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-Kline\(^1\) and Seth L. Shipman\(\,^1,2\)\(^*\)

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 (ncRNA) that result in more abundant reverse-transcribed DNA (RT-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.

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 and 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 recipients. These technical challenges likely contribute to low rates of precise editing and unintended editing that occurs in the absence of template DNA\(^2–4\). 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 concentration\(^5\). 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 transcription\(^6–9\). This RT-DNA is produced in cells from plasmids, transgenes or viruses and benefits from transcriptional amplification to create high cellular concentrations that overcome inefficiencies in genome editing. One retroelement class that has been useful in this regard is bacterial retrons\(^6,8,9\), which are elements involved in phage defense\(^10–13\). Retrons are attractive as tools for biotechnology due to their compact size, tightly defined sites of reverse transcription initiation and termination, lack of known host factor requirements and lack of transposable elements. Indeed, retron-generated RT-DNA has demonstrated utility in bacterial\(^14,15\) and eukaryotic\(^16\) genome editing.

Despite the potential of the retron as a component of molecular biotechnology, so far, it has been modified 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 editing efficiency. Further, most work with retrons has been performed in bacteria, with only one functional demonstration of RT-DNA production in yeast\(^17\) and only a brief description of reverse transcription in mammalian cells (NIH3T3 mouse cells)\(^18\). 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 Escherichia coli to show that extension of complementarity in the a1/a2 region of the retron ncRNA increases production of RT-DNA. This effect was 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), an ncRNA that is both the primer and template for the RT and one or more accessory proteins\(^15\) (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 to 163 base pairs (bp)\(^16\). The ncRNA can be subdivided into a region that is reverse transcribed (msd) and a region that remains RNA in the final molecule (msr), which are partially overlapping\(^12,20\).

One of the first described retrons was found in E. coli, Ec01 (previously ec86)\(^10\). In BL21 cells, this retron is both present and active and produces 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 ~eight- to tenfold enrichment of the RT-DNA/plasmid region over the plasmid alone, which is evidence of robust reverse transcription (Fig. 1d).

Given that 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 sought 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 a 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. 2a). Cells harboring these library vector sets were grown overnight and then diluted, and ncRNA expression was induced during growth for 5 h.

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 isolating single-stranded DNA from double-stranded DNA using a commercial column-based kit. We then sequenced the RT-DNAs and compared 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 prepare each RT-DNA without biasing toward any variant. This involved tailing purified RT-DNA with a string of poly-nucleotides using a template-independent polymerase (terminal deoxynucleotidyl transferase (TdT)) and generating a complementary strand via an adapter-containing, inverse anchored primer. Finally, we ligated a second adapter to this double-stranded DNA and proceeded to indexing and multiplexed sequencing (Extended Data Fig. 1a,b).

In this first library, we modified the msd stem length from 0 to 31 bp and found that stem length can have a large impact on RT-DNA production (Fig. 2b). The RT tolerated modifications of the msd stem length that deviate by a small amount from the WT length of 25 bp. However, variants with stem lengths of <12 and >30 bp produced less than half as much RT-DNA than the WT. Therefore, we used a stem length of between 12 and 30 bp 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 bp 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-bp loops, which are closest in size to the WT length of 4 bp. We found a substantial decline in RT-DNA production as loop length increased from 5 to ~14 bp, but we observed almost no continued decline beyond that point other than a single point at 28 bp, which inexplicably produced more RT-DNA than its neighboring loops. While we were limited by our synthesis and sequencing parameters to 70 bp, our conclusion is that loops shorter than 14 bp are ideal for RT-DNA production; however, loops that extend beyond 14 bp 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 on themselves that we hypothesized plays a role in initiating reverse transcription (Fig. 2d). 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 9-bp 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. 2c) or prepared 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 7 bp 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 increased RT-DNA production by the extended a1/a2 region would be a portable modification to other retrons and to eukaryotic systems. To facilitate expression of Eco1 in eukaryotic cells, we inverted the operon from its native arrangement. In the endogenous arrangement, the ncRNA is in the 5′-untranslated region (UTR) of the RT transcript, requiring internal ribosome entry for the RT from a ribosome-binding site (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, and the ncRNA structure and out-of-frame ATG codons, is expected to negatively affect RT translation. Moreover, altering the a1/a2 region in the native arrangement could have unintended effects on RT translation. In the inverted architecture, the RT is driven by an RNA polymerase II (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.

We first tested this arrangement for Eco1 in Saccharomyces cerevisiae by placing the RT ncRNA cassette under the expression of a 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 and compared 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 bp resulted in more abundant RT-DNA production (Fig. 3b and Extended Data Fig. 2a). We then extended this analysis to another retron, Eco2 (ref. 19). We found a similar effect; although the WT ncRNA produced detectable RT-DNA, a version extending the a1/a2 region from 13 to 29 bp produced significantly more RT-DNA (Fig. 3b and Extended Data Fig. 2a). In each case, we compared induced to uninduced cells, which likely underreports 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 HEK293T cells. 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). By 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 diminished, rather than enhanced, 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...
recombinering\(^{13}\). 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 (for example, \(\lambda\ Red\beta\)), 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 performed in modified bacterial strains with numerous nucleases and repair proteins knocked out, because editing occurs at a low rate in WT cells\(^1\). 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 yeast and mammalian expression, with the ncRNA in the 3’-UTR of the RT. We used a 12-bp stem for the msd, which retains near-WT RT-DNA production. We constructed two versions of the editing retron, one with the WT 12-bp a1/a2 region and another with an extended 22-bp 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 a previously described version\(^{13}\), 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 Ecol retron operon. Both ncRNAs resulted in appreciable editing after a single 16-h 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 loci\(^{26}\) for precise editing. We found that the engineered retron mediated editing at each additional locus 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 recombining 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 cells\(^3\). 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 a Streptococcus pyogenes Cas9 (SpCas9) guide RNA (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 version\(^3\). 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 Eco1RT, 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-bp 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 a 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 bp but differed in their a1/a2 sequence. We found that both versions outperformed the standard 12-bp form which had a length of 22 bp and extended (a1/a2 length: 22 bp) recombineering constructs to support qPCR; n = 6 biological replicates. e, Schematic of an RT-DNA/gRNA hybrid for genome editing in yeast. f, Percentage of colonies edited based on phenotype (pink colonies) at 24 and 48 h. Induction conditions are above each image. g, Representative images from each condition plotted in h at 24 h. Induction conditions are 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 = 6 biological replicates.

Fig. 4 | Improvements extend to applications in genome editing. a, Schematic of an RT-DNA template for recombineering. b, Fold enrichment of the Eco1-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 the three biological replicates, with black for the WT a1/a2 length and green for the extended a1/a2 length; one-way ANOVA with 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 qPCR; 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, Percentage of colonies edited based on phenotype (pink colonies) at 24 and 48 h. Circles show each of the 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 h at 24 h. Induction conditions are 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.
template, follow the same pattern as the phenotypic results, showing editing that depends on both the Cas9 nuclease and RT, and are 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 insertion–deletions (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 loci in yeast (TRP2, FAA1, CAN1 and LYP1). We found that for three of the 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 sought to test whether retron-produced RT-DNA could be used for precise editing of human cells as a step toward future therapeutic applications and 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 Eco1RT, and two versions where Cas9 and Eco1RT were separated by a P2A28 sequence in both possible orientations. These constructs were coexpressed with the best-performing ADE2-editing ncRNA/gRNA construct described above (extended v1, a1/a2 length of 27 bp). We found that a1/a2 extension improved editing rates at eight sites.

**Fig. 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 the 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. Top, integrated protein cassette that are compared in c–h. Bottom, 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 the three biological replicates; one-way ANOVA, effect of construct: \( P < 0.0001; n = 3 \) biological replicates. d–h, Experiments and plots identical to c, but for EMX1 (d), FANCF (e), HEK3 (f), HEK4 (g) and RNF2 (h) loci, respectively; for d–h, unpaired t-test: effect of Cas9 alone versus Cas9 and RT: \( P < 0.0001, P = 0.0001, P = 0.0002, P = 0.0543 \) and \( P = 0.0158 \), respectively; \( n = 3 \) biological replicates.
However, we found no evidence of either precise editing or indels, consistent with previous reports of inefficient ribozyme-mediated gRNA release in human cells. Therefore, we changed the expression of our retron ncRNA/gRNA to be driven by a Pol III H1 promoter, which was carried on a transiently transfected plasmid (Fig. 5b). Six genomic loci (HEK3, RNF2, EMXI, FANCF, HEK4 (ref. 7) and AAVS1 (ref. 8)) 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, and the second recoded the PAM nucleotides (NGG → NHH, H: non-G nucleotide). The reasoning for this was twofold. First, the multiple changes should both eliminate Cas9 cutting of the ncRNA/RT plasmid and recoding of the precisely recoded site. 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 h, transfected the ncRNA/gRNA plasmids and collected cells 3 d 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 precise 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 Fig. 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, 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 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. Here, we show that a single, unifying retron architecture is compatible with all of these host systems, simplifying comparisons and portability across kingdoms.

We also show, consistent with contemporaneous studies, 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 nuclelease-only activity and will ultimately help 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/humans) is the ratio of precise to imprecise editing. Yeast RT-DNA-based editing occurs at a ratio of ~74:1 precise edits: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.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41589-021-00927-y.

Received: 25 March 2021; Accepted: 19 October 2021; Published online: 23 December 2021

References

1. Luo, D. & Saltzman, W. M. Synthetic DNA delivery systems. Nat. Biotechnol. 18, 33–37 (2000).
2. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. eLife 3, e04766 (2014).
3. Paquet, D. et al. Efficient introduction of specific homorogynous and heterorogynous mutations using CRISPR/Cas9. Nature 533, 125–129 (2016).
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 (2018).
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. RMB Rep. 51, 437–443 (2018).
6. Farzadfard, F. & Lu, T. K. Synthetic biology. Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. Science 346, 1256272 (2014).
7. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature 566, 149–157 (2019).
8. Sharon, E. et al. Functional genetic variants revealed by massively parallel precise genome editing. Cell 175, 544–557 (2018).
9. Schubert, M. G. et al. High throughput functional variant screens via in-vivo production of single-stranded DNA. Proc. Natl Acad. Sci. 118, e20181118 (2021).
10. Millman, A. et al. Bacterial retron function in anti-phage defense. Cell 183, 1551–1561 (2020).
11. Bobonis, J. et al. Bacterial retrons encode tripartite toxin/antitoxin systems. Preprint at bioRxiv https://doi.org/10.1101/2020.06.22.160168 (2020).
12. Bobonis, J. et al. Phage proteins block and trigger retron toxin/antitoxin systems. Preprint at bioRxiv https://doi.org/10.1101/2020.06.22.160242 (2020).
13. Gao, L. et al. Diverse enzymatic activities mediate antiviral immunity in prokaryotes. Science 369, 1077–1084 (2020).
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).

15. Lampson, B. C., Inouye, M. & Inouye, S. Retrons, msDNA, and the bacterial genome. *Cytogenet. Genome Res.* **110**, 491–499 (2005).

16. Simon, A. J., Ellington, A. D. & Finkelstein, I. J. Retrons and their applications in genome engineering. *Nucleic Acids Res.* **47**, 11007–11019 (2019).

17. Inouye, S., Hsu, M. Y., Eagle, S. & Inouye, M. Reverse transcriptase associated with the biosynthesis of the branched RNA-linked msDNA in *Myxococcus xanthus*. *Cell* **56**, 709–717 (1989).

18. Lampson, B. C., 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 (1989).

19. Lampson, B. C. et al. Reverse transcriptase in a clinical strain of *Escherichia coli*: production of branched RNA-linked msDNA. *Science* **243**, 1033–1038 (1989).

20. Lim, D. & Maas, W. K. Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA–RNA compound in *E. coli B*. *Cell* **56**, 891–904 (1989).

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. *Proc. Natl Acad. Sci. USA* **89**, 5733–5739 (1992).

22. Chappell, S. A., Edelman, G. M. & Mauro, V. P. Ribosomal tethering and clustering as mechanisms for translation initiation. *Proc. Natl Acad. Sci. USA* **103**, 18077–18082 (2006).

23. Wannier, T. M. et al. Improved bacterial recombineering by parallelized protein discovery. *Proc. Natl Acad. Sci. USA* **117**, 13689–13698 (2020).

24. Aronshtam, A. & Marinus, M. G. Dominant negative mutator mutations in the mutL gene of *Escherichia coli*. *Nucleic Acids Res.* **24**, 2498–2504 (1996).

25. Nyerges, Á. et al. A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *Proc. Natl Acad. Sci. USA* **113**, 2502–2507 (2016).

26. Wang, H. H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).

27. Zhang, Y. et al. A gRNA–tRNA array for CRISPR–Cas9 based rapid multiplexed genome editing in *Saccharomyces cerevisiae*. *Nat. Commun.* **10**, 1053 (2019).

28. Liu, J. J. et al. CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* **566**, 218–223 (2019).

29. Knapp, D. et al. Decoupling tRNA promoter and processing activities enables specific Pol-II Cas9 guide RNA expression. *Nat. Commun.* **10**, 1490 (2019).

30. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).

31. Kong, X. et al. Precise genome editing without exogenous donor DNA via retron editing system in human cells. *Protein Cell* https://doi.org/10.1007/s13238-021-00862-7 (2021).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021
**Methods**

All biological replicates were collected from distinct samples and not from 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-λ1 genome and using Gibson assembly for integration into a backbone based on pSFSDUET1. The Eco1RT was cloned separately into the erythromycin-inducible vector pKR-O miniF to generate pSLS.402. Eco1 ncRNA variants were cloned behind a T7/λ1 promoter in a vector based on pSFSDUET1 with Bsal sites removed to facilitate Golden Gate cloning (pSLS.601) and is described further below. Eco1 RTs along with recombination ncRNAs driven by T7/λ1 promoters (pSLS.491 and pSLS.492) were synthesized by Twist in pET-21(+).

Bacterial experiments were performed in BL21-AI cells or a derivative of BL21-AI cells. cultures can harbor a T7 polymerase driven by a Pabλ arabinose-inducible promoter. A KO strain for the Eco1 operon (bLSL.114) was constructed from BL21-AI cells using a strategy based on Datsenko and Wanner3 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 ampiclon was electroporated into BL21-AI cells expressing lambda Red genes from pKD46, and clones were isolated by selection on 10 μg ml\(^{-1}\) chloramphenicol plates. After genotyping to confirm locus-specific insertion, the chloramphenicol cassette was excised by transient expression of FLP recombinase to leave only an FRT scar.

For yeast expression, 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.157 (ref. 1), an HIS3 yeast integrating plasmid for galactose-inducible Eco1RT and Cas9 expression (Gal1 promoter). A first set of variants of pZS.157, designed to compare the effect of WT versus extended a1/a2 region lengths on genome editing, were generated by PCR and expressed either as 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 six such plasmids: Eco1RT-linker-1Cas9 (pSCL.71), Cas9-linker-1-Eco1RT (pSCL.72), Eco1RT-linker-2-Cas9 (pSCL.94), Cas9-linker-2-Eco1RT (pSCL.95), Eco1RT-P2A-Cas9 (pSCL.102) and Cas9-P2A-Eco1RT (pSCL.103). The intervening sequences used were linker 1 (GGTSSGGGSGATGGAATG), linker 2 (GGGSGGGGGGGGGTGGGATAGGGGSGGGGG) and P2A (ATNNSFLKQAGDVEENPBG)36.

The second set of plasmids was built for the genome-editing experiments were based off of pZS.165 (ref. 1), a URA3 centromere plasmid for galactose (Ga77)-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: 12 bp), which, when reverse transcribed, encodes a 200-bp ADE2 repair template to introduce a stop codon (P27X2). The gBlock was synthesized with the msd region (inside). Results were analyzed by first taking the difference in replicate. Next, each biological replicate's C\(_\text{t}\) of the control condition (for example, uninduced). Fold change was calculated as 2\(^{-ΔΔCt}\) 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 of >1.
RT-DNA was prepared for sequencing by first treating with DBRI (OriGene) to remove the branched RNA and then extending the 3' end with a single nucleotide, dCTP, in a reaction with TdT. This reaction was performed in the absence of cobalt to keep the room temperature at the aim of adding only one nucleotide before inactivating the TdT at 70 °C. 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 at on the 3' end of the complementary strand. In one variation, the Eco1 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 paired-end Illumina MiSeq RT-DNA reads and to calculate the editing efficiency for the a1/a2 library or relative abundance (for example, 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 postinduction values to correct for differences in copy number due to 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 five (set to 100%).

Recombineering expression and analysis. In experiments using the retron ncRNA to edit bacterial genomes, the retron cassette was coexpressed with CasPcrRT and mutl. E32K from the plasmid pORTMAGE-Ec1 (ref. 1) for 16 h with shaking at 37 °C. After expression, a volume of 25 μl of culture was collected, mixed with 25 μl of water and incubated at 95 °C for 5 min. A volume of 0.3 μl of this boiled culture was used as a template in a 30-μl reactions with primers flanking the edit site, which additionally contained adapters for Illumina sequencing preparation. The unoccupied RT-DNA was reprecipitated by addition of equal volumes of isopropanol. The resulting RT-DNA was amplified with primers that also added BsaI sites so that the ncRNA variant was of the same length (150 bp). The vector to accept these parts (pSLS.601) was digested with the restriction enzymes BsaI and SfiI (NEB), and the synthesized parts were ligated into the vector. The reaction was included for 10 min. The samples were then centrifuged at maximum speed for 2 min, and the precipitated RT-DNA pellet was reprecipitated 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. Ecoc ncRNA variant parts were synthesized by Agilent. Variant parts were flanked by BsaI type IIS cut sites and specific primers that allowed for amplification of the sublibraries from a larger synthesis run. Random nucleotides were added to the 3' end of synthesized parts so that all sequences were the same length (150bp). 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-HFⅡv and T4 ligase to generate a pool of variant plasmids at high efficiency when electroporated into a cloning strain. Variant libraries were miniprepped from the cloning strain and electroporated into the expression strain. Primers for library construction are listed in Supplementary Table 3. Variant parts were included in Supplementary Data 2.

Variant library expression and analysis. Ecoc ncRNA variant libraries were grown overnight and then diluted 1:500 for expression. A sample of the culture preexpression was taken to quantify the variant plasmid library, mixed 1:1 with water, incubated at 95 °C for 5 min and frozen at −20 °C. Constructs were expressed (arabinose and IPTG for the ncRNA, erythromycin for the RT) as cells grew with shaking at 37 °C for 5 h, after which two samples were collected. One was collected to quantify the variant plasmid library. That sample was mixed 1:1 with water, incubated at 95 °C for 5 min and frozen at −20 °C identical to the preexpression 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 DNA was prepared for sequencing by first treating with DBRI (OriGene) to remove the branched RNA and then extending the 3' end with a single nucleotide, dCTP, in a reaction with TdT. This reaction was performed in the absence of cobalt to keep the room temperature at the aim of adding only one nucleotide before inactivating the TdT at 70 °C. 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 the 3' end of the complementary strand. In one variation, the Eco1 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 paired-end Illumina MiSeq RT-DNA reads and to calculate the editing efficiency for the a1/a2 library or relative abundance (for example, 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 postinduction values to correct for differences in copy number due to 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 five (set to 100%).

Yeast editing expression and analysis. For yeast genome-editing experiments, single colonies from strains containing variants of the Ecoc ncRNA–gRNA cassette (WT or extended a1/a2 length for WT versus 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 versus extended a1/a2 region experiments; Eco1RT–linker 1–Cas9, Cas9–linker 1–Eco1RT, Eco1RT–linker 2–Cas9, Cas9–linker 2–Eco1RT, and Cas9–P2A–Eco1RT to test single-promoter expression of Cas9–Eco1RT variants) were grown in SC-HIS-URA 2% raffinose for 24 h with shaking at 30 °C. Cultures were passed twice into SC-URA 2% galactose (1:30 dilutions) for 24 h for a total of 48 h of editing. At each timepoint (after 24 h of raffinose, 24 h of galactose, 48 h of galactose), an aliquot of the cultures was collected, diluted and plated on SC-URA low-ADE plates. Plates were incubated at 30 °C for 3–4 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 colonies to total colonies on each plate for each timepoint. This counting was performed by an experimenter blinded to the condition. The second was by deep sequencing of the target locus. For this, we collected cells from 250-μl aliquots of the culture for each timepoint in PCR strips and performed a genomics preparation as follows. The pellets were resuspended in 120 μl of lysis buffer (see above), heated at 100 °C for 15 min and cooled on ice. Protein precipitation buffer (60 μl, 7.5 M ammonium acetate) was added, and the samples were gently inverted and placed at −20 °C for 10 min. The samples were then centrifuged at maximum speed for 2 min, 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 −20 °C for 10 min followed by pelleting by centrifugation at maximum speed for 2 min. The pellets were washed twice with 200 μl of ice-cold 70% ethanol and dissolved in 40 μl of water. gDNA (0.5 μl) was used as template in 10-μl reactions with primers flanking the target site and additional primers containing an Illumina adapter sequence for Illumina sequencing preparation (see Supplementary Table 4 for oligonucleotide 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 percentage of precisely edited genomes.
transiently transfected with a plasmid constitutively expressing ncRNA/gRNA at a concentration of 5 µg of plasmid per T12.5 using Lipofectamine 3000 (see plasmid list described above and Supplementary Table 1). Cultures were passaged, and doxycycline was refreshed the following day for an additional 48 h. Three days after transfection, cells were collected for sequencing analysis.

To prepare samples for sequencing, cell pellets were processed, and gDNA was extracted using a QIAamp DNA mini kit according to the manufacturer’s instructions. DNA was eluted in 200 µl of ultra-pure, nuclease-free water. Then, 0.5 µl of the gDNA was used as template in 12.5-µl PCR reactions with primer pairs to amplify the locus of interest, which also contained adapters for Illumina sequencing preparation (see Supplementary Table 4 for oligonucleotide 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 were eluted in 12 µl 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 percentage of on-target precise and imprecise genomic edits.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**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 are available through the NCBI BioProject database under accession number PRJNA770365. Source data are provided with this paper.

**Code availability**
Custom code to process or analyze data from this study is available on GitHub at https://github.com/Shipman-Lab/retron_architectures.

**References**
32. Rogers, J. K. et al. Synthetic biosensors for precise gene control and real-time monitoring of metabolites. *Nucleic Acids Res.*, 43, 7648–7660 (2015).
33. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* 97, 6640–6645 (2000).
34. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* 2, 31–34 (2007).
35. Brachmann, C. B. 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* 14, 115–132 (1998).
36. Tian, S. & Das, R. Primerize-2D: automated primer design for RNA multidimensional chemical mapping. *Bioinformatics* 33, 1405–1406 (2017).
37. 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. *Nat. Biotechnol.* 34, 339–344 (2016).

**Acknowledgements**
This 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 K. Claiborn for editorial assistance.

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

**Additional information**
Extended data is available for this paper at https://doi.org/10.1038/s41589-021-00927-y.
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41589-021-00927-y.
Correspondence and requests for materials should be addressed to Seth L. Shipman.
Peer review information *Nature Chemical Biology* thanks Channabasavaiah Gurumurthy and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at www.nature.com/reprints.
Extended Data Fig. 1 | RT-DNA sequencing prep. 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. Related to Fig. 2.
Extended Data Fig. 2 | RT-DNA production in eukaryotic cells. 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. Related to Fig. 3.
Extended Data Fig. 3 | Precise genome editing rates across additional genomic loci in *E. coli*. 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*. Related to Fig. 4a–d.
Extended Data Fig. 4 | Imprecise editing profile of the yeast ADE2 locus. 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. Related to Fig. 4h.
Extended Data Fig. 5 | Genome editing rates across additional genomic loci in yeast. 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. Related to Fig. 4e–h.
Extended Data Fig. 6 | Imprecise editing rates across genomic loci in human cells. 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 Eco1 RT (as indicated below). Individual circles represent each of three biological replicates. Related to Fig. 5.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

☑ n/a Confirmed
☐ □ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐ □ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐ □ The statistical test(s) used AND whether they are one- or two-sided
☒ □ Only common tests should be described solely by name; describe more complex techniques in the Methods section.
☐ □ A description of all covariates tested
☐ □ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒ □ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐ □ For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
☒ □ Give P values as exact values whenever suitable.
☐ □ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐ □ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐ □ Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection: Illumina FASTQ. Generation version 1.0.0 was used for collection of sequencing data from a MiSeq instrument.

Data analysis: Custom code to process or analyze data from this study was implemented in Python 3.8.5, and is available on Github here: https://github.com/Shipman-Lab/retron_architectures.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. Github). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on 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.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | No sample size calculation was performed. Experiments were performed at n>=3 biological replicates, and no statistical testing was performed until all biological replicates were collected. |
|-------------|---------------------------------------------------------------------------------------------------------------|
| Data exclusions | No data was excluded. |
| Replication | All biological replicates were collected from distinct biological samples (n>=3), not the same sample measured repeatedly. In addition to biological replicates, we tested editing efficiency across a range of cell types and a range of loci within those cell types (n=3 biological replicates). All attempts at replication were successful. |
| Randomization | Initial cell sources go back to a common stock, so randomization is not relevant. Biological replicates are separate transfections/transformations/clones. Induced (experimental) and uninduced (control) cultures were drawn from the same starting material. |
| Blinding | Colony counting was performed blinded at the analysis step, by having one experimenter take images of yeast plates, which were then coded and counted by a different experimenter. For other experiments, blinding was not relevant to data collection and analysis because all samples (controls and experimental) were treated, collected and analyzed identically, using the same experimental workflow and computational analysis pipeline. |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | Methods |
|----------------------------------|---------|
| n/a | Involved in the study |
| ☑ | Antibodies |
| ☑ | Eukaryotic cell lines |
| ☑ | Palaeontology and archaeology |
| ☑ | Animals and other organisms |
| ☑ | Human research participants |
| ☑ | Clinical data |
| ☑ | Dual use research of concern |
| ☑ | Involved in the study |
| ☑ | ChiP-seq |
| ☑ | Flow cytometry |
| ☑ | MRI-based neuroimaging |

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s) | HEK293T: ATCC

Authentication | Low passage cells from ATCC were used without further authentication.

Mycoplasma contamination | not tested

Commonly misidentified lines (See ICCLAC register) | No commonly misidentified cell lines were used in this study.

Dual use research of concern

Policy information about dual use research of concern

Hazards
Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

| No | Yes |
|----|-----|
| ☐  | ☒  | Public health |
| ☒  |    | National security |
| ☒  | ☐  | Crops and/or livestock |
| ☐  | ☐  | Ecosystems |
| ☐  | ☐  | Any other significant area |

Hazard: A part of this research concerns gene editing, which could be used to modify cells in a heritable manner. For examples of agents subject to oversight, see the United States Government Policy for Institutional Oversight of Life Sciences Dual Use Research of Concern.

Experiments of concern

Does the work involve any of these experiments of concern:

| No | Yes |
|----|-----|
| ☐  | ☒  | Demonstrate how to render a vaccine ineffective |
| ☒  | ☐  | Confer resistance to therapeutically useful antibiotics or antiviral agents |
| ☒  | ☐  | Enhance the virulence of a pathogen or render a nonpathogen virulent |
| ☒  | ☐  | Increase transmissibility of a pathogen |
| ☒  | ☐  | Alter the host range of a pathogen |
| ☒  | ☐  | Enable evasion of diagnostic/detection modalities |
| ☐  | ☐  | Enable the weaponization of a biological agent or toxin |
| ☐  | ☐  | Any other potentially harmful combination of experiments and agents |

Precautions and benefits

Biosecurity precautions

| The ethical and dual-use concerns of gene editing are widely known and this manuscript does nothing to suggest uses beyond what is currently being considered or in practice. |

Biosecurity oversight

| n/a |

Benefits

| Gene editing has enormous potential benefits for fundamental research, industrial processes, and human therapeutics. |

Communication benefits

| Advancements in gene editing can make for better experiments, more controlled industrial uses, and safer therapeutics. |