **b.** One-way ANOVA, Sidak's multiple comparisons test (corrected): a1/a2 length 13, induced versus uninduced: $p < 0.0001$; a1/a2 length 29, induced versus uninduced: $p = 0.0012$; a1/a2 length 13 versus 29, induced: $p < 0.0001$; $N = 6$ biological replicates.

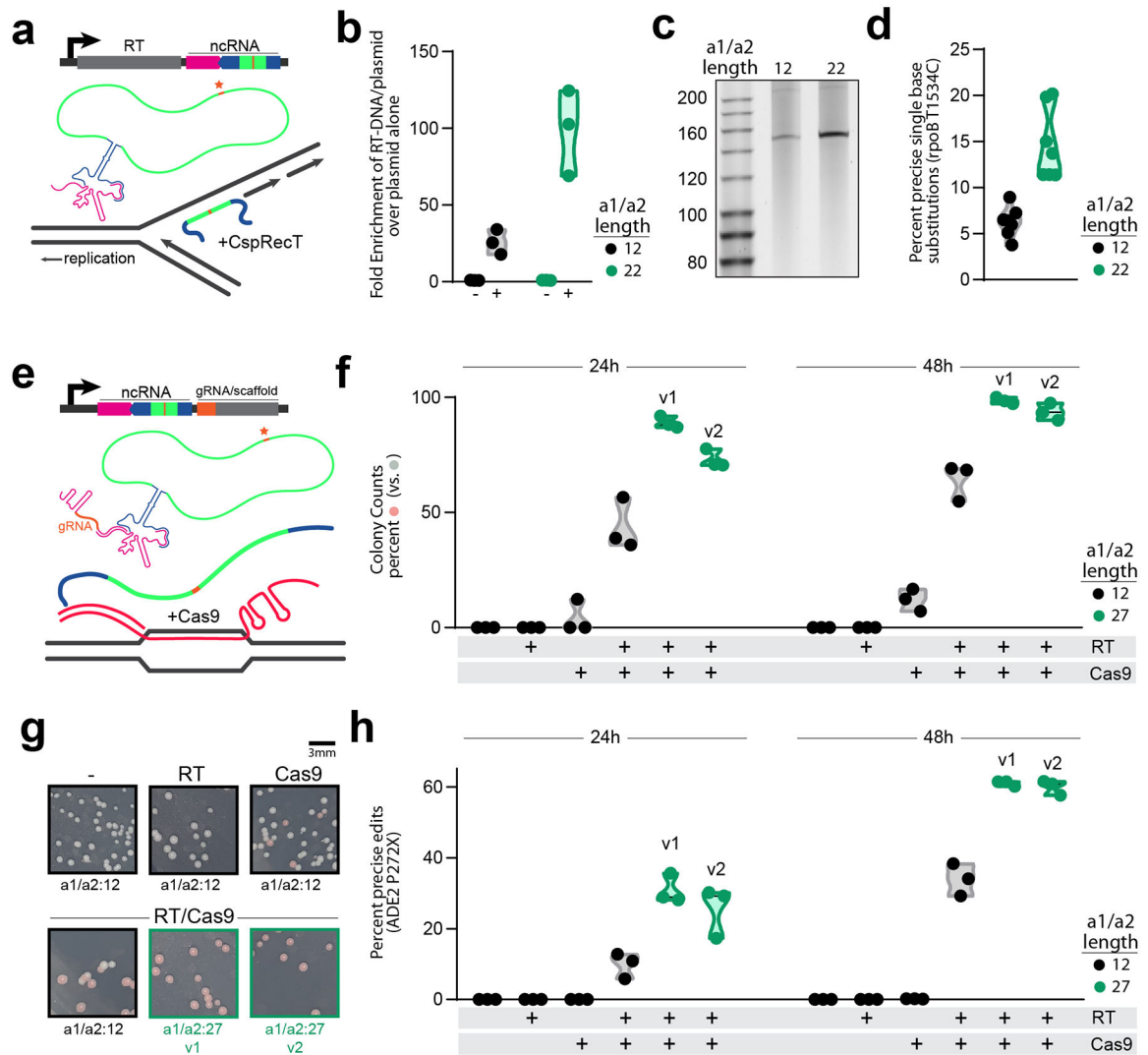


Figure 4. Improvements extend to applications in genome editing.

a. Schematic of an RT-DNA template for recombineering. **b.** Fold enrichment of the EcoI-based recombineering RT-DNA/plasmid template over the plasmid alone by qPCR in *E. coli*, with each construct shown relative to uninduced. Circles show each of three biological replicates, with black for the wt a1/a2 length and green for the extended a1/a2. One-way ANOVA, Sidak's multiple comparisons test (corrected): a1/a2 length 12, induced versus uninduced: $p = 0.1953$; a1/a2 length 22, induced versus uninduced: $p = 0.0001$; a1/a2 length 12 versus 22, induced: $p = 0.0008$; $N = 3$ biological replicates. **c.** PAGE gel showing purified RT-DNA for the wt (a1/a2 length: 12 bp) and extended (a1/a2 length: 22 bp) recombineering constructs to support quantitative PCR, $N = 1$. **d.** Percent of cells precisely edited, quantified by multiplexed sequencing, for the wt (black) and extended (green) recombineering constructs. unpaired t-test: a1/a2 length 12 versus 22: $p = 0.1953$; a1/a2 length 22, induced versus uninduced: $p = 0.0001$; a1/a2 length 12 versus 22, induced: $p = 0.0002$; $N = 6$ biological replicates. **e.** Schematic of an RT-DNA/gRNA hybrid for genome editing in yeast. **f.** Percent of colonies edited based on phenotype (pink colonies)

at 24 and 48 hours. Circles show each of three biological replicates, with black for the wt (a1/a2 length: 12 bp) and green for the extended a1/a2 (two extended versions, v1 and v2: a1/a2 length: 27 bp). Induction conditions are shown below the graph for the RT and Cas9. Two-way ANOVA: effect of condition (construct/induction), $p < 0.0001$; effect of time: $p < 0.0001$; $N = 3$ biological replicates. **g.** Representative images from each condition plotted in f, at 24h. Induction conditions above each image. **h.** Quantification of precise editing of the ADE2 locus in yeast by Illumina sequencing, plotted as in f. Two-way ANOVA: effect of condition (construct/induction), $p < 0.0001$; effect of time: $p < 0.0001$; $N = 3$ biological replicates.

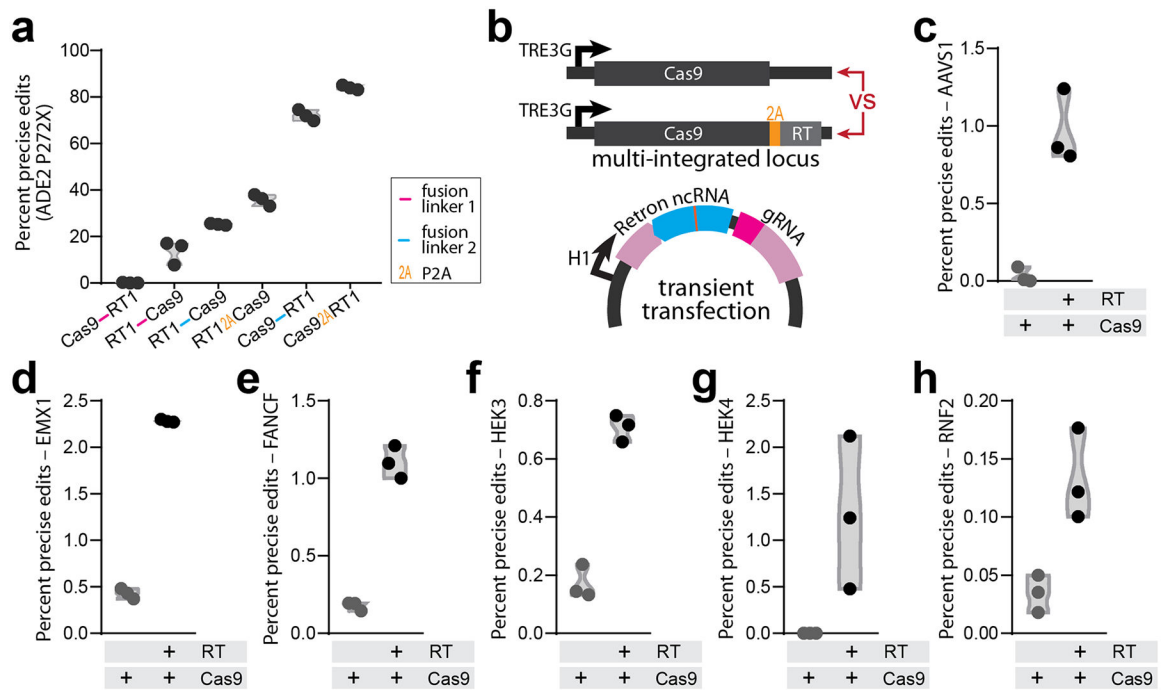


Figure 5. Precise editing by retrons extends to human cells.

a. Testing different single-promoter architectures for editing the ADE2 locus in *S. cerevisiae*. The arrangement of proteins is indicated below and the fusion linkers are listed in the Methods. Circles show each of three biological replicates. One-way ANOVA: effect of construct: $p < 0.0001$; $N = 3$ biological replicates. **b.** Schematic showing the elements for editing in human cells. Above are the integrated protein cassettes that are compared in panels **c-h**. Below is the plasmid for transient transfection of the site-specific ncRNA/gRNA. **c.** Quantification of precise editing of the AAVS1 locus in HEK293T cells by Illumina sequencing. Proteins present are shown below. Circles represent each of three biological replicates. Unpaired t-test: effect of Cas9 alone vs. Cas9 and RT: $p = 0.0026$; $N = 3$ biological replicates. **d-h.** Experiments and plots identical to **c**, but for EMX1, FANCF, HEK3, HEK4, and RNF2 loci, respectively. For d-h, unpaired t-test: effect of Cas9 alone vs. Cas9 and RT: $p < 0.0001$, $p = 0.0001$, $p = 0.0002$, $p = 0.0543$, and $p = 0.0158$, respectively; $N = 3$ biological replicates.