Engineering CRISPR Systems for Synthetic Biology

by

Kylie S Standage-Beier

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved March 2023 by the
Graduate Supervisory Committee:

Xiao Wang, Chair
David A. Brafman
Xiao-jun Tian
David R. Nielsen

ARIZONA STATE UNIVERSITY

May 2023

ABSTRACT

Clustered regularly interspace short palindromic repeats (CRISPR) and CRISPR associated (Cas) technologies have become integral to genome editing. Canonical CRISPR-Cas9 systems function as a ribonucleic acid (RNA)-guided nucleases. Single guide RNAs (sgRNA) can be easily designed to target Cas9's nuclease activity towards protospacer deoxyribonucleic acid (DNA) sequences. The relatively ease and efficiency of CRISPR-Cas9 systems has enabled numerous technologies and DNA manipulations. Genome engineering in human cell lines is centered around the study of genetic contribution to disease phenotypes. However, canonical CRISPR-Cas9 systems are largely reliant on double stranded DNA breaks (DSBs). DSBs can induce unintended genomic changes including deletions and complex rearrangements. Likewise, DSBs can induce apoptosis and cell cycle arrest confounding applications of Cas9-based systems for disease modeling. Base editors are a novel class of nicking Cas9 engineered with a cytidine or adenosine deaminase. Base editors can install single letter DNA edits without DSBs. However, detecting single letter DNA edits is cumbersome, requiring onerous DNA isolation and sequencing, hampering experimental throughput. This document describes the creation of a fluorescent reporter system to detect Cytosine-to-Thymine (C-to-T) base editing. The fluorescent reporter utilizes an engineered blue fluorescent protein (BFP) that is converted to green fluorescent protein (GFP) upon targeted C-to-T conversion. The BFP-to-GFP conversion enables the creation of a strategy to isolate edited cell populations, termed Transient Reporter for Editing Enrichment (TREE). TREE increases the ease of optimizing base editor designs and assists in editing cell types recalcitrant to DNA editing. More recently, Prime editing has been demonstrated to introduce user defined DNA edits without the need for DSBs and donor DNA. Prime editing requires specialized prime editing guide RNAs (pegRNAs). pegRNAs are however difficult to manually design. This document describes the creation of a software tool: Prime Induced Nucleotide Engineering Creator of New Edits (PINE-CONE). PINE-CONE rapidly designs pegRNAs based off basic edit information and will assist with synthetic biology and biomedical research.

DEDICATION

This is document is dedicated to all the friends and family that helped along the way.

It has been a long journey.

ACKNOWLEDGMENTS

This work was a highly collaborative project and would not have been possible without contributions from many colleagues. Nicholas Brookhouser, Stefan J. Tekel, Grace Schwarz, Toan Nguyen all provided unique contributions to the project and helped make TREE possible. Joshua Cutts and Wylie Standage-Beier provided useful comments regarding next generation sequencing analysis. Xiao Wang Lab Members Samat Bayakhmetov, Xingwen Chen and Qi Zhang for provided constructive comments in preparing this document.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The ability to edit DNA is important to understand genetic contribution to disease and engineer biological systems. “Engineering CRISPR Systems for Synthetic Biology” aims to describe our contribution to Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR Associated (Cas) based genome engineering methods. In this regard, Chapter 2 reviews genome engineering technologies. Chapter 3 describe our creation of a technology to detect and isolate cells with DNA edits. Chapter 4 describes our creation of a software tool to assist in genome editing experiments. Collectively the findings and tools in this dissertation will improve our ability to engineer the genome and learn about genetic contributions to disease.

Chapter 2, Genome Reprogramming for Synthetic Biology, introduces a diverse array of genome engineering technologies. The review includes a basic introduction to CRISPR and CRISPR Associated (Cas) technologies. Likewise, the chapter presents a broad range of other genome engineering strategies, such as engineering through evolution, via DNA recombinase proteins and the construction of synthetic genomes. The review provides insight into some of the tools available for genome engineering in a range of organisms.

The ability to make edits to the genome is crucial for the studying genetic contribution to disease. CRISPR-Cas9 systems are highly efficient and versatile in their ability to direct DNA breaks and subsequently DNA-edits. However, canonical CRISPR-Cas9 systems’ reliance on double stranded DNA breaks (DSBs) can lead to spurious mutations ranging from small insertion deletions to destabilized karyotype and large-scale chromosomal abnormalities(Kosicki et al., 2018). CRISPR-based technologies that avoid DSBs would be useful for studying genetic contribution of disease. To this end, Base editors are a class of engineered CRISPR-Cas9 systems capable of introducing C:G→T:A and A:T→G:C base pair mutations. Base editors can introduce disease relevant single nucleotide polymorphisms (SNPs) without the need for DSBs and be used to study genetic contribution to disease. Base editors accomplish this via fusion of a

Cas9 mutant that introduces single stranded DNA breaks (SSBs, $Cas9^{D10A}$) with a cytidine or adenosine deaminase(Gaudelli et al., 2017; Komor et al., 2016).

However single letter DNA editing can be difficult to detect and often require a specialized and lengthily experimental workflow to identify. To address this, chapter 3 describes our creation of Transient Reporter for Editing. Enrichment (TREE). TREE reporters provide a fluorescent signal that indicates if base editing has occurred inside a cell. We created a novel blue fluorescent protein (BFP) that converts to green fluorescence upon C:G→T:A base editing. Chapter 3 describes how the creation of a novel fluorescent reporter assists in genome editing.

Base editors are effective at introducing single letter edits within a narrow 'editing window'. However more recently, 'Prime editors' have been demonstrated to introduce a broad range of user defined edits on the human genome. Prime editors are composed of an alternative nicking Cas9 ($Cas9^{H840A}$) fused to a MMLV Reverse transcriptase (RT)(Anzalone et al., 2019). Prime editors are versatile but require specialized design of CRISPR-RNAs. In Chapter 4 we describe the creation a software tool for automated design of prime editing RNAs that we call Prime Induced Nucleotide Editing Creator of News Edits (PINE-CONE). This Python-based pipeline takes basic edit information, such as DNA letters and positions, and generates prime editing guides and strategies. We envision this software assisting in design and implementation of Prime editing experiments to understand genetic contribution to disease.

REFERENCES

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, *576*(7785), 149–157. https://doi.org/10.1038/s41586-019-1711-4

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*, *551*(7681), 464–471. https://doi.org/10.1038/nature24644

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, *533*(7603), 420–424. https://doi.org/10.1038/nature17946

Kosicki, M., Tomberg, K., & Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nature Biotechnology*, *36*(8), 765–771. https://doi.org/10.1038/nbt.4192

CHAPTER 2

GENOME REPROGRAMMING FOR SYNTHETIC BIOLOGY

ABSTRACT

The ability to go from a digitized DNA sequence to a predictable biological function is central to synthetic biology. Genome engineering tools facilitate rewriting and implementation of engineered DNA sequences. Recent development of new programmable tools to reengineer genomes has spurred myriad advances in synthetic biology. Tools such as clustered regularly interspace short palindromic repeats (CRISPR) enable RNA-guided rational redesign of organisms and implementation of synthetic gene systems. New directed evolution methods generate organisms with radically restructured genomes. These restructured organisms have useful new phenotypes for biotechnology, such as bacteriophage resistance and increased genetic stability. Advanced DNA synthesis and assembly methods have also enabled the construction of fully synthetic organisms, such a J. Craig Venter Institute (JCVI)-syn 3.0. Here we summarize the recent advances in programmable genome engineering tools.

INTRODUCTION

Synthetic biology seeks to develop new organisms through forward genetic engineering. We can develop tools to study complex gene regulatory networks *in silico*(Faucon et al., 2014), however, forward engineering of genetic systems enables us to identify and understand emergent and unexpected phenomena in biology(L.-Z. Wang et al., 2016; F. Wu et al., 2014). The ability to manipulate DNA is intrinsically linked to our ability to experimentally study and forward engineer regulatory gene networks. Genome engineering tools have allowed us to reprogram life to explore basic science and to engineer novel organisms for biotechnology. The field has progressed from basic molecular cloning to programmable methods for remodeling and constructing new organisms.

Implementations of stable synthetic gene circuits and reengineering of biosynthetic pathways requires reengineering of an organism(Brophy & Voigt, 2014). Even the scenario of episomal expression of synthetic gene constructs often also requires strains modified from wild-type counterparts. For instance, implementations of the genetic toggle switch requires removal of endogenous lacI repressor via genome editing(Gardner et al., 2000; Litcofsky et al., 2012). Library-based investigation of gene network engineering requires efficient genome integration methods(Ellis et al., 2009; M. Wu et al., 2013). A convergence of programmable editing, new directed evolution methods, rational protein engineering, and DNA synthesis have propelled synthetic biology forward. As the field of synthetic biology moves forward, so will the enabling technologies. Genome engineering will require increased specificity to move to therapeutic applications. Evolutionary methods will need to enable large-scale rewriting of organisms to find non-trivial solutions to challenging problems. Proteins that efficiently target DNA recombination will enable large-scale restructuring of organisms. DNA synthesis and assembly methods will enable production of large-synthetic constructs, including whole genomes.

In this review, we highlight newly developed technologies enabling the rational redesign of organisms. CRISPR derived technologies have revolutionized our ability to target DNA manipulation *in vivo*. We discuss the state-of-the-art CRISPR based methods for rewriting and implementing synthetic transgenes. We describe methods to conduct large-scale rewriting of

genomes, ranging from strategies to target genome reduction, to methods powered by genetic randomization and evolution. Furthermore, we discuss the development of methods to generate *de novo* organisms, such as the recently developed minimal synthetic *Mycoplasma mycoides* genome, JCVI-syn3.0(Hutchison et al., 2016).

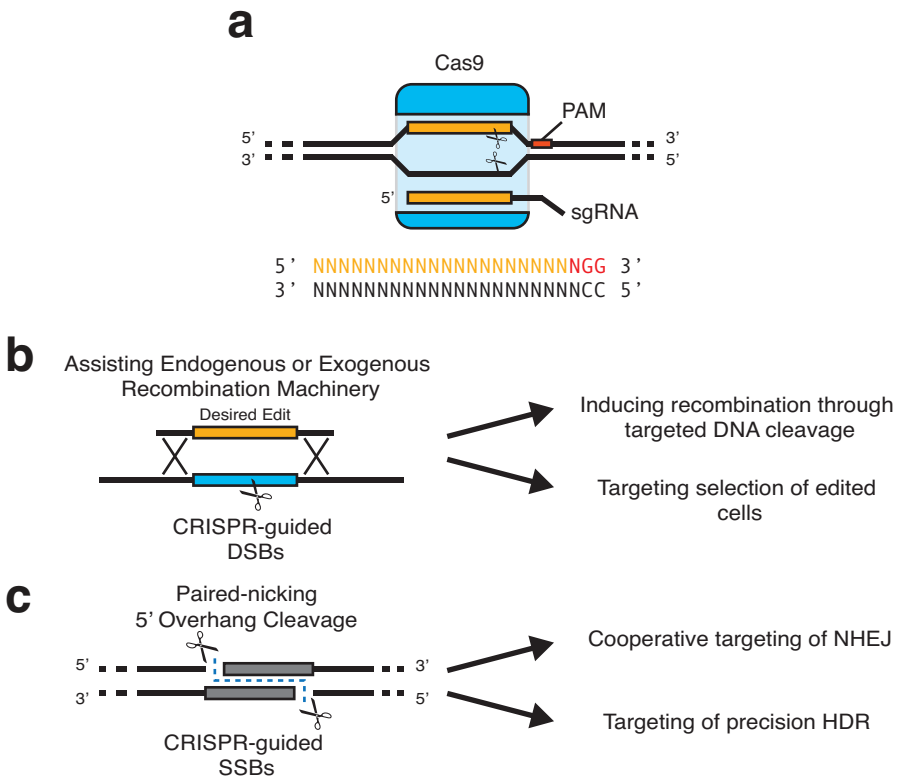


Figure 2-1**:** Programmable Editing of Genomes (**a**) A schematic of CRISPR-directed targeting with wildtype *S. pyogenes* Cas9. Cas9 (blue) is targeted to a DNA sequence based on the presence of a protospacer adjacent motif (PAM, red) nucleotides matching those in a short guide RNA (sgRNA). The target DNA sequence is written beneath. The strand matching the 20 nucleotide guide of the sgRNA is orange and the complementary strand is black. (**b**) CRISPR-guided double stranded DNA breaks (DSBs) involving a recombination template have various modalities. DSBs either induce host homology directed repair (HDR) or DSBs kill cells that have not acquired the desired edit, wherein the full CRISPR target site is not present in the desired edit. HDR-mediated editing can be either a function of one or both modalities. (**c**) Novel mutant versions of Cas9 that mediate single-stranded DNA cleavage have been developed to target recombination in a broad range of organisms.

RNA-Programmable Genome Engineering

CRISPR and CRISPR-associated (Cas) systems function as a prokaryotic and archaeal immune system(Brouns et al., 2008; Marraffini, 2015; Marraffini & Sontheimer, 2008; Mojica et al., 2009) CRISPR loci express a long non-coding RNA, which is subsequently processed by Cas proteins (eg. Cas9 of type II CRISPR systems)(Makarova et al., 2011) to form mature targeting CRISPR RNAs (crRNAs). These crRNAs target endonuclease activity of Cas9 (or other Cas proteins) to target DNAs. Watson-Crick base pairing between crRNA and target DNA combined with the presence of a protospacer adjacent motif (PAM) sequence on the target results in Cas9 catalyzed DNA cleavage (Fig. 2-1A) (Jinek et al., 2012). Researchers quickly saw the potential of CRISPR systems, in particular, those involving Cas9, as other systems require formation of large multiprotein complexes (eg. those of type I and III CRISPR systems)(Makarova et al., 2011).

Cong and colleagues along with Mali and coworkers co-published initial reports demonstrating the application of engineered CRISPR systems in human cells(Cong et al., 2013; Mali et al., 2013). They demonstrated that CRISPR RNAs can be engineered to target Cas9 nuclease activity to endogenous target sites. These reports spurred the development of CRISPR based technologies.

Initial reports suggested Cas9 activity was highly specific to target sites, requiring nearly 20-out-of-20 nucleotides matching between crRNA and target DNA, however a subsequent report showed CRISPR can readily induce off-target mutations(Fu et al., 2013). To circumvent, this numerous strategies have been developed to increase Cas9 specificity. Ran and colleagues developed a paired-nickase system for targeting non-homologous end joining (NHEJ), homology directed repair (HR) and non-HR mediated integration(Ran et al., 2013). Individual single-stranded DNA breaks (nicks) to the chromosome are repaired without mutagenesis. However, paired CRISPR-targeted nicks in close proximity and in a 5’ overhang orientation result in efficient mutagenesis (Fig. 2-1B and C). Alternatively, Tsai et al. and Guilinger et al. concurrently utilized protein engineering to increase specificity. They showed fusion of catalytically inactive Cas9 fused to a FokI endonuclease domain dramatically increase DNA cleavage specificity(Guilinger et al., 2014; Tsai et al., 2014). This system enables cooperative genome targeting, wherein double-

stranded DNA cleavage requires dimerization of FokI domains. This increases the specificity of genome editing over 140 fold over wild-type *Streptococcus pyogenes* Cas9. This approach is simultaneously versatile and highly specific. Fu and colleagues found an interesting alternative to increase accuracy of CRISPR editing methods: decreasing the length of short guide RNAs (sgRNAs) from 20 to 18 nucleotides increases targeting specificity(Fu et al., 2014). Furthermore truncating sgRNAs to 14 nucleotides enables targeted DNA binding of Cas9 while avoiding DNA cleavage(Kiani et al., 2015). This enables the development of multi-targeted Cas9 editing and regulatory fusions. These methods will likely be greatly useful with therapeutic application allowing simultaneous function of Cas9 in genetic circuitry and gene knockout(Kiani et al., 2014, 2015). Slaymaker and coworkers developed an enhanced specificity Cas9 (eSpCas9) through rational protein reengineering(Slaymaker et al., 2015). Mutation of various positively charged amino acids in the non-target DNA binding groove of *S. pyogenes* Cas9 confers higher specificity of Cas9 mediated cleavage. Similarly, Kleinstiver and colleagues described the engineering of a high fidelity version of Cas9 (spCas9-HF) through rational protein reengineering(Kleinstiver et al., 2016). Likewise, Kleinstiver and coworkers demonstrated in a separate piece the directed evolution of *S. pyogenes* and *Staphylococcus aureus* Cas9 to generate novel variants with altered PAM requirements(Kleinstiver, Prew, Tsai, Nguyen, et al., 2015; Kleinstiver, Prew, Tsai, Topkar, et al., 2015) This may be advantageous if alternative (non-NGG) PAMs are desired. Likewise, increasing the size of the requisite PAM region may be useful in the creation of therapeutically relevant highly specific Cas9s, which could ultimately be more likely to make it into therapeutic and disease modeling use.

Synthetic biology seeks to forward engineer novel cellular behaviors and phenotypes. This can be accomplished through both rewiring of endogenous gene networks or through integration of synthetic DNAs. He and coworkers demonstrate potential utility for CRISPR targeted genome integration in human cell lines. They show that double stranded DNA break (DSB) induced NHEJ can target chromosomal integration of fluorescent markers[29]. The system described by He and colleagues is capable of integrating 4.6 Kb of DNA with relative high efficiency (20%). This is accomplished by simultaneous cleavage of a genomic target along with

cleavage of a transfected donor plasmid. NHEJ results in incorporation of the synthetic reporter at the location of sgRNA targeted cleavage(He et al., 2016). This technique, along with other editing modalities will be useful in the implementation of genetic circuitry or differentiation state reporters.

After initial reports of CRISPR editing in human cells, numerous reports came out employing engineered CRISPR systems in other organisms. Jiang and colleagues demonstrated CRISPR-Cas9 systems can direct recombination between the genome of *Streptococcus pneumoniae* and exogenous editing templates. This enabled selection marker free editing of multiple genomic targets. Likewise, they showed CRISPR-Cas9 can assist the lambda Red recombineering system for *Escherichia coli* by selecting for desired edits of the genome(Jiang et al., 2013). More recently it was demonstrated that CRISPR-Cas9 systems could augment potential sizes of genome integrations. Building off work that demonstrated use of SceI meganuclease can work cooperatively with the lambda Red recombination machinery to integrate large, 7 Kb, synthetic constructs(Kuhlman & Cox, 2010), Bassalo and coworkers demonstrated cooperative use of lambda Red and CRISPR target DNA cleavage. They show CRISPR increases DNA editing efficiency up to 95%, wherein 50 out of 50 clones contain the correct integration. Furthermore they demonstrate integration of a 10 Kb isobutanol biosynthetic pathway, using CRISPR combined with lambda Red(Bassalo et al., 2016). Rapid implementation of full biosynthetic pathways, such as that for isobutanol described in Bassalo et al., was not possible through basic lambda Red mediated recombination. However, incorporation of CRISPR targeted DNA cleavage with lambda Red can target integration with ease and efficiency. The ability for single step integration of large-synthetic constructs is necessary for synthetic biologist to create complex new cellular functions. In *E. coli*, multiple works have demonstrated CRISPR working cooperatively with the lambda Red homologous recombination machinery, however, we took a different approach and developed a system to target endogenous *E. coli* homologous recombination. We demonstrated that nicking Cas9 mutant ($Cas9^{D10A}$) can be easily guided to genomic loci and, when dual-targeted, can efficiently direct large-scale recombination across the bacterial genome(Standage-Beier et al., 2015).

CRISPR-guided genome engineering has become an indispensable tool for non-model bacteria. For instance, Wang et al. and Li et al. demonstrate CRISPR as a tractable genome-engineering tool for *Clostridium beijerinckii* (Li et al., 2016; Y. Wang et al., 2015). *C. beijerinckii* is an industrially useful organism for the production of acetone, butanol, and ethanol that previously lacked easily programmable methods for genome engineering(Liao et al., 2015). Wang et al. show targeted gene deletion in C. *beijerinckii* using plasmid delivered 1 Kb homology sequences. Because homologous recombination rates are low in *C. beijerinckii*, the use of Cas9 enables high efficiency selection of edited clones(Y. Wang et al., 2015). Similarly, Li and colleagues show Cas9 nickase can target gene deletion via HR. They generate deletions ranging from 20 to 1149 bp in multiple clostridium species(Li et al., 2016). Mougiakos and colleagues provide an extensive review focusing on CRISPR's development from a bacterial immune system to a prokaryotic genome engineering technology(Mougiakos et al., 2016). Likewise, Choi and Lee provide a comprehensive description of the published methods using CRISPR systems for bacterial genome engineering(Choi & Lee, n.d.). Collectively CRISPR has functioned so reliably as a genome engineering platform in bacteria, it will expedite combinatoric reverse genetics studies and forward engineering of new organisms for synthetic biology(Jiang & Marraffini, 2015).

CRISPR-guided genome engineering's influence extends beyond bacteria into fungi also. *Saccharomyces cerevisiae* (yeast) is a model fungus with tremendous biotechnological potential. Likewise, the eukaryotic cell physiology of yeast has allowed it to serve as a 'proxy-organism' for the development of genome engineering tools(Doyon et al., 2008). It was not long after initial development of CRISPR systems that Dicarlo and coworkers demonstrated CRISPR-targeted double stranded DNA breaks can be used for simultaneous induction and selection of genome edits via homologous recombination(DiCarlo et al., 2013). Bao et al. demonstrated up to 3 simultaneous edits at a time in *S. cerevisiae*(Bao et al., 2015). Subsequent works have demonstrated multi-pathway assembly employing CRISPR systems with up to 6 exogenous DNA sequences combined simultaneously or conversely removal of large genetic fragments(Hao et al., 2016). Work from Jakocĭun̄as and colleagues demonstrated the power of CRISPR genome

engineering strategies in yeast, enabling the assembly of 15 parts simultaneously(Jakočiūnas et al., 2015). They demonstrate assembly of a multipart carotenoid pathway generating *S. cerevisiae* capable of producing red pigment. The authors also demonstrate engineering tyrosine production through simultaneous pathway assembly and deletion of competing metabolic processes(Jakočiūnas et al., 2015). Tsarmopopoulos and coworkers demonstrated in an interesting study that exogenous bacterial genomes can be edited inside *S. cerevisiae*(Tsarmpopoulos et al., 2016). Similarly Kannan et al. demonstrated CRISPR-guided editing in yeast combined with genome transplantation can be employed to study 16S rRNA structures in *M. mycoides*(Kannan et al., 2016). The methods described offer an efficient way to reengineer the minimal genomes and organisms recalcitrant to manipulation for basic science and forward genetic engineering. CRISPR systems have become an indispensable molecular instrument for the combinatoric rewriting and construction of new genetic systems.

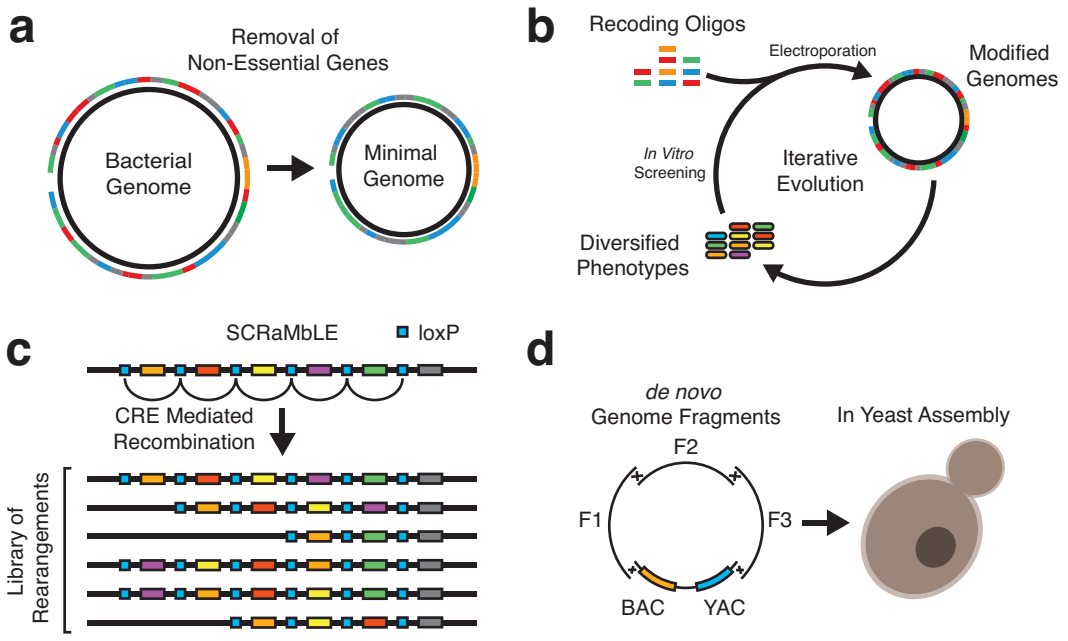


Figure 2-2: Large-scale Reengineering of organisms (**a**) Genome reduction methods, such as methods employing CRE recombinase, lambda Red recombineering, and CRISPR-nickases, have enabled large-scale reductions to the *E. coli* genome. Genome reduction methods look to investigate the emergent phenotypes by removal of large numbers of non-essential genes. These methods may identify novel organisms and phenotypes for synthetic biology. (**b**) Multiplex

automated genome engineering (MAGE) offers itself as a powerful tool for coupling DNA synthesis, targeted editing, and evolution. MAGE functions as an iterative process. Recoding oligo nucleotides are electroporated into *E. coli*, which are then screened for a desired phenotype. This process is repeated to maximize output from a biosynthetic pathway or to systematically replace DNA sequences (Adapted from(H. H. Wang et al., 2009)). (**c**) Synthetic chromosome rearrangement and modification by loxP-mediated evolution (SCRaMbLE) is a promising tool for investigated evolution and combinatorial genetics. loxP sites (blue squares) are placed around genes (various color rectangles), induction of Cre recombinases leads to recombination between loxP sites resulting in deletions, inversions, duplications, and translocations. Resulting clones from this method can be screened for desired phenotypes. (d) Forward genome construction methods such as yeast assembly enabled construction of large-subgenomic fragments. The efficiency of yeast homologous recombination enables connection of multiple fragments. Homolgoue fragments are connected via yeast HDR to a bacterial artificial chromosome (BAC, orange) and yeast artificial chromosome (YAC, blue) sequence. These circular fragments can measure up to 1 megabase and be propagated in *S. cerevisiae*. These assemblies can be transferred to recipient organisms via various methods.

Engineering Through Evolution

Evolution is the fundamental force that has driven the development of all life. Engineering through evolution has tremendous potential to enable researchers to identify non-intuitive and non-trivial solutions to biological problems(Pál et al., 2014). For instance, biosynthetic pathways and genetic circuitry may require evolutionary optimization to reach a desired function(Yokobayashi et al., 2002). Wang and colleagues demonstrated a method that can employ multiplex automated genome engineering (MAGE) to rewrite the *E. coli* genome(H. H. Wang et al., 2009). Using lambda Red recombineering machinery, this system integrated recoding oligonucleotides into the *E. coli* genome(Lajoie et al., 2012; Mosberg et al., 2010). The MAGE system facilitates efficient diversification and rewriting of the genome. This generates populations of cells with diversified phenotypes, which can be leveraged to identify and select organisms with desired traits (Fig.2-2B). For instance, in the initial implementations of MAGE the authors developed an *E. coli* strain capable of better producing lycopene, a commercially useful pigment. Following the first development of MAGE, the same group reported removal of all UAG stop codons in *E. coli* via combining MAGE and bacterial conjugation in a method called Conjugative Assembly Genome Engineering (CAGE)(Isaacs et al., 2011). Removal of all endogenous UAG stop codons renders *E. coli* resistant to various bacteriophages and frees the codon for researchers to study in vivo incorporation of new amino acids. Generating novel codon variants will be useful for engineering proteins with synthetic amino acids. Recoded organisms

are resistant to bacteriophages and are genetically orthogonal to their natural counterparts, making sharing traits by horizontal gene transfer highly unlikely(Lajoie et al., 2013). This suggests great biotechnological and *ex vitro* potential for recoded organisms.

More recently Farzadfard and Lu developed a novel genome-rewriting platform for *E. coli* called synthetic cellular recorders integrating biological events (SCRIBE)(Farzadfard & Lu, 2014). The SCRIBE system utilizes a reverse transcriptase along with a retron template RNA cassette to generate single stranded DNAs and the lambda Beta gene to facilitate DNA incorporation in lagging strand synthesis. This enables targeted and chemically controllable bacterial genome rewriting. The system was originally applied to generate *in vivo* analog memory in bacterial populations. The population of bacterial cells function as “recorders” where genome rewriting is linearly proportional to time of retron induction. It is foreseeable that this system becomes useful in the detection of specific compounds or pathogens and perhaps most interestingly as a tool for inducible genome editing and evolution of organisms. More recently, Perli and colleagues demonstrated human cell genetic recording(Perli et al., 2016). Self-targeting sgRNAs (stgRNAs) form indel mutations in response to environmental stimuli. This enables detection and quantification of inflammation response to lipopolysccharide (LPS). This suggests potential application of mammalian SCRIBE as a biological recorder and for the investigation of DNA sequence evolution. Fundamentally, SCRIBE systems demonstrate how genome engineering technologies can be utilized in synthetic biology as ‘recorders’ and analog memory units in gene circuitry.

Evolution of organisms also employs large-scale genetic rearrangements and genome minimization if advantageous. Work by Richard Lenski and colleagues on *E. coli* has demonstrated reductive genome evolution over years in laboratory culture(Barrick et al., 2009; Cooper et al., 2001; Elena & Lenski, 2003). Tools to target this have been developed to enable *de novo* generation of bacteria and other microorganisms with large-scale changes from their progenitors. Genome reduction strategies may be advantageous to synthetic biology by removing non-essential genetic and metabolic burden to cells (Fig. 2-2A). Lambda Red recombineering tools can be employed to target genome reduction. Posfai and colleagues have demonstrated

that 15% of the *E. coli* genome can be removed(Kolisnychenko et al., 2002). This leads to emergent phenotypes such as increased transformation efficiency and increased genome stability(Pósfai et al., 2006). This, like the minimal mycoplasma genome, will likely become a valuable tool for the generation of organisms with stable genetic content (i.e. lacking transposable elements, and perhaps reduced mutation rates) devoted to production of various bio-compounds. Alternatively, our group has demonstrated that CRISPR-guided nicking can target endogenous homolgous recombination. This enabled removal of 133 Kb, 3%, of the *E. coli* genome via a single plasmid transformation(Standage-Beier et al., 2015). Collectively these methods provide research and development tools for the creation of novel organisms. Concurrently, these tools are useful for the investigation of systems level reengineering of organisms(Csörgő et al., 2016).

Recombinase Based Engineering

Recombinases have functioned as an indispensible tool for efficient and precise genetic manipulation in a broad range of organisms. Bacterial suicide vectors often employ bacteriophage-derived recombinases to facilitate efficient site-specific integration(St-Pierre et al., 2013). These systems enable genome integration of large-synthetic constructs and can be easily designed into experimental workflows for restructuring organisms. Santos and colleagues demonstrated recombinase assisted genome engineering can generate *E. coli* capable of alginate metabolism and ethanol production with higher titers than typical plasmid based expression experimental regimes(Santos et al., 2013; Santos & Yoshikuni, 2014). Enyeart et al. demonstrated that targetron technology and Cre-lox recombinase systems can be used synergistically to restructure bacterial genomes(Enyeart et al., 2013). This system enables large-scale deletion (up to 120 Kb), targeted inversion (1.2 Mb), and translocation of targeted loci. Using this system they demonstrated a programmable and efficient way to remodel the genomes of *E. coli*, *S. aureus*, *Bacillus subtilis,* and *Shewanella oneidensis*(Enyeart et al., 2013). Krishnakumar and colleagues at the Craig Venter Institute developed a technology for large-scale bacterial genome restructuring. This system utilizes Cre-lox sites located on a donor vector and

the genome. This allows targeted replacement of large genomic fragments with synthetic fragments(Krishnakumar et al., 2014).

Recombinases methods extend beyond application in bacteria into mammalian cell lines and full organisms. Recombinases based methods can work cooperatively with evolutionary genome engineering methods(Dymond et al., 2011). They have enabled large-scale genome restructuring in *S. cerevisiae*. Dymond and colleagues developed synthetic chromosome rearrangement and modification by loxP-mediated evolution (SCRaMbLE)(Dymond et al., 2011). The authors systematically placed loxP sites in the 3' UTRs of genes on of the right arm of synthetic yeast chromsome IX (synIXR) and the left arm of semi-synthetic chromsome VI (semi-synVIL). They demonstrated induction of SCRaMbLE generates highly diverse genotypes with numerous genomic deletions, duplications, and transpositions (Fig. 2-2C). The SCRaMbLE method offers itself as a tool for studying higher order combinatorial genetics and for large-scale reengineering of eukaryotic genomes. Cre recombinase is an auspicious protein, and because of comprehensive biochemical understanding from years of research, it is also being investigated as a tool for gene therapy. For example, Karpinksi and colleagues recently created Brec1 recombinase(Karpinski et al., 2016). Brec1 was created by directed evolution of Cre recombinase to target Human immunodeficiency virus (HIV) LTRs. It is shown that expression of Brec1 in patient-derived HIV+ cells leads to proviral excision and curing of the virus.

Fully Synthetic Organisms

Perhaps one of the most promising aspects of genome engineering coupled with synthetic biology is the *de novo* design and construction of new organisms. The methods to accomplish this, have in large part been undertaken by Synthetic Genomics, who have developed a series of novel methods for identifying the minimal set of genes needed for a genome and *in vitro* and *in vivo* assembly of large synthetic DNA molecules(Gibson et al., 2009). To facilitate large-scale *in vitro* DNA assembly, Daniel Gibson et al developed Gibson DNA Assembly to assemble DNA molecules nearly half a megabase(Gibson et al., 2009). To accomplish this, T5 exonuclease removes nucleotides from substrate DNA molecules. This reveals single-stranded

DNA homologies and allows hybridization between separate DNA molecules. Meanwhile, Phusion DNA polymerase adds nucleotides counter to the exonuclease and Taq ligase catalyzes formation of phosphodiester linkages thus connecting DNA molecules.

To create the synthetic genome, Lartigue and colleagues developed a method to transfer whole genomes between bacteria via digestion of cells in agarose plugs and polyethylene glycol (PEG) mediated transformation(Lartigue et al., 2007). More recently, Lartigue and coworkers described and updated method for genome transplantation wherein whole bacterial chromosomes are transplanted to yeast. *S. cerevisiae* is an extremely effective host for homologous recombination experiments and propagation of large DNA molecules(Lartigue et al., 2009). In this study they demonstrated the utility of yeast HR by removing a type III restriction enzyme gene that renders M. mycoides cells resistant to introduction of exogenous DNA molecules. In addition, Karas et al recently described a protocol for direct cell-to-cell transfer of genomes(Karas et al., 2014). This method utilizes PEG mediated cellular fusion, thus reducing the likelihood of chromosome damage caused during the DNA purification process.

To go from synthetic DNA sequences to a full genome, Gibson and coworkers utilized a hierarchical DNA assembly scheme with a mixture of *in vitro* and in yeast DNA construction (Fig. 2-2D). The 1-megabase synthetic *M. Mycoides* genome (JCVI-syn1.0) was transplanted into recipient *Mycoplasma capricolum* cells. This represented the first assembly scheme going entirely from synthetic DNA sequence to full genome of an organism(Gibson et al., 2010). More recently Hutchinson and coworkers revealed a new minimal genome, JCVI-syn3.0. Here they reduced the size of the 1 megabase *M. mycoides* genome to 531 kilobases(Hutchison et al., 2016). This substantial reduction of genome sizes was accomplished by genome redesign informed by TN5 transposon mutagenesis studies of JCVI-syn1.0. Collectively the tools for genome construction developed by the J. Craig Venter Institute provide a framework for synthetic biologist to go from digitized DNA sequence to full genomes.

Table 2-1: Example applications of genome engineering to obtain certain products:

Method:	Host:	Description:	Product:	Ref:
MAGE	*Escherichia coli*	Multiplex automated genome engineering: an automated recombineering work flow for directed evolution	Lycopene	(H. H. Wang et al., 2009)
CAGE	*Escherichia coli*	Conjugative assembly genome engineering for hierarchical assembly of genomic mutations	UAG codon replacement*	(Isaacs et al., 2011)
RAGE	*Escherichia coli*	Recombinase assisted genome engineering for integration of heterologous pathways into the *E. coli* genome	Ethanol from brown macroalgae	(Santos et al., 2013)
CRISPR/lambda Red	*Escherichia coli*	Integration of large synthetic constructs into the *E. coli* genome	Isobutanol	(Bassalo et al., 2016)
CasEMBLR	*Saccharomyces cerevisiae*	Multiplex assembly of biosynthetic pathways on the yeast genome	Carotenoids and tyrosine	(Jakočiūnas et al., 2015)
mCRISTAR	*Saccharomyces cerevisiae*	Combined CRISPR and TAR cloning for construction and refactoring of pathways for application in heterologous organisms	Tetarimycin A	(Kang et al., 2016)

*Product will be useful for engineering of synthetic proteins

CONCLUSION

As circuits progress and become more complex, editing and genome redesign schemes will require more and more power. Systematic rational and combinatoric design coupled with evolutionary based engineering methods will enable production of microbes with larger synthetic gene networks(Pál et al., 2014; Smanski et al., 2014; Temme et al., 2012). Technologies that allow us to go from digitized DNA sequence to biological implementation are central to synthetic biology. CRISPR systems with high specificity will become the most tractable for implementing biological devices in human cell systems(Sander & Joung, 2014). CRISPR based genome engineering methods make rapid construction of biosynthetic pathways possible (See Table 2-1). Genome engineering tools like MAGE enable optimization of biosynthetic pathways and will become more prevalent as automated workflows become more commonplace. Coupling of MAGE with CAGE has made removal of all UAG stop codons from the *E. coli* genome possible (See Table 2-1). Freeing of various codons is a step towards engineering organisms with orthogonal genetic code from their outside counter parts. This is an important hurdle synthetic biology faces in eventual application outside the laboratory. Large-scale genome reduction and construction of minimal genomes will enable creation of designer bacterial strains with reduced metabolic burden and increased genetic stability(Pósfai et al., 2006).

REFERENCES

Bao, Z., Xiao, H., Liang, J., Zhang, L., Xiong, X., Sun, N., Si, T., & Zhao, H. (2015). Homology-integrated CRISPR-Cas (HI-CRISPR) system for one-step multigene disruption in Saccharomyces cerevisiae. *ACS Synthetic Biology*, *4*(5), 585–594. https://doi.org/10.1021/sb500255k

Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schneider, D., Lenski, R. E., & Kim, J. F. (2009). Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature*, *461*(7268), 1243–1247. https://doi.org/10.1038/nature08480

Bassalo, M. C., Garst, A. D., Halweg-Edwards, A. L., Grau, W. C., Domaille, D. W., Mutalik, V. K., Arkin, A. P., & Gill, R. T. (2016). Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli. *ACS Synthetic Biology*, *5*(7), 561–568. https://doi.org/10.1021/acssynbio.5b00187

Brophy, J. A. N., & Voigt, C. A. (2014). Principles of genetic circuit design. *Nature Methods*, *11*(5), 508–520. https://doi.org/10.1038/nmeth.2926

Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., Dickman, M. J., Makarova, K. S., Koonin, E. V., & Oost, J. van der. (2008). Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science*, *321*(5891), 960–964. https://doi.org/10.1126/science.1159689

Choi, K. R., & Lee, S. Y. (n.d.). CRISPR Technologies for Bacterial Systems: Current Achievements and Future Directions. *Biotechnology Advances*, *34*(7), 1180–1209. https://doi.org/10.1016/j.biotechadv.2016.08.002

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, *339*(6121), 819–823. https://doi.org/10.1126/science.1231143

Cooper, V. S., Schneider, D., Blot, M., & Lenski, R. E. (2001). Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of Escherichia coli B. *Journal of Bacteriology*, *183*(9), 2834–2841. https://doi.org/10.1128/JB.183.9.2834-2841.2001

Csörgő, B., Nyerges, Á., Pósfai, G., & Fehér, T. (2016). System-level genome editing in microbes. *Current Opinion in Microbiology*, *33*, 113–122. https://doi.org/10.1016/j.mib.2016.07.005

DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J., & Church, G. M. (2013). Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Research*, *41*(7), 4336–4343. https://doi.org/10.1093/nar/gkt135

Doyon, Y., McCammon, J. M., Miller, J. C., Faraji, F., Ngo, C., Katibah, G. E., Amora, R., Hocking, T. D., Zhang, L., Rebar, E. J., Gregory, P. D., Urnov, F. D., & Amacher, S. L. (2008). Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nature Biotechnology*, *26*(6), 702–708. https://doi.org/10.1038/nbt1409

Dymond, J. S., Richardson, S. M., Coombes, C. E., Babatz, T., Muller, H., Annaluru, N., Blake, W. J., Schwerzmann, J. W., Dai, J., Lindstrom, D. L., Boeke, A. C., Gottschling, D. E., Chandrasegaran, S., Bader, J. S., & Boeke, J. D. (2011). Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature*, *477*(7365), 471–476. https://doi.org/10.1038/nature10403

Elena, S. F., & Lenski, R. E. (2003). Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation. *Nature Reviews Genetics*, *4*(6), 457–469. https://doi.org/10.1038/nrg1088

Ellis, T., Wang, X., & Collins, J. J. (2009). Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nature Biotechnology*, *27*(5), 465–471. https://doi.org/10.1038/nbt.1536

Enyeart, P. J., Chirieleison, S. M., Dao, M. N., Perutka, J., Quandt, E. M., Yao, J., Whitt, J. T., Keatinge-Clay, A. T., Lambowitz, A. M., & Ellington, A. D. (2013). Generalized bacterial genome editing using mobile group II introns and Cre-lox. *Molecular Systems Biology*, *9*(1), 685. https://doi.org/10.1038/msb.2013.41

Farzadfard, F., & Lu, T. K. (2014). Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science*, *346*(6211), 1256272. https://doi.org/10.1126/science.1256272

Faucon, P. C., Pardee, K., Kumar, R. M., Li, H., Loh, Y.-H., & Wang, X. (2014). Gene networks of fully connected triads with complete auto-activation enable multistability and stepwise stochastic transitions. *PloS One*, *9*(7), e102873. https://doi.org/10.1371/journal.pone.0102873

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., & Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*, *31*(9), 822–826. https://doi.org/10.1038/nbt.2623

Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M., & Joung, J. K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnology*, *32*(3), 279–284. https://doi.org/10.1038/nbt.2808

Gardner, T. S., Cantor, C. R., & Collins, J. J. (2000). Construction of a genetic toggle switch in Escherichia coli. *Nature*, *403*(6767), 339–342. https://doi.org/10.1038/35002131

Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R.-Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E. A., Young, L., Qi, Z.-Q., Segall-Shapiro, T. H., … Venter, J. C. (2010). Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science*, *329*(5987), 52–56. https://doi.org/10.1126/science.1190719

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345. https://doi.org/10.1038/nmeth.1318

Guilinger, J. P., Thompson, D. B., & Liu, D. R. (2014). Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nature Biotechnology*, *32*(6), 577–582. https://doi.org/10.1038/nbt.2909

Hao, H., Wang, X., Jia, H., Yu, M., Zhang, X., Tang, H., & Zhang, L. (2016). Large fragment deletion using a CRISPR/Cas9 system in Saccharomyces cerevisiae. *Analytical Biochemistry*, *509*, 118–123. https://doi.org/10.1016/j.ab.2016.07.008

He, X., Tan, C., Wang, F., Wang, Y., Zhou, R., Cui, D., You, W., Zhao, H., Ren, J., & Feng, B. (2016). Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucleic Acids Research*, *44*(9), e85. https://doi.org/10.1093/nar/gkw064

Hutchison, C. A., Chuang, R.-Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., Gill, J., Kannan, K., Karas, B. J., Ma, L., Pelletier, J. F., Qi, Z.-Q., Richter, R. A., Strychalski, E. A., Sun, L., Suzuki, Y., Tsvetanova, B., Wise, K. S., Smith, H. O., … Venter, J. C. (2016). Design and synthesis of a minimal bacterial genome. *Science*, *351*(6280), aad6253. https://doi.org/10.1126/science.aad6253

Isaacs, F. J., Carr, P. A., Wang, H. H., Lajoie, M. J., Sterling, B., Kraal, L., Tolonen, A. C., Gianoulis, T. A., Goodman, D. B., Reppas, N. B., Emig, C. J., Bang, D., Hwang, S. J., Jewett, M. C., Jacobson, J. M., & Church, G. M. (2011). Precise Manipulation of Chromosomes in Vivo Enables Genome-Wide Codon Replacement. *Science*, *333*(6040), 348–353. https://doi.org/10.1126/science.1205822

Jakočiūnas, T., Rajkumar, A. S., Zhang, J., Arsovska, D., Rodriguez, A., Jendresen, C. B., Skjødt, M. L., Nielsen, A. T., Borodina, I., Jensen, M. K., & Keasling, J. D. (2015). CasEMBLR: Cas9-Facilitated Multiloci Genomic Integration of in Vivo Assembled DNA Parts in Saccharomyces cerevisiae. *ACS Synthetic Biology*, *4*(11), 1226–1234. https://doi.org/10.1021/acssynbio.5b00007

Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology*, *31*(3), 233–239. https://doi.org/10.1038/nbt.2508

Jiang, W., & Marraffini, L. A. (2015). CRISPR-Cas: New Tools for Genetic Manipulations from Bacterial Immunity Systems. *Annual Review of Microbiology*, *69*(1), 209–228. https://doi.org/10.1146/annurev-micro-091014-104441

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829

Kang, H.-S., Charlop-Powers, Z., & Brady, S. F. (2016). Multiplexed CRISPR/Cas9- and TAR-Mediated Promoter Engineering of Natural Product Biosynthetic Gene Clusters in Yeast. *ACS Synthetic Biology*, *5*(9), 1002–1010. https://doi.org/10.1021/acssynbio.6b00080

Kannan, K., Tsvetanova, B., Chuang, R.-Y., Noskov, V. N., Assad-Garcia, N., Ma, L., Hutchison III, C. A., Smith, H. O., Glass, J. I., Merryman, C., Venter, J. C., & Gibson, D. G. (2016). One step engineering of the small-subunit ribosomal RNA using CRISPR/Cas9. *Scientific Reports*, *6*, 30714. https://doi.org/10.1038/srep30714

Karas, B. J., Jablanovic, J., Irvine, E., Sun, L., Ma, L., Weyman, P. D., Gibson, D. G., Glass, J. I., Venter, J. C., Hutchison Iii, C. A., Smith, H. O., & Suzuki, Y. (2014). Transferring whole genomes from bacteria to yeast spheroplasts using entire bacterial cells to reduce DNA shearing. *Nature Protocols*, *9*(4), 743–750. https://doi.org/10.1038/nprot.2014.045

Karpinski, J., Hauber, I., Chemnitz, J., Schäfer, C., Paszkowski-Rogacz, M., Chakraborty, D., Beschorner, N., Hofmann-Sieber, H., Lange, U. C., Grundhoff, A., Hackmann, K., Schrock, E., Abi-Ghanem, J., Pisabarro, M. T., Surendranath, V., Schambach, A., Lindner, C., van Lunzen, J., Hauber, J., & Buchholz, F. (2016). Directed evolution of a recombinase that excises the provirus of most HIV-1 primary isolates with high specificity. *Nature Biotechnology*, *34*(4), 401–409. https://doi.org/10.1038/nbt.3467

Kiani, S., Beal, J., Ebrahimkhani, M. R., Huh, J., Hall, R. N., Xie, Z., Li, Y., & Weiss, R. (2014). CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nature Methods*, *11*(7), 723–726. https://doi.org/10.1038/nmeth.2969

Kiani, S., Chavez, A., Tuttle, M., Hall, R. N., Chari, R., Ter-Ovanesyan, D., Qian, J., Pruitt, B. W., Beal, J., Vora, S., Buchthal, J., Kowal, E. J. K., Ebrahimkhani, M. R., Collins, J. J., Weiss, R., & Church, G. (2015). Cas9 gRNA engineering for genome editing, activation and repression. *Nature Methods*, *12*(11), 1051–1054. https://doi.org/10.1038/nmeth.3580

Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., & Joung, J. K. (2016). High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, *529*(7587), 490–495. https://doi.org/10.1038/nature16526

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Topkar, V. V., Zheng, Z., & Joung, J. K. (2015). Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nature Biotechnology*, *33*(12), 1293–1298. https://doi.org/10.1038/nbt.3404

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., Gonzales, A. P. W., Li, Z., Peterson, R. T., Yeh, J.-R. J., Aryee, M. J., & Joung, J. K. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, *523*(7561), 481–485. https://doi.org/10.1038/nature14592

Kolisnychenko, V., Plunkett, G., Herring, C. D., Feher, T., Posfai, J., Blattner, F. R., & Posfai, G. (2002). Engineering a Reduced Escherichia coli Genome. *Genome Research*, *12*(4), 640–647. https://doi.org/10.1101/gr.217202

Krishnakumar, R., Grose, C., Haft, D. H., Zaveri, J., Alperovich, N., Gibson, D. G., Merryman, C., & Glass, J. I. (2014). Simultaneous non-contiguous deletions using large synthetic DNA and site-specific recombinases. *Nucleic Acids Research*, *42*(14), e111. https://doi.org/10.1093/nar/gku509

Kuhlman, T. E., & Cox, E. C. (2010). Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Research*, *38*(6), e92. https://doi.org/10.1093/nar/gkp1193

Lajoie, M. J., Gregg, C. J., Mosberg, J. A., Washington, G. C., & Church, G. M. (2012). Manipulating replisome dynamics to enhance lambda Red-mediated multiplex genome engineering. *Nucleic Acids Research*, *40*(22), e170. https://doi.org/10.1093/nar/gks751

Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H.-R., Haimovich, A. D., Kuznetsov, G., Mercer, J. A., Wang, H. H., Carr, P. A., Mosberg, J. A., Rohland, N., Schultz, P. G., Jacobson, J. M., Rinehart, J., Church, G. M., & Isaacs, F. J. (2013). Genomically Recoded Organisms Expand Biological Functions. *Science*, *342*(6156), 357–360. https://doi.org/10.1126/science.1241459

Lartigue, C., Glass, J. I., Alperovich, N., Pieper, R., Parmar, P. P., Hutchison, C. A., Smith, H. O., & Venter, J. C. (2007). Genome Transplantation in Bacteria: Changing One Species to Another. *Science*, *317*(5838), 632–638. https://doi.org/10.1126/science.1144622

Lartigue, C., Vashee, S., Algire, M. A., Chuang, R.-Y., Benders, G. A., Ma, L., Noskov, V. N., Denisova, E. A., Gibson, D. G., Assad-Garcia, N., Alperovich, N., Thomas, D. W., Merryman, C., Hutchison, C. A., Smith, H. O., Venter, J. C., & Glass, J. I. (2009). Creating Bacterial Strains from Genomes That Have Been Cloned and Engineered in Yeast. *Science*, *325*(5948), 1693–1696. https://doi.org/10.1126/science.1173759

Li, Q., Chen, J., Minton, N. P., Zhang, Y., Wen, Z., Liu, J., Yang, H., Zeng, Z., Ren, X., Yang, J., Gu, Y., Jiang, W., Jiang, Y., & Yang, S. (2016). CRISPR-based genome editing and expression control systems in Clostridium acetobutylicum and Clostridium beijerinckii. *Biotechnology Journal*, *11*(7), 961–972. https://doi.org/10.1002/biot.201600053

Liao, C., Seo, S.-O., Celik, V., Liu, H., Kong, W., Wang, Y., Blaschek, H., Jin, Y.-S., & Lu, T. (2015). Integrated, systems metabolic picture of acetone-butanol-ethanol fermentation by Clostridium acetobutylicum. *Proceedings of the National Academy of Sciences*, *112*(27), 8505–8510. https://doi.org/10.1073/pnas.1423143112

Litcofsky, K. D., Afeyan, R. B., Krom, R. J., Khalil, A. S., & Collins, J. J. (2012). Iterative plug-and-play methodology for constructing and modifying synthetic gene networks. *Nature Methods*, *9*(11), 1077–1080. https://doi.org/10.1038/nmeth.2205

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J. M., Wolf, Y. I., Yakunin, A. F., van der Oost, J., & Koonin, E. V. (2011). Evolution and classification of the CRISPR-Cas systems. *Nature Reviews. Microbiology*, *9*(6), 467–477. https://doi.org/10.1038/nrmicro2577

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., & Church, G. M. (2013). RNA-Guided Human Genome Engineering via Cas9. *Science*, *339*(6121), 823–826. https://doi.org/10.1126/science.1232033

Marraffini, L. A. (2015). CRISPR-Cas immunity in prokaryotes. *Nature*, *526*(7571), 55–61. https://doi.org/10.1038/nature15386

Marraffini, L. A., & Sontheimer, E. J. (2008). CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. *Science*, *322*(5909), 1843–1845. https://doi.org/10.1126/science.1165771

Mojica, F. J. M., Diez-Villasenor, C., Garcia-Martinez, J., & Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, *155*(3), 733–740. https://doi.org/10.1099/mic.0.023960-0

Mosberg, J. A., Lajoie, M. J., & Church, G. M. (2010). Lambda red recombineering in Escherichia coli occurs through a fully single-stranded intermediate. *Genetics*, *186*(3), 791–799. https://doi.org/10.1534/genetics.110.120782

Mougiakos, I., Bosma, E. F., de Vos, W. M., van Kranenburg, R., & van der Oost, J. (2016). Next Generation Prokaryotic Engineering: The CRISPR-Cas Toolkit. *Trends in Biotechnology*, *34*(7), 575–587. https://doi.org/10.1016/j.tibtech.2016.02.004

Pál, C., Papp, B., & Pósfai, G. (2014). The dawn of evolutionary genome engineering. *Nature Reviews Genetics*, *15*(7), 504–512. https://doi.org/10.1038/nrg3746

Perli, S. D., Cui, C. H., & Lu, T. K. (2016). Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science*, *353*(6304), aag0511. https://doi.org/10.1126/science.aag0511

Pósfai, G., Plunkett, G., Fehér, T., Frisch, D., Keil, G. M., Umenhoffer, K., Kolisnychenko, V., Stahl, B., Sharma, S. S., Arruda, M. de, Burland, V., Harcum, S. W., & Blattner, F. R. (2006). Emergent Properties of Reduced-Genome Escherichia coli. *Science*, *312*(5776), 1044–1046. https://doi.org/10.1126/science.1126439

Ran, F. A., Hsu, P. D., Lin, C.-Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., Scott, D. A., Inoue, A., Matoba, S., Zhang, Y., & Zhang, F. (2013). Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell*, *155*(2), 479–480. https://doi.org/10.1016/j.cell.2013.09.040

Sander, J. D., & Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnology*, *32*(4), 347–355. https://doi.org/10.1038/nbt.2842

Santos, C. N. S., Regitsky, D. D., & Yoshikuni, Y. (2013). Implementation of stable and complex biological systems through recombinase-assisted genome engineering. *Nature Communications*, *4*, 2503. https://doi.org/10.1038/ncomms3503

Santos, C. N. S., & Yoshikuni, Y. (2014). Engineering complex biological systems in bacteria through recombinase-assisted genome engineering. *Nature Protocols*, *9*(6), 1320–1336. https://doi.org/10.1038/nprot.2014.084

Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., & Zhang, F. (2015). Rationally engineered Cas9 nucleases with improved specificity. *Science*, *351*(6268), 84–88. https://doi.org/10.1126/science.aad5227

Smanski, M. J., Bhatia, S., Zhao, D., Park, Y., Woodruff, L. B. A., Giannoukos, G., Ciulla, D., Busby, M., Calderon, J., Nicol, R., Gordon, D. B., Densmore, D., & Voigt, C. A. (2014). Functional optimization of gene clusters by combinatorial design and assembly. *Nature Biotechnology*, *32*(12), 1241–1249. https://doi.org/10.1038/nbt.3063

Standage-Beier, K., Zhang, Q., & Wang, X. (2015). Targeted Large-Scale Deletion of Bacterial Genomes Using CRISPR-Nickases. *ACS Synthetic Biology*, *4*(11), 1217–1225. https://doi.org/10.1021/acssynbio.5b00132

St-Pierre, F., Cui, L., Priest, D. G., Endy, D., Dodd, I. B., & Shearwin, K. E. (2013). One-Step Cloning and Chromosomal Integration of DNA. *ACS Synthetic Biology*, *2*(9), 537–541. https://doi.org/10.1021/sb400021j

Temme, K., Zhao, D., & Voigt, C. A. (2012). Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(18), 7085–7090. https://doi.org/10.1073/pnas.1120788109

Tsai, S. Q., Wyvekens, N., Khayter, C., Foden, J. A., Thapar, V., Reyon, D., Goodwin, M. J., Aryee, M. J., & Joung, J. K. (2014). Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature Biotechnology*, *32*(6), 569–576. https://doi.org/10.1038/nbt.2908

Tsarmpopoulos, I., Gourgues, G., Blanchard, A., Vashee, S., Jores, J., Lartigue, C., & Sirand-Pugnet, P. (2016). In-Yeast Engineering of a Bacterial Genome Using CRISPR/Cas9. *ACS Synthetic Biology*, *5*(1), 104–109. https://doi.org/10.1021/acssynbio.5b00196

Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., & Church, G. M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, *460*(7257), 894–898. https://doi.org/10.1038/nature08187

Wang, L.-Z., Wu, F., Flores, K., Lai, Y.-C., & Wang, X. (2016). Build to understand: Synthetic approaches to biology. *Integrative Biology*, *8*(4), 394–408. https://doi.org/10.1039/C5IB00252D

Wang, Y., Zhang, Z.-T., Seo, S.-O., Choi, K., Lu, T., Jin, Y.-S., & Blaschek, H. P. (2015). Markerless chromosomal gene deletion in Clostridium beijerinckii using CRISPR/Cas9 system. *Journal of Biotechnology*, *200*, 1–5. https://doi.org/10.1016/j.jbiotec.2015.02.005

Wu, F., Menn, D. J., & Wang, X. (2014). Quorum-Sensing Crosstalk-Driven Synthetic Circuits: From Unimodality to Trimodality. *Chemistry & Biology*, *21*(12), 1629–1638. https://doi.org/10.1016/j.chembiol.2014.10.008

Wu, M., Su, R.-Q., Li, X., Ellis, T., Lai, Y.-C., & Wang, X. (2013). Engineering of regulated stochastic cell fate determination. *Proceedings of the National Academy of Sciences*, 201305423. https://doi.org/10.1073/pnas.1305423110

Yokobayashi, Y., Weiss, R., & Arnold, F. H. (2002). Directed evolution of a genetic circuit. *Proceedings of the National Academy of Sciences*, *99*(26), 16587–16591. https://doi.org/10.1073/pnas.252535999

CHAPTER 3

TRANSIENT REPORTER FOR EDITING ENRICHMENT

ABSTRACT

Current approaches to identify cell populations that have been modified with deaminase base editing technologies are inefficient and rely on downstream sequencing techniques. In this study, we utilized a blue fluorescent protein (BFP) that converts to green fluorescent protein (GFP) upon a C-to-T substitution as an assay to report directly on base editing activity within a cell. Using this assay, we optimize various base editing transfection parameters and delivery strategies. Moreover, we utilize this assay in conjunction with flow cytometry to develop a transient reporter for editing enrichment (TREE) to efficiently purify base-edited cell populations. Compared to conventional cell enrichment strategies that employ reporters of transfection (RoT), TREE significantly improved the editing efficiency at multiple independent loci, with efficiencies approaching 80%. We also employed the BFP-to-GFP conversion assay to optimize base editor vector design in human pluripotent stem cells (hPSCs), a cell type that is resistant to genome editing and in which modification via base editors has not been previously reported. Finally, using these optimized vectors in the context of TREE allowed for the highly efficient editing of hPSCs. We envision TREE as a readily adoptable method to facilitate base editing applications in synthetic biology, disease modeling, and regenerative medicine.

INTRODUCTION

The rapid advancement of CRISPR/Cas-based technologies has allowed for the modification (i.e. deletion, mutation, and insertion) of human cells at precise genomic locations (Hsu et al., 2014; Komor et al., 2017; Wright et al., 2016). For applications in which precise editing of a single nucleotide is desired, the CRISPR/Cas machinery can be used to introduce site-specific double-stranded breaks (DSB) followed by homology-directed repair (HDR) using an exogenous DNA template (Brookhouser et al., 2017). However, HDR is inefficient in mammalian cells, especially in recalcitrant cells such as human pluripotent stem cells (hPSCs), and repair of DSB is predominantly achieved through non-homologous end joining (NHEJ) (Grobarczyk et al., 2015; Huang et al., 2015; Li et al., 2015; Miyaoka et al., 2014; Reinhardt et al., 2013). In addition, NHEJ results in insertion or deletion of nucleotides (indels), resulting in undesired disruption (e.g. frameshift mutations, premature stop codons, deletion) of the targeted genes.

As an alternative to standard gene editing approaches that require a DSB, several groups have reported the development of deaminase base editors that do not rely on HDR to introduce single nucleotide genomic changes (Komor et al., 2018). Broadly speaking, these base editors consist of a D10A nickase Cas endonuclease fused to deaminase enzymes capable of converting cytosine to thymine (Komor et al., 2016) and adenine to guanine (Gaudelli et al., 2017) without the need for DSB and homology repair templates. Moreover, genome modification through the use of base editors has been shown to result in formation of fewer indels when compared to HDR-based methods (Eid et al., 2018; Gehrke et al., 2018).

Despite the advantages that deaminase base editors offer, identification and isolation of cell populations that have been successfully edited remains challenging. Specifically, there is no readily detectable phenotype to distinguish edited from unedited cells. In turn, isolation of edited cell populations requires single cell isolation followed by downstream sequencing verification (Germini et al., 2018). Some progress has been made to help enrich for edited cells, such as co-transfecting plasmids with a fluorescent reporter and using flow cytometry to isolate reporter-positive cells. Similarly, base editors fused to fluorescent proteins have been used to enrich for

edited cell populations (Koblan et al., 2018). However, these techniques are only reporters of transfection (RoT) and do not report on base editing activity within a cell population.

In this work, we sought to develop an assay to allow for the real-time, fluorescent-based identification and isolation of base-edited cell populations. To develop this method, we were motivated by previous work that employed a genomically integrated green fluorescent protein (GFP) that is converted to blue fluorescent protein (BFP) upon CRISPR/Cas9-driven HDR (Glaser et al., 2016). Here, we engineered a BFP variant that undergoes conversion to GFP after targeted modification with a cytidine deaminase-based DNA base editor. We applied our BFP-to-GFP conversion assay to optimize various base editing transfection parameters and delivery strategies. We then utilized this BFP-to-GFP assay in conjunction with flow cytometry to develop a technique called transient reporter for editing enrichment (TREE) which allows for the fluorescent-based isolation of base edited cell populations. As such, we applied TREE to enrich for cell populations that had been edited at various genomic loci, including sites that are refractory to modification. Significantly, we demonstrate how TREE can provide for enrichment of edited human pluripotent stem cells (hPSCs), a cell type that is resistant to traditional CRISPR/Cas9 HDR-based approaches and in which modification via base editors has not been previously reported. Overall, because TREE can be facilely implemented to isolate edited cell populations, it will significantly enhance and enable the use of base editors for numerous downstream applications including those in synthetic biology, protein engineering, disease modeling, and regenerative medicine.

MATERIALS AND METHODS

Plasmid construction. Unless otherwise noted, all molecular cloning PCR reactions were performed using Phusion® High-Fidelity DNA polymerase (New England Biolabs, Ipswich, MA, USA) using the using the manufacturer's recommended protocols. All restriction enzyme (New England Biolabs) digests were performed according to the manufacturer's instructions. Ligation reactions were performed with T4 DNA Ligase (New England Biolabs) according to the manufacturer's instructions. PCR primers and oligonucleotides were synthesized by Integrated DNA Technologies (Coralville, IA, USA). All PCR products and intermediate plasmid products were confirmed via Sanger sequencing (DNASU Sequencing Core Facility and Genewiz). Complete plasmid sequences will be made available upon request.

For construction of the pEF-BFP plasmid, we utilized PCR to add the H-66 and PAM site mutations into a GFP cassette (Addgene #11154). PCR products containing these mutations were digested with SapI/EcoRI and SapI/NotI and ligated into a EcoRI/NotI digested EF1α expression vector (Addgene #11154).

For construction of the pDT-sgRNA vector, sgRNAs were synthesized as pairs of oligonucleotides (Appendix A Table A-1). Subsequently, 5' phosphates were added to each oligonucleotide pair by incubating 1 µg oligonucleotide in 50 µL reactions containing 1X T4 DNA Ligase Buffer (New England Biolabs) and 10 units of T4 Polynucleotide Kinase at 37°C overnight. Oligonucleotides were then duplexed by heating the kinase reactions to 90°C on an aluminum heating block for 5 minutes followed by slowly returning the reaction to room temperature over 1 hour. Following duplexing, guides were cloned into a modified pSB1C3 vector containing a U6 promoter, inverted BbsI restriction enzyme digestion sites, and a *S. pyogenes* recognized sgRNA hairpin. For construction of pMT-sgRNA, pairs of sgRNAs (Appendix A Table A-1) were PCR amplified with primers adding EcoRI/SapI restriction enzyme digestion sites or SapI/XbaI restriction enzyme digestion sites. Purified PCR products were then digested with the respective restriction enzymes and ligated into EcoRI/XbaI digested pUC19 vector (Addgene #50005). The resultant vector contained pairs of sgRNA expression cassettes. To add additional sgRNA expression cassettes, pairs of sgRNAs were PCR amplified with primers that add HindIII/SapI or

SapI/HindIII restriction enzyme digestion sites. These products were then digested with HindII/SapI and ligated into HindIII digested and dephosphorylated pDT-sgRNA vector.

For insertion of the EF1α promoter into the pCMV-BE4-Gam (Addgene #100806) and pCMV-AncBE4max (Addgene #112094), EF1α was PCR amplified from an EF1α expression vector (Addgene #11154) adding SpeI/NotI restriction enzyme digestion sites. After purification and digestion, these PCR products were ligated into SpeI/NotI digested and dephosphorylated pCMV-BE4-Gam or pCMV-AncBE4max vectors.

Cell culture. All media component were purchased from ThermoFisher Scientific (Waltham, Massachusetts, USA) unless indicated otherwise. HEK293 cells were cultured on poly-L-ornithine (4 μg/mL; Sigma Aldrich, St. Louis MO, USA) coated plates in the following media: 1X high glucose DMEM, 10% (v/v) fetal bovine serum, 1% (v/v) L-glutamine penicillin/streptomycin. Culture medium was every other day and cells were passaged with Accutase (ThermoFisher) every 5 days. HPSCs were cultured on Matrigel™ (BD Biosciences, San Jose, CA, USA) coated plates in Essential 8™ Medium (E8) (ThermoFisher). Culture medium was changed everyday and cells were passaged with Accutase every 4-5 days. After passaging, the medium was supplemented with 5 μM Rho kinase inhibitor (ROCKi; Y-27632 [BioGems, Westlake Village, CA, USA]) for 24 hours to aid in single cell survival.

Isolation of episomal DNA. After 48 hours following transfection, cells were dissociated from the tissue plates with Accutase, washed twice with PBS, and resuspended in RNAse-A containing solution. Cells were then lysed via alkaline lysis and the resultant debris was precipitated via centrifugation at $1.2x10^4$ x g for 10 minutes. Supernatant DNA was isolated by column DNA purification using the manufacture recommended protocol (Sigma Aldrich: NA0160).

Generation of HEK293-BFP line. The HEK293T-BFP cell line was generated via homology independent target integration (HITI) (56). Briefly, the BFP coding sequence was PCR amplified with primers adding EcoRI restriction enzyme digestion sites. The resultant PCR product was EcoRI digested, phosphorylated, and ligated into an EcoRI/SmaI digested vector containing an EF1α promoter, puromycin resistance cassette, and HITI protospacer sequence (pEF-BFP-Puro^R). The pEF-BFP-Puro^R vector was co-transfected in HEK293s with pX330

(Addgene #42230) and a custom sgRNA vector (pHSG(BG)-1C3) targeting the *C1ORF228* locus. Transfections were conducted in a 24-well plate with 300 ng pX330, 400 ng pEF-BFP-Puro^{R}, 50 ng sgRNA vector, 1.5 µL Lipofectamine 3000 (ThermoFisher Scientific), and 1 µL P3000 transfection reagent. Cells were passaged at 72 hours post-transfection into a single well of a 6-well plate and selected with 0.5 µg/mL puromycin for 2 weeks.

RNP complex formation. For purification of recombinant BE3 (rBE3) protein, BL21 Star DE3 cells (ThermoFisher) were transformed with pET42b-BE3 (Addgene #87437). Protein expression was induced for 18 hours in 2 L baffled flasks at 16C with 0.5 mM isopropyl β-d-1-thiogalactopyranoside (IPTG). Cells were then harvested by centrifugation followed by lysis by sonication in lysis buffer [50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, 10 mM imidazole, 1% Triton X-100, 1 mM DTT, and 1 mg/ml lysozyme]. The lysate was cleared by centrifugation at 10,000 g for 30 minutes at 4C. The supernatant was incubated with 2 mL Ni-NTA beads (Qiagen, Germantown, MD, USA) equilibrated in lysis buffer for 1 h at 4 °C, followed by washing with 5 mL wash buffer [50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, and 20 mM imidazole] 3 times. BE3 protein was eluted with 1 mL elution buffer [50 mM Tris-HCl (pH 7.6), 250 mM NaCl, and 0.2 M imidazole]. The purified BE3 protein was exchanged and concentrated with storage buffer [20 mM HEPES (pH 7.5), 150 mM KCl, 1 mM DTT, and 10% glycerol] using an Ultracel 100K cellulose column (Millipore, Burlington MA, USA). The concentration of the protein was determined by SDS-PAGE using bovine serum albumin (BSA) standards.

Synthetic sgRNAs were synthesized as 2'-O-methylated sgRNAs (Synthego, Menlo Park, CA, USA). sgRNA was resuspended in ddH_20 to a concentration of 100 uM. Concentrated rBE3 (~1 µM) was supplemented with 10 mM $MgCl_2$, followed by addition of a 3:1 molar ratio of sgRNA. The solution was incubated at room temperature for 15 minutes to allow BE3-sgRNA complex formation.

Cell transfections. For plasmid-based transfections HEK293 cells were transfected in 12 well tissue culture plates at 40% confluence with the following reagents per well: 600 ng pCMV-BE4-Gam, 200 ng sgRNA vector [sg(BG), sg(NT), pDT-sgRNA, or pMT-sgRNA], 1.5 uL Lipofectamine 3000 Transfection Reagent (ThermoFisher), and 2 uL P3000 reagent (Thermo

Fisher). For RNP-based transfections, complexed BE3-RNPs were incubated with transfection reagents for 10-15 minutes and added dropwise to each well at a final concentration 250nM in 250µl volume total. HPSCs were transfected with 900 ng of base editing vector (pCMV-BE4-Gam, pCMV-AncBE4max, pEF1α-BE4-Gam, or pEF1α-AncBE4max), 300 ng pEF1α-BFP, 300 ng pDT-sgRNA, and 4 uL Lipofectamine Stem Transfection Reagent (ThermoFisher). All cells were harvested for sorting and/or analysis 48 hours post-transfection

Fluorescence microscopy. All imaging was performed on a Nikon Ti-Eclipse inverted microscope with and LED-based Lumencor SOLA SE Light Engine using a Semrock band pass filter. GFP was visualized with an excitation at 472 nm and emission at 520 nm. BFP was visualized with the DAPI fluorescence channel with excitation at 395 nm and emission at 460 nm.

Flow cytometry. Cells were dissociated with Accutase for 10 min at 37°C, and passed through a 40 µm cell strainer. Cells were then washed twice with flow cytometry buffer (BD Biosciences) and resuspended at a maximum concentration of 5 x 10^6 cells per 100 µL. Flow cytometry analysis was performed on an ACCURI C6 (BD Biosciences). Flow cytometry sorting was performed on a FACSAria IIu. Flow cytometry files were analyzed using with FACSDiva software (BD Biosciences), FlowJo (FlowJo LLC, Ashland, OR, USA), and custom Matlab (MathWorks, Natick, MA, USA) script.

Quantification of base editing efficiency. For HEK293 cells, genomic DNA (gDNA) was extracted from sorted and unsorted cells using NucleoSpin kit (Macherey Nagel, Bethlehem, PA, USA). PCR was performed with 500 ng DNA in a 50 µL reaction with Phusion ® High Fidelity DNA polymerase (New England Biolabs) using the primers listed in Appendix A Table A-2 and PCR protocols listed in Appendix A Table A-3. HPSCs were directly sorted into a 50 µL master mix consisting of 1X Phire Hot Start II DNA Polymerase (ThermoFisher), 1 uM forward primer, and 1 uM reverse primer. PCR was performed using the following conditions: 98°C for 5 minutes, followed by 40 cycles at 98 °C for 5 seconds, 56 °C for 5 seconds, and 72°C for 20 seconds, followed by a final 5 min 72°C extension. All products sizes were confirmed on an agarose gel prior to Sanger sequencing. Sanger sequencing was performed using column purified PCR products and the reverse primers listed in Appendix A Table A-2**.** Base editing efficiencies were

analyzed from Sanger sequence chromatograms using EditR (Kluesner et al., 2018) using the parameters listed in Appendix A Table A-4.

Off-target analysis**.** For the data presented in **Figure 4**, analysis was performed for top the off-target loci for sgRNAs for genomic sites 1-3 as predicted using the GUIDE-seq (Tsai et al., 2015). sg(BG) genomic off-targets were predicted *in silico* via CCTop using default parameters for *S. pyogenes* Cas9 against human genome reference sequence hg38 (Stemmer et al., 2015). Quantification of base editing efficiency at these off-target sites was performed in a similar manner to that at on-target sites. The PCR primers used to analyze these off-target sites are presented in Appendix A Table A-5.

Clonal isolation of edited HEK293 cells. HEK293 cells were transfected in 12 well tissue culture plates at 40% confluence with the following reagents per well: 600 ng pCMV-BE4-Gam, 200 ng pMT-sgRNA, 1.5 uL Lipofectamine 3000 Transfection Reagent, and 2 uL P3000 reagent. After 48 hours, cells were dissociated with Accutase for 5 min at 37°C, triturated, and passed through a 40 μm cell strainer. Cells were then washed twice with flow cytometry buffer (BD Biosciences) and resuspended at a maximum concentration of 5 x 10^6 cells per 100 μL. Single GFP+ cells were sorted into a single well of a 96 well plate and expanded to a 24 well plate prior to analysis.

Next-generation sequencing (NGS) of PCR amplicons. After gDNA isolation, PCR was performed using the NGS primers listed in Appendix A Table A-6 and PCR protocols listed in Appendix A Table A-7**.** PCR amplification was carried out using Phusion ® High Fidelity DNA polymerase (New England Biolabs) as described above. The products were column purified using the QIAquick PCR purification kit (Qiagen). Samples were sequenced on an Illumina MiSeq by GENWIZ. Reads were trimmed for high quality sequences via BBDuk adapter/quality filtering tool of the BBtools suite. Reads below a threshold quality score of 31 were removed using the following command (bbduk.sh in="$i" out="$x"_trim.fastq.gz trimq=30 minlen=250), where “i” is the sample file and “x” is the base name of the respective input sample file. Trimmed FASTQ files were analyzed for C-to-T editing outcomes via custom python script (Pyhton Software Foundation).

Statistical analysis. Unless otherwise noted, all data are displayed as mean ± standard deviation (S.D). Pairwise comparisons were made using Student's t-test and multiple comparisons were made using ANOVA statistical methods.

RESULTS

BFP-to-GFP conversion allows for detection of base-editing activity

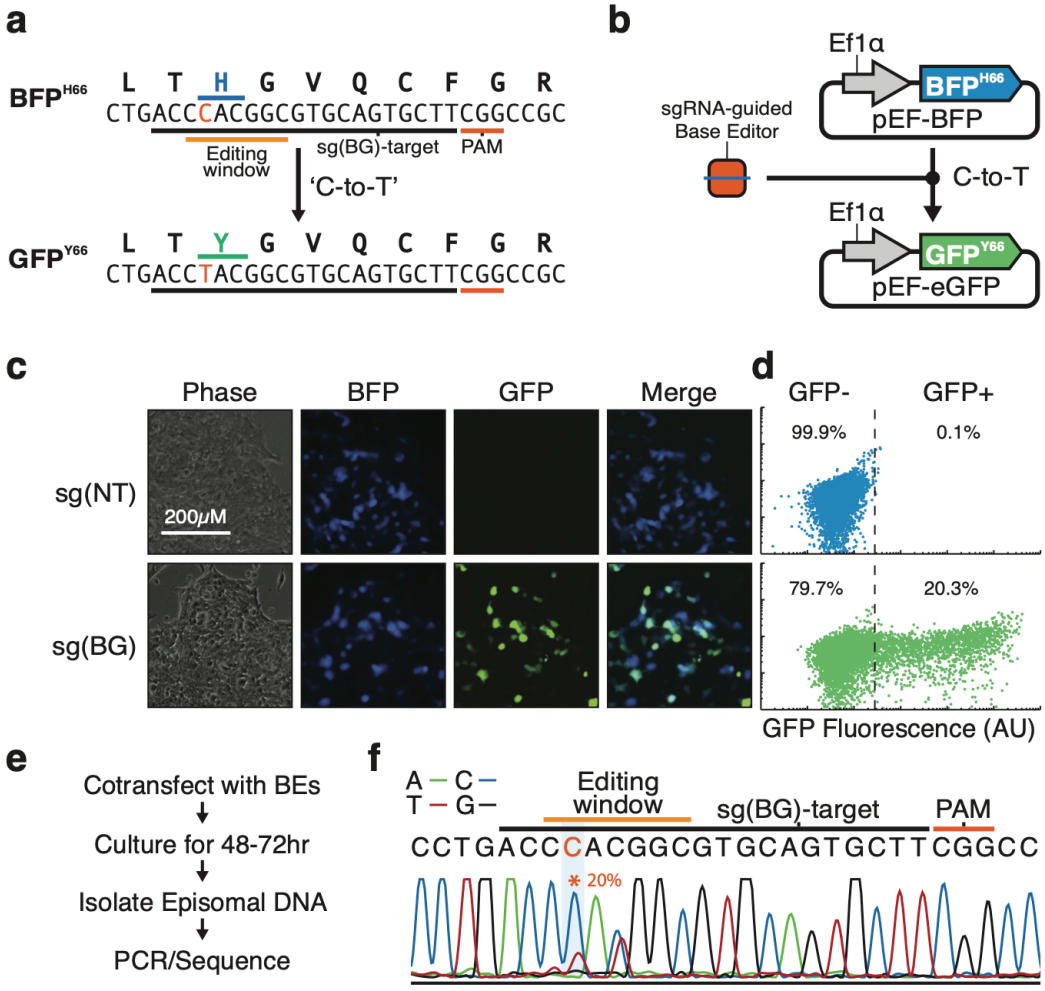


Figure 3-1. Fluorescent conversion assay. Conversion of blue fluorescent protein (BFP) to green fluorescent protein (GFP) enables detection of base-editing activity in cells. **(a)** A mutant BFP was designed to convert to GFP upon a C-to-T nucleotide conversion. The protospacer sequence (underlined black) for the sgRNA, sg (BG), targeting the 'CAC' codon (underlined blue) resulting in a C-to-T conversion to 'TAC' (underlined green) and the corresponding amino acid change of histidine (blue) to tyrosine (green) at the 66th amino acid position in BFP. A protospacer adjacent motif (PAM, underlined red) was placed in the position to orient the base editing window (underlined orange) around the C nucleotide (red) to facilitate BFP^{H66} to GFP^{Y66} conversion. All alternative C-to-T conversions in the editing window resulted in silent mutations of the coding sequence. **(b)** The BFP mutant was cloned into a vector, pEF-BFP, with a human EF1α promoter

driving expression. Targeting pEF-BFP with a cytidine deaminase base editor results in a C-to-T conversion causing a shift in the fluorescent emission spectra from BFP to GFP. **(c)** Representative fluorescent microscopy images of HEK293 cells transfected with pEF-BFP, pCMV-BE4-Gam, and sg(NT) (top row) or sg(BG) (bottom row). **(d)** Representative flow cytometry plots of HEK293 cells transfected with pEF-BFP, pCMV-BE4-Gam, and sg(NT) (top) or sg(BG) (bottom). Y-axis is a non-fluorescent control channel. **(e)** Schematic for isolation and detection of editing of episomal DNA after transfection. **(f)** Representative Sanger sequencing chromatogram of amplicons of episomal DNA isolated from HEK293 cells transfected with pEF-BFP, pCMV-BE4-Gam, and sg(BG). The presence of T-nucleotide (red trace) at the target nucleotide (red asterisk) demonstrates the C-to-T base conversion responsible for the amino acid change of histidine to tyrosine at the 66th amino acid position and subsequent shift of the BFP emission spectra of the resultant protein to a GFP variant.

To establish that blue fluorescent protein (BFP) to green fluorescent protein (GFP) conversion could be used as the basis for an assay to detect genomic base editing, we utilized a BFP mutant that converts to a GFP upon a C-to-T nucleotide conversion (Figure 3-1A). Briefly, this BFP mutant (BFP^{H66}) contains a histidine at the 66th amino acid position encoded by a 'CAC' codon. The C-to-T conversion of that codon to 'TAC' or 'TAT' will result in an amino acid change from a histidine to a tyrosine. In turn, this amino acid change will cause a shift of the emission spectra of the resultant protein generating a GFP variant (GFP^{Y66}) (Heim et al., 1994). Because the optimal nucleotide base editing window is typically 12-18 nucleotides upstream from the PAM, we also placed a *S. pyogenes* Cas9 PAM 'NGG' in a position that would enable base editing to occur at the target 'CAC' codon. To verify the utility of this fluorescent protein to report on base editing activity, we cloned the BFP coding sequence into a vector with a human EF1α promoter to drive expression (pEF-BFP; Figure 3-1B). In addition, we designed a sgRNA vector [sg(BG)] that would direct the base editing machinery to the target 'CAC' codon resulting in a C-to-T conversion and the subsequent amino acid change of histidine to tyrosine at the 66th amino acid position (Figure 3-1A). HEK293 cells were co-transfected with pEF-BFP, a base editing vector (pCMV-BE4-Gam), and sg(BG) or a control non-targeting sgRNA [sg(NT)]. Fluorescent microscopy (Figure 3-1C) and flow cytometry (Figure 3-1D) revealed that targeting pEF-BFP with sg(BG) resulted in the generation of BFP/GFP double positive cells. However, targeting pEF-BFP with sg(NT) did not result in the generation of any BFP/GFP positive cells. To confirm GFP expression was a consequence of direct editing of the target codon in pEF-BFP, we implemented a strategy to isolate and detect editing of episomal DNA after transfection (Figure 3-1E). Sanger sequencing of isolated pEF-BFP DNA established that editing had occurred at the target 'CAC' codon in pEF-

BFP resulting in a change to 'TAC' or 'TAT' reflected in the GFP emission (Figure 3-1F). Overall, these results confirm that the GFP-to-BFP conversion corresponds to C-to-T conversion at targeted base editing sites.

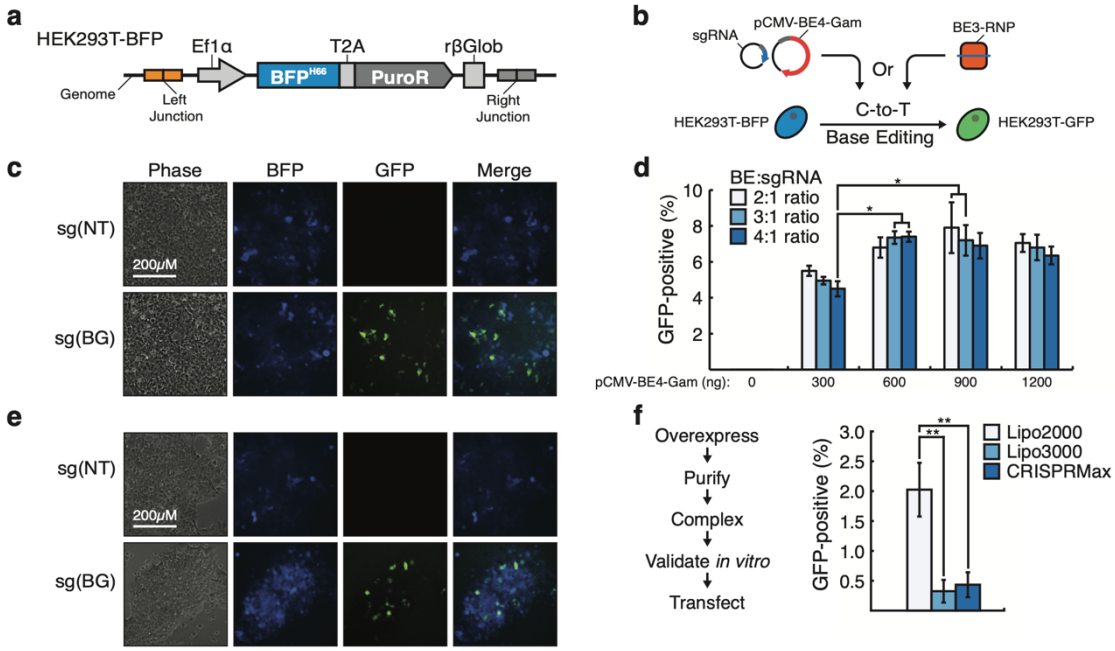


Figure 3-2: BFP-to-GFP conversion correlates with base-editing at an endogenous locus. **(a)** A pEF-BFP-Puro^R vector was integrated into the *C1ORF228* locus using homology-independent target integration to generate the HEK293-BFP cell line. **(b)** Editing efficiency (percentage GFP-positive cells) of targeting in HEK293-BFP cell line with various amounts of pCMV-BE4-Gam plasmid and ratios with the sg(BG) vector. n = 3, * = p<0.05. **(c)** Representative fluorescent microscopy images of HEK293-BFP cells transfected with 600 ng pCMV-BE4-Gam and 200 ng sg(NT) (top row) or sg(BG) (bottom row). Scale bar = 200 μm. **(d)** Relative editing efficiencies (GFP-positive cells) of HEK293-BFP cells transfected with various amounts of pCMV-BE4-Gam and ratios with the sg(BG) vector. n = 3, * = p<0.05. **(e)** Representative fluorescent microscopy images of HEK293-BFP cells transfected with BE3-sg(BG) or –sg(NT) RNP complexes. **(f)** Schematic for RNP complex generation and transfection. BE3 was overexpressed, purified, complexed and validated *in vitro*, and transfected. Relative editing efficiencies (GFP-positive cells) of HEK293-BFP cells transfected with RNP complexes using various delivery reagents. n = 3, * = p<0.05.

Next, we wanted to establish that the BFP-to-GFP conversion would correlate with base-editing efficiency at endogenous loci. To that end, we employed a HEK-293 cell line (herein referred to as HEK293-BFP) in which BFP^{H66} was stably integrated into a known genomic location (C1ORF228; Figure 3-2A). We then used this line to enable the analysis of the efficiency of base editing genomic loci (Figure 3-2B). To first assess plasmid-based base editing, we co-

transfected pCMV-BE4-Gam and sg(BG) plasmid DNA in HEK293-BFP cells. Targeting with sg(BG), but not sg(NT), resulted in generation of detectable GFP+ cells, indicating successful base editing at the targeted genomic loci (**Figure 2C**). Moreover, we were able to use this assay to systematically evaluate genomic base editing efficiencies using a range of pCMV-BE4-Gam plasmid amounts at varying ratios with the sg(BG) vector (**Figure 2D**). This analysis revealed that base editing plasmid concentration and base editor to sgRNA ratios could enhance genomic base editing efficiencies approximately 2-fold. Because ribonucleoprotein (RNP) complex-based strategies have been previously shown as an attractive alternative to plasmid-based Cas9 genome engineering (S. Kim et al., 2014; Kouranova et al., 2016; Zuris et al., 2015), we also utilized BFP-to-GFP conversion as an assay to optimize RNP-driven base editing. As such, we generated RNPs through the *in vitro* complexing of purified base editing protein with sg(BG) or sg(NT) (**Figure 2E**)**.** Our initial analysis revealed that RNP delivery using the same transfection reagent that was used for plasmid delivery of the base editor (i.e. Lipofectamine^{TM}3000) did not result in substantial BFP-to-GFP conversion (**Figure 2F**). In turn, we utilized BFP-to-GFP to evaluate various commercially available transfection reagents to optimize RNP delivery for base editing applications. From this analysis, we were able to determine that Lipofectamine^{TM}2000 allowed for a greater than 4-fold increase in genomic base editing efficiency compared to other commercially available reagents such as Lipofectamine^{TM} and CRISPRMAX (**Figure 2F**). Despite this, RNP-driven delivery was about 4-fold less efficient in genomic base editing compared to plasmid delivery. Thus, for the remainder of this study we proceeded with plasmid delivery of base editors. Nonetheless, this collective data demonstrates that BFP-to-GFP conversion correlates to base editing efficiency at genomic loci. Moreover, this approach allows for the facile and systematic optimization of base editing in human cells using plasmid- and RNP-based approaches.

Development of transient reporter for editing enrichment (TREE) to identify and efficiently isolate base-edited cell populations

Conventional base editing approaches that use reporters of transfection (herein abbreviated as RoT) only report on the efficiency of plasmid delivery to a cell but not directly on

the efficiency of base editing within these cells. As such, we hypothesized that we could employ BFP-to-GFP conversion, which directly correlates to base editing activity within a cell, as a transient reporter for editing enrichment (TREE) to allow for the identification and enrichment of cells in which targeted genomic base editing had occurred. To facilitate this, we engineered a dual-targeting sgRNA (pDT-sgRNA) vector that contains both sg(BG) and a sgRNA for a genomic target site [sg(TS)] (**Figure 3A**). Moreover, the pDT-sgRNA vector was designed to allow for the facile cloning of new target sites via BbsI restriction enzyme digestion and ligation of sg(TS) oligonucleotides (**Figure 3A**). Accordingly, we designed pDT-sgRNA vectors with sequences targeting three genomic locations (Sites 1-3). To utilize TREE for enrichment of cells that have been edited at specific loci, we co-transfected these pDT-sgRNA vectors with pEF-BFP and pCMV-BE4-Gam into HEK293 cells using the optimized base editing parameters identified using the BFP-to-GFP conversion assay (**Figure 3B**). Flow cytometry was then used to isolate GFP-positive and -negative cells. For comparison, we used a conventional RoT as a strategy to enrich for edited cell populations (**Figure 3C**). Specifically, after co-transfecting HEK293 cells with pEF-GFP and sg(TS) plasmids, we used flow cytometry to sort for GFP-positive and -negative cell populations. Flow cytometry analysis of cells in which TREE was applied confirmed the presence of BFP and GFP-positive cell populations indicative of active base editing (**Figure 3D**). Importantly, in these cell populations there was also a significant percentage of cells that were BFP-positive but GFP-negative, suggesting that isolating cell populations exclusively based upon a reporter of transfection would significantly limit the enrichment of edited cells. To confirm this, we performed Sanger sequencing of the targeted genomic sites in GFP-positive, GFP-negative, and unsorted cell populations isolated from TREE and RoT approaches (Figure 3-3E and Appendix A Figure A-1). As expected, GFP-positive cells isolated using both TREE- and RoT-based strategies were enriched for edited cells when compared to GFP-negative and unsorted cell populations. We found that base editing efficiency at these three target loci in HEK293 cells using RoT-based approaches was similar to those reported previously (Appendix A Table A-5) (15). Importantly, this analysis also revealed across all three targeted sites that GFP-positive cells

isolated via TREE had a statistically significant higher frequency of base editing than GFP-positive cells isolated using traditional RoT approaches.

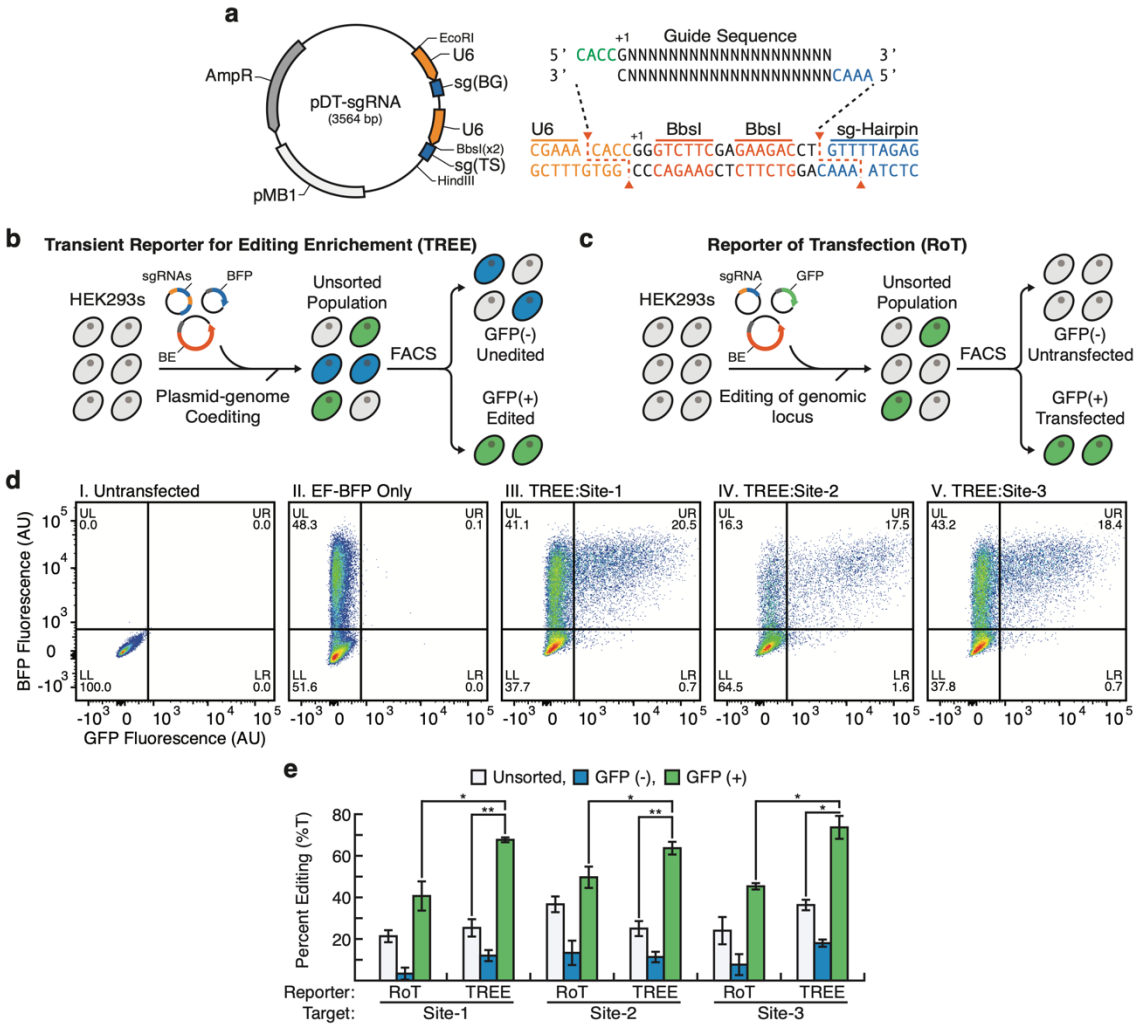


Figure 3-3. Enrichment of base-edited cell populations using transient reporter of editing efficiency (TREE). **(a)** Plasmid map of pDT-sgRNA vector that contains sg(BG) and sg(TS). Expression for both sgRNA cassettes is driven by separate U6 promoters (orange arrows). The BbsI restriction sites allow for direct restriction enzyme-based cloning of new target sites. **(b)** Schematic for enrichment of edited cells using TREE. HEK293 cells are co-transfected with pEF-BFP, pCMV-BE4-Gam, and pDT-sgRNA vectors. After 48 hours post-transfection, flow cytometry is used to sort cell populations into GFP-positive and –negative fractions. **(c)** Schematic for enrichment of edited cells using reporter of transfection (RoT). HEK293 cells are co-transfected with pEF-GFP, pCMV-BE4-Gam, and sg(TS) vectors. After 48 hours post-transfection, flow cytometry is used to sort cell populations into GFP-positive and –negative fractions. **(d)** Representative flow cytometry plots of (i) untransfected HEK293 cells and (ii) HEK293 cells transfected with pEF-BFP only as well as HEK293 cells in which TREE was applied targeting (iii) Site-1, (iv) Site-2, and (v) Site-3. **(e)** Quantification of base editing efficiency at Site-1, Site-2, and Site-3 in GFP-positive, GFP-negative, and unsorted cell populations isolated using TREE- or RoT-based enrichment strategies. n = 3; * = p<0.05, ** = p<0.01.

Because of the success of targeting these loci, we investigated if TREE could be utilized to target additional genomic sites that display very low editing efficiency when traditional RoT approaches are applied. One such example is the *APOE* locus, a well-established risk factor associated with altered probability of sporadic Alzheimer's disease onset (Hauser & Ryan, 2013). Human *APOE* has three major isoforms, ApoE2, ApoE3, and ApoE4, which differ by two amino acid substitutions at positions 112 and 158 in exon 4—ApoE2 (Cys112, Cys158), ApoE3 (Cys112, Arg158), ApoE4 (Arg112, Arg158). Attempts to use base editing to convert ApoE3 to ApoE2 by targeting the APOE(R158) locus revealed undetectable levels of editing in unsorted cell populations despite similar transfection efficiencies when other genomic sites (Sites 1-3) were targeted (Appendix A Figure A-2A). In addition, our attempts to use RoT-based methods in HEK293 cells to convert ApoE3 to ApoE2 by targeting the APOE(R158) locus revealed very low levels of editing in GFP+ isolated cells (Appendix A Figure A-2B), further establishing the APOE(R158) locus as recalcitrant to genomic editing. We then used TREE-based methods to edit this same loci in HEK293 cells by co-transfecting pEF-BFP, pCMV-BE4-Gam, and pDT-sgRNA with a sg(TS) targeting the APOE(R158) locus. As expected, flow cytometry analysis demonstrated that the transfection efficiency when TREE was used to target the APOE(R158) locus was similar to when TREE was used to target other genomic sites (Appendix A Figure A-2C). In addition, despite these similarities in transfection efficiencies, there was no detectable editing in the unsorted cell populations using TREE to target the APOE(R158) locus, thereby confirming the difficulty in editing this genomic location (Appendix A Figure A-2D). However, unlike in GFP-positive isolated using RoT methods, GFP-positive cells purified using TREE methods displayed a high level of base editing at the APOE(R158) locus (Appendix A Figure A-2E). Together, these results demonstrate that TREE can not only provide for a higher level of enrichment of base-edited cell populations compared to conventional RoT strategies but also can allow for isolation of base-edited cells at genomic loci that were not previously achievable with traditional RoT approaches.

Finally, we wanted to confirm that the fluorescent signal associated with cells isolated by TREE was transient. To that end, we measured the long-term fluorescence of GFP-positive cells

purified after TREE-based editing (Appendix A Figure A-3A). Notably, analysis of these cells by fluorescent microscopy (Appendix A Figure A-3B) and flow cytometry (Appendix A Figure A-3C) revealed no long-term detectable GFP signal, verifying that the TREE fluorescent output is indeed transient in nature.

Multiplex base-editing using TREE

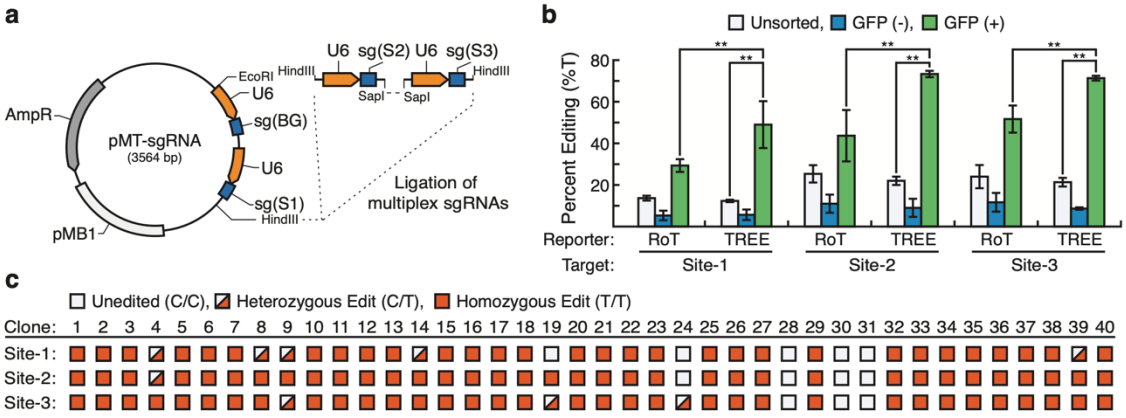


Figure 3-4. TREE allows for multiplex base editing. **(a)** Plasmid map of pMT-sgRNA vector that contains sg(BG) in addition to sgRNA for multiple target sites. Expression for all sgRNA cassettes is driven by separate U6 promoters (orange arrows). The HindIII restriction site allows for additional sgRNAs for target sites to be cloned in through restriction enzyme-based cloning. **(b)** Quantification of multiplex base editing efficiency at Site-1, Site-2, and Site-3 in GFP-positive, GFP-negative, and unsorted cell populations using TREE- or RoT-based enrichment strategies. n=3; * = p<0.05, ** = p<0.01. (**c**) Clonal analysis of editing at multiple genomic loci using TREE. 40 GFP-positive clones were isolated via single-cell sorting. Editing was detected via PCR and Sanger sequencing. Blank icon indicates no editing observed, half-red icon indicates heterozygous C and T at the target site, and solid red icon indicates homozygous T edits at the genomic site.

We further investigated if TREE could be utilized in conjunction with multiplexed genome engineering strategies. To accomplish this, we generated a multi-targeted vector (pMT-sgRNA) that contains sg(BG) as well as sgRNA for genomic targets Sites 1-3 (Figure 3-4A). In a similar manner to when TREE was employed to target a single locus, we utilized TREE to simultaneously target multiple genomic sites by co-transfecting HEK293 cells with pMT-sgRNA, pEF-BFP, and pCMV-BE4-Gam. In parallel, we used a RoT-based approach by co-transfecting HEK293 cells with pMT-sgRNA, pEF-GFP, and pCMV-BE4-Gam. After 48 hours, GFP-negative and GFP-positive cells were isolated using flow cytometry (Appendix A Figure A-4A). Along similar lines to single locus targeting, Sanger sequencing of the multiplex targeted genomic sites

in GFP-positive cell populations isolated from TREE and RoT approaches revealed that TREE allowed for statistically significant higher frequency of base editing than RoT approaches (Figure 3-4B and Appendix A Figure A-4B). Importantly, this analysis revealed that there was no statistically significant difference in editing efficiency when TREE was used to target these sites individually or a multiplexed manner (Appendix A Figure A-4C). Finally, we wanted to determine if TREE increased the likelihood of C-to-T conversions at off-target loci. Therefore, in GFP-positive cell populations isolated from TREE and RoT approaches we PCR-amplified and Sanger sequenced the top predicted off-target loci for the sgRNA sequences used for multiplexed editing. Overall, this analysis revealed there was no distinguishable increase in C-to-T conversions in either GFP-positive cells isolated with TREE- or RoT-based strategies when compared to that of untransfected cells (Appendix A Figure A-5).

Sanger sequencing that was performed on bulk sorted GFP-positive cells suggested that multiplex editing in conjunction with TREE could result in multiplexed editing in the same cell. To confirm that this indeed occurred, we again used our multi-targeting vector (pMT-sgRNA) in conjunction with TREE to simultaneously target genomic Sites 1-3 in HEK293 cells. We then sorted single GFP-positive cells into a 96 well plate. After expansion, Sanger sequencing of the multiplexed genomic sites was performed on a total of 40 clones. This analysis revealed that 36 out of the 40 clones had base editing at more than one genomic site (Figure 3-4C). Remarkably, this analysis revealed that almost 80% of the isolated clones (31 out of 40) had biallelic conversions at all three genomic loci.

One of the caveats of all base-editing approaches, regardless of whether or not RoT- or TREE-based enrichment strategies are employed, is that base editors can potentially edit non-target Cs that are located in an 6 nucleotide window (termed the editing window) within the protospacer (Tan et al., 2019). Consequently, this could potentially limit the application of base editing approaches in which conversion of non-target Cs result in a non-silent mutation or other phenotypic changes. To that end, we wanted to determine if any of our clones contained edits exclusively at the target C and not any other Cs within the editing window. Indeed, we identified a number of clones in which at genomic Site 2 and Site 3 modification only occurred at the target C

(Appendix A Figure A-6). Interestingly, we did not identify any clones in which at genomic Site 1 such exclusive modification of the target C occurred. We speculate that because another C occurs immediately adjacent to this target C, that such exclusive modification will require the use of recently published site-specific editors that allow for single nucleotide changes free from off-targeting conversions within the editing window (Tan et al., 2019).

TREE allows for highly efficient editing in human pluripotent stem cells (hPSCs)

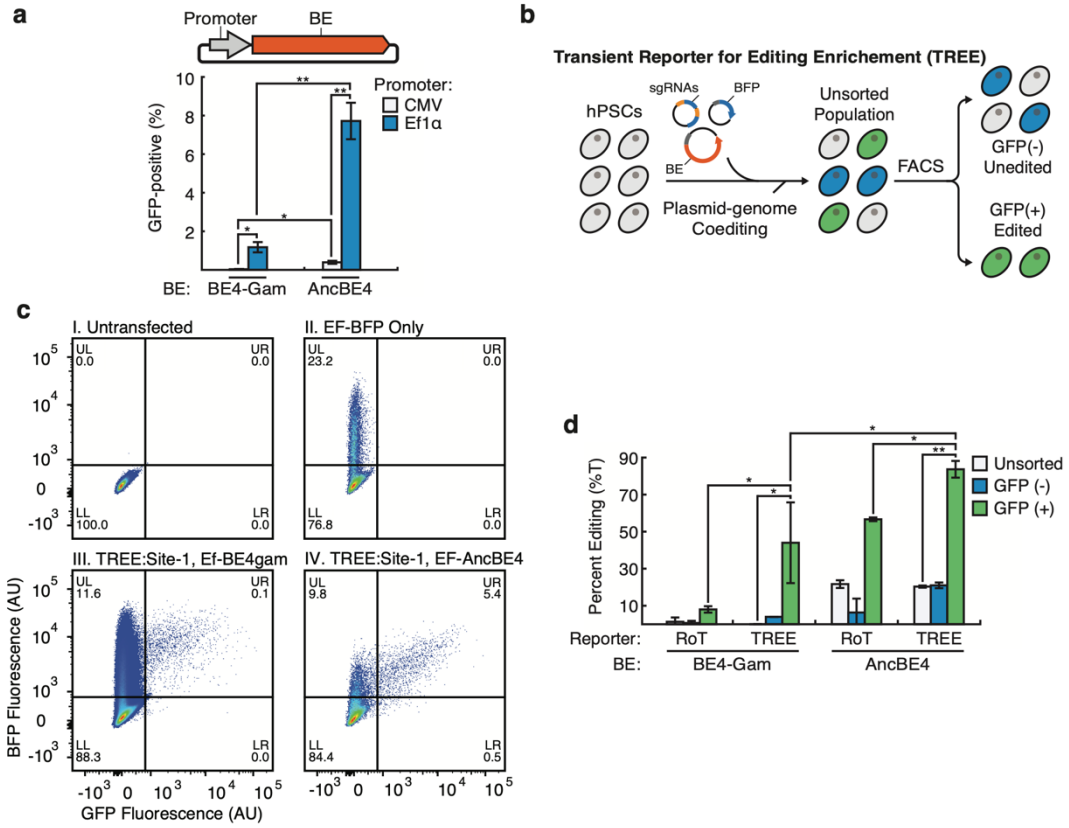


Figure 3-5. Highly efficient editing in human pluripotent stem cells (hPSCs) using TREE-based methods. **(a)** Quantification of base editing efficiency (percentage GFP-positive cells) when hPSCs were co-transfected with pEF-BFP, sg(BG), and various base editing vectors. n=3; * = p<0.05, ** = p<0.01. **(b)** Schematic for enrichment of edited hPSC using TREE. HPSCs were co-transfected with pEF-BFP, pEF-BE4-Gam / pEF-AncBE4, and pDT-sgRNA vectors. After 48 hours post-transfection, flow cytometry was used to sort cell populations into GFP-positive and –negative fractions. **(c)** Representative flow cytometry plots of (i) untransfected hPSCs cells and (ii) hPSCs transfected with pEF-BFP only as well as hPSCs cells in which TREE was applied targeting Site-1 utilizing (iii) pEF-BE4-Gam or (iv) pEF-AncBE4. **(d)** Quantification of base editing efficiency at Site-1 in GFP-positive, GFP-negative, and unsorted cell populations isolated using TREE- or RoT-based enrichment strategies in which pEF-BE-Gam or pEF-AncBe4 was employed. n = 3; * = p<0.05, ** = p<0.01.

Single base pair modification in hPSCs via CRISPR/Cas9-induced DSB followed by HDR suffers from low efficiencies (5-9). In addition, to date genomic modification of hPSCs using deaminase-based DNA base editor has yet to be reported. Therefore, we wanted to investigate if TREE could be utilized to efficiently edit specific loci in hPSCs. Hence, we co-transfected pEF-BFP and pCMV-BE4-Gam into hPSCs using a transfection reagent (Lipofectamine™ Stem) that had been previously used by others for the efficient delivery of Cas9-related plasmids to hPSCs (Giacalone et al., 2018; Hauser & Ryan, 2013). Surprisingly, we did not observe any GFP-positive cells in these cell populations (Figure 3-5A and Appendix A Figure A-7A). In addition, we observed very few GFP-positive cells when a more recently published, higher efficiency base editor was used (16; Appendix A Figure A-7A). Because previous reports have suggested that the CMV promoter is inefficient for transgene expression in pluripotent stem cells (Chung et al., 2002; Norrman et al., 2010; XIA et al., 2007), we replaced the promoter driving base editor expression with EF1α. Indeed, when co-transfected hPSCs with pEF-BE4-Gam or pEF-AncBE4 as well as pEF-BFP and sg(BG), a significant number of GFP-positive cells were observed (Figure 3-5A). Using the pEF-AncBE4 vector, we also optimized editing efficiency in hPSCs by using a range of base editor amount at varying ratio with sg(BG) (Appendix A Figure A-7B). Similar to our optimization experiments with HEK293 cells, this analysis revealed that base editing efficiencies were significantly affected by these parameters. Interestingly, the most optimal parameters in hPSCs differed from those identified in HEK293 cells (Figure 3-2D) highlighting the utility of this assay to evaluate these variables. Using these optimized base editing vector designs, we applied TREE to target a genomic loci in hPSCs by co-transfecting pEF-BE-Gam/pEF-AncBE4, pEF-BFP, and pDT-sgRNA (with a sg(TS) targeting site 1) (Figure 3-5B). In turn, flow cytometry was used to isolate GFP-positive and –negative cell populations (Figure 3-5C). Subsequently, Sanger sequencing was performed on the targeted genomic site in GFP-positive, GFP-negative, and unsorted cell populations isolated from TREE and RoT approaches in which pEF-BE4-Gam and pEF-AncBE4 was used (Appendix A Figure A-7C). This analysis demonstrated that GFP-positive hPSCs isolated via TREE had a statistically significant higher frequency of base editing than GFP-positive hPSCs isolated using traditional RoT approaches (Figure 3-5D). In addition, TREE

employed with the pEF-AncBE4 vector allowed for the efficient modification of the difficult to edit APOE(R158) locus (Appendix A Figure A-7D and A-7E).

Similar to our work with HEK293 cells, we wanted to confirm that the fluorescent output of TREE was transient in nature. In that regard, we measured the fluorescence of GFP-positive hPSCs isolated after TREE-based editing. Flow cytometry analysis revealed that after 2 weeks of culture there was no detectable GFP signal (Appendix A Figure A-8), demonstrating that the fluorescent signal associated with hPSCs purified by TREE was transient.

Collectively, although this data demonstrates that TREE can be utilized for the efficient base editing of hPSCs, one of the caveats of all base editing approaches is the C-to-T conversion of non-target Cs within the editing window (Y. B. Kim et al., 2017; Tan et al., 2019). Indeed, the Sanger sequencing analysis of GFP-positive populations isolated from TREE revealed editing of such non-target Cs when either Site 1 (Appendix A Figure A-7C) or the APOE(R158) (Appendix A Figure A-7E) locus was targeted in hPSCs. As such, to determine whether TREE allowed allelic outcomes in which targeting only occurred at the desired C, we performed next-generation deep sequencing (NGS) of PCR amplicons of Site 1 and APOR(R158) in GFP-positive cells purified using TREE or RoT methods. Indeed, this analysis revealed at both these loci allelic outcomes in which base editing occurred exclusively at the target C, free from confounding C-to-T conversions at other sites within the targeting window (Appendix A Figure A-9**)**. In sum, this data demonstrates the broad utility of TREE to allow for the efficient editing in hPSCs.

DISCUSSION

Since the first deaminase base editor was developed by Komor et al. in 2016 (Komor et al., 2016), multiple additional base-editing technologies have been rapidly developed with various endonucleases, deaminases, targeting windows, and PAM specificities (Eid et al., 2018). Application of these emerging base editors to new cell types requires a slow, iterative process in which various base editing parameters are tested and editing efficiency is assessed through downstream sequencing methods. Additionally, as we demonstrate, transfection efficiency does

not precisely correlate with editing efficiency, so reporters of transfection do not provide accurate information about the efficacy of various base editing strategies. Here, we describe how BFP-to-GFP conversion and TREE can be utilized to rapidly optimize various factors that influence base editing efficiency, including base editor plasmid concentration and design as well as base editor to sgRNA ratios. In fact, we show that these parameters are cell line-specific, demonstrating the advantage of TREE to allow for the high-throughput evaluation of base editing approaches. In the future, we can utilize TREE in the context of high-throughput screening to identify small molecules to further enhance base editing efficiency in a manner similar to that which has been previously achieved with CRISPR-mediated HDR approaches (Riesenberg & Maricic, 2018; Yu et al., 2015).

It has been shown that CRISPR/Cas9 genome engineering is compatible with a variety of delivery methods (e.g. lipid-mediated transfection, electroporation) and expression systems (e.g. plasmid DNA, Cas9-gRNA ribonucleoprotein complexes [RNP]), each with advantages and disadvantages that have been reviewed extensively elsewhere (Lino et al., 2018; Liu et al., 2017). In this study, we employed lipid-based delivery reagents that have been previously employed by others for the CRISPR/Cas9-based editing of HEK293 cells (Lipofectamine 3000; (X. Liang et al., 2015)) and hPSCs (Lipofectamine Stem; (Giacalone et al., 2018; Maguire et al., 2019)). Given TREE's ease of use and readily detectable fluorescent output we anticipate that TREE can be employed with whatever transfection method that is preferred by the end user. For instance, we demonstrated that our base editing assay was compatible with both plasmid and RNP approaches. Although we observed that the overall genomic base editing efficiency of RNP-based expression was lower than that of lipid-based expression, we provide proof-of-principle that TREE can be employed in future applications where the advantages of RNPs are desirable.

One potential limitation of the use of the plasmid DNA expression systems in the context of TREE approaches is random integration of all or part of the plasmid DNA into the genome of targeted cells. It should be noted that it has been reported by others that the stable integration of circular plasmid DNA into the host genome is infrequent, especially for cells such as hPSCs where it has been reported on the order of 1 per 1 x 10^5 cells (Eiges et al., 2001;

Haridhasapavalan et al., 2019; Moore et al., 2010). Indeed, as it relates to potential integration of the pEF-BFP plasmid, we demonstrate that the fluorescent output of TREE is transient in both HEK293 cells and hPSCs, suggesting that this plasmid does not integrate into the genome. As it relates to the integration of the base editing and sgRNA plasmids, it has been shown by others in CRISPR/Cas9 genome engineering that the Cas9 and sgRNA plasmids can be integrated at on- and off-target sites (S. Kim et al., 2014). However, we speculate that because base editors do not introduce DSBs the integration of these plasmids into the genome would be infrequent. In fact, we did not observe any integration of these plasmids when Sanger sequencing or NGS was performed at the on- or off-target sites. Moving forward, undesirable insertions of plasmid DNA sequences at target sites can be detected using PCR-based methods followed by Sanger sequencing or NGS of the resultant amplicons. On the other hand, similar insertions at off-target or random genomic sites are difficult to detect and will require the use of more comprehensive techniques such as whole genome sequencing.

Human cell models are critical for elucidating the mechanisms of disease progression as well as identifying and testing potential therapeutic interventions. Because a high percentage of human diseases are due to single nucleotide polymorphisms (SNPs)(Ganesan et al., 2019), base editors can allow for the precise engineering of *in vitro* models of human disease. Here we provide proof-of-principle that TREE can be employed to edit disease-relevant loci. Specifically, we demonstrate that TREE enables for the enrichment of cells that had been edited at the APOE(R158) locus, a gene associated with altered risk of Alzheimer's disease onset (Hauser & Ryan, 2013). Notably, conventional RoT-based methods did not allow for significant enrichment of edited cells at this same refractory locus. In addition, because many human diseases are multigenetic disorders that are a result of complex gene interactions, we also investigated the ability of TREE to be utilized in multiplexed genome engineering applications. By using a multi-targeted vector, we demonstrated that compared to RoT-based methods TREE resulted in a significantly higher level of cells enriched for simultaneous editing at multiple independent loci. In fact, we demonstrated that through analysis of single cell clones that 90% of the clones had simultaneous base editing at more than one genomic site and almost 80% of the clones had

biallelic conversions at all three targeted loci. In this vein, TREE provides a highly efficient method for generating cell-based models of multigenic diseases.

Many immortalized cell lines, such as HEK293s, are aneuploid with unknown mutations and dosage at key disease-relevant genes. Alternatively, hPSCs, which have a normal euploid karyotype and the potential to differentiate into all cell types of the mature adult body, represent an attractive alternative to immortalized cell lines for disease modeling and drug screening applications (Benam et al., 2015; Horvath et al., 2016; Niu & Wang, 2015; Xu & Zhong, 2013). In particular, the ability to use gene editing technologies to generate isogenic hPSC lines that differ only with respect to disease mutations has great potential as it relates to precisely defining genotype to phenotype relationships (Yu et al., 2015). The RNA-guided CRISPR-Cas9 system has the potential to allow for precise genetic modifications in hPSCs through the introduction of site-specific DSBs. Although previous reports demonstrate that introduction of DSB via CRISPR/Cas9 significantly improves the ability to obtain knock out cell lines from hPSCs by the NHEJ pathway (Grobarczyk et al., 2015), single base modification using CRISPR/Cas9-induced DSB followed by HDR is extremely inefficient (1-2% of sequenced colonies in which one allele is targeted and <1% where both alleles are targeted; (Grobarczyk et al., 2015; Huang et al., 2015; Li et al., 2015; Miyaoka et al., 2014; Reinhardt et al., 2013). Recently, it has been reported that co-delivery of Cas9, sgRNA, and a puromycin selection cassette followed by transient puromycin selection can increase the HDR-mediated genome engineering in hPSCs (Steyer et al., 2018; Supharattanasitthi et al., 2019) However, these strategies rely on the introduction of DSBs, which in pluripotent stem cells can lead to large deletions and complex chromosomal rearrangements (Kosicki et al., 2018), significant cytotoxicity (Haapaniemi et al., 2018), and increased acquisition of p53 mutations (Ihry et al., 2018). In addition, it has been shown that the use of antibiotic selection, even in a transient manner, may lead to the selection of hPSCs, with chromosomal abnormalities (G. Liang & Zhang, 2013; Omole & Fakoya, 2018). Yet, to our knowledge, base editors, which do not have these same limitations as CRISPR/Cas9-induced DSB followed by HDR, have not previously been used with hPSCs. In fact, our initial attempts to apply base editors in the context of both RoT- and TREE-based approaches with hPSCs did not allow for observable

modification of target loci. Instead, by replacing the standard CMV promoter in the base editing plasmids with an EF1α promoter, we were able to achieve modification of genomic sites using both RoT- and TREE-centered methods. However, TREE allowed for significantly higher enrichment of edited hPSCs when compared to RoT isolation strategies. We contend that the use of TREE with hPSCs will significantly advance the use of these cells in disease modeling, drug screening, and cell-based therapies.

Despite their tremendous potential in a variety of downstream applications, base editing approaches have a few of caveats that should be noted, regardless of whether RoT- or TREE-based enrichment strategies are employed. First, as is the case with all Cas9-directed genome editing approaches, is the potential for genome modification at off-target loci (Kuscu et al., 2014; Tsai et al., 2015). In this work, GFP-positive cells isolated via TREE did not display untargeted C-to-T conversions at the off-target genomic loci examined. Recently, it has been reported that base editors can induce site-specific inosine formation on RNA (Zhou et al., 2019). Accordingly, in the future, the effect of TREE-based approaches on unwanted RNA modifications should be examined. Another limitation of base editing methods is modification of additional C nucleotides that are in close proximity to the target C (Tan et al., 2019). In fact, some base editors can cause C-to-T conversions at any Cs in up to a 9 nucleotide window within the protospacer (Komor et al., 2016; Nishida et al., 2016; Tan et al., 2019; Zong et al., 2017). Such C-to-T modifications could be especially problematic if they result in amino acid alterations during translation, induce epigenetic changes, or cause other phenotypic changes in targeted cells. To that end, through clonal isolation and next generation sequencing (NGS) analysis we identified that such exclusive modifications of the target C were achieved in both edited HEK293 cells or hPSCs that were enriched using TREE-based methods. It should be noted, though, that at genomic Site-1, where a C lies adjacent to the target C, allelic outcomes in which modification only occurred at the target C were rare events. Moving forward, modified base editors that have a narrow editing window (Y. B. Kim et al., 2017; Tan et al., 2019) could be easily employed with TREE to target such genomic loci that contain multiple Cs in close proximity to the target C.

In summary, we demonstrate that TREE allows not only for the optimization of base editing strategies in the context of variety of cell types and genomic locations but also the enrichment for cell populations to be utilized in variety of downstream applications. With the rate at which the genome editing field has been progressing over the past few years, TREE is a readily adoptable method that will expedite and improve tractability of single-nucleotide genome engineering methods.

REFERENCES

Benam, K. H., Dauth, S., Hassell, B., Herland, A., Jain, A., Jang, K.-J., Karalis, K., Kim, H. J., MacQueen, L., Mahmoodian, R., Musah, S., Torisawa, Y., van der Meer, A. D., Villenave, R., Yadid, M., Parker, K. K., & Ingber, D. E. (2015). Engineered in vitro disease models. *Annual Review of Pathology*, *10*, 195–262. https://doi.org/10.1146/annurev-pathol-012414-040418

Brookhouser, N., Raman, S., Potts, C., & Brafman, David. A. (2017). May I Cut in? Gene Editing Approaches in Human Induced Pluripotent Stem Cells. *Cells*, *6*(1), 5. https://doi.org/10.3390/cells6010005

Chung, S., Andersson, T., Sonntag, K.-C., Björklund, L., Isacson, O., & Kim, K.-S. (2002). Analysis of different promoter systems for efficient transgene expression in mouse embryonic stem cell lines. *Stem Cells (Dayton, Ohio)*, *20*(2), 139–145. https://doi.org/10.1634/stemcells.20-2-139

Eid, A., Alshareef, S., & Mahfouz, M. M. (2018). CRISPR base editors: Genome editing without double-stranded breaks. *The Biochemical Journal*, *475*(11), 1955–1964. https://doi.org/10.1042/BCJ20170793

Eiges, R., Schuldiner, M., Drukker, M., Yanuka, O., Itskovitz-Eldor, J., & Benvenisty, N. (2001). Establishment of human embryonic stem cell-transfected clones carrying a marker for undifferentiated cells. *Current Biology: CB*, *11*(7), 514–518. https://doi.org/10.1016/s0960-9822(01)00144-0

Ganesan, K., Kulandaisamy, A., Priya, S. B., & Gromiha, M. M. (2019). HuVarBase: A human variant database with comprehensive information at gene and protein levels. *PLOS ONE*, *14*(1), e0210475. https://doi.org/10.1371/journal.pone.0210475

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*, *551*(7681), 464–471. https://doi.org/10.1038/nature24644

Gehrke, J. M., Cervantes, O., Clement, M. K., Wu, Y., Zeng, J., Bauer, D. E., Pinello, L., & Joung, J. K. (2018). An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nature Biotechnology*, *36*(10), 977–982. https://doi.org/10.1038/nbt.4199

Germini, D., Tsfasman, T., Zakharova, V. V., Sjakste, N., Lipinski, M., & Vassetzky, Y. (2018). A Comparison of Techniques to Evaluate the Effectiveness of Genome Editing. *Trends in Biotechnology*, *36*(2), 147–159. https://doi.org/10.1016/j.tibtech.2017.10.008

Giacalone, J. C., Sharma, T. P., Burnight, E. R., Fingert, J. F., Mullins, R. F., Stone, E. M., & Tucker, B. A. (2018). CRISPR-Cas9-Based Genome Editing of Human Induced Pluripotent Stem Cells. *Current Protocols in Stem Cell Biology*, *44*, 5B.7.1-5B.7.22. https://doi.org/10.1002/cpsc.46

Glaser, A., McColl, B., & Vadolas, J. (2016). GFP to BFP Conversion: A Versatile Assay for the Quantification of CRISPR/Cas9-mediated Genome Editing. *Molecular Therapy. Nucleic Acids*, *5*(7), e334. https://doi.org/10.1038/mtna.2016.48

Grobarczyk, B., Franco, B., Hanon, K., & Malgrange, B. (2015). Generation of Isogenic Human iPS Cell Line Precisely Corrected by Genome Editing Using the CRISPR/Cas9 System. *Stem Cell Reviews and Reports*, *11*(5), 774–787. https://doi.org/10.1007/s12015-015-9600-1

Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., & Taipale, J. (2018). CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nature Medicine*, *24*(7), 927–930. https://doi.org/10.1038/s41591-018-0049-z

Haridhasapavalan, K. K., Borgohain, M. P., Dey, C., Saha, B., Narayan, G., Kumar, S., & Thummer, R. P. (2019). An insight into non-integrative gene delivery approaches to generate transgene-free induced pluripotent stem cells. *Gene*, *686*, 146–159. https://doi.org/10.1016/j.gene.2018.11.069

Hauser, P. S., & Ryan, R. O. (2013). Impact of apolipoprotein E on Alzheimer's disease. *Current Alzheimer Research*, *10*(8), 809–817. https://doi.org/10.2174/15672050113109990156

Heim, R., Prasher, D. C., & Tsien, R. Y. (1994). Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proceedings of the National Academy of Sciences*, *91*(26), 12501–12504. https://doi.org/10.1073/pnas.91.26.12501

Horvath, P., Aulner, N., Bickle, M., Davies, A. M., Nery, E. D., Ebner, D., Montoya, M. C., Östling, P., Pietiäinen, V., Price, L. S., Shorte, S. L., Turcatti, G., von Schantz, C., & Carragher, N. O. (2016). Screening out irrelevant cell-based models of disease. *Nature Reviews Drug Discovery*, *15*(11), Article 11. https://doi.org/10.1038/nrd.2016.175

Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*, *157*(6), 1262–1278. https://doi.org/10.1016/j.cell.2014.05.010

Huang, X., Wang, Y., Yan, W., Smith, C., Ye, Z., Wang, J., Gao, Y., Mendelsohn, L., & Cheng, L. (2015). Production of Gene-Corrected Adult Beta Globin Protein in Human Erythrocytes Differentiated from Patient iPSCs After Genome Editing of the Sickle Point Mutation. *Stem Cells (Dayton, Ohio)*, *33*(5), 1470–1479. https://doi.org/10.1002/stem.1969

Ihry, R. J., Worringer, K. A., Salick, M. R., Frias, E., Ho, D., Theriault, K., Kommineni, S., Chen, J., Sondey, M., Ye, C., Randhawa, R., Kulkarni, T., Yang, Z., McAllister, G., Russ, C., Reece-Hoyes, J., Forrester, W., Hoffman, G. R., Dolmetsch, R., & Kaykas, A. (2018). P53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nature Medicine*, *24*(7), 939–946. https://doi.org/10.1038/s41591-018-0050-6

Kim, S., Kim, D., Cho, S. W., Kim, J., & Kim, J.-S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Research*, *24*(6), 1012–1019. https://doi.org/10.1101/gr.171322.113

Kim, Y. B., Komor, A. C., Levy, J. M., Packer, M. S., Zhao, K. T., & Liu, D. R. (2017). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nature Biotechnology*, *35*(4), 371–376. https://doi.org/10.1038/nbt.3803

Kluesner, M. G., Nedveck, D. A., Lahr, W. S., Garbe, J. R., Abrahante, J. E., Webber, B. R., & Moriarity, B. S. (2018). EditR: A Method to Quantify Base Editing from Sanger Sequencing. *The CRISPR Journal*, *1*(3), 239–250. https://doi.org/10.1089/crispr.2018.0014

Koblan, L. W., Doman, J. L., Wilson, C., Levy, J. M., Tay, T., Newby, G. A., Maianti, J. P., Raguram, A., & Liu, D. R. (2018). Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nature Biotechnology*, *36*(9), 843–846. https://doi.org/10.1038/nbt.4172

Komor, A. C., Badran, A. H., & Liu, D. R. (2017). CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell*, *168*(1–2), 20–36. https://doi.org/10.1016/j.cell.2016.10.044

Komor, A. C., Badran, A. H., & Liu, D. R. (2018). Editing the Genome Without Double-Stranded DNA Breaks. *ACS Chemical Biology*, *13*(2), 383–388. https://doi.org/10.1021/acschembio.7b00710

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, *533*(7603), 420–424. https://doi.org/10.1038/nature17946

Kosicki, M., Tomberg, K., & Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nature Biotechnology*, *36*(8), 765–771. https://doi.org/10.1038/nbt.4192

Kouranova, E., Forbes, K., Zhao, G., Warren, J., Bartels, A., Wu, Y., & Cui, X. (2016). CRISPRs for Optimal Targeting: Delivery of CRISPR Components as DNA, RNA, and Protein into Cultured Cells and Single-Cell Embryos. *Human Gene Therapy*, *27*(6), 464–475. https://doi.org/10.1089/hum.2016.009

Kuscu, C., Arslan, S., Singh, R., Thorpe, J., & Adli, M. (2014). Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature Biotechnology*, *32*(7), 677–683. https://doi.org/10.1038/nbt.2916

Li, H. L., Fujimoto, N., Sasakawa, N., Shirai, S., Ohkame, T., Sakuma, T., Tanaka, M., Amano, N., Watanabe, A., Sakurai, H., Yamamoto, T., Yamanaka, S., & Hotta, A. (2015). Precise correction of the dystrophin gene in duchenne muscular dystrophy patient induced pluripotent stem cells by TALEN and CRISPR-Cas9. *Stem Cell Reports*, *4*(1), 143–154. https://doi.org/10.1016/j.stemcr.2014.10.013

Liang, G., & Zhang, Y. (2013). Genetic and epigenetic variations in iPSCs: Potential causes and implications for application. *Cell Stem Cell*, *13*(2), 149–159. https://doi.org/10.1016/j.stem.2013.07.001

Liang, X., Potter, J., Kumar, S., Zou, Y., Quintanilla, R., Sridharan, M., Carte, J., Chen, W., Roark, N., Ranganathan, S., Ravinder, N., & Chesnut, J. D. (2015). Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *Journal of Biotechnology*, *208*, 44–53. https://doi.org/10.1016/j.jbiotec.2015.04.024

Lino, C. A., Harper, J. C., Carney, J. P., & Timlin, J. A. (2018). Delivering CRISPR: A review of the challenges and approaches. *Drug Delivery*, *25*(1), 1234–1257. https://doi.org/10.1080/10717544.2018.1474964

Liu, C., Zhang, L., Liu, H., & Cheng, K. (2017). Delivery Strategies of the CRISPR-Cas9 Gene-Editing System for Therapeutic Applications. *Journal of Controlled Release : Official Journal of the Controlled Release Society*, *266*, 17–26. https://doi.org/10.1016/j.jconrel.2017.09.012

Maguire, J. A., Cardenas-Diaz, F. L., Gadue, P., & French, D. L. (2019). Highly Efficient CRISPR-Cas9-Mediated Genome Editing in Human Pluripotent Stem Cells. *Current Protocols in Stem Cell Biology*, *48*(1), e64. https://doi.org/10.1002/cpsc.64

Miyaoka, Y., Chan, A. H., Judge, L. M., Yoo, J., Huang, M., Nguyen, T. D., Lizarraga, P. P., So, P.-L., & Conklin, B. R. (2014). Isolation of single-base genome-edited human iPS cells without antibiotic selection. *Nature Methods*, *11*(3), 291–293. https://doi.org/10.1038/nmeth.2840

Moore, J. C., Atze, K., Yeung, P. L., Toro-Ramos, A. J., Camarillo, C., Thompson, K., Ricupero, C. L., Brenneman, M. A., Cohen, R. I., & Hart, R. P. (2010). Efficient, high-throughput transfection of human embryonic stem cells. *Stem Cell Research & Therapy*, *1*(3), 23. https://doi.org/10.1186/scrt23

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K. Y., Shimatani, Z., & Kondo, A. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science*, *353*(6305), aaf8729. https://doi.org/10.1126/science.aaf8729

Niu, N., & Wang, L. (2015). In vitro human cell line models to predict clinical response to anticancer drugs. *Pharmacogenomics*, *16*(3), 273–285. https://doi.org/10.2217/pgs.14.170

Norrman, K., Fischer, Y., Bonnamy, B., Sand, F. W., Ravassard, P., & Semb, H. (2010). Quantitative Comparison of Constitutive Promoters in Human ES cells. *PLOS ONE*, *5*(8), e12413. https://doi.org/10.1371/journal.pone.0012413

Omole, A. E., & Fakoya, A. O. J. (2018). Ten years of progress and promise of induced pluripotent stem cells: Historical origins, characteristics, mechanisms, limitations, and potential applications. *PeerJ*, *6*, e4370. https://doi.org/10.7717/peerj.4370

Reinhardt, P., Schmid, B., Burbulla, L. F., Schöndorf, D. C., Wagner, L., Glatza, M., Höing, S., Hargus, G., Heck, S. A., Dhingra, A., Wu, G., Müller, S., Brockmann, K., Kluba, T., Maisel, M., Krüger, R., Berg, D., Tsytsyura, Y., Thiel, C. S., … Sterneckert, J. (2013). Genetic correction of a LRRK2 mutation in human iPSCs links parkinsonian neurodegeneration to ERK-dependent changes in gene expression. *Cell Stem Cell*, *12*(3), 354–367. https://doi.org/10.1016/j.stem.2013.01.008

Riesenberg, S., & Maricic, T. (2018). Targeting repair pathways with small molecules increases precise genome editing in pluripotent stem cells. *Nature Communications*, *9*(1), Article 1. https://doi.org/10.1038/s41467-018-04609-7

Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J., & Mateo, J. L. (2015). CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PloS One*, *10*(4), e0124633. https://doi.org/10.1371/journal.pone.0124633

Steyer, B., Bu, Q., Cory, E., Jiang, K., Duong, S., Sinha, D., Steltzer, S., Gamm, D., Chang, Q., & Saha, K. (2018). Scarless Genome Editing of Human Pluripotent Stem Cells via Transient Puromycin Selection. *Stem Cell Reports*, *10*(2), 642–654. https://doi.org/10.1016/j.stemcr.2017.12.004

Supharattanasitthi, W., Carlsson, E., Sharif, U., & Paraoan, L. (2019). CRISPR/Cas9-mediated one step bi-allelic change of genomic DNA in iPSCs and human RPE cells in vitro with dual antibiotic selection. *Scientific Reports*, *9*(1), 174. https://doi.org/10.1038/s41598-018-36740-2

Tan, J., Zhang, F., Karcher, D., & Bock, R. (2019). Engineering of high-precision base editors for site-specific single nucleotide replacement. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-018-08034-8

Tsai, S. Q., Zheng, Z., Nguyen, N. T., Liebers, M., Topkar, V. V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A. J., Le, L. P., Aryee, M. J., & Joung, J. K. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature Biotechnology*, *33*(2), 187–197. https://doi.org/10.1038/nbt.3117

Wright, A. V., Nuñez, J. K., & Doudna, J. A. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell*, *164*(1–2), 29–44. https://doi.org/10.1016/j.cell.2015.12.035

XIA, X., ZHANG, Y., ZIETH, C. R., & ZHANG, S.-C. (2007). Transgenes Delivered by Lentiviral Vector Are Suppressed in Human Embryonic Stem Cells in a Promoter-Dependent Manner. *Stem Cells and Development*, *16*(1), 167–176.

Xu, X., & Zhong, Z. (2013). Disease modeling and drug screening for neurological diseases using human induced pluripotent stem cells. *Acta Pharmacologica Sinica*, *34*(6), Article 6. https://doi.org/10.1038/aps.2013.63

Yu, C., Liu, Y., Ma, T., Liu, K., Xu, S., Zhang, Y., Liu, H., La Russa, M., Xie, M., Ding, S., & Qi, L. S. (2015). Small molecules enhance CRISPR genome editing in pluripotent stem cells. *Cell Stem Cell*, *16*(2), 142–147. https://doi.org/10.1016/j.stem.2015.01.003

Zhou, C., Sun, Y., Yan, R., Liu, Y., Zuo, E., Gu, C., Han, L., Wei, Y., Hu, X., Zeng, R., Li, Y., Zhou, H., Guo, F., & Yang, H. (2019). Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature*, *571*(7764), 275–278. https://doi.org/10.1038/s41586-019-1314-0

Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., Qiu, J.-L., Wang, D., & Gao, C. (2017). Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nature Biotechnology*, *35*(5), 438–440. https://doi.org/10.1038/nbt.3811

Zuris, J. A., Thompson, D. B., Shu, Y., Guilinger, J. P., Bessen, J. L., Hu, J. H., Maeder, M. L., Joung, J. K., Chen, Z.-Y., & Liu, D. R. (2015). Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nature Biotechnology*, *33*(1), 73–80. https://doi.org/10.1038/nbt.3081

CHAPTER 4

DESIGN AUTOMATION OF CRISPR RNAS

ABSTRACT

CRISPR-based technologies are paramount in genome engineering and synthetic biology. Prime editing (PE) is a technology capable of installing genomic edits without double-stranded DNA breaks (DSBs) or donor DNA. Prime editing guide RNAs (pegRNAs) simultaneously encode both guide and edit template sequences. They are more design intensive than CRISPR single guide RNAs (sgRNAs). As such, application of PE technology is hindered by the limited throughput of manual pegRNA design. To that end, we designed a software tool, Prime Induced Nucleotide Engineering Creator of New Edits (PINE-CONE) that enables high-throughput automated design of pegRNAs and prime editing strategies. PINE-CONE translates edit coordinates and sequences into pegRNA designs, accessory guides, and oligonucleotides for facile cloning workflows. To demonstrate PINE-CONE's utility in studying disease-relevant genotypes, we rapidly design a library of pegRNAs targeting Alzheimer's Disease single nucleotide polymorphisms (SNPs). Overall, PINE-CONE will accelerate the application of PEs in synthetic biology and biomedical research.

INTRODUCTION

Automation accelerates our ability to engineer living systems. As a result, synthetic biology has adopted design software and standardization to improve forward engineering(Bartley et al., 2019; Nielsen et al., 2016). Experimental automation increases throughput, enables expanded assembly of genetic circuits and interrogation of genetics on a scale not attainable by manual efforts(HamediRad et al., 2019; Iverson et al., 2016). CRISPR-based technologies are highly amenable to design automation and are functional in a broad range of organisms(Stemmer et al., 2015). This has made CRISPR-systems indispensible for the fields of synthetic biology and genome engineering(Standage-Beier & Wang, 2017). Canonical CRISPR technologies target DNA via inducing double stranded DNA breaks (DSBs) and are often subsequently repaired via non-homologous end joining (NHEJ), or by homology directed repair (HDR) with exogenous DNA templates. However, DSBs can induce off-target mutations, apoptosis and destabilize karyotype(Fu et al., 2013; Ihry et al., 2018; Kosicki et al., 2018). To address these shortcomings, new technologies have fused the programmability of CRISPR associated (Cas) proteins to enzymes capable of mediating DNA manipulations without DSBs including Cas9-fused recombinases, transposases and deaminases(Chaikind et al., 2016; Chen & Wang, 2019; Rees & Liu, 2018; Standage-Beier et al., 2019). Deaminase fused-Cas9 base editing (BE) technologies have enabled single base pair chromosomal editing without the introduction of deleterious DSBs. Base editing consists of cytosine base editors (CBE) which mediate the change of C-to-T (or G-to-A) and adenine base editors (ABE) with facilitate the conversion of A-to-G (or T-to-C)(Gaudelli et al., 2017; Komor et al., 2016). To date, BEs have been used to interrogate genotype-to-phenotype relationships, engineer animal model of disease, and develop cell therapies(Anzalone et al., 2020). However, base editors can only facilitate the four transition mutations, are restricted to single nucleotide modifications within the editing window, and cannot facilitate insertion or removal of nucleotides.

CRISPR-Cas9 systems have been used with reverse transcriptases to facilitate highly efficient user programmed editing. For instance, CRISPey (Retron precISe Parallel Editing via homology) enables highly efficient editing in yeast(Sharon et al., 2018). Alternatively, prime

editors (PE) are a recently developed gene editing technology that is capable of introducing all 12 possible single nucleotide changes as well as small insertions and deletions without the need for DSBs or donor DNA templates(Anzalone et al., 2019). PEs are a fusion protein composed of a nicking Cas9 mutant fused to reverse transcriptase domain (Moloney Murine Leukemia Virus Reverse Transcriptase; MMLV-RT; Figure 4-1a). The PE protein is targeted to the editing site by a prime editing guide (pegRNA) which encodes three components: (i) a guide sequence, (ii) a primer binding sequence (PBS), and (iii) a reverse transcription template (RTT), which encodes the intended edit. The pegRNA directs the PE to the target locus, where Cas9 mediates a single-stranded DNA break (SSB) on the PAM-strand. The PBS of the pegRNA then hybridizes with the 3' end of the nicked DNA strand resulting in a double-stranded DNA-RNA heteroduplex, with the edit on one strand and the wildtype (WT) sequence on the opposite strand. The nick on the 3' end of the target DNA serves as the initiation point of polymerization by MMLV-RT, with the RTT sequence used as the template (Figure 4-1a)(Anzalone et al., 2019). Endogenous DNA mismatch repair is then capable of incorporating the edit into the opposing strand resulting in the final editing product.

Various PE-based strategies have been developed, including PE1, PE2, PE3, and PE3b(Anzalone et al., 2019) (Figure 4-1b). Compared to PE1, PE2 utilizes an engineered MMLV-RT that significantly increases editing efficiency. PE3-based strategies utilize a pegRNA in combination with an accessory sgRNA targeting a SSB 40-90 base pairs (bp) downstream of the edit locus. Although PE3 results in higher targeting efficiencies it has been shown to result in increased indel formation. Finally, PE3b utilizes an accessory sgRNA that induces a nick on the complementary (WT) strand in the edit/WT heteroduplex. This favors mismatch repair to incorporate the edit into both strands of the target locus, which avoids transient DSBs and significantly reduces indel formation.

Although PE addresses many of the limitations of other CRISPR-based methods, the critical determinant in PE is the facile design of pegRNAs. Compared to the straightforward design of sgRNAs, pegRNA design requires proper placement of guide, PBS, RTT and edit sequences. As such, the multifactorial design of pegRNAs results in higher complexity and limits

manual design. To that end, we developed a freely accessible software tool, called Prime Induced Nucleotide Engineering Creator of New Edits (PINE-CONE) (https://github.com/xiaowanglab/PINE-CONE) that allows for the high-throughput design of pegRNAs. Overall, this tool will enable scientists from diverse fields to easily navigate their PE-based experiments by automating design of pegRNAs.

RESULTS AND DISCUSSION

PINE-CONE is a software capable of turning basic edit information into pegRNA designs and accessory primers for PE workflows. The interface allows users to select from various organisms, such as human (hg38) and yeast (S288C) (Supplemental Table B-1). PINE-CONE uses the organism selection to retrieve DNA sequence data from online reference genome's web-based API (Supplementary Figure B-1a). Because many laboratory strains and cell lines differ from their canonical reference genomes, PINE-CONE is also capable of running on locally stored sequence information (via the 'Manual .txt' selection). Consequently, information obtained in the lab, such as by DNA sequencing, can be used to inform pegRNA design.

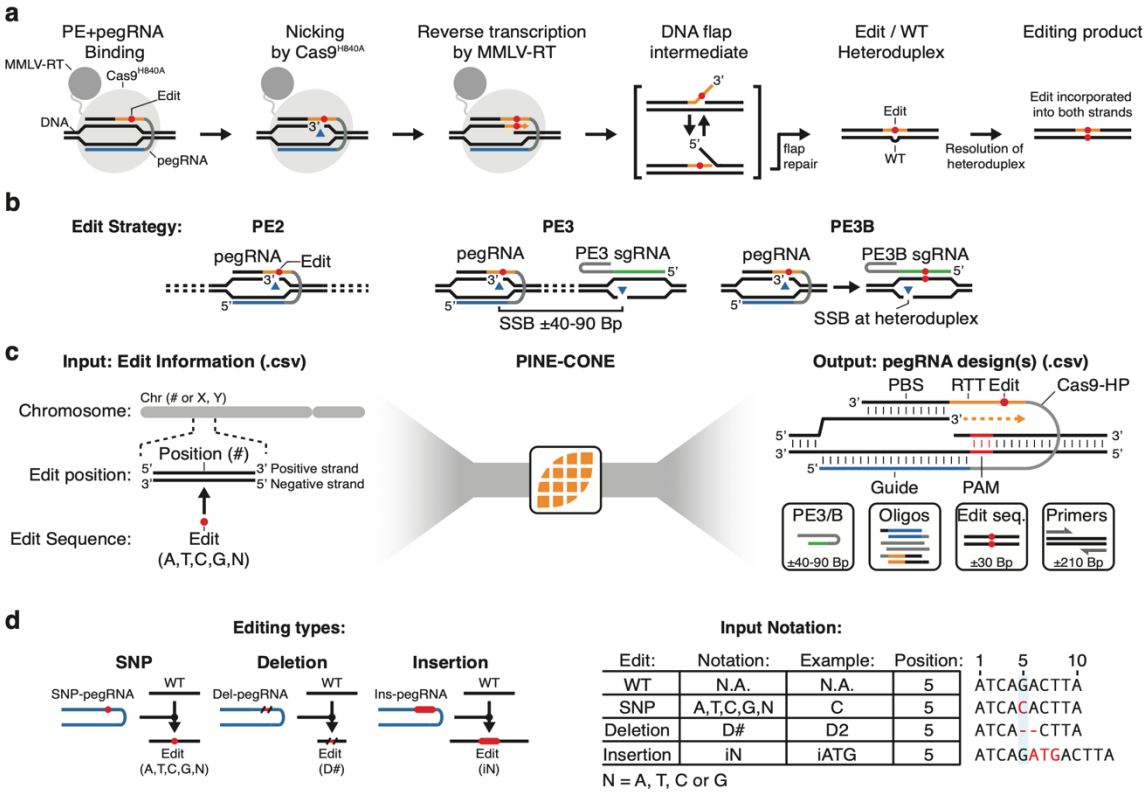


Figure 4-1: PINE-CONE automated design of Prime Editing Guide RNAs (pegRNAs). (**a**) Prime Editor (PE) utilizes a nicking Cas9 (Cas9^{H840A}) fused to a Moloney Murine Leukemia Virus Reverse Transcriptase (MMLV-RT). The PE fusion is targeted to a specific locus via a prime editing Guide RNA (blue line), where the locus is subsequently nicked (blue triangle), exposing a 3' OH. The MMLV-RT initiates reverse transcription from the free 3' OH group using the pegRNA as the template for the edit (red circle). The flap intermediate and Edit/WT DNA heteroduplex is resolved via endogenous DNA repair. This results in the intended editing product incorporated into both DNA strands. (**b**) Prime editing strategies include: PE2, PE3 and PE3B. PE2 utilzes a single pegRNA. PE3 utilizes a pegRNA matching the target locus and a separate sgRNA that targets downstream of the edit site. PE3b employs a sgRNA that is designed to nick the complement (WT) strand of the Edit/WT heteroduplex. (**c**) PINE-CONE takes edit information including chromosome, nucleotide position and intended editing product as input. pegRNAs are designed using DNA sequence data and include a guide directing Cas9 (blue), a Cas9-Hairpin (Cas9-HP, gray), Reverse transcription template (RTT, orange), edit sequence (red circle), and primer binding sequence (PBS). The output file includes PE3 or PE3B guides, oligonucleotides for cloning, intended edit DNA sequences and PCR primers. (**d**) PINE-CONE is capable of designing multiple types of edits including single point mutations (*e.g.* SNPs), deletions and insertions. Mutations are encoded in the input file in the format shown in the "input notation" panel.

Edit input information is provided by a simple comma separated variable (.CSV) file (Supplemental Figure B-1b). The Input file includes edit chromosome, coordinates, sequence and optional basic pegRNA parameters (RTT and PBS length). The input information is used by

PINE-CONE to design pegRNAs encoding the intended edit along with PE3/B, cloning oligos intended edit sequences, and PCR primers (Figure 4-1c). PINE-CONE retrieves the wildtype (WT) target DNA sequences. In turn, edit information is used to design the intended editing products. Broadly speaking, PINE-CONE designs guides based off proximity to edit or by specificity. Specifically, to account for potential off-target effects, specificity scoring has been integrated into PINE-CONE's pegRNA design. When designing pegRNAs against a reference genome, PINE-CONE retrieves "MIT Specificity scores". In turn, PINE-CONE ranks these scores and uses the highest specificity guides available for pegRNA design. RTT sequence lengths are (i) defined by the user via the input file ("RT (Bp)", Supplemental Figure B-1b), or (ii) if the RTT input is blank, determined by PINE-CONE with a viable size (10-33 Bp). Similarly, PBS design are (i) defined by the user using a preferred PBS length or (ii) if the PBS section is left blank, optimized by PINE-CONE using GC-content as the deterministic criteria as previous studies have shown that high-GC contents favor short PBSs while low GC-content favors longer PBSs(Kim et al., 2020). Because pegRNAs often require design of multiple guides, RTT and PBS lengths for experimental optimization(Anzalone et al., 2019), for most target loci PINE-CONE designs at least 2 pegRNAs. In addition, the user can enter multiple rows with systematic changes to RTT and/or PBS lengths to the same edit generating multiple pegRNA variants. Critically, PINE-CONE is capable of designing pegRNAs for a range of edits from single nucleotide edits, such as single nucleotide polymorphisms (SNPs/replacements), or deletions (Del, D) and insertions (Ins, i) (Figure 4-1d). Finally, PINE-CONE can design pegRNAs and accessory sgRNAs for various PE-based strategies including PE2, PE3 and PE3B(Anzalone et al., 2019) (Figure 4-1b**)**.

PINE-CONE's outputs design results in a '.CSV' format and encodes edit information, pegRNA PE3 or 3B sgRNAs, cloning oligonucleotides and PCR primers. Edit information is encoded in 'WT-to-Edit' format along with WT and Edit DNA sequences. PE3 or 3B sgRNA protospacer and target cleavage distance(s) are provided. The output file also includes oligonucleotides necessary for pegRNA and sgRNA cloning workflows (Supplemental Figure B-2, Supplemental Table B-2). Cloning of peg and sgRNAs uses straightforward restriction enzyme cloning and is compatible with an available CRISPR RNA expression vector. Since PCR and

sequencing are often necessary in genome editing workflows, PINE-CONE also designs PCR and sequencing primers flanking the edit locus. Importantly, PINE-CONE designs primers with annealing temperatures that correlates with a commercially available TM calculator (Supplemental Figure B-3). PINE-CONE is capable of plotting valid pegRNA loci in the form of a Circos plot for Human (hg38) and yeast (S288C) reference genomes.

For our initial validation of PINE-CONE's functionality, we used PINE-CONE to design pegRNAs for targets in which pegRNAs had been previously experimentally validated by *Anzalone et al*., *Kim et al.* and *Schene et al.* (Anzalone et al., 2019; Kim et al., 2020; Schene et al., 2020). These targets included a broad spectrum of single nucleotide substitutions as well as small deletions. Overall, this analysis revealed that PINE-CONE generated pegRNA sequences with matching guides, PBS, and RTT sequences to previously published designs (Supplemental Figure B-4).

Next, we employed PINE-CONE for the design of *de novo* pegRNA constructs that would be useful for disease modeling applications (Figure 4-2a). To test PINE-CONE's ability to improve design automation, we assessed its ability to generate pegRNAs to target 24 diverse single nucleotide polymorphisms (SNPs) that have been previously been identified to be associated with increased risk of Alzheimer's Disease (AD) (Figure 4-2b)(Kunkle et al., 2019; Raman et al., 2020). Initially, we systematically assessed the effect of RTT length on pegRNA targeting by analyzing a cumulative 625 pegRNA designs at these 24 loci. We found longer RTT lengths expanded the editing window of PEs with 30 bp RTT sequences targeting up to 87% of loci (21/24). Concurrently, we analyzed the prevalence of valid PE3 and PE3B accessory sgRNA targets. PINE-CONE successfully designed PE3 sgRNAs for 79% of loci (19/24) and PE3B sgRNAs for 58% of loci (14/24) (Figure 4-2c). Circos style plots generated by PINE-CONE indicate valid pegRNA loci across numerous chromosomal contexts (Figure 4-2d)(Krzywinski et al., 2009). Finally, we analyzed the type of base conversions within our *in silico* experiment and found pegRNAs target a series of transition mutations accomplishable by BEs and PEs (38%). However, the majority (62%) of mutations consist of base transversions accomplishable solely through use of PEs (Figure 4-2e). This highlights the expanded editing scope of PEs and the

ability of PINE-CONE to allow for pegRNA design automation.

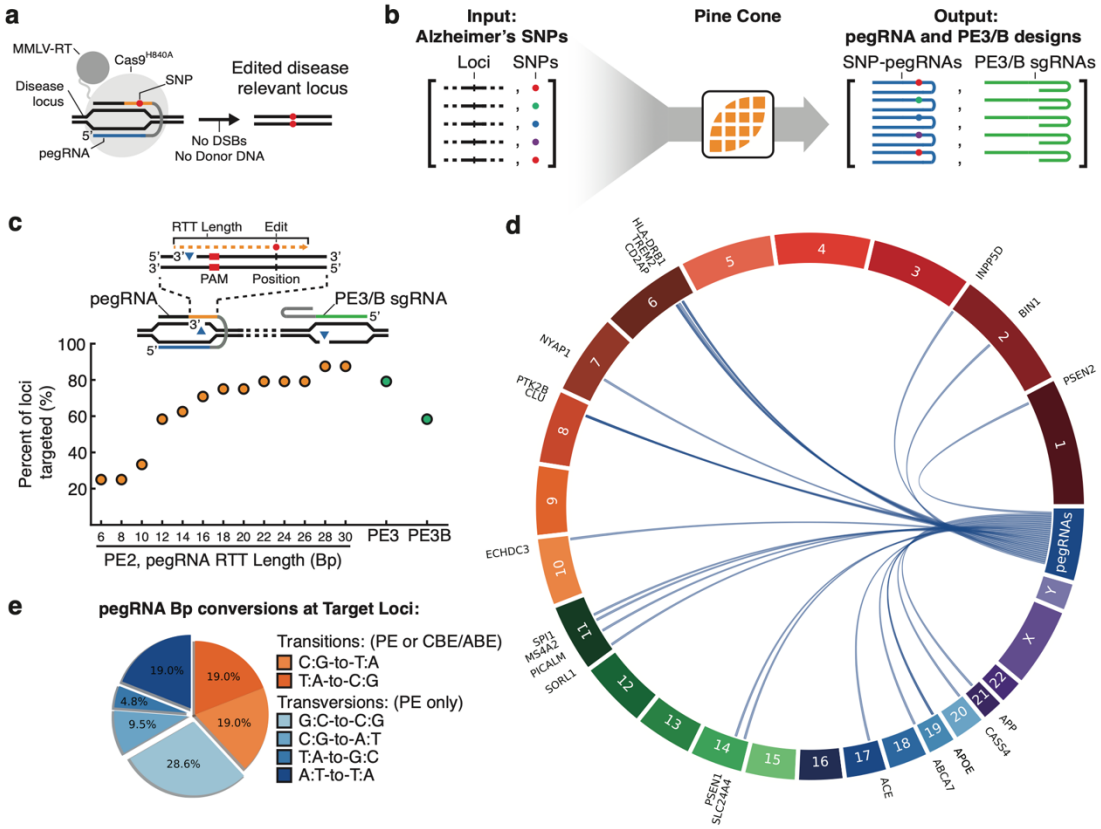


Figure 4-2: PINE-CONE design of pegRNAs of Alzheimer's disease (AD)-related single nucleotide polymorphisms (SNPs). (**a**) Prime Editing mediated introduction of SNPs. A pegRNA targeting a disease locus encodes an edit, which is then incorporated into the target locus without the need for double-stranded DNA breaks (DSBs) or introduction of linearized donor DNA. (**b**) PINE-CONE rapidly analyzed and designed a library of pegRNAs and PE3 or PE3B sgRNAs for 24 AD-related loci. (**c**) The percent of loci targeted by pegRNAs was systematically analyzed for various RTT lengths. Inset schematic indicates valid edits that fall with the reverse transcription range. Longer RTT lengths expand the Prime Editing window and thus increase the number of targets up to 87% of loci with RTTs of 30 Bp (21/24). In parallel, PINE-CONE generated designs of PE3 sgRNAs for 79% of loci (19/24) and PE3B sgRNAs for 58% of loci (14/24). (**d**) PINE-CONE generated Circos mapping of pegRNAs to target loci indicates PINE-CONE successfully designs pegRNAs across numerous chromosome contexts. (**e**) A pie chart of PINE-CONE designed edits at the 24 AD-relevant loci. Transition mutations are accomplishable by cytosine base editors (CBEs), adenosine base editors (ABEs) and Prime Editors (PEs) (38%, orange). The majority of mutations consist of base transversion mutations (62%, blue).

Finally, to validate design of pegRNAs in an alternative organism *in silico,* we tested PINE-CONE on *Saccharomyces cerevisiae* S288C, an important host for biotechnology and synthetic biology. We utilized PINE-CONE to rapidly design pegRNAs for a series of loci including a series of autotrophic marker genes. PINE-CONE was able to design pegRNAs that would

induce a range of genetic modifications including introducing premature stop codons, short deletions resulting in frameshift knockouts, and insertions of LoxP site flanking target loci (Supplemental Figure B-5). Collectively, this demonstrates PINE-CONE is capable of automating pegRNAs design in multiple organism contexts for a variety of applications.

Tools for rapid design and implementation of genome engineering techniques are important for their broad adoption. As such, multiple pegRNA design tools have recently become available. For instance, pegFinder designs pegRNAs via *in silico* alignment of WT and intended edit products(Chow et al., 2020). Multicrispr is a R package for a wide range of CRISPR-based strategies including pegRNA designs(Bhagwat et al., 2020). PrimeDesign is capable of designing pegRNAs for genome-wide and saturation mutagenesis(Hsu et al., 2020). These tools are effective and each offers unique functionality, however they require prior generation of the intended editing product with flanking DNA sequences. Consequently, this may reduce throughput and increase the likelihood of user imparted errors. We sought to develop a tool that enables direct integration of nucleotide coordinates and straightforward editing nomenclature. To that end, PINE-CONE automates pegRNA design for multiple species, offers flexible RTT and PBS specification, and requires only numerical DNA positional information and simple editing notation.

In summary, PINE-CONE can design a range of edits and systematically analyzing pegRNA designs. Specifically, we demonstrated design of a series of pegRNA libraries in multiple contexts for both disease study and synthetic biology. Altogether, PINE-CONE increases ease of pegRNA design and significantly accelerates PE-based workflows.

METHODS

PINE-CONE was written in Python with the user interface (UI) constructed using Tkinter. PINE-CONE source code, executables and example files are provided for download at the Xiao Lab GitHub (https://github.com/xiaowanglab/PINE-CONE). A callable python script version for integration into genome-wide design pipelines is also available.

DNA sequences of each reference genome are accessed via API hosted by UCSC genome browser (Supplemental Table B-1). For API based retrieval of genomic DNA sequence, PINE-CONE limits searches to 1 search per 0.5 seconds to avoid high frequency requests. After DNA retrieval, PINE-CONE conducts a bidirectional PAM search based off of user preferred RTT length. If RTT length is undefined, PINE-CONE will identify a viable RTT length given the availability of PAMs. Guide sequences are defined from available PAM motifs and will retrieve MIT specificity scores from UCSC browser web api if selected (via the “high specificity” preference). PINE-CONE utilizes the highest specificity guides available. The PBS is either (i) of a user defined length or (ii) PINE-CONE will design the PBS based off of GC-content. Guide, Cas9 hairpin, RTT and PBS are combined to create pegRNA sequences. PINE-CONE attempts to design at least 2 guides per target locus.

PINE-CONE’s ability to design pegRNAs was first tested by comparing sequence output of loci tested in *Anzalone et al*. *Kim et al.* and *Schene et al.*(Anzalone et al., 2019; Kim et al., 2020; Schene et al., 2020).Coordinates edited were determined by Basic Local Alignment Tool (BLAT) analysis of edit locations. The subsequent coordinates and edit nucleotides were provided to PINE-CONE in an input .csv matching the format (Supplemental Figure B-1). Output pegRNA sequences were aligned back to experimentally validated pegRNAs assuming use of the same scaffold sequence. This input file is available for download at the Xiao Lab GitHub.

For design of pegRNAs to target AD-related loci, PINE-CONE curated list of alleles identified by genome wide association study (reference(Kunkle et al., 2019)) and via physiological importance (reference(Raman et al., 2020)). For systematic analysis of RTT length. Loci were downloaded from UCSC Genome Browser hg38 and were analyzed locally by PINE-CONE. The list of valid RTT lengths was then used to query pegRNA designs against the human genome (hg38).

To test *in silico* pegRNA design on *Saccharomyces cerevisiae* (S288C), the coordinates for marker loci were downloaded from SGD and used to either introduce stop codons at points in the ORF, deletions 4-5 Bp in length or via insertion loxP sites by 2 pegRNAs flanking the coding sequence.

REFERENCES

Anzalone, A. V., Koblan, L. W., & Liu, D. R. (2020). Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nature Biotechnology*, *38*(7), Article 7. https://doi.org/10.1038/s41587-020-0561-9

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, *576*(7785), 149–157. https://doi.org/10.1038/s41586-019-1711-4

Bartley, B. A., Choi, K., Samineni, M., Zundel, Z., Nguyen, T., Myers, C. J., & Sauro, H. M. (2019). pySBOL: A Python Package for Genetic Design Automation and Standardization. *ACS Synthetic Biology*, *8*(7), 1515–1518. https://doi.org/10.1021/acssynbio.8b00336

Bhagwat, A. M., Graumann, J., Wiegandt, R., Bentsen, M., Welker, J., Kuenne, C., Preussner, J., Braun, T., & Looso, M. (2020). multicrispr: GRNA design for prime editing and parallel targeting of thousands of targets. *Life Science Alliance*, *3*(11). https://doi.org/10.26508/lsa.202000757

Chaikind, B., Bessen, J. L., Thompson, D. B., Hu, J. H., & Liu, D. R. (2016). A programmable Cas9-serine recombinase fusion protein that operates on DNA sequences in mammalian cells. *Nucleic Acids Research*, *44*(20), 9758–9770. https://doi.org/10.1093/nar/gkw707

Chen, S. P., & Wang, H. H. (2019). An Engineered Cas-Transposon System for Programmable and Site-Directed DNA Transpositions. *The CRISPR Journal*, *2*(6), 376–394. https://doi.org/10.1089/crispr.2019.0030

Chow, R. D., Chen, J. S., Shen, J., & Chen, S. (2020). A web tool for the design of prime-editing guide RNAs. *Nature Biomedical Engineering*, 1–5. https://doi.org/10.1038/s41551-020-00622-8

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., & Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*, *31*(9), 822–826. https://doi.org/10.1038/nbt.2623

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*, *551*(7681), 464–471. https://doi.org/10.1038/nature24644

HamediRad, M., Chao, R., Weisberg, S., Lian, J., Sinha, S., & Zhao, H. (2019). Towards a fully automated algorithm driven platform for biosystems design. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-13189-z

Hsu, J. Y., Anzalone, A. V., Grünewald, J., Lam, K. C., Shen, M. W., Liu, D. R., Joung, J. K., & Pinello, L. (2020). PrimeDesign software for rapid and simplified design of prime editing guide RNAs. *BioRxiv*, 2020.05.04.077750. https://doi.org/10.1101/2020.05.04.077750

Ihry, R. J., Worringer, K. A., Salick, M. R., Frias, E., Ho, D., Theriault, K., Kommineni, S., Chen, J., Sondey, M., Ye, C., Randhawa, R., Kulkarni, T., Yang, Z., McAllister, G., Russ, C., Reece-Hoyes, J., Forrester, W., Hoffman, G. R., Dolmetsch, R., & Kaykas, A. (2018). P53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nature Medicine*, *24*(7), 939–946. https://doi.org/10.1038/s41591-018-0050-6

Iverson, S. V., Haddock, T. L., Beal, J., & Densmore, D. M. (2016). CIDAR MoClo: Improved MoClo Assembly Standard and New E. coli Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. *ACS Synthetic Biology*, *5*(1), 99–103. https://doi.org/10.1021/acssynbio.5b00124

Kim, H. K., Yu, G., Park, J., Min, S., Lee, S., Yoon, S., & Kim, H. H. (2020). Predicting the efficiency of prime editing guide RNAs in human cells. *Nature Biotechnology*, 1–9. https://doi.org/10.1038/s41587-020-0677-y

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, *533*(7603), 420–424. https://doi.org/10.1038/nature17946

Kosicki, M., Tomberg, K., & Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nature Biotechnology*, *36*(8), 765–771. https://doi.org/10.1038/nbt.4192

Krzywinski, M. I., Schein, J. E., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*. https://doi.org/10.1101/gr.092759.109

Kunkle, B. W., Grenier-Boley, B., Sims, R., Bis, J. C., Damotte, V., Naj, A. C., Boland, A., Vronskaya, M., van der Lee, S. J., Amlie-Wolf, A., Bellenguez, C., Frizatti, A., Chouraki, V., Martin, E. R., Sleegers, K., Badarinarayan, N., Jakobsdottir, J., Hamilton-Nelson, K. L., Moreno-Grau, S., … Pericak-Vance, M. A. (2019). Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nature Genetics*, *51*(3), Article 3. https://doi.org/10.1038/s41588-019-0358-2

Nielsen, A. A. K., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E. A., Ross, D., Densmore, D., & Voigt, C. A. (2016). Genetic circuit design automation. *Science*, *352*(6281). https://doi.org/10.1126/science.aac7341

Raman, S., Brookhouser, N., & Brafman, D. A. (2020). Using human induced pluripotent stem cells (hiPSCs) to investigate the mechanisms by which Apolipoprotein E (APOE) contributes to Alzheimer's disease (AD) risk. *Neurobiology of Disease*, *138*, 104788. https://doi.org/10.1016/j.nbd.2020.104788

Rees, H. A., & Liu, D. R. (2018). Base editing: Precision chemistry on the genome and transcriptome of living cells. *Nature Reviews Genetics*, *19*(12), Article 12. https://doi.org/10.1038/s41576-018-0059-1

Schene, I. F., Joore, I. P., Oka, R., Mokry, M., van Vugt, A. H. M., van Boxtel, R., van der Doef, H. P. J., van der Laan, L. J. W., Verstegen, M. M. A., van Hasselt, P. M., Nieuwenhuis, E. E. S., & Fuchs, S. A. (2020). Prime editing for functional repair in patient-derived disease models. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-19136-7

Sharon, E., Chen, S.-A. A., Khosla, N. M., Smith, J. D., Pritchard, J. K., & Fraser, H. B. (2018). Functional Genetic Variants Revealed by Massively Parallel Precise Genome Editing. *Cell*, *175*(2), 544-557.e16. https://doi.org/10.1016/j.cell.2018.08.057

Standage-Beier, K., Brookhouser, N., Balachandran, P., Zhang, Q., Brafman, D. A., & Wang, X. (2019). RNA-Guided Recombinase-Cas9 Fusion Targets Genomic DNA Deletion and Integration. *The CRISPR Journal*, *2*(4), 209–222. https://doi.org/10.1089/crispr.2019.0013

Standage-Beier, K., & Wang, X. (2017). Genome reprogramming for synthetic biology. *Frontiers of Chemical Science and Engineering*, *11*(1), 37–45. https://doi.org/10.1007/s11705-017-1618-2

Stemmer, M., Thumberger, T., Keyer, M. del S., Wittbrodt, J., & Mateo, J. L. (2015). CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PLOS ONE*, *10*(4), e0124633. https://doi.org/10.1371/journal.pone.0124633

CHAPTER 5

CONCLUSION

Genome engineering encompasses a broad range of DNA manipulations. Here we provide a review of canonical double stranded DNA break (DSB) based CRISPR technologies (Chapter 2). DSB-based approaches have broad utility in many organisms as DSB can initiate various modes of recombination including non-homologous end joining (NHEJ) or homology directed repair (HDR). Alternately, depending on the species and organism, the lethality of DSBs can also function selectively for incorporation of edited DNA sequences.

Engineering human cells line for the study of disease and potential therapeutic purpose is of high interest, however a significant amount of evidence indicates DSB have deleterious effects. For instance, double stranded DNA breaks can initiate unwanted recombination events including large deletions and complex rearrangements(Kosicki et al., 2018). Concurrently CRISPR-directed double-stranded DNA breaks can upregulate p53, a crucial tumor suppressor gene, leading to cell cycle arrest and apoptosis(Haapaniemi et al., 2018).

To address these drawbacks of DSB-based technologies, RNA-guided Base Editors (described in Chapter 1) can introduce ‘C-to-T’ and ‘A-to-G’ DNA edits without requisite DSBs(Gaudelli et al., 2017; Komor et al., 2016). Single letter DNA edits, however, can be difficult to detect and require lengthy amplification and sequencing techniques to assay editing efficiency. Low editing efficiency obfuscates base editors’ ease-of-use especially at loci recalcitrant to editing. Likewise, initial efforts to optimize transfection efficiency largely did not improve editing efficiencies at difficult to edit loci (Chapter 3).

We reasoned a fluorescent reporter system for base editing would be of value in addressing the inability to easily detect edited cells (Chapter 3). To this end we engineered a blue florescent protein (BFP) that converts to green fluorescent protein (GFP) upon C-to-T base editing. We successfully demonstrate fluorescent conversion detects base editing activity. We then applied this fluorescent conversion assay to various delivery methods. To further expand this technology, we developed a plasmid-genome co-targeting strategy that enabled fluorescent activated cell sorting (FACS) isolation of edited cell populations. We find our Transient Reporter

for Editing Enrichment (TREE) substantially outperforms traditional Reporters of Transfection (RoT). The high editing efficiency of TREE enables multiplex targeting and editing of 3 loci simultaneously, suggesting TREE's potential use in the study of polygenic diseases. Furthermore, using we identify an optimized promoter-base combination for application in human pluripotent stem cells (Chapter 3).

The efficiency and ease-of-use attained by TREE enabled the follow up strategy 'Base-edited isogenic hPSC (human pluripotent stem cells) line generation using a transient reporter for editing enrichment' (BIG-TREE). BIG-TREE enables generation of isogenic cell line with high specificity edits at disease relevant loci. Likewise, we adapt TREE as a technique for generating gene knockouts via the introduction of premature stop codons(Brookhouser, Tekel, et al., 2020).

The initial TREE was specific to C-to-T base editing and did not report on A-to-G base editing. To address this limitation, we developed Cas9-Mediated Adenosine Transient Reporter for Editing Enrichment (Abbreviated XMAS-TREE). We found the XMAS-TREE reporter is effective at editing numerous loci and isolating base edited cell and generating edited clonal lines(Brookhouser, Nguyen, et al., 2020).To facilitate broad adoption of TREE-based methods we published a protocol describing application of TREE in hPSCs. Wherein we describe TREE-based strategies, design and cloning of TREE-compatible single guide RNAs (sgRNAs), and isolation and culture edited cell lines(Tekel et al., 2021).

Base editors are effective at introducing DNA transition mutations within a narrow editing window, however they are limited to this small subset of edits. Base editors cannot readily accomplish transversion mutations, insertions, or deletions. More recently, Prime editors can introduce a broad range of user-defined DNA edits. Prime editors accomplish this via a reverse transcriptase (RT) fused to a nicking version of Cas9 ($Cas9^{H840A}$) along with a specialized prime editing guide RNA (pegRNA). Upon targeting the 3' end of the pegRNA binds to the target sequence and primes the RT to reverse transcribe the pegRNA template at the locus. Consequently, part of the pegRNA is converted into a DNA edit(Anzalone et al., 2019). Prime editing is an auspicious technology, however pegRNAs can be difficult to manually design requiring accurate placement of guide, reverse transcriptase template (RTT) and primer binding

sequences (PBS). This hinders adoption by researchers less experienced with DNA manipulations and lowers throughput of pegRNA design. To address this, we created a software tool Prime Induced Nucleotide Engineering Creator of New Edits (PINE-CONE, Chapter 4). PINE-CONE takes basic edit information and generates prime editing strategies. PINE-CONE designs pegRNAs, supporting sgRNAs and oligonucleotides for cloning and sequencing. We find PINE-CONE rapidly analyses pegRNA designs not easily accomplished through manual effort. Using basic information of single nucleotide polymorphisms (SNPs) associated with Alzheimer's Disease, PINE-CONE rapidly designs pegRNAs for a diverse range of disease relevant genomic loci. Collectively we believe the literary and scientific contributions presented in this document represent an important addition to genome engineering techniques for synthetic biology and the study of disease genetics.

REFERENCES

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, *576*(7785), 149–157. https://doi.org/10.1038/s41586-019-1711-4

Brookhouser, N., Nguyen, T., Tekel, S. J., Standage-Beier, K., Wang, X., & Brafman, D. A. (2020). A Cas9-mediated adenosine transient reporter enables enrichment of ABE-targeted cells. *BMC Biology*, *18*(1), 193. https://doi.org/10.1186/s12915-020-00929-7

Brookhouser, N., Tekel, S. J., Standage-Beier, K., Nguyen, T., Schwarz, G., Wang, X., & Brafman, D. A. (2020). BIG-TREE: Base-Edited Isogenic hPSC Line Generation Using a Transient Reporter for Editing Enrichment. *Stem Cell Reports*, *14*(2), 184–191. https://doi.org/10.1016/j.stemcr.2019.12.013

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*, *551*(7681), 464–471. https://doi.org/10.1038/nature24644

Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., & Taipale, J. (2018). CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nature Medicine*, *24*(7), 927–930. https://doi.org/10.1038/s41591-018-0049-z

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, *533*(7603), 420–424. https://doi.org/10.1038/nature17946

Kosicki, M., Tomberg, K., & Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nature Biotechnology*, *36*(8), 765–771. https://doi.org/10.1038/nbt.4192

Tekel, S. J., Brookhouser, N., Standage-Beier, K., Wang, X., & Brafman, D. A. (2021). Cytosine and adenosine base editing in human pluripotent stem cells using transient reporters for editing enrichment. *Nature Protocols*, *16*(7), 3596–3624. https://doi.org/10.1038/s41596-021-00552-y

REFERENCES

Anzalone, A. V., Koblan, L. W., & Liu, D. R. (2020). Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. Nature Biotechnology, 38(7), Article 7. https://doi.org/10.1038/s41587-020-0561-9

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature, 576(7785), 149–157. https://doi.org/10.1038/s41586-019-1711-4

Bao, Z., Xiao, H., Liang, J., Zhang, L., Xiong, X., Sun, N., Si, T., & Zhao, H. (2015). Homology-integrated CRISPR-Cas (HI-CRISPR) system for one-step multigene disruption in Saccharomyces cerevisiae. ACS Synthetic Biology, 4(5), 585–594. https://doi.org/10.1021/sb500255k

Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schneider, D., Lenski, R. E., & Kim, J. F. (2009). Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature, 461(7268), 1243–1247. https://doi.org/10.1038/nature08480

Bartley, B. A., Choi, K., Samineni, M., Zundel, Z., Nguyen, T., Myers, C. J., & Sauro, H. M. (2019). pySBOL: A Python Package for Genetic Design Automation and Standardization. ACS Synthetic Biology, 8(7), 1515–1518. https://doi.org/10.1021/acssynbio.8b00336

Bassalo, M. C., Garst, A. D., Halweg-Edwards, A. L., Grau, W. C., Domaille, D. W., Mutalik, V. K., Arkin, A. P., & Gill, R. T. (2016). Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli. ACS Synthetic Biology, 5(7), 561–568. https://doi.org/10.1021/acssynbio.5b00187

Benam, K. H., Dauth, S., Hassell, B., Herland, A., Jain, A., Jang, K.-J., Karalis, K., Kim, H. J., MacQueen, L., Mahmoodian, R., Musah, S., Torisawa, Y., van der Meer, A. D., Villenave, R., Yadid, M., Parker, K. K., & Ingber, D. E. (2015). Engineered in vitro disease models. Annual Review of Pathology, 10, 195–262. https://doi.org/10.1146/annurev-pathol-012414-040418

Bhagwat, A. M., Graumann, J., Wiegandt, R., Bentsen, M., Welker, J., Kuenne, C., Preussner, J., Braun, T., & Looso, M. (2020). multicrispr: GRNA design for prime editing and parallel targeting of thousands of targets. Life Science Alliance, 3(11). https://doi.org/10.26508/lsa.202000757

Brookhouser, N., Nguyen, T., Tekel, S. J., Standage-Beier, K., Wang, X., & Brafman, D. A. (2020). A Cas9-mediated adenosine transient reporter enables enrichment of ABE-targeted cells. BMC Biology, 18(1), 193. https://doi.org/10.1186/s12915-020-00929-7

Brookhouser, N., Raman, S., Potts, C., & Brafman, David. A. (2017). May I Cut in? Gene Editing Approaches in Human Induced Pluripotent Stem Cells. Cells, 6(1), 5. https://doi.org/10.3390/cells6010005

Brookhouser, N., Tekel, S. J., Standage-Beier, K., Nguyen, T., Schwarz, G., Wang, X., & Brafman, D. A. (2020). BIG-TREE: Base-Edited Isogenic hPSC Line Generation Using a Transient Reporter for Editing Enrichment. Stem Cell Reports, 14(2), 184–191. https://doi.org/10.1016/j.stemcr.2019.12.013

Brophy, J. A. N., & Voigt, C. A. (2014). Principles of genetic circuit design. Nature Methods, 11(5), 508–520. https://doi.org/10.1038/nmeth.2926

Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., Dickman, M. J., Makarova, K. S., Koonin, E. V., & Oost, J. van der. (2008). Small CRISPR RNAs

Guide Antiviral Defense in Prokaryotes. Science, 321(5891), 960–964. https://doi.org/10.1126/science.1159689

Chaikind, B., Bessen, J. L., Thompson, D. B., Hu, J. H., & Liu, D. R. (2016). A programmable Cas9-serine recombinase fusion protein that operates on DNA sequences in mammalian cells. Nucleic Acids Research, 44(20), 9758–9770. https://doi.org/10.1093/nar/gkw707

Chen, S. P., & Wang, H. H. (2019). An Engineered Cas-Transposon System for Programmable and Site-Directed DNA Transpositions. The CRISPR Journal, 2(6), 376–394. https://doi.org/10.1089/crispr.2019.0030

Choi, K. R., & Lee, S. Y. (n.d.). CRISPR Technologies for Bacterial Systems: Current Achievements and Future Directions. Biotechnology Advances, 34(7), 1180–1209. https://doi.org/10.1016/j.biotechadv.2016.08.002

Chow, R. D., Chen, J. S., Shen, J., & Chen, S. (2020). A web tool for the design of prime-editing guide RNAs. Nature Biomedical Engineering, 1–5. https://doi.org/10.1038/s41551-020-00622-8

Chung, S., Andersson, T., Sonntag, K.-C., Björklund, L., Isacson, O., & Kim, K.-S. (2002). Analysis of different promoter systems for efficient transgene expression in mouse embryonic stem cell lines. Stem Cells (Dayton, Ohio), 20(2), 139–145. https://doi.org/10.1634/stemcells.20-2-139

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science, 339(6121), 819–823. https://doi.org/10.1126/science.1231143

Cooper, V. S., Schneider, D., Blot, M., & Lenski, R. E. (2001). Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of Escherichia coli B. Journal of Bacteriology, 183(9), 2834–2841. https://doi.org/10.1128/JB.183.9.2834-2841.2001

Csörgő, B., Nyerges, Á., Pósfai, G., & Fehér, T. (2016). System-level genome editing in microbes. Current Opinion in Microbiology, 33, 113–122. https://doi.org/10.1016/j.mib.2016.07.005

DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J., & Church, G. M. (2013). Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Research, 41(7), 4336–4343. https://doi.org/10.1093/nar/gkt135

Doyon, Y., McCammon, J. M., Miller, J. C., Faraji, F., Ngo, C., Katibah, G. E., Amora, R., Hocking, T. D., Zhang, L., Rebar, E. J., Gregory, P. D., Urnov, F. D., & Amacher, S. L. (2008). Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. Nature Biotechnology, 26(6), 702–708. https://doi.org/10.1038/nbt1409

Dymond, J. S., Richardson, S. M., Coombes, C. E., Babatz, T., Muller, H., Annaluru, N., Blake, W. J., Schwerzmann, J. W., Dai, J., Lindstrom, D. L., Boeke, A. C., Gottschling, D. E., Chandrasegaran, S., Bader, J. S., & Boeke, J. D. (2011). Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. Nature, 477(7365), 471–476. https://doi.org/10.1038/nature10403

Eid, A., Alshareef, S., & Mahfouz, M. M. (2018). CRISPR base editors: Genome editing without double-stranded breaks. The Biochemical Journal, 475(11), 1955–1964. https://doi.org/10.1042/BCJ20170793

Eiges, R., Schuldiner, M., Drukker, M., Yanuka, O., Itskovitz-Eldor, J., & Benvenisty, N. (2001). Establishment of human embryonic stem cell-transfected clones carrying a marker for

undifferentiated cells. Current Biology: CB, 11(7), 514–518. https://doi.org/10.1016/s0960-9822(01)00144-0

Elena, S. F., & Lenski, R. E. (2003). Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation. Nature Reviews Genetics, 4(6), 457–469. https://doi.org/10.1038/nrg1088

Ellis, T., Wang, X., & Collins, J. J. (2009). Diversity-based, model-guided construction of synthetic gene networks with predicted functions. Nature Biotechnology, 27(5), 465–471. https://doi.org/10.1038/nbt.1536

Enyeart, P. J., Chirieleison, S. M., Dao, M. N., Perutka, J., Quandt, E. M., Yao, J., Whitt, J. T., Keatinge-Clay, A. T., Lambowitz, A. M., & Ellington, A. D. (2013). Generalized bacterial genome editing using mobile group II introns and Cre-lox. Molecular Systems Biology, 9(1), 685. https://doi.org/10.1038/msb.2013.41

Farzadfard, F., & Lu, T. K. (2014). Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. Science, 346(6211), 1256272. https://doi.org/10.1126/science.1256272

Faucon, P. C., Pardee, K., Kumar, R. M., Li, H., Loh, Y.-H., & Wang, X. (2014). Gene networks of fully connected triads with complete auto-activation enable multistability and stepwise stochastic transitions. PloS One, 9(7), e102873. https://doi.org/10.1371/journal.pone.0102873

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., & Sander, J. D. (2013a). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature Biotechnology, 31(9), 822–826. https://doi.org/10.1038/nbt.2623

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., & Sander, J. D. (2013b). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature Biotechnology, 31(9), 822–826. https://doi.org/10.1038/nbt.2623

Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M., & Joung, J. K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nature Biotechnology, 32(3), 279–284. https://doi.org/10.1038/nbt.2808

Ganesan, K., Kulandaisamy, A., Priya, S. B., & Gromiha, M. M. (2019). HuVarBase: A human variant database with comprehensive information at gene and protein levels. PLOS ONE, 14(1), e0210475. https://doi.org/10.1371/journal.pone.0210475

Gardner, T. S., Cantor, C. R., & Collins, J. J. (2000). Construction of a genetic toggle switch in Escherichia coli. Nature, 403(6767), 339–342. https://doi.org/10.1038/35002131

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017a). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature, 551(7681), 464–471. https://doi.org/10.1038/nature24644

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017b). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature, 551(7681), 464–471. https://doi.org/10.1038/nature24644

Gehrke, J. M., Cervantes, O., Clement, M. K., Wu, Y., Zeng, J., Bauer, D. E., Pinello, L., & Joung, J. K. (2018). An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. Nature Biotechnology, 36(10), 977–982. https://doi.org/10.1038/nbt.4199

Germini, D., Tsfasman, T., Zakharova, V. V., Sjakste, N., Lipinski, M., & Vassetzky, Y. (2018). A Comparison of Techniques to Evaluate the Effectiveness of Genome Editing. Trends in Biotechnology, 36(2), 147–159. https://doi.org/10.1016/j.tibtech.2017.10.008

Giacalone, J. C., Sharma, T. P., Burnight, E. R., Fingert, J. F., Mullins, R. F., Stone, E. M., & Tucker, B. A. (2018). CRISPR-Cas9-Based Genome Editing of Human Induced Pluripotent Stem Cells. Current Protocols in Stem Cell Biology, 44, 5B.7.1-5B.7.22. https://doi.org/10.1002/cpsc.46

Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R.-Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E. A., Young, L., Qi, Z.-Q., Segall-Shapiro, T. H., … Venter, J. C. (2010). Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. Science, 329(5987), 52–56. https://doi.org/10.1126/science.1190719

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods, 6(5), 343–345. https://doi.org/10.1038/nmeth.1318

Glaser, A., McColl, B., & Vadolas, J. (2016). GFP to BFP Conversion: A Versatile Assay for the Quantification of CRISPR/Cas9-mediated Genome Editing. Molecular Therapy. Nucleic Acids, 5(7), e334. https://doi.org/10.1038/mtna.2016.48

Grobarczyk, B., Franco, B., Hanon, K., & Malgrange, B. (2015). Generation of Isogenic Human iPS Cell Line Precisely Corrected by Genome Editing Using the CRISPR/Cas9 System. Stem Cell Reviews and Reports, 11(5), 774–787. https://doi.org/10.1007/s12015-015-9600-1

Guilinger, J. P., Thompson, D. B., & Liu, D. R. (2014). Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nature Biotechnology, 32(6), 577–582. https://doi.org/10.1038/nbt.2909

Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., & Taipale, J. (2018). CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. Nature Medicine, 24(7), 927–930. https://doi.org/10.1038/s41591-018-0049-z

HamediRad, M., Chao, R., Weisberg, S., Lian, J., Sinha, S., & Zhao, H. (2019). Towards a fully automated algorithm driven platform for biosystems design. Nature Communications, 10(1), Article 1. https://doi.org/10.1038/s41467-019-13189-z

Hao, H., Wang, X., Jia, H., Yu, M., Zhang, X., Tang, H., & Zhang, L. (2016). Large fragment deletion using a CRISPR/Cas9 system in Saccharomyces cerevisiae. Analytical Biochemistry, 509, 118–123. https://doi.org/10.1016/j.ab.2016.07.008

Haridhasapavalan, K. K., Borgohain, M. P., Dey, C., Saha, B., Narayan, G., Kumar, S., & Thummer, R. P. (2019). An insight into non-integrative gene delivery approaches to generate transgene-free induced pluripotent stem cells. Gene, 686, 146–159. https://doi.org/10.1016/j.gene.2018.11.069

Hauser, P. S., & Ryan, R. O. (2013). Impact of apolipoprotein E on Alzheimer's disease. Current Alzheimer Research, 10(8), 809–817. https://doi.org/10.2174/15672050113109990156

He, X., Tan, C., Wang, F., Wang, Y., Zhou, R., Cui, D., You, W., Zhao, H., Ren, J., & Feng, B. (2016). Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. Nucleic Acids Research, 44(9), e85. https://doi.org/10.1093/nar/gkw064

Heim, R., Prasher, D. C., & Tsien, R. Y. (1994). Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proceedings of the National Academy of Sciences, 91(26), 12501–12504. https://doi.org/10.1073/pnas.91.26.12501

Horvath, P., Aulner, N., Bickle, M., Davies, A. M., Nery, E. D., Ebner, D., Montoya, M. C., Östling, P., Pietiäinen, V., Price, L. S., Shorte, S. L., Turcatti, G., von Schantz, C., & Carragher, N. O. (2016). Screening out irrelevant cell-based models of disease. Nature Reviews Drug Discovery, 15(11), Article 11. https://doi.org/10.1038/nrd.2016.175

Hsu, J. Y., Anzalone, A. V., Grünewald, J., Lam, K. C., Shen, M. W., Liu, D. R., Joung, J. K., & Pinello, L. (2020). PrimeDesign software for rapid and simplified design of prime editing guide RNAs. BioRxiv, 2020.05.04.077750. https://doi.org/10.1101/2020.05.04.077750

Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell, 157(6), 1262–1278. https://doi.org/10.1016/j.cell.2014.05.010

Huang, X., Wang, Y., Yan, W., Smith, C., Ye, Z., Wang, J., Gao, Y., Mendelsohn, L., & Cheng, L. (2015). Production of Gene-Corrected Adult Beta Globin Protein in Human Erythrocytes Differentiated from Patient iPSCs After Genome Editing of the Sickle Point Mutation. Stem Cells (Dayton, Ohio), 33(5), 1470–1479. https://doi.org/10.1002/stem.1969

Hutchison, C. A., Chuang, R.-Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., Gill, J., Kannan, K., Karas, B. J., Ma, L., Pelletier, J. F., Qi, Z.-Q., Richter, R. A., Strychalski, E. A., Sun, L., Suzuki, Y., Tsvetanova, B., Wise, K. S., Smith, H. O., … Venter, J. C. (2016). Design and synthesis of a minimal bacterial genome. Science, 351(6280), aad6253. https://doi.org/10.1126/science.aad6253

Ihry, R. J., Worringer, K. A., Salick, M. R., Frias, E., Ho, D., Theriault, K., Kommineni, S., Chen, J., Sondey, M., Ye, C., Randhawa, R., Kulkarni, T., Yang, Z., McAllister, G., Russ, C., Reece-Hoyes, J., Forrester, W., Hoffman, G. R., Dolmetsch, R., & Kaykas, A. (2018a). P53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. Nature Medicine, 24(7), 939–946. https://doi.org/10.1038/s41591-018-0050-6

Ihry, R. J., Worringer, K. A., Salick, M. R., Frias, E., Ho, D., Theriault, K., Kommineni, S., Chen, J., Sondey, M., Ye, C., Randhawa, R., Kulkarni, T., Yang, Z., McAllister, G., Russ, C., Reece-Hoyes, J., Forrester, W., Hoffman, G. R., Dolmetsch, R., & Kaykas, A. (2018b). P53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. Nature Medicine, 24(7), 939–946. https://doi.org/10.1038/s41591-018-0050-6

Isaacs, F. J., Carr, P. A., Wang, H. H., Lajoie, M. J., Sterling, B., Kraal, L., Tolonen, A. C., Gianoulis, T. A., Goodman, D. B., Reppas, N. B., Emig, C. J., Bang, D., Hwang, S. J., Jewett, M. C., Jacobson, J. M., & Church, G. M. (2011). Precise Manipulation of Chromosomes in Vivo Enables Genome-Wide Codon Replacement. Science, 333(6040), 348–353. https://doi.org/10.1126/science.1205822

Iverson, S. V., Haddock, T. L., Beal, J., & Densmore, D. M. (2016). CIDAR MoClo: Improved MoClo Assembly Standard and New E. coli Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. ACS Synthetic Biology, 5(1), 99–103. https://doi.org/10.1021/acssynbio.5b00124

Jakočiūnas, T., Rajkumar, A. S., Zhang, J., Arsovska, D., Rodriguez, A., Jendresen, C. B., Skjødt, M. L., Nielsen, A. T., Borodina, I., Jensen, M. K., & Keasling, J. D. (2015). CasEMBLR: Cas9-Facilitated Multiloci Genomic Integration of in Vivo Assembled DNA Parts in Saccharomyces cerevisiae. ACS Synthetic Biology, 4(11), 1226–1234. https://doi.org/10.1021/acssynbio.5b00007

Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature Biotechnology, 31(3), 233–239. https://doi.org/10.1038/nbt.2508

Jiang, W., & Marraffini, L. A. (2015). CRISPR-Cas: New Tools for Genetic Manipulations from Bacterial Immunity Systems. Annual Review of Microbiology, 69(1), 209–228. https://doi.org/10.1146/annurev-micro-091014-104441

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science, 337(6096), 816–821. https://doi.org/10.1126/science.1225829

Kang, H.-S., Charlop-Powers, Z., & Brady, S. F. (2016). Multiplexed CRISPR/Cas9- and TAR-Mediated Promoter Engineering of Natural Product Biosynthetic Gene Clusters in Yeast. ACS Synthetic Biology, 5(9), 1002–1010. https://doi.org/10.1021/acssynbio.6b00080

Kannan, K., Tsvetanova, B., Chuang, R.-Y., Noskov, V. N., Assad-Garcia, N., Ma, L., Hutchison III, C. A., Smith, H. O., Glass, J. I., Merryman, C., Venter, J. C., & Gibson, D. G. (2016). One step engineering of the small-subunit ribosomal RNA using CRISPR/Cas9. Scientific Reports, 6, 30714. https://doi.org/10.1038/srep30714

Karas, B. J., Jablanovic, J., Irvine, E., Sun, L., Ma, L., Weyman, P. D., Gibson, D. G., Glass, J. I., Venter, J. C., Hutchison Iii, C. A., Smith, H. O., & Suzuki, Y. (2014). Transferring whole genomes from bacteria to yeast spheroplasts using entire bacterial cells to reduce DNA shearing. Nature Protocols, 9(4), 743–750. https://doi.org/10.1038/nprot.2014.045

Karpinski, J., Hauber, I., Chemnitz, J., Schäfer, C., Paszkowski-Rogacz, M., Chakraborty, D., Beschorner, N., Hofmann-Sieber, H., Lange, U. C., Grundhoff, A., Hackmann, K., Schrock, E., Abi-Ghanem, J., Pisabarro, M. T., Surendranath, V., Schambach, A., Lindner, C., van Lunzen, J., Hauber, J., & Buchholz, F. (2016). Directed evolution of a recombinase that excises the provirus of most HIV-1 primary isolates with high specificity. Nature Biotechnology, 34(4), 401–409. https://doi.org/10.1038/nbt.3467

Kiani, S., Beal, J., Ebrahimkhani, M. R., Huh, J., Hall, R. N., Xie, Z., Li, Y., & Weiss, R. (2014). CRISPR transcriptional repression devices and layered circuits in mammalian cells. Nature Methods, 11(7), 723–726. https://doi.org/10.1038/nmeth.2969

Kiani, S., Chavez, A., Tuttle, M., Hall, R. N., Chari, R., Ter-Ovanesyan, D., Qian, J., Pruitt, B. W., Beal, J., Vora, S., Buchthal, J., Kowal, E. J. K., Ebrahimkhani, M. R., Collins, J. J., Weiss, R., & Church, G. (2015). Cas9 gRNA engineering for genome editing, activation and repression. Nature Methods, 12(11), 1051–1054. https://doi.org/10.1038/nmeth.3580

Kim, H. K., Yu, G., Park, J., Min, S., Lee, S., Yoon, S., & Kim, H. H. (2020). Predicting the efficiency of prime editing guide RNAs in human cells. Nature Biotechnology, 1–9. https://doi.org/10.1038/s41587-020-0677-y

Kim, S., Kim, D., Cho, S. W., Kim, J., & Kim, J.-S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Research, 24(6), 1012–1019. https://doi.org/10.1101/gr.171322.113

Kim, Y. B., Komor, A. C., Levy, J. M., Packer, M. S., Zhao, K. T., & Liu, D. R. (2017). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nature Biotechnology, 35(4), 371–376. https://doi.org/10.1038/nbt.3803

Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., & Joung, J. K. (2016). High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. Nature, 529(7587), 490–495. https://doi.org/10.1038/nature16526

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Topkar, V. V., Zheng, Z., & Joung, J. K. (2015). Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. Nature Biotechnology, 33(12), 1293–1298. https://doi.org/10.1038/nbt.3404

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., Gonzales, A. P. W., Li, Z., Peterson, R. T., Yeh, J.-R. J., Aryee, M. J., & Joung, J. K. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature, 523(7561), 481–485. https://doi.org/10.1038/nature14592

Kluesner, M. G., Nedveck, D. A., Lahr, W. S., Garbe, J. R., Abrahante, J. E., Webber, B. R., & Moriarity, B. S. (2018). EditR: A Method to Quantify Base Editing from Sanger Sequencing. The CRISPR Journal, 1(3), 239–250. https://doi.org/10.1089/crispr.2018.0014

Koblan, L. W., Doman, J. L., Wilson, C., Levy, J. M., Tay, T., Newby, G. A., Maianti, J. P., Raguram, A., & Liu, D. R. (2018). Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. Nature Biotechnology, 36(9), 843–846. https://doi.org/10.1038/nbt.4172

Kolisnychenko, V., Plunkett, G., Herring, C. D., Feher, T., Posfai, J., Blattner, F. R., & Posfai, G. (2002). Engineering a Reduced Escherichia coli Genome. Genome Research, 12(4), 640–647. https://doi.org/10.1101/gr.217202

Komor, A. C., Badran, A. H., & Liu, D. R. (2017). CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell, 168(1–2), 20–36. https://doi.org/10.1016/j.cell.2016.10.044

Komor, A. C., Badran, A. H., & Liu, D. R. (2018). Editing the Genome Without Double-Stranded DNA Breaks. ACS Chemical Biology, 13(2), 383–388. https://doi.org/10.1021/acschembio.7b00710

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature, 533(7603), 420–424. https://doi.org/10.1038/nature17946

Kosicki, M., Tomberg, K., & Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. Nature Biotechnology, 36(8), 765–771. https://doi.org/10.1038/nbt.4192

Kouranova, E., Forbes, K., Zhao, G., Warren, J., Bartels, A., Wu, Y., & Cui, X. (2016). CRISPRs for Optimal Targeting: Delivery of CRISPR Components as DNA, RNA, and Protein into Cultured Cells and Single-Cell Embryos. Human Gene Therapy, 27(6), 464–475. https://doi.org/10.1089/hum.2016.009

Krishnakumar, R., Grose, C., Haft, D. H., Zaveri, J., Alperovich, N., Gibson, D. G., Merryman, C., & Glass, J. I. (2014). Simultaneous non-contiguous deletions using large synthetic DNA and site-specific recombinases. Nucleic Acids Research, 42(14), e111. https://doi.org/10.1093/nar/gku509

Krzywinski, M. I., Schein, J. E., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. Genome Research. https://doi.org/10.1101/gr.092759.109

Kuhlman, T. E., & Cox, E. C. (2010). Site-specific chromosomal integration of large synthetic constructs. Nucleic Acids Research, 38(6), e92. https://doi.org/10.1093/nar/gkp1193

Kunkle, B. W., Grenier-Boley, B., Sims, R., Bis, J. C., Damotte, V., Naj, A. C., Boland, A., Vronskaya, M., van der Lee, S. J., Amlie-Wolf, A., Bellenguez, C., Frizatti, A., Chouraki, V., Martin, E. R., Sleegers, K., Badarinarayan, N., Jakobsdottir, J., Hamilton-Nelson, K. L., Moreno-Grau, S., … Pericak-Vance, M. A. (2019). Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nature Genetics, 51(3), Article 3. https://doi.org/10.1038/s41588-019-0358-2

Kuscu, C., Arslan, S., Singh, R., Thorpe, J., & Adli, M. (2014). Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. Nature Biotechnology, 32(7), 677–683. https://doi.org/10.1038/nbt.2916

Lajoie, M. J., Gregg, C. J., Mosberg, J. A., Washington, G. C., & Church, G. M. (2012). Manipulating replisome dynamics to enhance lambda Red-mediated multiplex genome engineering. Nucleic Acids Research, 40(22), e170. https://doi.org/10.1093/nar/gks751

Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H.-R., Haimovich, A. D., Kuznetsov, G., Mercer, J. A., Wang, H. H., Carr, P. A., Mosberg, J. A., Rohland, N., Schultz, P. G., Jacobson, J. M., Rinehart, J., Church, G. M., & Isaacs, F. J. (2013). Genomically Recoded Organisms Expand Biological Functions. Science, 342(6156), 357–360. https://doi.org/10.1126/science.1241459

Lartigue, C., Glass, J. I., Alperovich, N., Pieper, R., Parmar, P. P., Hutchison, C. A., Smith, H. O., & Venter, J. C. (2007). Genome Transplantation in Bacteria: Changing One Species to Another. Science, 317(5838), 632–638. https://doi.org/10.1126/science.1144622

Lartigue, C., Vashee, S., Algire, M. A., Chuang, R.-Y., Benders, G. A., Ma, L., Noskov, V. N., Denisova, E. A., Gibson, D. G., Assad-Garcia, N., Alperovich, N., Thomas, D. W., Merryman, C., Hutchison, C. A., Smith, H. O., Venter, J. C., & Glass, J. I. (2009). Creating Bacterial Strains from Genomes That Have Been Cloned and Engineered in Yeast. Science, 325(5948), 1693–1696. https://doi.org/10.1126/science.1173759

Li, H. L., Fujimoto, N., Sasakawa, N., Shirai, S., Ohkame, T., Sakuma, T., Tanaka, M., Amano, N., Watanabe, A., Sakurai, H., Yamamoto, T., Yamanaka, S., & Hotta, A. (2015). Precise correction of the dystrophin gene in duchenne muscular dystrophy patient induced pluripotent stem cells by TALEN and CRISPR-Cas9. Stem Cell Reports, 4(1), 143–154. https://doi.org/10.1016/j.stemcr.2014.10.013

Li, Q., Chen, J., Minton, N. P., Zhang, Y., Wen, Z., Liu, J., Yang, H., Zeng, Z., Ren, X., Yang, J., Gu, Y., Jiang, W., Jiang, Y., & Yang, S. (2016). CRISPR-based genome editing and expression control systems in Clostridium acetobutylicum and Clostridium beijerinckii. Biotechnology Journal, 11(7), 961–972. https://doi.org/10.1002/biot.201600053

Liang, G., & Zhang, Y. (2013). Genetic and epigenetic variations in iPSCs: Potential causes and implications for application. Cell Stem Cell, 13(2), 149–159. https://doi.org/10.1016/j.stem.2013.07.001

Liang, X., Potter, J., Kumar, S., Zou, Y., Quintanilla, R., Sridharan, M., Carte, J., Chen, W., Roark, N., Ranganathan, S., Ravinder, N., & Chesnut, J. D. (2015). Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. Journal of Biotechnology, 208, 44–53. https://doi.org/10.1016/j.jbiotec.2015.04.024

Liao, C., Seo, S.-O., Celik, V., Liu, H., Kong, W., Wang, Y., Blaschek, H., Jin, Y.-S., & Lu, T. (2015). Integrated, systems metabolic picture of acetone-butanol-ethanol fermentation by Clostridium acetobutylicum. Proceedings of the National Academy of Sciences, 112(27), 8505–8510. https://doi.org/10.1073/pnas.1423143112

Lino, C. A., Harper, J. C., Carney, J. P., & Timlin, J. A. (2018). Delivering CRISPR: A review of the challenges and approaches. Drug Delivery, 25(1), 1234–1257. https://doi.org/10.1080/10717544.2018.1474964

Litcofsky, K. D., Afeyan, R. B., Krom, R. J., Khalil, A. S., & Collins, J. J. (2012). Iterative plug-and-play methodology for constructing and modifying synthetic gene networks. Nature Methods, 9(11), 1077–1080. https://doi.org/10.1038/nmeth.2205

Liu, C., Zhang, L., Liu, H., & Cheng, K. (2017). Delivery Strategies of the CRISPR-Cas9 Gene-Editing System for Therapeutic Applications. Journal of Controlled Release : Official Journal of the Controlled Release Society, 266, 17–26. https://doi.org/10.1016/j.jconrel.2017.09.012

Maguire, J. A., Cardenas-Diaz, F. L., Gadue, P., & French, D. L. (2019). Highly Efficient CRISPR-Cas9-Mediated Genome Editing in Human Pluripotent Stem Cells. Current Protocols in Stem Cell Biology, 48(1), e64. https://doi.org/10.1002/cpsc.64

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J. M., Wolf, Y. I., Yakunin, A. F., van der Oost, J., & Koonin, E. V. (2011). Evolution and classification of the CRISPR-Cas systems. Nature Reviews. Microbiology, 9(6), 467–477. https://doi.org/10.1038/nrmicro2577

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., & Church, G. M. (2013). RNA-Guided Human Genome Engineering via Cas9. Science, 339(6121), 823–826. https://doi.org/10.1126/science.1232033

Marraffini, L. A. (2015). CRISPR-Cas immunity in prokaryotes. Nature, 526(7571), 55–61. https://doi.org/10.1038/nature15386

Marraffini, L. A., & Sontheimer, E. J. (2008). CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. Science, 322(5909), 1843–1845. https://doi.org/10.1126/science.1165771

Miyaoka, Y., Chan, A. H., Judge, L. M., Yoo, J., Huang, M., Nguyen, T. D., Lizarraga, P. P., So, P.-L., & Conklin, B. R. (2014). Isolation of single-base genome-edited human iPS cells without antibiotic selection. Nature Methods, 11(3), 291–293. https://doi.org/10.1038/nmeth.2840

Mojica, F. J. M., Diez-Villasenor, C., Garcia-Martinez, J., & Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology, 155(3), 733–740. https://doi.org/10.1099/mic.0.023960-0

Moore, J. C., Atze, K., Yeung, P. L., Toro-Ramos, A. J., Camarillo, C., Thompson, K., Ricupero, C. L., Brenneman, M. A., Cohen, R. I., & Hart, R. P. (2010). Efficient, high-throughput transfection of human embryonic stem cells. Stem Cell Research & Therapy, 1(3), 23. https://doi.org/10.1186/scrt23

Mosberg, J. A., Lajoie, M. J., & Church, G. M. (2010). Lambda red recombineering in Escherichia coli occurs through a fully single-stranded intermediate. Genetics, 186(3), 791–799. https://doi.org/10.1534/genetics.110.120782

Mougiakos, I., Bosma, E. F., de Vos, W. M., van Kranenburg, R., & van der Oost, J. (2016). Next Generation Prokaryotic Engineering: The CRISPR-Cas Toolkit. Trends in Biotechnology, 34(7), 575–587. https://doi.org/10.1016/j.tibtech.2016.02.004

Nielsen, A. A. K., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E. A., Ross, D., Densmore, D., & Voigt, C. A. (2016). Genetic circuit design automation. Science, 352(6281). https://doi.org/10.1126/science.aac7341

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K. Y., Shimatani, Z., & Kondo, A. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science, 353(6305), aaf8729. https://doi.org/10.1126/science.aaf8729

Niu, N., & Wang, L. (2015). In vitro human cell line models to predict clinical response to anticancer drugs. Pharmacogenomics, 16(3), 273–285. https://doi.org/10.2217/pgs.14.170

Norrman, K., Fischer, Y., Bonnamy, B., Sand, F. W., Ravassard, P., & Semb, H. (2010). Quantitative Comparison of Constitutive Promoters in Human ES cells. PLOS ONE, 5(8), e12413. https://doi.org/10.1371/journal.pone.0012413

Omole, A. E., & Fakoya, A. O. J. (2018). Ten years of progress and promise of induced pluripotent stem cells: Historical origins, characteristics, mechanisms, limitations, and potential applications. PeerJ, 6, e4370. https://doi.org/10.7717/peerj.4370

Pál, C., Papp, B., & Pósfai, G. (2014). The dawn of evolutionary genome engineering. Nature Reviews Genetics, 15(7), 504–512. https://doi.org/10.1038/nrg3746

Perli, S. D., Cui, C. H., & Lu, T. K. (2016). Continuous genetic recording with self-targeting CRISPR-Cas in human cells. Science, 353(6304), aag0511. https://doi.org/10.1126/science.aag0511

Pósfai, G., Plunkett, G., Fehér, T., Frisch, D., Keil, G. M., Umenhoffer, K., Kolisnychenko, V., Stahl, B., Sharma, S. S., Arruda, M. de, Burland, V., Harcum, S. W., & Blattner, F. R. (2006). Emergent Properties of Reduced-Genome Escherichia coli. Science, 312(5776), 1044–1046. https://doi.org/10.1126/science.1126439

Raman, S., Brookhouser, N., & Brafman, D. A. (2020). Using human induced pluripotent stem cells (hiPSCs) to investigate the mechanisms by which Apolipoprotein E (APOE) contributes to Alzheimer's disease (AD) risk. Neurobiology of Disease, 138, 104788. https://doi.org/10.1016/j.nbd.2020.104788

Ran, F. A., Hsu, P. D., Lin, C.-Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., Scott, D. A., Inoue, A., Matoba, S., Zhang, Y., & Zhang, F. (2013). Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Cell, 155(2), 479–480. https://doi.org/10.1016/j.cell.2013.09.040

Rees, H. A., & Liu, D. R. (2018). Base editing: Precision chemistry on the genome and transcriptome of living cells. Nature Reviews Genetics, 19(12), Article 12. https://doi.org/10.1038/s41576-018-0059-1

Reinhardt, P., Schmid, B., Burbulla, L. F., Schöndorf, D. C., Wagner, L., Glatza, M., Höing, S., Hargus, G., Heck, S. A., Dhingra, A., Wu, G., Müller, S., Brockmann, K., Kluba, T., Maisel, M., Krüger, R., Berg, D., Tsytsyura, Y., Thiel, C. S., … Sterneckert, J. (2013). Genetic correction of a LRRK2 mutation in human iPSCs links parkinsonian neurodegeneration to ERK-dependent changes in gene expression. Cell Stem Cell, 12(3), 354–367. https://doi.org/10.1016/j.stem.2013.01.008

Riesenberg, S., & Maricic, T. (2018). Targeting repair pathways with small molecules increases precise genome editing in pluripotent stem cells. Nature Communications, 9(1), Article 1. https://doi.org/10.1038/s41467-018-04609-7

Sander, J. D., & Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nature Biotechnology, 32(4), 347–355. https://doi.org/10.1038/nbt.2842

Santos, C. N. S., Regitsky, D. D., & Yoshikuni, Y. (2013). Implementation of stable and complex biological systems through recombinase-assisted genome engineering. Nature Communications, 4, 2503. https://doi.org/10.1038/ncomms3503

Santos, C. N. S., & Yoshikuni, Y. (2014). Engineering complex biological systems in bacteria through recombinase-assisted genome engineering. Nature Protocols, 9(6), 1320–1336. https://doi.org/10.1038/nprot.2014.084

Schene, I. F., Joore, I. P., Oka, R., Mokry, M., van Vugt, A. H. M., van Boxtel, R., van der Doef, H. P. J., van der Laan, L. J. W., Verstegen, M. M. A., van Hasselt, P. M., Nieuwenhuis, E. E. S., & Fuchs, S. A. (2020). Prime editing for functional repair in patient-derived disease models. Nature Communications, 11(1), Article 1. https://doi.org/10.1038/s41467-020-19136-7

Sharon, E., Chen, S.-A. A., Khosla, N. M., Smith, J. D., Pritchard, J. K., & Fraser, H. B. (2018). Functional Genetic Variants Revealed by Massively Parallel Precise Genome Editing. Cell, 175(2), 544-557.e16. https://doi.org/10.1016/j.cell.2018.08.057

Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., & Zhang, F. (2015). Rationally engineered Cas9 nucleases with improved specificity. Science, 351(6268), 84–88. https://doi.org/10.1126/science.aad5227

Smanski, M. J., Bhatia, S., Zhao, D., Park, Y., Woodruff, L. B. A., Giannoukos, G., Ciulla, D., Busby, M., Calderon, J., Nicol, R., Gordon, D. B., Densmore, D., & Voigt, C. A. (2014). Functional optimization of gene clusters by combinatorial design and assembly. Nature Biotechnology, 32(12), 1241–1249. https://doi.org/10.1038/nbt.3063

Standage-Beier, K., Brookhouser, N., Balachandran, P., Zhang, Q., Brafman, D. A., & Wang, X. (2019). RNA-Guided Recombinase-Cas9 Fusion Targets Genomic DNA Deletion and Integration. The CRISPR Journal, 2(4), 209–222. https://doi.org/10.1089/crispr.2019.0013

Standage-Beier, K., & Wang, X. (2017). Genome reprogramming for synthetic biology. Frontiers of Chemical Science and Engineering, 11(1), 37–45. https://doi.org/10.1007/s11705-017-1618-2

Standage-Beier, K., Zhang, Q., & Wang, X. (2015). Targeted Large-Scale Deletion of Bacterial Genomes Using CRISPR-Nickases. ACS Synthetic Biology, 4(11), 1217–1225. https://doi.org/10.1021/acssynbio.5b00132

Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J., & Mateo, J. L. (2015). CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. PloS One, 10(4), e0124633. https://doi.org/10.1371/journal.pone.0124633

Steyer, B., Bu, Q., Cory, E., Jiang, K., Duong, S., Sinha, D., Steltzer, S., Gamm, D., Chang, Q., & Saha, K. (2018). Scarless Genome Editing of Human Pluripotent Stem Cells via Transient Puromycin Selection. Stem Cell Reports, 10(2), 642–654. https://doi.org/10.1016/j.stemcr.2017.12.004

St-Pierre, F., Cui, L., Priest, D. G., Endy, D., Dodd, I. B., & Shearwin, K. E. (2013). One-Step Cloning and Chromosomal Integration of DNA. ACS Synthetic Biology, 2(9), 537–541. https://doi.org/10.1021/sb400021j

Supharattanasitthi, W., Carlsson, E., Sharif, U., & Paraoan, L. (2019). CRISPR/Cas9-mediated one step bi-allelic change of genomic DNA in iPSCs and human RPE cells in vitro with dual antibiotic selection. Scientific Reports, 9(1), 174. https://doi.org/10.1038/s41598-018-36740-2

Tan, J., Zhang, F., Karcher, D., & Bock, R. (2019). Engineering of high-precision base editors for site-specific single nucleotide replacement. Nature Communications, 10(1), Article 1. https://doi.org/10.1038/s41467-018-08034-8

Tekel, S. J., Brookhouser, N., Standage-Beier, K., Wang, X., & Brafman, D. A. (2021). Cytosine and adenosine base editing in human pluripotent stem cells using transient reporters for editing enrichment. Nature Protocols, 16(7), 3596–3624. https://doi.org/10.1038/s41596-021-00552-y

Temme, K., Zhao, D., & Voigt, C. A. (2012). Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. Proceedings of the National Academy of Sciences of the United States of America, 109(18), 7085–7090. https://doi.org/10.1073/pnas.1120788109

Tsai, S. Q., Wyvekens, N., Khayter, C., Foden, J. A., Thapar, V., Reyon, D., Goodwin, M. J., Aryee, M. J., & Joung, J. K. (2014). Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nature Biotechnology, 32(6), 569–576. https://doi.org/10.1038/nbt.2908

Tsai, S. Q., Zheng, Z., Nguyen, N. T., Liebers, M., Topkar, V. V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A. J., Le, L. P., Aryee, M. J., & Joung, J. K. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nature Biotechnology, 33(2), 187–197. https://doi.org/10.1038/nbt.3117

Tsarmpopoulos, I., Gourgues, G., Blanchard, A., Vashee, S., Jores, J., Lartigue, C., & Sirand-Pugnet, P. (2016). In-Yeast Engineering of a Bacterial Genome Using CRISPR/Cas9. ACS Synthetic Biology, 5(1), 104–109. https://doi.org/10.1021/acssynbio.5b00196

Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., & Church, G. M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. Nature, 460(7257), 894–898. https://doi.org/10.1038/nature08187

Wang, L.-Z., Wu, F., Flores, K., Lai, Y.-C., & Wang, X. (2016). Build to understand: Synthetic approaches to biology. Integrative Biology, 8(4), 394–408. https://doi.org/10.1039/C5IB00252D

Wang, Y., Zhang, Z.-T., Seo, S.-O., Choi, K., Lu, T., Jin, Y.-S., & Blaschek, H. P. (2015). Markerless chromosomal gene deletion in Clostridium beijerinckii using CRISPR/Cas9 system. Journal of Biotechnology, 200, 1–5. https://doi.org/10.1016/j.jbiotec.2015.02.005

Wright, A. V., Nuñez, J. K., & Doudna, J. A. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. Cell, 164(1–2), 29–44. https://doi.org/10.1016/j.cell.2015.12.035

Wu, F., Menn, D. J., & Wang, X. (2014). Quorum-Sensing Crosstalk-Driven Synthetic Circuits: From Unimodality to Trimodality. Chemistry & Biology, 21(12), 1629–1638. https://doi.org/10.1016/j.chembiol.2014.10.008

Wu, M., Su, R.-Q., Li, X., Ellis, T., Lai, Y.-C., & Wang, X. (2013). Engineering of regulated stochastic cell fate determination. Proceedings of the National Academy of Sciences, 201305423. https://doi.org/10.1073/pnas.1305423110

Xia, X., Zhang, Y., Zieth, C. R., & Zhang, S.-C. (2007). Transgenes Delivered by Lentiviral Vector Are Suppressed in Human Embryonic Stem Cells in a Promoter-Dependent Manner. Stem Cells and Development, 16(1), 167–176.

Xu, X., & Zhong, Z. (2013). Disease modeling and drug screening for neurological diseases using human induced pluripotent stem cells. Acta Pharmacologica Sinica, 34(6), Article 6. https://doi.org/10.1038/aps.2013.63

Yokobayashi, Y., Weiss, R., & Arnold, F. H. (2002). Directed evolution of a genetic circuit. Proceedings of the National Academy of Sciences, 99(26), 16587–16591. https://doi.org/10.1073/pnas.252535999

Yu, C., Liu, Y., Ma, T., Liu, K., Xu, S., Zhang, Y., Liu, H., La Russa, M., Xie, M., Ding, S., & Qi, L. S. (2015). Small molecules enhance CRISPR genome editing in pluripotent stem cells. Cell Stem Cell, 16(2), 142–147. https://doi.org/10.1016/j.stem.2015.01.003

Zhou, C., Sun, Y., Yan, R., Liu, Y., Zuo, E., Gu, C., Han, L., Wei, Y., Hu, X., Zeng, R., Li, Y., Zhou, H., Guo, F., & Yang, H. (2019). Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. Nature, 571(7764), 275–278. https://doi.org/10.1038/s41586-019-1314-0

Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., Qiu, J.-L., Wang, D., & Gao, C. (2017). Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. Nature Biotechnology, 35(5), 438–440. https://doi.org/10.1038/nbt.3811

Zuris, J. A., Thompson, D. B., Shu, Y., Guilinger, J. P., Bessen, J. L., Hu, J. H., Maeder, M. L., Joung, J. K., Chen, Z.-Y., & Liu, D. R. (2015). Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. Nature Biotechnology, 33(1), 73–80. https://doi.org/10.1038/nbt.3081

APPENDIX A

SUPPLEMENTAL: TRANSIENT REPORTER FOR EDITING ENRICHMENT

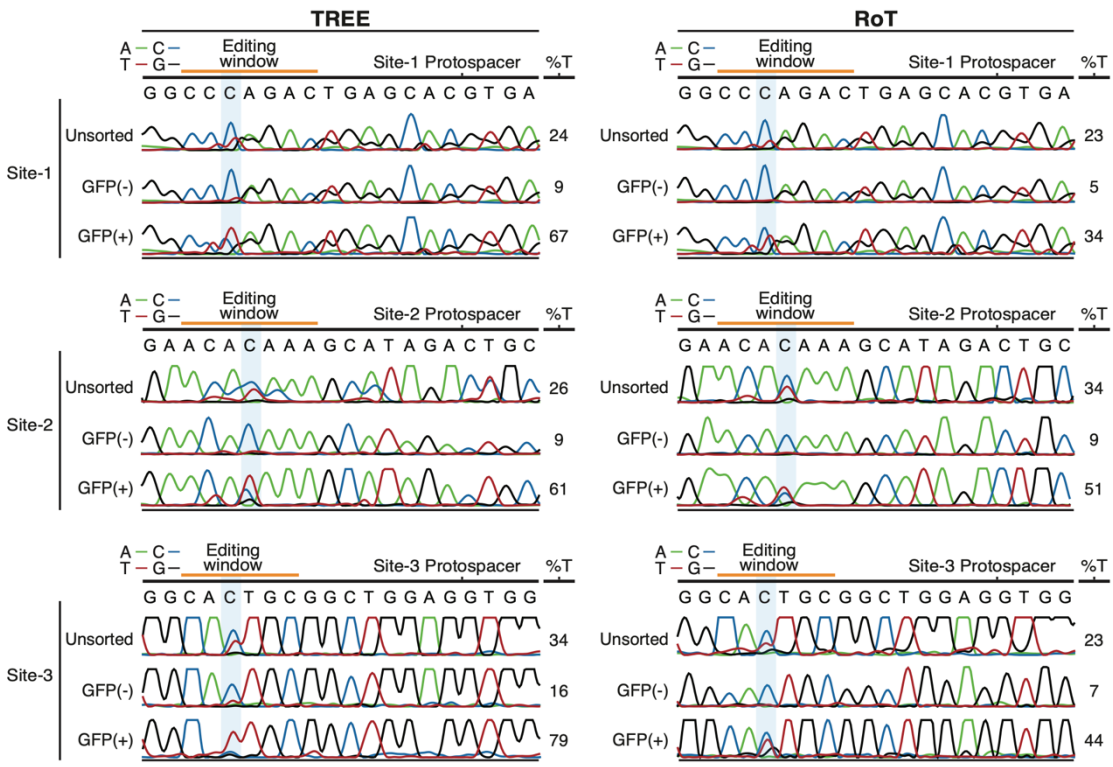


Figure A-1. Representative Sanger sequencing chromatographs of edited HEK293 cells enriched using TREE- and RoT-based approaches. Sanger sequencing chromatographs of Site-1, Site-2, and Site-3 of GFP-positive, GFP-negative, and unsorted cell populations isolated with TREE- and RoT-based approaches.

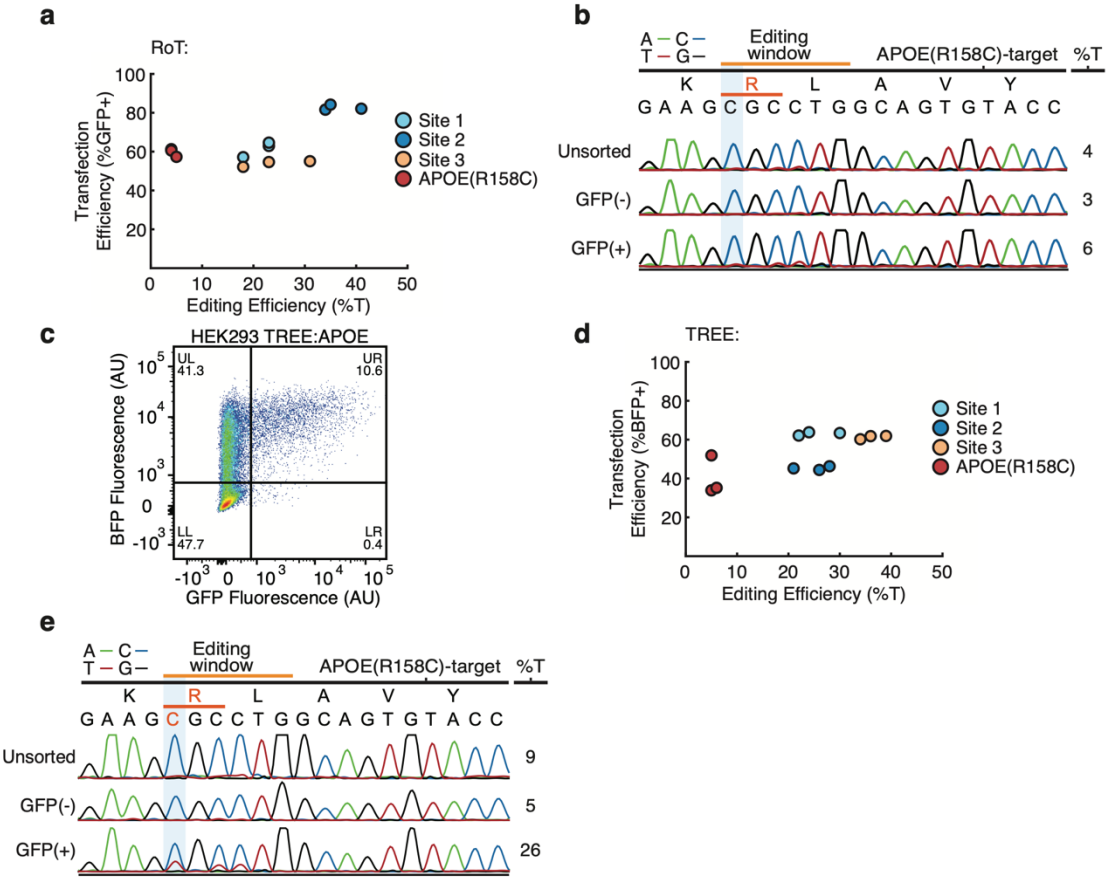


Figure A-2. TREE allows for base editing of refractory APOE(R158) locus in HEK293 cells. (a) HEK293 cells were transfected with pEF-GFP, pCMV-BE4-Gam, and sg(TS). Comparison of transfection efficiency (percentage of GFP-positive cells) and editing efficiency (percentage of C-to-T conversion at target nucleotide) in unsorted cell populations at Site-1, Site-2, Site-2, and APOE(R158) locus. (b) Representative Sanger sequencing chromatographs of APOE(R158) locus in GFP-positive, GFP-negative, and unsorted cell populations isolated with RoT-based methods. (c) Representative flow cytometry plot of HEK293 cells in which TREE was applied targeting the APOE(R158) locus. (d) HEK293 cells were transfected with pEF-BFP, pCMV-BE4-Gam, and pDT-sgRNA. Comparison of transfection efficiency (percentage of BFP-positive cells) and editing efficiency (percentage of C-to-T conversion at target nucleotide) in unsorted cell populations at Site-1, Site-2, Site-3, and APOE(R158) locus. (e) Representative Sanger sequencing chromatographs of APOE(R158) locus in GFP-positive, GFP-negative, and unsorted cell populations isolated with TREE-based methods.

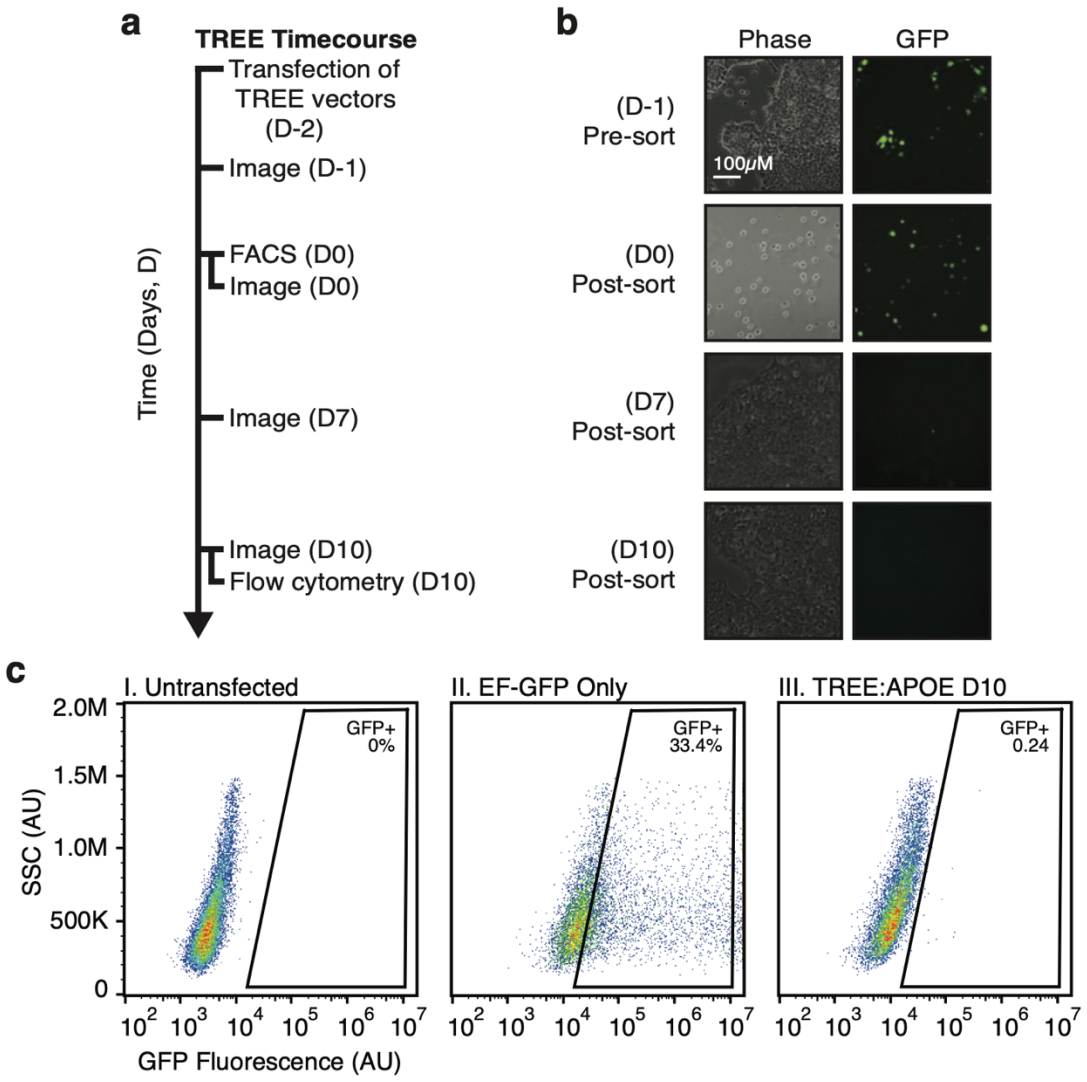


Figure A-3. TREE fluorescent output in HEK293 cells is transient. **(a)** HEK293 cells were transfected with pEF-BFP, pCMV-BE4-Gam, and pDT-sgRNA and GFP-positive cells were isolated by flow cytometry. Replated GFP-positive cells were analyzed by fluorescent microscopy and flow cytometry at various time points post-sorting. **(b)** Representative fluorescent microscopy images of cells prior to cell sorting (D-1, Pre-sort) and various time points (D0, D7, D10) after sorting. **(c)** Representative flow cytometry plots of (i) untransfected HEK293 cells, (ii) pEF-GFP transfected HEK293 cells, and (iii) TREE-enriched GFP-positive HEK293 cells 10 days after sorting.

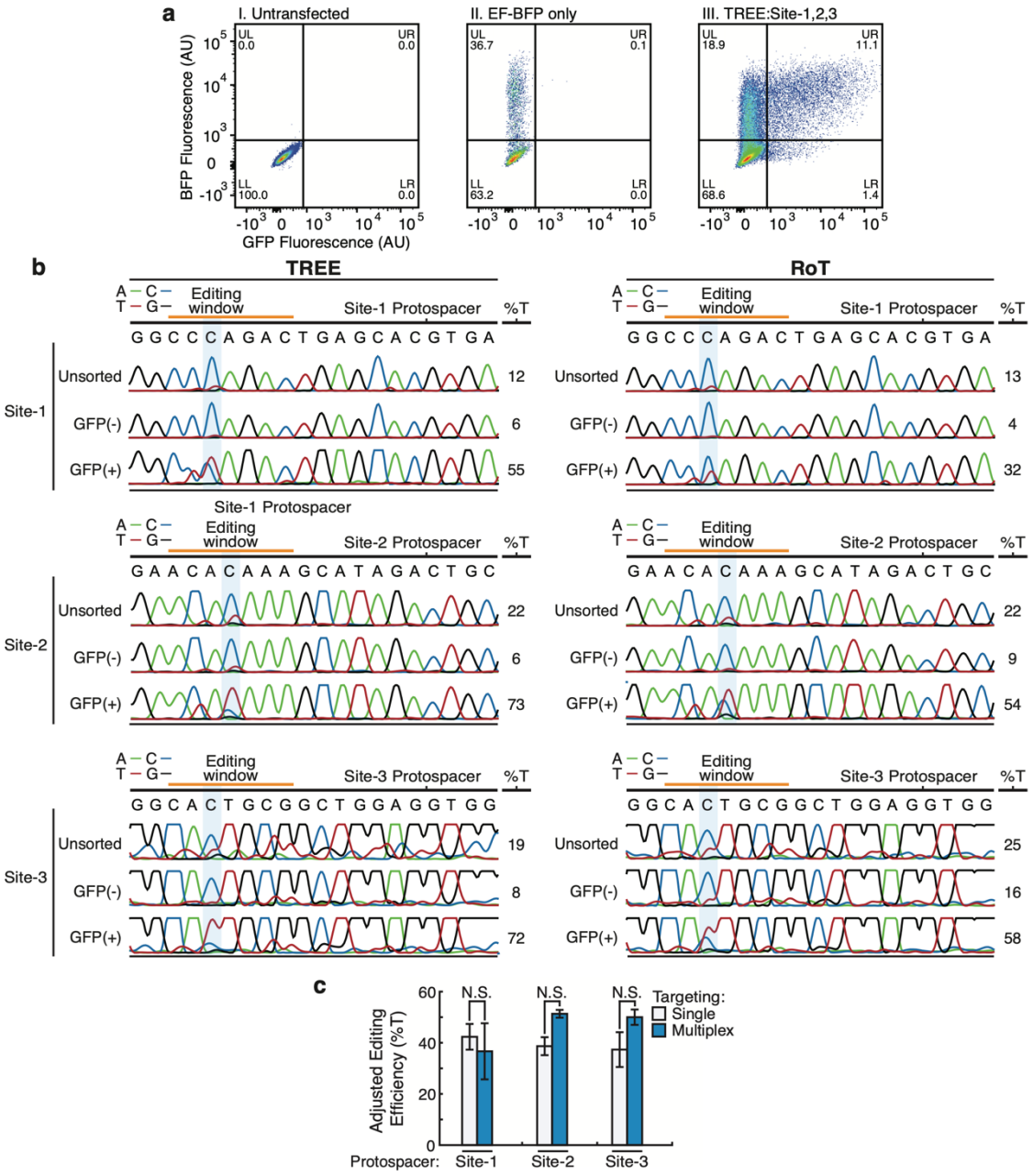


Figure A-4. Analysis of multiplexed edited HEK293 cells using TREE- and RoT-based methods. **(a)** Representative flow cytometry plot of HEK293 cells in which multiplex TREE was applied simultaneously targeting Site-1, Site-2, and Site-3. **(b)** Representative Sanger sequencing chromatographs of the Site-1, Site-2, and Site-3 loci in GFP-positive, GFP-negative, and unsorted cell populations isolated with TREE multiplex-based methods. **(c)** Comparison of base editing efficiencies at Site-1, Site-2, and Site-3 in GFP-positive, GFP-negative, and unsorted cell populations using TREE-based methods to target these sites individually or in a multiplexed manner. n=3; N.S. = not significant. (**d**) Sanger sequences of HEK293T clones isolated with single-nucleotide changes within the editing window

BG-OT1:	C	C	C	T	A	C	A	T	C	G	T	G	C	A	G	T	G	C	T	T
Untransfected:			94			95														
TREE:			95			95														
BG-OT2:	C	C	C	C	A	A	G	T	A	G	T	G	C	A	G	T	G	C	T	T
Untransfected:			97	90																
TREE:			97	94																
BG-OT3:	A	A	C	C	A	A	G	A	T	G	T	G	C	A	G	T	G	C	T	T
Untransfected:			95	93																
TREE:			91	96																
BG-OT4:	A	A	C	C	A	G	C	G	C	C	T	G	C	A	G	T	G	C	T	T
Untransfected:			95	96			96													
TREE:			95	97			95													
BG-OT5:	C	C	C	C	A	T	G	G	C	T	T	G	C	T	G	T	G	C	T	T
Untransfected:			97	91																
TREE:			98	96																
Site1-OT1:	C	A	C	C	C	A	G	A	C	T	G	A	G	C	A	C	G	T	G	C
Untransfected:			98	97	98															
RoT:			97	98	98															
TREE:			98	96	97															
Site1-OT2:	G	A	C	A	C	A	G	A	C	C	G	G	G	C	A	C	G	T	G	A
Untransfected:			98		95															
RoT:			98		97															
TREE:			98		97															
Site1-OT3:	A	G	C	T	C	A	G	A	C	T	G	A	G	C	A	A	G	T	G	A
Untransfected:			98		98															
RoT:			97		97															
TREE:			89		97															
Site1-OT4:	A	G	A	C	C	A	G	A	C	T	G	A	G	C	A	A	G	A	G	A
Untransfected:				96	96															
RoT:				94	96															
TREE:				92	93															
Site1-OT5:	G	A	G	C	C	A	G	A	A	T	G	A	G	C	A	C	G	T	G	A
Untransfected:				96	95															
RoT:				96	95															
TREE:				95	94															
Site2-OT1:	G	A	A	C	A	C	A	A	T	G	C	A	T	A	G	A	T	T	G	C
Untransfected:				80		92														
RoT:				97		96														
TREE:				94		94														
Site2-OT2:	G	C	A	G	T	C	T	A	T	G	C	T	T	T	A	T	G	T	T	T
Untransfected:						90														
RoT:						88														
TREE:						89														

Continued on next page.

Site3-OT1:	T	G	C	A	C	T	G	C	G	G	C	C	G	G	A	G	G	A	G	G
Untransfected:			91		94			92												
RoT:			93		93			93												
TREE:			89		90			92												
Site3-OT2:	G	G	C	T	C	T	G	C	G	G	C	T	G	G	A	G	G	G	G	G
Untransfected:			90		84			93												
RoT:			95		91			95												
TREE:			94		82			94												
Site3-OT3:	G	G	C	A	C	G	A	C	G	G	C	T	G	G	A	G	G	T	G	G
Untransfected:			92		96			95												
RoT:			93		88			93												
TREE:			91		91			93												
Site3-OT4:	G	G	C	A	T	C	A	C	G	G	C	T	G	G	A	G	G	T	G	G
Untransfected:			92			97		96												
RoT:			94			90		94												
TREE:			94			91		92												
Site3-OT4:	G	G	C	G	C	T	G	C	G	G	C	G	G	G	A	G	G	T	G	G
Untransfected:			98		94			95												
RoT:			96		94			98												
TREE:			97		82			98												

Figure A-5. Analysis of off-target sites in HEK293 cells using TREE- and RoT-based methods. GFP-positive cell populations isolated from TREE and RoT approaches we PCR-amplified and subject to Sanger sequencing on the top predicted off-target loci for the sgRNA sequences for sg(BG) and genomic Sites 1-3. The C nucleotides in red text are potential Cs that can undergo C-to-T conversion within the editing window in the protospacer.

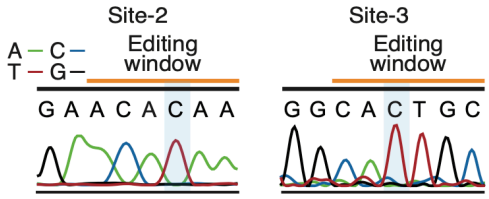


Figure A-6. Identification of exclusively editing events in clonal HEK293 cells**.** Representative Sanger sequencing chromatographs of clonal cell populations that contain edits exclusively at the target C and not any other Cs within the editing window.

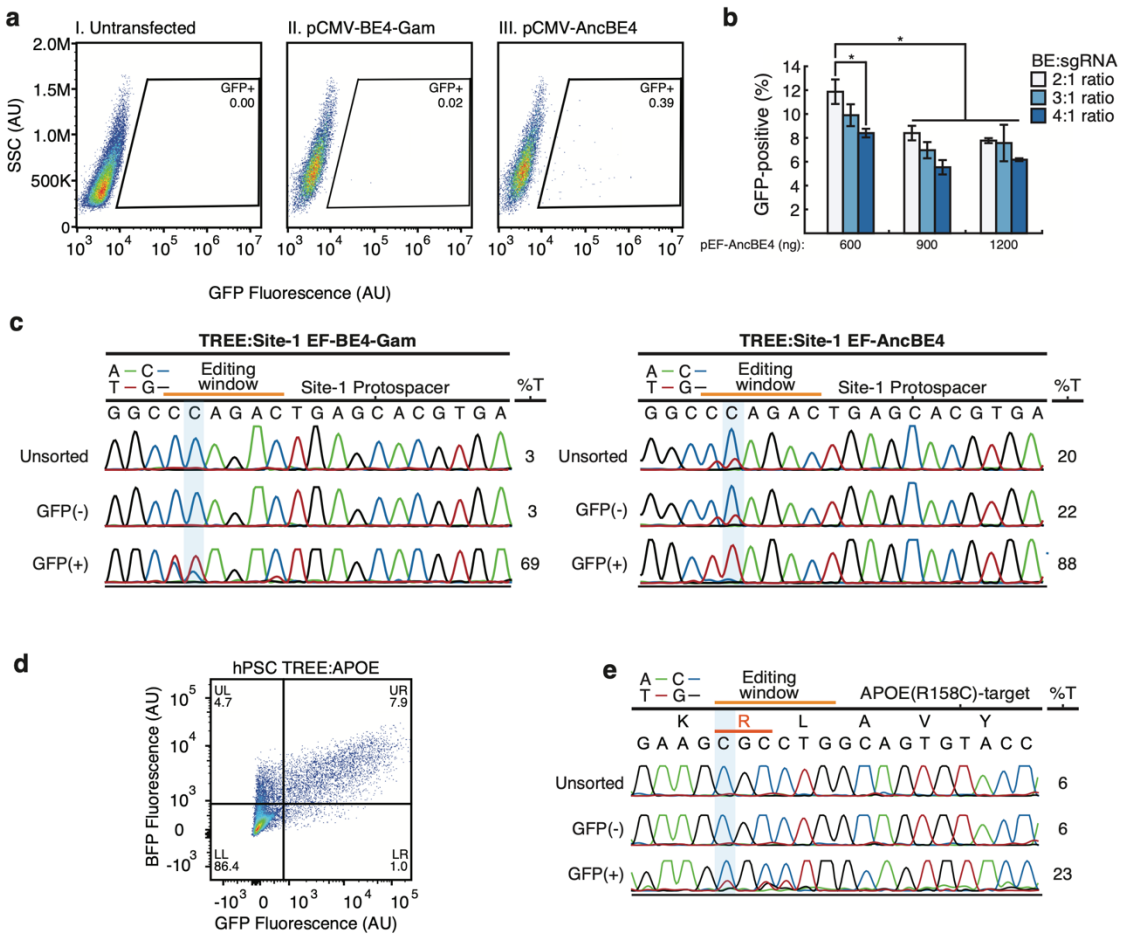


Figure A-7. TREE allows for base editing in hPSCs. **(a)** Representative flow cytometry plots in which TREE was employed in hPSCs utilizing (i) pCMV-BE4-Gam or (ii) pCMV-AncBE4. **(b)** Editing efficiency (percentage GFP-positive cells) of targeting in hPSCs line with various amounts of pEF-AncBE4 plasmid and ratios with the sg(BG) vector. n = 3, * = p<0.05. **(c)** Representative Sanger sequencing chromatographs of Site-1 in GFP-positive, GFP-negative, and unsorted cell populations isolated with TREE- and RoT-based methods in which pEF-BE4-Gam or pEF-AncBE4 was utilized. **(d)** Representative flow cytometry plot of hPSCs cells in which TREE was applied targeting the APOE(R158) locus. **(e)** Representative Sanger sequencing chromatographs of APOE(R158) locus in GFP-positive, GFP-negative, and unsorted cell populations isolated with TREE-based methods.

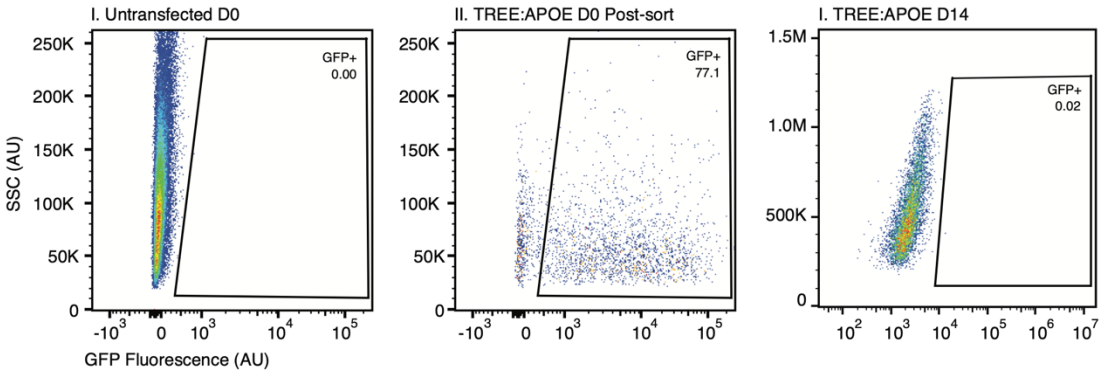


Figure A-8. TREE fluorescent output in hPSCs is transient. Representative flow cytometry plots of (i) untransfected hPSCs, (ii) TREE-enriched GFP-positive hPSCs 0 days (iii) 14 days after sorting.

a

	Untransfected Site-1		TREE:Site-1	
	Editing window / PAM	% of Reads	Editing window / PAM	% of Reads
WT	GGCCCAGACTGAGCACGTGATGG	99.7	GGCCCAGACTGAGCACGTGATGG	5.2
18	GGTCCAGACTGAGCACGTGATGG	0.1	GGTCCAGACTGAGCACGTGATGG	0.0
17	GGCTCAGACTGAGCACGTGATGG	0.1	GGCTCAGACTGAGCACGTGATGG	0.1
16	GGCCTAGACTGAGCACGTGATGG	0.1	GGCCTAGACTGAGCACGTGATGG	0.3
12	GGCCCAGATTGAGCACGTGATGG	0.0	GGCCCAGATTGAGCACGTGATGG	0.0
18,17	GGTTCAGACTGAGCACGTGATGG	0.0	GGTTCAGACTGAGCACGTGATGG	0.0
18,16	GGTCTAGACTGAGCACGTGATGG	0.0	GGTCTAGACTGAGCACGTGATGG	0.0
18,12	GGTCCAGATTGAGCACGTGATGG	0.0	GGTCCAGATTGAGCACGTGATGG	0.0
17,16	GGCTTAGACTGAGCACGTGATGG	0.0	GGCTTAGACTGAGCACGTGATGG	85.2
17,12	GGCTCAGATTGAGCACGTGATGG	0.0	GGCTCAGATTGAGCACGTGATGG	0.1
16,12	GGCCTAGATTGAGCACGTGATGG	0.0	GGCCTAGATTGAGCACGTGATGG	0.2
18,17,16	GGTTTAGACTGAGCACGTGATGG	0.0	GGTTTAGACTGAGCACGTGATGG	6.2
17,16,12	GGCTTAGATTGAGCACGTGATGG	0.0	GGCTTAGATTGAGCACGTGATGG	2.4
18,16,12	GGTCTAGATTGAGCACGTGATGG	0.0	GGTCTAGATTGAGCACGTGATGG	0.0
18,17,12	GGTTCAGATTGAGCACGTGATGG	0.0	GGTTCAGATTGAGCACGTGATGG	0.0
18,17,16,12	GGTTTAGATTGAGCACGTGATGG	0.0	GGTTTAGATTGAGCACGTGATGG	0.1

b

	Untransfected APOE		TREE:APOE	
	Editing window / PAM	% of Reads	Editing window / PAM	% of Reads
WT	GAAGCGCCTGGCAGTGTACCAGG	93.4	GAAGCGCCTGGCAGTGTACCAGG	64.2
16	GAAGTGCCTGGCAGTGTACCAGG	2.5	GAAGTGCCTGGCAGTGTACCAGG	5.2
14	GAAGCGTCTGGCAGTGTACCAGG	0.1	GAAGCGTCTGGCAGTGTACCAGG	1.1
13	GAAGCGCTTGGCAGTGTACCAGG	0.1	GAAGCGCTTGGCAGTGTACCAGG	1.0
16,14	GAAGTGTCTGGCAGTGTACCAGG	1.1	GAAGTGTCTGGCAGTGTACCAGG	1.2
16,13	GAAGTGCTTGGCAGTGTACCAGG	0.4	GAAGTGCTTGGCAGTGTACCAGG	1.0
14,13	GAAGCGTTTGGCAGTGTACCAGG	0.5	GAAGCGTTTGGCAGTGTACCAGG	7.0
16,14,13	GAAGTGTTTGGCAGTGTACCAGG	1.9	GAAGTGTTTGGCAGTGTACCAGG	19.2

Figure A-9. Next generation sequencing (NGS) analysis of allelic outcomes at target sites in hPSCs. NGS analysis for the target site when TREE-based methods were applied to edit Site-1 or the APOE(R158) in hPSCs. The number to left of the allelic outcome indicates the position relative to the PAM. Abbreviation: WT = wild-type unedited locus.

Table A-1. List of sgRNA sequences used in this study.

Site	Sequence (5'→3')
Site-1	GGCCCAGACTGAGCACGTGA
Site-2	GAACACAAAGCATAGACTGC
Site-3	GGCACTGCGGCTGGAGGTGG
APOE(R158)	GAAGCGCCTGGCAGTGTACC
BFP(H66Y)	GACCCACGGCGTGCAGTGCTT
C1ORF228	GTGCTGTTAGCACCCTGGAAA

Table A-2. List of primers used in this study to amplify on-target sites for Sanger sequencing.

Primer	Forward Sequence (5'→3')	Reverse Sequence (5'→3')
Site-1	ATGTGGGCTGCCTAGAAAGG	CCCAGCCAAACTTGTCAACC
Site-2	CCAGCCCCATCTGTCAAACT	TGAATGGATTCCTTGGAAACAATGA
Site-3	TGGTCTTCTTTCCCCTCCCCTGCCCTCC	GGCCTGGAGGCGGGGGCTCAGAGA
APOE(R158)	GGACGAGACCATGAAGGAGTTGAAGGC	CCACCTGCTCCTTCACCTCGTCCAG

Table A-3. PCR conditions for each target site to be analyzed by Sanger sequencing.

Target	Initial denature time and temperature	Denature time and temperature	Annealing time and temperature	Extension time and temperature	Final extension time and temperature
		40 cycles			
Site-1	98°C, 45 seconds	98°C, 10 seconds	54°C, 5 seconds	72°C , 20 seconds	72°C, 10 minutes
Site-2	98°C, 45 seconds	98 °C, 10 seconds	56°C, 5 seconds	72°C , 20 seconds	72°C, 10 minutes
Site-3	98°C, 45 seconds	98°C, 10 seconds	56°C, 5 seconds	72°C , 20 seconds	72°C, 10 minutes
APOE(R158)	98°C, 45 seconds	98°C, 10 seconds	62°C, 5 seconds	72°C , 20 seconds	72°C, 10 minutes

Table A-4. Parameters for EditR analysis.

Target Site	Sequencing Direction	Protospacer	5’ bound	3’ bound
Site-1	Forward	GGCCCAGACTGAGCACGTGA	GGCCTGGGTCAA	TTCCTTTCCTCTG
	Reverse	TCACGTGCTCAGTCTGGGCC	GAGGAAAGGAAGCCCTGCT	CAGGCCAGGGCTGGA
Site-2	Forward	GAACACAAAGCATAGACTGC	CCCGCTGGCCCTGT	TCAGGCTGGCCCGC
	Reverse	GCAGTCTATGCTTTGTGTTC	CCAGCCCGCTGGCCCTGTA	AGCTATTCAGGCT
Site-3	Forward	GTGGCACTGCGGCTGGAGGT	GATGACAGGCAGGGGCA	CAGCACCAGA
	Reverse	ACCTCCAGCCGCAGTGCC	CCGCGGTGCCCCTGCCT	AAGCGGAGACTCTGGTGC
APOE(R158)	Forward	GAAGCGCCTGGCAGTGTACC	CTGCGCAAGCTGCG	TCGGCGCCCTCGCG
	Reverse	GGTACACTGCCAGGCGCTTC	GGATGGCGCTGA	GCCTCGCCTCCCACC

Table A-5. List of primers used in this study to amplify off-target sites.

Primer	Forward Sequence (5'→3')	Reverse Sequence (5'→3')
BG-OT1	GATGCGCTTCCGGAAGACC	GCTTCTTGAGCTTCTCAGCG
BG-OT2	GGTAGCATGTTCAGGCACCAG	CATCCCTAGTACCGAATCCCATATAGC
BG-OT3	CATCCTCCCACCTAAGCCTTTCAA	TTGAGTTAATAGCATTATAACAATTTCCACA
BG-OT4	ACTCCTTACAACCGGAAGGCAAAC	TGGACGTGGTGAAGCCCGTGGTG
BG-OT5	TAGGTCTCTAGGGGGCCTCTG	AGGCTGCCCAACAGCCCCACT
Site1-OT1	TCCCCTGTTGACCTGGAGAA	CACTGTACTTGCCCTGACCA
Site1-OT2	TGAGATGTGGGCAGAAGGG	TTGGTGTTGACAGGGAGCAA
Site1-OT3	GTCCAAAGGCCCAAGAACCT	TGAGAGGGAACAGAAGGGCT
Site1-OT4	GCTCATCTTAATCTGCTCAGCC	TCCTAGCACTTTGGAAGGTCG
Site1-OT5	AAAGGAGCAGCTCTTCCTGG	GTCTGCACCATCTCCCACAA
Site2-OT1	GTGTGGAGAGTGAGTAAGCCA	ACGGTAGGATGATTTCAGGCA
Site2-OT2	TTTTTTGGTACTCGAGTGTTATTCAG	CACAAAGCAGTGTAGCTCAGG
Site3-OT1	GGCATGGCTTCTGAGACTCA	CCCCTTGCACTCCCTGTCTTT
Site3-OT2	GAAGAGGCTGCCCATGAGAG	TTTGGCAATGGAGGCATTGG
Site3-OT3	GGTCTGAGGCTCGAATCCTG	CTGTGGCCTCCATATCCCTG
Site3-OT4	TTTCCACCAGAACTCAGCCC	CCTCGGTTCCTCCACAACAC
Site3-OT5	GCAGGGGAGGGATAAAGCAG	CACGGGAAGGACAGGAGAAG

Abbreviations: BG-OT = Off-targets associated with sg(BG), Site1-OT = Off-targets associated with sg(Site-1), Site2-OT = Off-targets associated with sg(Site-2), Site3-OT = Off-targets associated with Sg(Site-3).

Table A-6. List of primers used in this study for NGS analysis.

Primer	Forward Sequence (5'→3')	Reverse Sequence (5'→3')
Site-1	ATGTGGGCTGCCTAGAAAGG	CCCAGCCAAACTTGTCAACC
APOE(R158)	GGACGAGACCATGAAGGAGTTGAAGGC	CCACCTGCTCCTTCACCTCGTCCAG

Table A-7. PCR conditions next generation sequencing analysis.

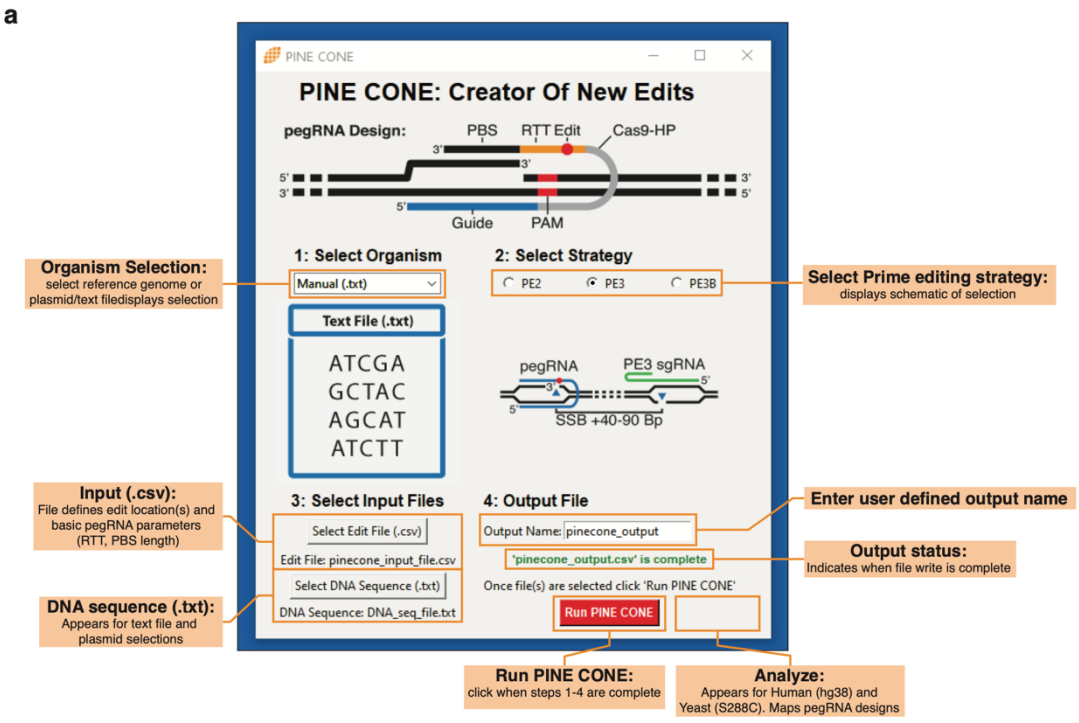
Target	Initial denature time and temperature	Denature time and temperature	Annealing time and temperature	Extension time and temperature	Final extension time and temperature
		40 cycles			
Site-1	98˚C, 45 seconds	98˚C, 10 seconds	54˚C, 5 seconds	72˚C, 20 seconds	72˚C, 10 minutes
APOE(R158)	98˚C, 45 seconds	98˚C, 10 seconds	62˚C, 5 seconds	72˚C, 20 seconds	72˚C, 10 minutes

Table A-8. Comparison of base editing efficiency using RoT-based approaches at the same target loci in this manuscript, Komor et al, and Koblan et al.

	Figure 3E Standage-Beier et al.			**Figure 5C Komar et. al Sci Adv. 2017 Aug 30;3(8)**			**Figure 1C Koblan et. al Nat Biotechnol. 2018 Oct;36(9):843-846**		
	Reporter of Transfection			No Reporter			Reporter of Transfection		
	Unsorted	Reporter-	Reporter+	Unsorted	Reporter-	Reporter+	Unsorted	Reporter-	Reporter+
Site-1 (HEK 3)	21.3±2.9	3.3±2.8	40.7±7.0	~45	N/A	N/A	~38	N/A	~55
Site-2 (HEK 2)	36.6±3.8	13.3±5.9	49.7±5.1	~35	N/A	N/A	~20	N/A	~38
Site-3 (HEK 4)	24.0±6.6	7.6±5.0	45.3±1.5	~45	N/A	N/A	~25	N/A	~40

APPENDIX B

SUPPLEMENTAL: DESIGN AUTOMATION OF CRISPR RNAS



b

pegRNA Name: Names are arbitrary
Chromosome: Edit locus chromsome
Position: Absolute nucleotide coordinate of edit locus
Edit: Sequence of intended edit
RTT Length: longer RTT's increase editing range and are necessary for insertions
PBS Length: Primer sequence length
Notes: Section is copied over to output for record keeping

	A	B	C	D	E	F	G
1	**Name**	**Chromosome**	**Position (Bp)**	**Edit (A,T,C,G)**	**RT (Bp)**	**PBS (Bp)**	**Notes:**
2	HEK3-SNP	9	107422356	A	13	13	Example SNP-pegRNA
3	HEK3-Del	9	107422356	D5	14	13	Example Del-pegRNA
4	HEK3-Ins	9	107422356	iATGGAC	22	13	Example Ins-pegRNA
5							

Figure B-1: PINE-CONE User Interface (UI) and Input (.CSV) format. (**a**) PINE-CONE's UI includes 4 primary steps. (1) Selection of organism determines if sequence is retrieved from a reference genome or is accessed locally via "Manual (.txt)" or "Plasmid (.txt)" selections. (2) Selection of editing strategy determines the design format of 'accessory' sgRNAs. (3) Input file selection or select locally an accompanying DNA sequence text file (.txt). (4) User defined arbitrary name of output file. This file is written in the same path as PINE-CONE. Following input of information, click "Run PINE-CONE". This will initiate API requests and design of pegRNAs. A green status is displayed once the output file is done writing. If pegRNAs are designed for Human (hg38) or Yeast (S288C) an 'analyze' will appear mapping the location of pegRNAs to their respective reference genomes. (**b**) Input file format (.CSV) includes an arbitrary pegRNA name, chromosome, position, edit sequence, RTT length and PBS length preference and user notes. The figure includes an example single base edit, deletion and insertion.

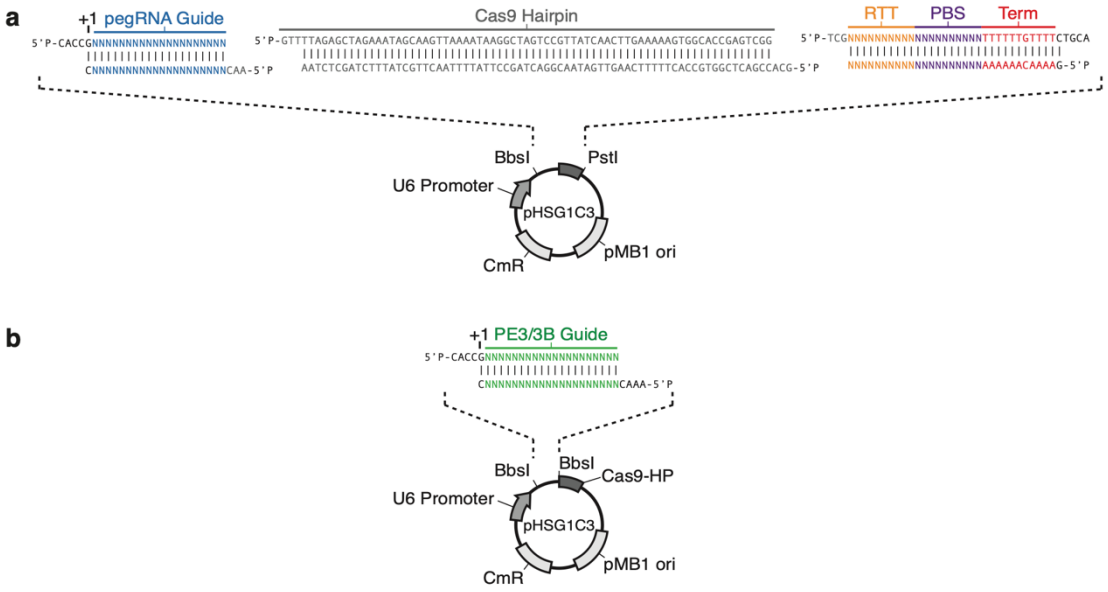


Figure B-2: PINE-CONE oligonucleotide output for cloning of pegRNA and PE3/PE3B sgRNAs. (**a**) Cloning strategy for pegRNAs using pHSG1C3. pHSG1C3 enables cloning of pegRNAs via digestion with BbsI and PstI. 5' phosphorylated (5'P) Oligonucleotide duplexes encoding a guide (blue, +1 is the transcriptional start of the U6 promoter), Cas9 hairpin (gray) and RTT/PBS and U6 terminator (Orange, Purple Blue respectively) are ligated into BbsI and PstI digested pHSG1C3. (**b**) PE3 or PE3b guides (green) are cloned into BbsI digested pHSG1C3. Oligonucleotide formatting can also be found in Supplemental table 2.

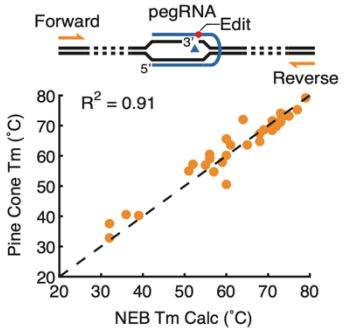


Figure B-3: PINE-CONE design of PCR primers flanking the edit site. PINE-CONE designed primers Tm correlate with a commercially available New England Bio Labs (NEB) Tm calculator.

Figure B-4: pegRNAs designs by PINE-CONE match previously validated pegRNAs. A series of loci previously targeted and validated by *Anzalone et al. Schene et al.* and *Kim et al.* (References (Anzalone et al., 2019)[,19,20]) were provided as input to PINE-CONE. Alignments of *in silico* designed pegRNAs and designs with PINE-CONE designs (top) and published bottom. +1 is the 'G' transcriptional start from the U6 promoter added by PINE-CONE.

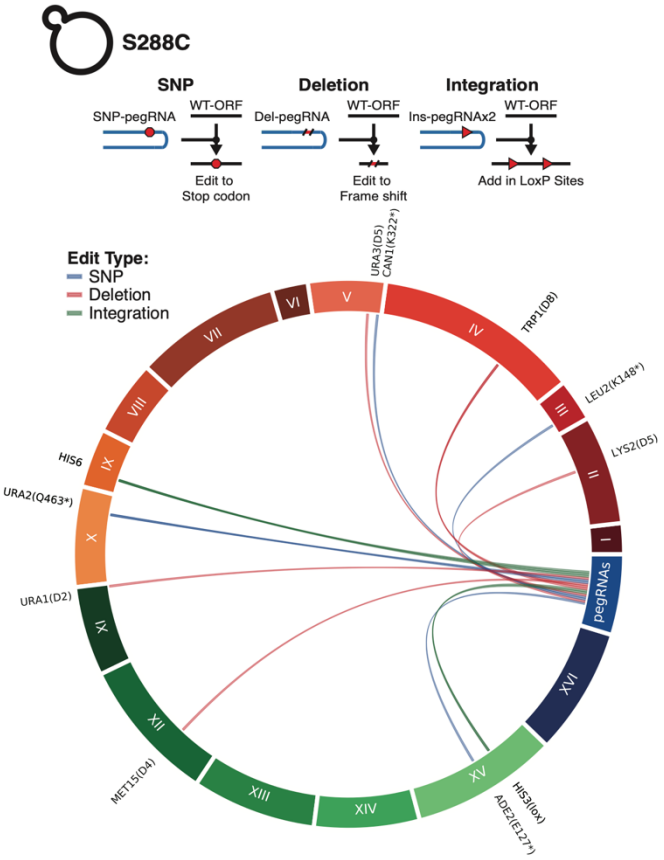


Figure B-5:PINE-CONE design of pegRNAs targeting various marker loci in yeast (S288C). (**a**) Strategies for targeting open reading frames (ORFs) on the yeast genome. Knockouts of various ORFs via point mutation introduction of stop codons, introduction of short deletions resulting in frame shift mutations and insertion of LoxP sites flanking the ORF. (**b**) PINE-CONE Mapping of pegRNA targeting various marker loci with point mutations (blue chords), targeted deletion (Red chords), or integration of LoxP sites (green chords).

Table B-1: PINE-CONE organism selection.

Organism:	Species:	Ref. Genome:	Host Link:
Human	*H.sapiens*	hg38	https://genome.ucsc.edu/index.html
Yeast	*S.cerevisiae*	S288C	https://www.yeastgenome.org/
Mouse	*M.musculus*	mm10	https://genome.ucsc.edu/index.html
Rat	*R.norvegicus*	rn6	https://genome.ucsc.edu/index.html
Zebrafish	*D.rerio*	danRer11	https://genome.ucsc.edu/index.html
Roundworm	*C.elegans*	ce11	https://genome.ucsc.edu/index.html
Fruitfly	*D.melanogaster*	dm6	https://genome.ucsc.edu/index.html

PINE-CONE is capable of retrieving reference genome sequences for various multiple species.

Table B-2: PINE-CONE Cloning Oligo Format.

Oligo:	Sequence: (5'→3')	Notes:
pegRNA Guide Top	CACCGNNNNNNNNNNNNNNNNNNN	'N' Sequence varies between pegRNAs
pegRNA Guide Bottom	AACNNNNNNNNNNNNNNNNNNC	'N' Sequence varies between pegRNAs
pegRNA Cas9 Hairpin Top	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGG	Constant for all pegRNA outputs
pegRNA Cas9 Hairpin Bottom	GCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAA	Constant for all pegRNA outputs
pegRNA RTT-PBS Top	TCGNNNNNNNNNNNNNNNNNNNTTTTTTGTTTTCTGCA	Sequence and length of 'N' varies for different pegRNAs
pegRNA RTT-PBS Bottom	GAAAACAAAAAANNNNNNNNNNNNNNNNNNNN	Sequence and length of 'N' varies for different pegRNAs
PE3/3B sgRNA guide Top	CACCGNNNNNNNNNNNNNNNNNNN	'N' Sequence varies between PE3 and PE3B sgRNAs
PE3/3B sgRNA guide Bottom	AAACNNNNNNNNNNNNNNNNNNC	N' Sequence varies between PE3 and PE3B sgRNAs

General format of oligonucleotides for cloning pegRNAs and PE3 or PE3B sgRNA guides into pHSG1C3. N = A, T, C or G. Cas9 hairpin oligos are constant across pegRNA designs and thus are not included directly in the PINE-CONE output to avoid redundancy.

APPENDIX C

STATEMENT REGARDING PREVIOUSLY PUBLISHED WORK

Chapter 2, “Genome Reprogramming for Synthetic Biology”, was previously published as “Genome reprogramming for synthetic biology” in the peer-reviewed journal *Frontiers of Chemical Science & Engineering*.

Chapter 3, “Transient Reporter for Editing Enrichment”, was previously published as “A transient reporter for editing enrichment (TREE) in human cells” in the peer-reviewed journal *Nucleic Acids Research*.

Chapter 4, “Design automation of CRISPR RNAs”, was previously published as “Prime editing guide RNA design using PINE-CONE” in the peer-reviewed journal *ACS Synthetic Biology.*

The co-authors of the aforementioned manuscripts allow their use in this dissertation. The author of this dissertation is first listed on the above publications and wishes to emphasize that all authors have provided unique contributions in the preparations of these documents.