scientific reports

OPEN

Check for updates

Targeted insertion of large genetic payloads using cas directed LINE-1 reverse transcriptase

Femila Manoj¹, Laura W. Tai², Katelyn Sun Mi Wang³ & Thomas E. Kuhlman^{3⊠}

A difficult genome editing goal is the site-specific insertion of large genetic constructs. Here we describe the GENEWRITE system, where site-specific targetable activity of Cas endonucleases is coupled with the reverse transcriptase activity of the ORF2p protein of the human retrotransposon LINE-1. This is accomplished by providing two RNAs: a guide RNA targeting Cas endonuclease activity and an appropriately designed payload RNA encoding the desired insertion. Using *E. coli* as a simple platform for development and deployment, we show that with proper payload design and co-expression of helper proteins, GENEWRITE can enable insertion of large genetic payloads to precise locations, although with off-target effects, using the described approach. Based upon these results, we describe a potential strategy for implementation of GENEWRITE in more complex systems.

Discovered as a bacterial immune system against foreign genetic elements such as phages, CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) Associated Proteins (Cas) are endonucleases that target and cleave DNA sequences based upon their homology with a "guide RNA"^{1,2}. Consequently, by providing an engineered "single-guide" RNA (sgRNA), Cas enzymes can be targeted to cleave any desired sequence. This flex-ibility in gene editing by CRISPR-Cas endonucleases has revolutionized genome editing^{3–6} in a wide variety of organisms^{7–23} and in its application to clinical therapeutics²⁵.

Despite their flexibility and ease of use, the repertoire of genome editing modalities that CRISPR/Cas systems allow remains limited. Knockout or point mutants can be generated relatively easily by targeting Cas cleavage to coding or control regions of the genome. The cell must repair such cuts to survive, and errors introduced by the nonhomologous end joining (NHEJ) repair machinery can lead to inactivation of control regions or introduction of missense or point mutations to coding sequences^{26–29}. An additional editing modality is to introduce novel sequences to the genome through Homology Directed Repair (HDR), where a DNA fragment with ends homologous to the sequences flanking the cut site and containing the desired sequence to be inserted is introduced to the cell along with the Cas-sgRNA ribonucleoprotein (RNP) complexes. After cleavage, the fragment is then used to repair the cut by the cell's homologous recombination repair machinery, resulting in its integration. However, HDR remains inefficient and difficult to accomplish, particularly for gene-sized or larger [$\geq ~1$ kilobase pair (kbp)] fragments^{30–33}. A primary reason for this difficulty is that for HDR to be successful, non-homologous end joining (NHEJ) DNA repair, the primary repair mechanism for DNA repair in advanced eukaryotic cells^{34–37}, must be suppressed^{38–40}.

Here we introduce a method for the active insertion of lengthy genetic sequences into host DNA we call GENEWRITE: Genome Engineering With RNA-Integrating Targetable Endonucleases. This is accomplished by coupling the targetable endonuclease activity of Cas enzymes to the reverse transcriptase activity of the human retrotransposon LINE-1 through translationally fusing Cas and LINE-1 reverse transcriptase proteins (Fig. 1A). A number of recent reports have described approaches coupling the targetability of Cas enzymes with the activity of other transposons or reverse transcriptases. These include Tn7-like transposons whose genomic insertion is accomplished through an associated CRISPR-effector, from the cyanobacterium *Scytonema hofmanni*⁴¹ and *Vibrio cholerae*⁴². Insertion of these 2–3 kbp bacterial transposons is programmable to specific genomic locations in *E. coli* through a guide RNA similar to other Cas enzymes. Another approach, prime editing⁴³, fuses a catalytically impaired Cas9 fused to an engineered Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase^{44–47}, using a "prime editing guide RNA" (pegRNA) to target short insertions, deletions, and all types of point mutations into human cells. GENEWRITE offers functionality that is distinct from each of these examples. While prime editing similarly uses a reverse transcriptase to insert RNA-encoded sequences into the genome, insertions

¹Microbiology Program, University of California Riverside, Riverside, CA 92521, USA. ²Department of Biology, University of California Riverside, Riverside, CA 92521, USA. ³Department of Physics and Astronomy, University of California Riverside, Riverside, CA 92521, USA. ^{Sem}email: thomas.kuhlman@ucr.edu



Figure 1. GENEWRITE components and strategy. (**A**) ORF2p and GENEWRITE domain structure. Wildtype ORF2p consists of endonuclease (EN, blue), Z (Z, orange), reverse transcriptase (RT), and cysteine-rich RNA binding domains (Cys, yellow). The GENEWRITE protein replaces the EN domain with a Cas protein (Cas9 or Cas12a/Cpf1, blue) and includes an N-terminal EGL13 nuclear localization signal (NLS, green), C-terminal c-Myc NLS (NLS, green), and 6xHis tag for in vitro purification (His, gray). (**B**) GENEWRITE components. The system consists of the GENEWRITE protein and a DNA target for insertion. A guide sgRNA complementary to the desired cut site (red) and a payload RNA encoding the desired insertion with a 3' end designed to hybridize to the insertion target (green). Optionally, as described in the text, NHEJ proteins, ORF1p protein, and 5' homology on the payload RNA to the target site can be included to increase insertion efficiency. (**C**) The sgRNA directs Cas cleavage to the integration site. (**D**) After Cas-induced cleavage, the 3' end of the payload RNA hybridizes with the cut site priming TPRT (**E**). After mRNA removal and second strand synthesis by host enzymes, the cut site is resolved (**F**).

.....

performed with prime editing are typically limited to short 10–40 bp epitopes. Conversely, we illustrate the site-specific reverse transcription and insertion of ~ 1.5 kbp payload RNAs, larger than that offered by prime editing.

Previous studies have shown that reverse transcription by the LINE-1 protein ORF2p can be directed to pre-existing nicks and cuts in targeted DNA sequences in vitro⁴⁸, and we have previously shown that LINE-1 is functional in *E. coli*, particularly when complemented by expression of enzymes for NHEJ repair⁴⁹. Here we use *E. coli* expressing *B. subtilis* NHEJ enzymes as a simple platform to optimize design and delivery of GENEWRITE for future application to more complex systems. The strategy used here for integration using GENEWRITE is shown in Fig. 1B–F. *E. coli* cells expressing GENEWRITE are transformed with two RNAs: a guide sgRNA to target Cas cleavage to the desired integration site, and a payload RNA carrying the coding sequence of the desired integration. The 3' end of the payload RNA is designed to be homologous to the bottom DNA strand downstream from the cut such that RNA–DNA hybridization occurs to prime reverse transcription. Host enzymes complete second strand synthesis and payload RNA removal, and the insert is sealed into the site⁴⁹. We illustrate that GENEWRITE can be used to effectively target insertion of large, gene-sized payloads to specific locations, although not without off-target effects.

Results

GENEWRITE rationale and design. The human retrotransposon LINE-1 (Long Interspersed Nuclear Element, or L1) encodes the two proteins ORF1p and ORF2p, and both proteins are required for efficient retrotransposition in humans. A primary function of ORF1p appears to be chaperone activity⁵⁰, while ORF2p includes endonuclease (EN) and reverse transcriptase (RT) domains. To retrotranspose, ORF2p EN nicks TArich target DNA, and the 3' end of the LINE-1 mRNA hybridizes with DNA adjacent to the nick to initiate reverse transcription through a process called target primed reverse transcription (TPRT)⁵¹. In most active L1 elements, this hybridization is facilitated through the presence of a ~ 100 bp long poly(A) tract, which is also thought to be the primary binding target of ORF2p to its encoding mRNA⁵².

LINE-1 and its accessory proteins naturally exist in human cells, making it an appealing target for optimization as a genome editing tool. To attempt to further enhance specifically targeted reverse transcribed insertions by ORF2p in vivo, we removed the promiscuous ORF2p EN domain by deleting amino acids 1–347. The remaining fragment, from amino acids 348–1275, which includes the Z, RT, and cysteine-rich RNA-binding domains, we dub ORF2pZRT. Finally, the GENEWRITE protein consists of a translational fusion of ORF2pZRT to targetable Cas endonucleases (Cas9 or Cas12a/Cpf1) with a flexible linker. In addition, the GENEWRITE protein includes N and C-terminal nuclear localization signals (NLS) and a C-terminal 6xHis tag to enable purification (Fig. 1A). A previous similar attempt at replacing ORF2p EN with Cas9 and using *Alu*-like payload RNA to target ORF2p RT to specific loci in human cells proved unsuccessful⁵³. As described below, we have made several refinements to the GENEWRITE system relative to this attempt, including the use of *Escherichia coli* as a simpler in vivo platform in which to test and optimize. We additionally show that the 10 base pair homology between target and payload used in this previous study is likely inadequate for priming of TPRT.

High expression of GENEWRITE protein is Lethal to E. coli. We designed and synthesized the GENEWRITE protein under control of a T7 promoter, which was cloned into the plasmid pUC57-*kan*. We transformed this plasmid into *E. coli* strain BL21-AI, along with either empty plasmid pZA31, or pZA31 carrying *ykoU* and *ykoV B. subtilis* NHEJ enzymes expressed from $P_{Ltet01}^{49,54}$. In strain BL21-AI, GENEWRITE expression is inducible by the addition of L-arabinose. Curiously, while expression of Cas9/12a, ORF2pZRT, or both Cas9/12a and ORF2pZRT in individual *E. coli* cells does not affect growth, strong expression of the GENEWRITE Cas-ORF2pZRT fusion protein induced through the addition of arabinose is lethal to *E. coli*. This lethality is partially relieved by simultaneous expression of *B. subtilis* NHEJ enzymes. This suggests lethality may be a consequence of genomic breaks generated by GENEWRITE, perhaps driven by high affinity of ORF2p to arbitrary RNAs in vivo⁵⁵. Consequently, the results described below rely upon low, leakage levels of expression of GENEWRITE without induction.

GENEWRITE is effective at insertions into high-copy number targets in E. coli. We expected the strategy outlined in Fig. 1B–F to be difficult to successfully execute for a number of reasons, including the expected difficulty of co-transforming individual cells with appropriate amounts of both sgRNA and payload RNA, as well as previously documented preference of ORF2p to act primarily upon its *cis*-encoding RNA⁵⁶. Hence, as an initial integration target, we chose the high copy number plasmid pUC57-*kan* [~500–1000 /cell] from which the GENEWRITE protein itself is expressed to maximize chances of success. For experiments described here, the ~1200 bp payload RNA consisted of an *aadA* spectinomycin resistance gene driven by a strong, constitutive *lacI*Q1 promoter⁵⁷ and Shine-Dalgarno ribosomal binding site (RBS). Consequently, after the GENEWRITE protocol, cells were spread on plates containing spectinomycin to select for potentially successful integrants.

Based upon our current understanding of TPRT, design of the payload RNA 3' hybridization region is critical. To determine the optimal length of the hybridization region, we generated an array of six identical payload RNAs with hybridization length variable from 0 to 50 bp in 10 bp increments. Based on prior reports of the essentiality of a 3' poly(A) tract for ORF2p binding and reverse transcription⁵², we generated a second array of payload RNAs, identical to the first, but also including the 30 bp poly(A) tract found in the SINE element AluYA5⁵⁸.

We transformed the pUC57-targeting sgRNA along with each payload RNA into *E. coli* weakly expressing GENEWRITE-Cas9, either with or without simultaneous expression of *B. subtilis* NHEJ enzymes. The results are shown in Fig. 2. For those payload RNAs containing a poly(A) tract, we observed very few spectinomycin resistant colonies, for both with or without simultaneous co-expression of NHEJ. Conversely, without the poly(A) tract, we obtained hundreds of spectinomycin resistant colonies when complemented with co-expression of NHEJ (Fig. 2A). Site-specific integration was verified by PCR using primers that amplified across the 5' and 3' integration junctions (Fig. 2B); 63 out of 96 colonies screened yielded a positive signal for a success rate of ~72% (Fig. 2C). Sequencing of eight purified plasmids revealed some small deletions at the 5' end of the insertion (Supplementary Fig. S1). From these experiments, we conclude that optimal design of the payload RNA includes 40–50 bp of 3' homology to the intended target facilitated by NHEJ DNA repair, and with no poly(A) tract. This difference in essentiality of the poly(A) tract to TPRT between *E. coli* and humans may be the result of mRNA 3' poly(A) tails stabilizing RNAs in eukaryotes, while poly(A) tails designate mRNAs for degradation in bacteria^{59–62}.

We performed a series of controls and further investigations using the payload designed to target pUC57*kan* with 40 bp 3' hybridization region (Fig. 2D): (1) as expected, the sgRNA is required for efficient targeting and integration; (2) NLS sequences at the N- and C-termini do not significantly interfere with function; (3) the Cas12a/Cpf1 GENEWRITE variant is functional, although with lower efficiency than the Cas9 variant, consistent with previous findings that blunt-end cuts fragments serve as better TPRT substrates than those with 3' or 5' overhangs⁴⁸; and (4) simultaneous co-expression of unfused Cas9 and ORF2pZRT, rather than the translationally-fused GENEWRITE protein, is not functional. However, LINE-1 reverse transcriptase has been shown to function even when encoded and expressed separately from the endonuclease through association via the naturally occurring cryptic Z domain, raising the possibility of potentially using naturally expressed LINE-1 in the human genome as an editing tool.

GENEWRITE can insert payloads into low and single copy targets. For the next target, we attempted insertion into the much lower copy number pZA31 plasmid hosting the NHEJ genes [$\sim 20-30$ copies/ cell⁵⁴]. Using 40 bp of 3' homology to target as described above, we obtained no colonies when transforming payload and sgRNAs into cells expressing GENEWRITE but deficient in NHEJ. However, we obtained ~ 50 colonies on average when transforming into cells expressing both GENEWRITE and NHEJ proteins. PCR screening of putative positive colonies generated a positive signal in 10 out of 50 colonies (Fig. 3A), yielding a success rate of 20%.



Figure 2. GENEWRITE site-specific insertion in a high-copy number plasmid. sgRNA and payload RNAs were designed to integrate an *aadA* spectinomycin resistance gene into the plasmid pUC57-*kan*. (**A**) Number of spectinomycin resistant colonies as a function of payload RNA 3'hybridization length. Dark red: -NHEJ + poly(A); Dark cyan: -NHEJ -poly(A); Bright red: + NHEJ + poly(A); Bright cyan: + NHEJ -poly(A). Data points are the average of three replicates, error bars are SD. (**B**) PCR verification of integration. Lanes 1, 4, 13: NEB 1 kb Plus Ladder. 300 and 400 bp bands are indicated. Lanes 5–12: amplicons resulting from four spectinomycin resistant colonies. For each pair of lanes, the leftmost lane is PCR across the 5' junction (300 bp amplicon expected), rightmost is PCR across 3' junction (400 bp amplicon expected). (**C**) Representative screening of 16 randomly selected colonies by PCR across the 5' integration junction. (**D**) Controls and effect of various GENEWRITE components, relative to that of intact GENEWRITE-Cas9 with 40 bp homology payload RNA (first column). 2nd column: replacement of Cas9 with Cas12a/Cpf1; 3rd column: effect of removal of NLS and His tags; 4th column: simultaneous expression of unfused Cas9 and ORF2pZRT; 5th column: GENEWRITE-Cas9 transformed with only guide RNA but no payload; 6th column: GENEWRITE-Cas9 transformed with only guide.

.....

We finally attempted to use GENEWRITE to site-specifically insert a payload into single copy chromosomal loci. We attempted insertions at three loci we have previously shown to accept insertions at high efficiency using recombineering-like methods: the *nth* locus near the terminus of replication; the *atpI* locus near the origin of replication; and the *ybbD* locus midway on the right replichore [Fig. $3B^{63,64}$]. In these cases, repeated attempts at the GENEWRITE protocol as described above were unsuccessful. Prior reports⁶⁵ and our own studies of retrotransposition of native LINE-1 in *E. coli* (Supplementary Fig. S2) suggest that homology between the 5' end of the payload and insertion location may also aid in targeting. Consequently, we attempted two strategies: (1) inclusion of 20 bp of 5' homology between the payload and the targeted insertion site; and (2) simultaneous co-expression of ORF1p. Each of these strategies alone was unsuccessful. Only when targeting *nth* by including 20 bp of 5' and 40 bp of 3' payload homology to the target, along with simultaneous co-expression of ORF1p, did we obtain significant numbers of colonies after transformation (~ 20 colonies on average). Under these



Figure 3. GENEWRITE site-specific insertion in a low-copy number targets. (**A**) Insertion in low copy number (15-30 copies /cell) plasmid pZA31. Chosen primers bind at start of payload promoter and within the adjacent pZA31 sequence after 3' end of payload. Colony PCR was performed with 10% DMSO to eliminate extraneous non-specific amplification. Expected amplicon is ~ 1300 bp. (**B**) Attempted chromosomal insertion sites. Coordinates of sites are $x_{chromosome} = 0.0098$ (*atp1*), 0.5476 (*ybbD*), and -0.943 (*nth*), where x = 0 corresponds to *oriC* and $x = \pm 1$ corresponds to *terC*. (**C**) PCR amplification across *nth* integration location. Primers bind to chromosomal regions adjacent to targeted integration site. Amplicon expected from successful integration is ~ 1600 bp. We conservatively identify the last colony as negative for integration.

conditions, PCR screening of 50 positive colonies (Fig. 3C, Supplementary Fig. S3) demonstrate a success rate of 60%. However, repeated attempts at insertion the *atpI* and *ybbD* sites with ORF1p co-expression and 20 bp of 5' and 40 bp of 3' payload homology to target have so far proven unsuccessful. As with 3' homology, further optimization of the amount of 5' homology to the target included in the payload may improve the efficiency of insertion at low copy number targets. Inclusion of homology in the 5' end of the payload, with the same sequence as the sgRNA, suggests the possibility of using a single RNA as both guide and payload. However, attempting to include the necessary secondary structure and using the 5' end of the payload itself as the sgRNA for the Cas component proved unsuccessful, and we found it was necessary to co-transform two separate sgRNA and payload RNAs for successful targeted integration.

Off-target effects and application to complex organisms. Whole genome sequencing of cells subjected to the GENEWRITE expression exhibit larger numbers of high frequency mutations relative to a negative control, with mutations scattered throughout the genome (Fig. 4). Moreover, a large fraction of the plasmids purified and sequenced from GENEWRITE-exposed cells have curiously had the coding sequences of both GENEWRITE and NHEJ proteins excised from their host plasmids (Supplementary Fig. S4), suggesting that GENEWRITE may also be effective in excising coding regions of inappropriately-highly expressed genes.

Discussion

We have shown that a fusion of a Cas endonuclease and LINE-1 ORF2p reverse transcriptase, which we call GENEWRITE, is capable of integration of large genetic payloads in the *E. coli* genome through appropriate design of the homology regions of guide and payload RNAs. We also find that assistance from NHEJ DNA repair enzymes and LINE-1 ORF1p protein may help increase the efficiency and specificity of the insertion (results summarized in Supplementary Table S3). We have not yet tested the limits on size of GENEWRITE payloads, but LINE-1 itself is ~ 5 kbp long and hence similarly sized payloads may be accessible.

The above-described results were obtained using a simplistic method where each component is delivered separately: the RNAs through electroporation, and the GENEWRITE protein through constitutive, low-level expression from a plasmid. We find GENEWRITE to be remarkably successful given this simplistic approach, despite ORF2p's *cis*-preference for its encoding RNA⁵⁶ and propensity to produce inserts with 5' truncations^{65–68}. However, using this method, we find significant off-target effects, including an increase in the rate of off-target mutations relative to a control, and the excision of highly expressed DNA segments from the genome. We speculate that these off-target effects and the lethality of the GENEWRITE protein to *E. coli* may be coupled:





high affinity of ORF2p to arbitrary RNAs may force non-sgRNAs into the Cas component, serving as a guide for endonuclease activity and generating off-target DNA breaks. Hence, we suggest that rather than direct in vivo expression, deployment of the GENEWRITE system to more complex mammalian cells may be better accomplished through in vitro assembly and transfection of RNP particles as is frequently performed with traditional CRISPR-Cas editing. Furthermore, it remains to be seen if GENEWRITE insertions will produce truncated insertions when applied to more advanced systems.

Methods

Reagents. Primers used for PCR and RNA synthesis were synthesized by Integrated DNA Technologies (IDT, Coralville, IA). Kits used include QIAprep Spin Miniprep Kit (QIAGEN; Germantown, MD; Catalog Number 27106), QIAquick PCR Purification Kit (QIAGEN; Catalog Number 28106), DNeasy UltraClean Microbial Kit (QIAGEN; Catalog Number 12224-50), Megascript T7 Transcription Kit (ThermoFisher Scientific; Waltham, MA; Catalog Number AMB13345), TURBO DNase (ThermoFisher Scientific; Catalog Number AM2239), and NEBNext Ultra II Library Prep kit (New England Biosciences; Ipswich, MA; Catalog Number E7645S). PCR was performed with Phusion High-Fidelity PCR Master Mix with HF Buffer (NEB; M0531L).

Biological resources and media. *E. coli* BL21-AI (ThermoFisher Scientific; Catalog Number C607003, GenBank accession number CP047231) was used for all experiments. Overnight seed cultures were grown in Super Optimal Broth with Catabolite Repression [SOC; SOB+0.5% w/v glucose] medium with appropriate antibiotics. Electrocompetent cells were prepared by growth in Super Optimal Broth (SOB) with appropriate antibiotics.

Plasmid design and construction. All GENEWRITE proteins and variants were designed in Vector NTI software (Thermo Fisher Scientific) and synthesized de novo and cloned into pUC57-*kan* by GENEWIZ Gene synthesis (GENEWIZ); the exception is ORF2pZRT, which was cloned into pUC57-*amp* by GENEWIZ. A list of all constructs used in this study is found in Supplementary Table S1.

Bacillus subtilis NHEJ enzymes [Ku (encoded by the gene *ykoV*) and LigD (encoded by the gene *ykoU*)] were expressed from the anhydrotetracycline-inducible P_{LtetO1} promoter⁵⁴ on the plasmid pZA31⁴⁹. Cells not expressing NHEJ were transformed with empty pZA31 as a control.

sgRNA and payload RNA synthesis. DNAs encoding sgRNAs were prepared using primers including a T7 promoter driving a 20 bp guide sequence. The 3' end of this primer was designed with a 14 bp overhang homologous to a 77 bp scaffold oligo containing sequence encoding the necessary sgRNA secondary structure and used to prime amplification of the sgRNA-encoding DNA. Sequences of all oligos used in the study are available in Supplementary Table S2.

Payload RNAs were prepared using primers including a T7 promoter driving sequence encoding a strong, constitutive PlacIQ1 promoter, a Shine-Dalgarno ribosomal binding site, and 20 bp of sequence homologous to the spectinomycin resistance gene *aadA*. Reverse primers were designed with 20 bp homology to the 3' end of *aadA* and included indicated lengths of sequence homologous to the intended integration site. Payload RNA-encoding DNA was amplified from the plasmid pTKRED⁶⁹ using these primers.

RNAs were generated using the above DNA templates using T7 MEGAscript (Thermo Fisher Scientific), with incubation at 37 °C for 16 h. Samples were then digested with TURBO DNase (Thermo Fisher Scientific) for 1 h at 37 °C and purified by phenol–chloroform extraction and isopropanol precipitation.

Preparation of electrocompetent cells and transformation. Electrocompetent cells were prepared by preparation of a seed culture by overnight growth in SOC at 37 °C in a shaking water bath (New Brunswick C76). Seed cultures were diluted 10:1 in fresh SOB and grown at 37 °C in a shaking water bath until OD600 ~ 0.6, at which point 0.1% L-arabinose was added. Importantly, after L-arabinose was completely dissolved, cells were immediately harvested by centrifugation at 4 °C; extended induction of GENEWRITE with L-arabinose is lethal. This was followed by 3 × washing in ice-cold 10% v/v glycerol. 100 µl of cells thus prepared were mixed with an excess of payload RNA and sgRNA (5 µg and 10 µg, respectively); these quantities yielded success but have not been optimized. The mixture was electroporated (BIO-RAD Gene Pulser) using standard settings for *E. coli*. 1 ml of SOC + 1 mM IPTG and 100 ng/ml aTc were added, and cells were allowed to recover overnight. Transformants were spread on LB agar plates containing 100 µg/ml spectinomycin, 0.5% w/v glucose, 1 mM IPTG, and 100 ng/ml aTc were not be a 37 °C air incubator.

Genome sequencing. Genomic DNA was obtained from cultures prepared in 2 ml Lysogeny Broth (LB) by purification using the QIAGEN DNeasy UltraClean Microbial Kit. Resulting samples were submitted to the UCR Genomic Core at the Institute for Integrative Genomic Biology for processing and sequencing. Samples were sheared using a Covaris S220 Ultrasonicator and libraries prepared using an NEBNext Ultra II Library Prep kit. After preparation, libraries were analyzed using qPCR and an Agilent 2100 Bioanalyzer. The resulting libraries were pooled and sequenced using an Illumina MiSeq sequencer with 150 bp paired-end reads. Sequencing data were analyzed with Geneious Prime.

Data availability

All data and materials are available upon reasonable request to thomas.kuhlman@ucr.edu. Sequence data are available at the following accession numbers: Cas9-GENEWRITE: MZ493917; Cpf1-GENEWRITE: MZ493918; Whole genome sequencing data: BioProject PRJNA743056.

Received: 29 September 2021; Accepted: 29 November 2021 Published online: 08 December 2021

References

- Singh, D. et al. Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1 (Cas12a). Proc. Natl. Acad. Sci. 115, 5444–5449. https://doi.org/10.1073/pnas.1718686115 (2018).
- Singh, D., Sternberg, S. H., Fei, J., Doudna, J. A. & Ha, T. Real-time observation of DNA recognition and rejection by the RNAguided endonuclease Cas9. Nat. Commun. 7, 12778. https://doi.org/10.1038/ncomms12778 (2016).
- Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–821. https://doi.org/10.1126/science.1225829 (2012).
- 4. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157, 1262–1278. https://doi.org/10.1016/j.cell.2014.05.010 (2014).
- Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823. https://doi.org/10.1126/science. 1231143 (2013).
- 6. Mali, P. et al. RNA-guided human genome engineering via Cas9. Science 339, 823–826. https://doi.org/10.1126/science.1232033 (2013).
- DiCarlo, J. E. et al. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucl. Acids Res. 41, 4336–4343. https://doi.org/10.1093/nar/gkt135 (2013).
- Zhang, G. C. *et al.* Construction of a quadruple auxotrophic mutant of an industrial polyploid saccharomyces cerevisiae strain by using RNA-guided Cas9 nuclease. *Appl. Environ. Microbiol.* 80, 7694–7701. https://doi.org/10.1128/aem.02310-14 (2014).
- Liu, J. J. et al. Metabolic engineering of probiotic Saccharomyces boulardii. Appl. Environ. Microbiol. 82, 2280–2287. https://doi. org/10.1128/aem.00057-16 (2016).
- 10. Vyas, V. K., Barrasa, M. I. & Fink, G. R. A Candida albicans CRISPR system permits genetic engineering of essential genes and gene families. *Sci. Adv.* **1**, e1500248. https://doi.org/10.1126/sciadv.1500248 (2015).
- Ng, H. & Dean, N. Dramatic improvement of CRISPR/Cas9 editing in *Candida albicans* by increased single guide RNA expression. mSphere 2. https://doi.org/10.1128/mSphere.00385-16 (2017).
- 12. Hwang, W. Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. 31, 227-229. https://doi.org/10.1038/nbt.2501 (2013).
- Gratz, S. J. et al. Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics 194, 1029–1035. https:// doi.org/10.1534/genetics.113.152710 (2013).
- Bassett, A. R., Tibbit, C., Ponting, C. P. & Liu, J. L. Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. Cell Rep. 4, 220–228. https://doi.org/10.1016/j.celrep.2013.06.020 (2013).
- 15. Yan, H. *et al.* An engineered orco mutation produces aberrant social behavior and defective neural development in ants. *Cell* **170**, 736-747.e739. https://doi.org/10.1016/j.cell.2017.06.051 (2017).
- Trible, W. *et al.* orco Mutagenesis causes loss of antennal lobe Glomeruli and impaired social behavior in ants. *Cell* **170**, 727-735. e710. https://doi.org/10.1016/j.cell.2017.07.001 (2017).
- Kistler, K. E., Vosshall, L. B. & Matthews, B. J. Genome engineering with CRISPR-Cas9 in the mosquito Aedes aegypti. *Cell Rep.* 11, 51–60. https://doi.org/10.1016/j.celrep.2015.03.009 (2015).
- 18. Friedland, A. E. *et al.* Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nat. Methods* **10**, 741–743. https://doi.org/10.1038/nmeth.2532 (2013).
- Jiang, W. et al. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucl. Acids Res 41, e188. https://doi.org/10.1093/nar/gkt780 (2013).
- Wang, H. et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell 153, 910–918. https://doi.org/10.1016/j.cell.2013.04.025 (2013).

- Soni, D. et al. Deubiquitinase function of A20 maintains and repairs endothelial barrier after lung vascular injury. Cell Death Discov. 4, 60. https://doi.org/10.1038/s41420-018-0056-3 (2018).
- 22. Guo, X. & Li, X. J. Targeted genome editing in primate embryos. Cell Res. 25, 767-768. https://doi.org/10.1038/cr.2015.64 (2015).
- Baltimore, D. *et al.* Biotechnology. A prudent path forward for genomic engineering and germline gene modification. *Science* 348, 36–38. https://doi.org/10.1126/science.aab1028 (2015).
- Hirakawa, M. P., Krishnakumar, R., Timlin, J. A., Carney, J. P. & Butler, K. S. Gene editing and CRISPR in the clinic: Current and future perspectives. *Biosci. Rep.* 40, BSR20200127. https://doi.org/10.1042/BSR20200127 (2020).
- Xu, L. et al. CRISPR-edited stem cells in a patient with HIV and acute lymphocytic leukemia. N. Engl. J. Med. 381, 1240–1247. https://doi.org/10.1056/NEJMoa1817426 (2019).
- Guo, T. et al. Harnessing accurate non-homologous end joining for efficient precise deletion in CRISPR/Cas9-mediated genome editing. Genome Biol. 19, 170. https://doi.org/10.1186/s13059-018-1518-x (2018).
- Bothmer, A. *et al.* Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nat. Commun.* 8, 13905. https://doi.org/10.1038/ncomms13905 (2017).
- Brinkman, E. K. et al. Kinetics and fidelity of the repair of Cas9-induced double-strand DNA breaks. Mol. Cell 70, 801-813.e806. https://doi.org/10.1016/j.molcel.2018.04.016 (2018).
- van Overbeek, M. et al. DNA repair profiling reveals nonrandom outcomes at Cas9-mediated breaks. Mol. Cell 63, 633–646. https:// doi.org/10.1016/j.molcel.2016.06.037 (2016).
- Nambiar, T. S. *et al.* Stimulation of CRISPR-mediated homology-directed repair by an engineered RAD18 variant. *Nat. Commun.* 10, 3395. https://doi.org/10.1038/s41467-019-11105-z (2019).
- Aird, E. J., Lovendahl, K. N., St. Martin, A., Harris, R. S. & Gordon, W. R. Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. *Commun. Biol.* 1, 54. https://doi.org/10.1038/s42003-018-0054-2 (2018).
- Liu, M. et al. Methodologies for improving HDR efficiency. Front. Genet. 9, 691–691. https://doi.org/10.3389/fgene.2018.00691 (2019).
- Rozov, S. M., Permyakova, N. V. & Deineko, E. V. The problem of the low rates of CRISPR/Cas9-mediated knock-ins in plants: Approaches and solutions. *Int. J. Mol. Sci.* 20, 3371. https://doi.org/10.3390/ijms20133371 (2019).
- Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu. Rev. Biochem. 79, 181–211. https://doi.org/10.1146/annurev.biochem.052308.093131 (2010).
- Lieber, M. R., Ma, Y., Pannicke, U. & Schwarz, K. Mechanism and regulation of human non-homologous DNA end-joining. *Nat. Rev. Mol. Cell Biol.* 4, 712–720 (2003).
- Davis, A. J. & Chen, D. J. DNA double strand break repair via non-homologous end-joining. *Transl. Cancer Res.* 2, 130–143. https:// doi.org/10.3978/j.issn.2218-676X.2013.04.02 (2013).
- Iliakis, G. Backup pathways of NHEJ in cells of higher eukaryotes: cell cycle dependence. Radiother. Oncol. J. Eur. Soc. Therap. Radiol. 92, 310–315. https://doi.org/10.1016/j.radonc.2009.06.024 (2009).
- 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 Rep.* 51, 437–443. https://doi.org/10.5483/BMBRep.2018.51.9.187 (2018).
- Li, G. et al. Small molecules enhance CRISPR/Cas9-mediated homology-directed genome editing in primary cells. Sci. Rep. 7, 8943. https://doi.org/10.1038/s41598-017-09306-x (2017).
- Chu, V. T. et al. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat. Biotechnol. 33, 543–548. https://doi.org/10.1038/nbt.3198 (2015).
- Strecker, J. et al. RNA-guided DNA insertion with CRISPR-associated transposases. Science 365, 48–53. https://doi.org/10.1126/ science.aax9181 (2019).
- Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S. & Sternberg, S. H. Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. *Nature* 571, 219–225. https://doi.org/10.1038/s41586-019-1323-z (2019).
- Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature 576, 149–157. https://doi.org/10.1038/s41586-019-1711-4 (2019).
- Baranauskas, A. et al. Generation and characterization of new highly thermostable and processive M-MuLV reverse transcriptase variants. Protein Eng. Des. Sel. PEDS 25, 657–668. https://doi.org/10.1093/protein/gzs034 (2012).
- Gerard, G. F. et al. The role of template-primer in protection of reverse transcriptase from thermal inactivation. Nucl. Acids Res. 30, 3118–3129. https://doi.org/10.1093/nar/gkf417 (2002).
- Arezi, B. & Hogrefe, H. Novel mutations in Moloney Murine Leukemia Virus reverse transcriptase increase thermostability through tighter binding to template-primer. *Nucl. Acids Res.* 37, 473–481. https://doi.org/10.1093/nar/gkn952 (2009).
- Kotewicz, M. L., Sampson, C. M., D'Alessio, J. M. & Gerard, G. F. Isolation of cloned Moloney murine leukemia virus reverse transcriptase lacking ribonuclease H activity. *Nucl. Acids Res.* 16, 265–277. https://doi.org/10.1093/nar/16.1.265 (1988).
- Cost, G. J., Feng, Q., Jacquier, A. & Boeke, J. D. Human L1 element target-primed reverse transcription in vitro. *EMBO J.* 21, 5899–5910. https://doi.org/10.1093/emboj/cdf592 (2002).
- Lee, G. et al. Testing the retroelement invasion hypothesis for the emergence of the ancestral eukaryotic cell. Proc. Natl. Acad. Sci. https://doi.org/10.1073/pnas.1807709115 (2018).
- Martin, S. L. The ORF1 protein encoded by LINE-1: Structure and function during L1 retrotransposition. J. Biomed. Biotechnol. 2006, 45621. https://doi.org/10.1155/JBB/2006/45621 (2006).
- Lambowitz, A. M. & Belfort, M. Mobile bacterial group II introns at the crux of eukaryotic evolution. *Microbiol. Spectr.* 3, MDNA3-0050-2014 (2015).
- Doucet, A. J., Wilusz, J. E., Miyoshi, T., Liu, Y. & Moran, J. V. A 3' poly(A) tract is required for LINE-1 retrotransposition. *Mol. Cell* 60, 728–741. https://doi.org/10.1016/j.molcel.2015.10.012 (2015).
- Ade, C. M. et al. Evaluating different DNA binding domains to modulate L1 ORF2p-driven site-specific retrotransposition events in human cells. Gene 642, 188–198. https://doi.org/10.1016/j.gene.2017.11.033 (2018).
- 54. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucl. Acids Res.* 25, 1203–1210 (1997).
- Piskareva, O., Ernst, C., Higgins, N. & Schmatchenko, V. The carboxy-terminal segment of the human LINE-1 ORF2 protein is involved in RNA binding(). FEBS Open Bio 3, 433–437. https://doi.org/10.1016/j.fob.2013.09.005 (2013).
- Wei, W. et al. Human L1 retrotransposition: cis preference versus trans complementation. Mol. Cell. Biol. 21, 1429–1439. https:// doi.org/10.1128/mcb.21.4.1429-1439.2001 (2001).
- Calos, M. P. & Miller, J. H. The DNA sequence change resulting from the IQ1 mutation, which greatly increases promoter strength. *Mol. Gen. Genet.* 183, 559–560 (1981).
- Konkel, M. K. *et al.* Sequence analysis and characterization of active human alu subfamilies based on the 1000 genomes pilot project. *Genome Biol. Evol.* 7, 2608–2622. https://doi.org/10.1093/gbe/evv167 (2015).
- Dreyfus, M. & Régnier, P. The Poly(A) Tail of mRNAs: Bodyguard in Eukaryotes, Scavenger in Bacteria. Cell 111, 611–613. https:// doi.org/10.1016/S0092-8674(02)01137-6 (2002).
- Sarkar, N. Polyadenylation of mRNA in bacteria. *Microbiology (Reading)* 142(Pt 11), 3125–3133. https://doi.org/10.1099/13500 872-142-11-3125 (1996).

- Hajnsdorf, E. & Kaberdin, V. R. RNA polyadenylation and its consequences in prokaryotes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 373. https://doi.org/10.1098/rstb.2018.0166 (2018).
- Mohanty, B. K. & Kushner, S. R. Bacterial/archaeal/organellar polyadenylation. Wiley Interdiscip. Rev. RNA 2, 256–276. https:// doi.org/10.1002/wrna.51 (2011).
- Kuhlman, T. E. & Cox, E. C. Gene location and DNA density determine transcription factor distributions in *Escherichia coli. Mol. Syst. Biol.* 8, 610. https://doi.org/10.1038/msb.2012.42msb201242 (2012).
- Tas, H., Nguyen, C. T., Patel, R., Kim, N. H. & Kuhlman, T. E. An integrated system for precise genome modification in *Escherichia coli. PLoS ONE* 10, e0136963. https://doi.org/10.1371/journal.pone.0136963 (2015).
- Zingler, N. et al. Analysis of 5' junctions of human LINE-1 and Alu retrotransposons suggests an alternative model for 5'-end attachment requiring microhomology-mediated end-joining. *Genome Res.* 15, 780–789. https://doi.org/10.1101/gr.3421505 (2005).
- Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. https://doi.org/10.1038/35057062 (2001).
- Babushok, D. V., Ostertag, E. M., Courtney, C. E., Choi, J. M. & Kazazian, H. H. L1 integration in a transgenic mouse model. Genome Res. 16, 240–250. https://doi.org/10.1101/gr.4571606 (2006).
- Chen, J.-M., Férec, C. & Cooper, D. N. Mechanism of Alu integration into the human genome. *Genom. Med.* 1, 9–17. https://doi. org/10.1007/s11568-007-9002-9 (2007).
- Kuhlman, T. E. & Cox, E. C. Site-specific chromosomal integration of large synthetic constructs. Nucl. Acids Res. 38, e92. https:// doi.org/10.1093/nar/gkp1193 (2010).

Author contributions

Conceptualization, T.E.K.; methodology, T.E.K., investigation, F.M, L.W.T, K.S.M.W, and T.E.K.; data curation, F.M. and T.E.K.; writing—original draft preparation, T.E.K.; writing—review and editing, F.M., L.W.T., K.S.M.W, and T.E.K.; supervision, T.E.K. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by startup funds from the University of California and the University of California Riverside Office of Research and Economic Development.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/ 10.1038/s41598-021-03130-0.

Correspondence and requests for materials should be addressed to T.E.K.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021