Precise Correction of A1AT E342K by Modified NGA PAM Prime Editing and Determination of Prime Editing Inhibition by TREX2

Precise Correction of A1AT E342K by Modified NGA PAM Prime Editing and Determination of Prime Editing Inhibition by TREX2

Genesis Lung

A Thesis in the Field of Bioengineering and Nanotechnology

for the Degree of Master of Liberal Arts in Extension Studies

Harvard University

November 2021

Abstract

The purpose of this research is to evaluate the potential of the CRISPR-derived class of prime editors for precise correction of the E342K mutation in Alpha-1 Antitrypsin Deficiency (A1AD) and to discover genetic dependencies of the prime editing mechanism. A1AD is thus far a disease with unmet need that remains challenging to target with genome editors which would benefit from an expanded set of editing tools. Several versions of modified prime editors targeting NGA and NGC PAMs were tested with a library of pegRNAs and nickRNAs generated through PrimeDesign, a program published by the lab which originated prime editors. In doing so, we have shown that A1AT E342K is accessible for precise correction via two NGA PAMs. As prime editing technology is new, the molecular interactions with cellular repair are as of yet only putative and require elucidation that could serve to further bolster success of this and other therapeutic applications. An siRNA library of DNA repair factors was compiled and evaluated for the ability to up- or down- regulate prime editing efficiency. Through this screen, we have discovered a gene, TREX2, which influences prime editing and have validated this through an orthogonal overexpression experiment.

Dedication

I dedicate this thesis to my late high school biology teacher Dr. Susan Offner, who first set me on the path to becoming a scientist by showing me how to use DNAMaster after school, my lifelong friend Kelly Ziyi Miao with whom I began that after school journey as my research partner for my very first scientific paper and with whom I made a pact with to go to grad school, and to my wife Knottlynn Tran, who has been my source of happiness in life and indelible rock through the challenges of working on a thesis through a pandemic.

Acknowledgments

I'd like to thank Dr. Michael S. Packer, who has been a wonderful thesis director, and prior to that, a patient and understanding mentor, and more importantly a friend. This work is made possible foremost largely through his scientific advice in innumerable ways and his communication with other teams to provide assistance. I'd also like to thank Dr. Holly Rees and Shelby Tonelli, who both worked on prime editing in different projects and contributed immensely to suggesting possible angles of exploration with data and peer review as well as by providing reagents and performing protein analysis work. The team over at Prime Medicine, with Drs. Jon Levy, Ariel Yeh, and Andrew Anzalone was also helpful with their expertise in prime medicine and with the sharing of concurrent research results and critical analysis of my data to inform decisions for this work. I also thank the Liu Group at Harvard for their contributions to prime editing technology and their communication with Beam. I'd like to also thank the whole team at Beam Therapeutics, who have contributed in varying essential capacities to generate materials with which the experiments here were conducted. I thank John Evans and Dr. Guiseppe "Pino" Ciaramella and the rest of the leadership team at Beam Therapeutics for sponsoring my thesis at Beam. I am grateful to Lo-I Cheng, who helped with cloning plasmids during experimentally intensive times, Valentina McEneany, who helped me deliver sequencing samples to Eton on days when I forgot them in the incubator, and Jason St. Laurent, who along with Valentina produced mRNA from plasmids for me. I am indebted to Thomas Fernandez and Dr. Yvonne Aratyn-Schaus, who developed the

fibroblast-HPH coculture lentivirus-transduced model. I also thank Dr. Nicole Gaudelli and Dieter Kyle Lam, whose company and occasional help in the lab were indispensable. I thank Dr. Steven Denkin for helping me review my thesis proposal and thesis draft and helping to provide direction. I also thank the Harvard Thesis committee for their understanding during this pandemic. I thank Horizon Discovery, OriGene, Eton Biosciences, Thermofisher and the scientific community for their services and products which were used in this work. There are many people who have touched this work in some way, and I apologize for my memory if I have not been able to include them all.

Table of Contents

List of Tables

List of Figures

Chapter I.

Introduction

Since the discovery of genetics and the advent of the Human Genome Project, more than 75,000 pathogenic human genetic variants have been identified which contribute to a variety of diseases. Genetic editing aims to correct such mutations at the source rather than treating symptoms. The earliest gene editing technologies such as meganucleases, Zinc Finger Nucleases (ZFNs), Transcription Activator-Like Effector Nucleases (TALENs), and CRISPR-Cas9 are able to modify genomes with varying rates of efficiency and control over editing outcomes.

Meganucleases are highly specific nucleases first isolated from yeast which target long sequences, but are difficult to engineer for custom sites (Takeuchi et al., 2014). ZFNs consist of a pair of zinc finger arrays fused to a FokI nuclease domain. Upon binding of the pair to both flanking sequences of a target site the FokI dimer can induce a double-stranded break. ZFNs are more modular than meganucleases but often exhibit sequence context dependence and can require significant protein engineering (Caroll 2011). TALENs are bacterial transcriptional activators fused to FokI in the same way as ZFNs, and are highly specific but are challenging to deliver due to size and sequence complexity (Wang et al., 2016). CRISPR Cas9 is able to make double stranded breaks via modular guide RNA sequences that are homologous to host DNA around protospacer-adjacent motif (PAM) recognition sites (Adli 2018).

Regardless of the specific mode of action, all of the aforementioned genome editing tools work via the introduction of double-stranded breaks. Double stranded breaks

rely on host cell recognition of the breaks and the initiation of either non-homologous end joining or homology-directed repair. For non-homologous end joining (NHEJ), the blunt ends of a strand of DNA that has been cleaved at two ends are paired and ligated. This process is error-prone and can result in deletion of intervening sequences or insertion of bases (Adli 2018). In homology-directed repair (HDR), a donor template with homology arms invades the two blunt ends up and downstream of the cut site to insert or otherwise alter a sequence (Adli 2018).

In the context of therapeutic genome editing, double-stranded breaks are problematic for a number of reasons. First, double-stranded breaks activate the P53 pathway and can increase the risk of senescence and apoptosis (Menon and Povirk, 2014). Secondly, even upon repair the rate of erroneous editing can lead to unpredictable deletions and translocations (Brunet and Jasin, 2018). In recent years base editing technology invented by the means of fusing deaminase enzymes to Cas9 and its orthologs have been able to create a system which induces transition mutations of cytosine-uridine-thymine (C>U>T) and adenine-inosine-guanine (A>I>G) (Komor et al., 2016) (Gaudelli et al., 2017). While this technology is not reliant on double stranded breaks, it is limited to specific classes of transition mutations within the protospacer of the Cas9 recognition sequence. Base editing additionally operates by stochastic chemistry with the possibility of editing neighboring bystander nucleotides, with the requirement of a very short distance between the PAM and target edit site.

The Liu Group at Harvard recently published the paper “Search-and-replace genome editing without double-strand breaks or donor DNA” in Nature (Anzalone et al. 2019). The use of a nickase Cas9 (nCas9) fused to an engineered reverse transcriptase

with a modified prime editing guide RNA (pegRNA) shows great potential for precise editing with multiplex transversions, deletions, and insertions achieved at various lengths up to 44bp. Of particular interest, the system achieved 6-8% insertion editing in post-mitotic non-dividing cells, which is a large hurdle for HDR systems. However, the paper surmises many of the underlying mechanics, which remain to be demystified. The authors have proposed that prime editing works through a series of hypothetical intermediates: (1) recognition of the target site and opening of the upstream guide protospacer region of the pegRNA, (2) nicking of the noncomplementary strand, (3) priming with the downstream end of the pegRNA primer binding sequence (PBS) region that extends out of the Cas9, (4) complementary extension of the nicked strand in a complementary fashion to the reverse transcription template (RT) region of the pegRNA adjacent to the PBS region by the reverse transcriptase, (5) competitive equilibrium annealing of the extended single-stranded flap to the complementary genomic strand against the downstream unextended flap, (6) preferential cleavage of the displaced unedited flap by host nucleases, and (7) DNA repair resolution of strand mismatches (Figure 1, Anzalone et al., 2019).

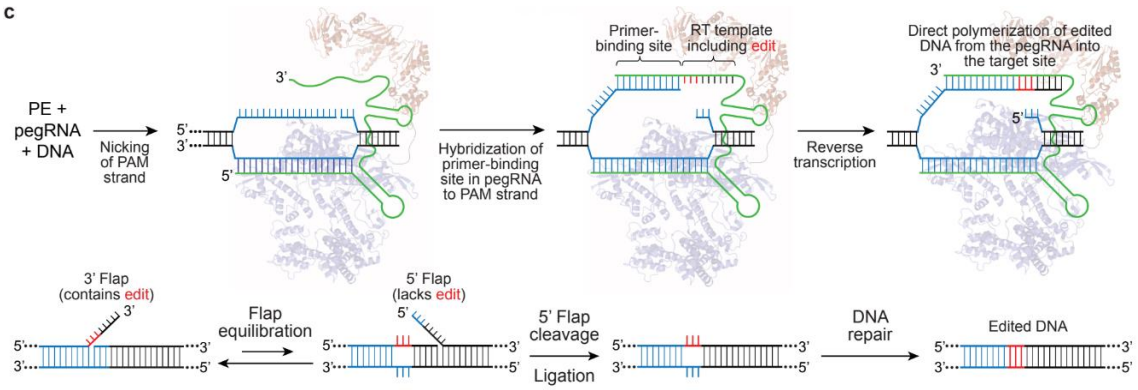


Figure 1. Proposed mechanism for prime editing from Anzalone et al. figure 1c.

The spacer from the pegRNA invades one strand, followed by enzymatic nicking and the attachment of the PBS to the nicked flap, with extension of the pegRNA RTT by reverse transcriptase, and including with flap equilibration and cellular repair processes which include the desired edit.

Various strategies for improving prime editing outcomes have included several iterative versions. (1) PE1, which uses the basic Cas9 fused to a wildtype reverse transcriptase a pegRNA. (2) PE2, which incorporates mutations conferring stability and activity into the reverse transcriptase. (3) PE3, which introduces a guide RNA to nick the complementary unedited genomic strand and induce favoring repair to incorporate the mutation. (4) PE3b, which introduces a nicking guide RNA similarly to the PE3 strategy but where the nicking guide protospacer is complementary only to the newly edited strand and thus nicking occurs only after editing. (5) PE3_PAM and (6) PE3b_PAM, where the PAM strategy represents a modification to the reverse transcription template that ablates the PAM recognition sequence upon extension so that the prime editing enzyme will not repeatedly bind and nick to undo its own editing.

This work aimed to apply prime editing to correct a disease-causing mutation of high unmet need that has previously been challenging to address with CRISPR HDR and base editing due to inefficiencies in DNA repair mechanisms or target site limitations, respectively. In order to accomplish this goal we first developed prime editing reagents capable of correcting this mutation in immortalized cell lines. We then evaluated these reagents in a primary cell type which exhibited a different transcriptome and cellular physiology from immortalized cell lines. Anticipating challenges in the translation of this technology between cell types, a parallel aim was to interrogate which DNA repair genes could influence rates of prime editing. To address this question, we relied upon a genetic knockdown screen accompanied by orthogonal validation methods such as overexpression and knockout.

A1AD

Alpha-1 Antitrypsin Deficiency (A1AD) is a genetic disorder caused by mutations in the SERPINA1 gene which encodes alpha-1 antitrypsin (A1AT). A1AT is predominantly produced in the liver and secreted into circulation where it serves as an essential anti-protease. This function is particularly vital in the lungs where uncontrolled elastase from migrating neutrophils will otherwise damage the structure of elastin-rich lung tissue (Stoller and Aboussouan, 2011).

The mechanism by which A1AT neutralizes elastase is through its highly unstable "mousetrap" structure which undergoes a conformational change and clamps onto elastase after binding (Stoller and Aboussouan, 2011). Due to its instability, mutations in A1AT can lead to drastic misfolding of the protein. Misfolded A1AT is unable to be secreted, and thus causes two concurrent effects. First, as mentioned above the lack of

A1AT causes lung damage. Second, the hepatocytes in the liver accumulate misfolded protein leading to fibrosis and cirrhosis (Stoller and Aboussouan, 2011).

A1AT mutant alleles include PiZ, PiS, and rare Null alleles. The genotype of both alleles in a patient contribute to a codominant phenotype with a spectrum of A1AT serum levels at varying percentages of normal concentration (Brode et al., 2012). The low risk of lung disease in PiMZ heterozygotes sets a clinical "normal" threshold of 11uM serum A1AT (Figure 2). The most common genotype in clinical A1AD is the PiZZ genotype representing 10-15% of deficiency cases with an estimated 60,000 individuals in the U.S., including many undiagnosed. The PiZ allele contains a G to A single nucleotide polymorphism (SNP) resulting in the amino acid substitution E342K (Figure 3).

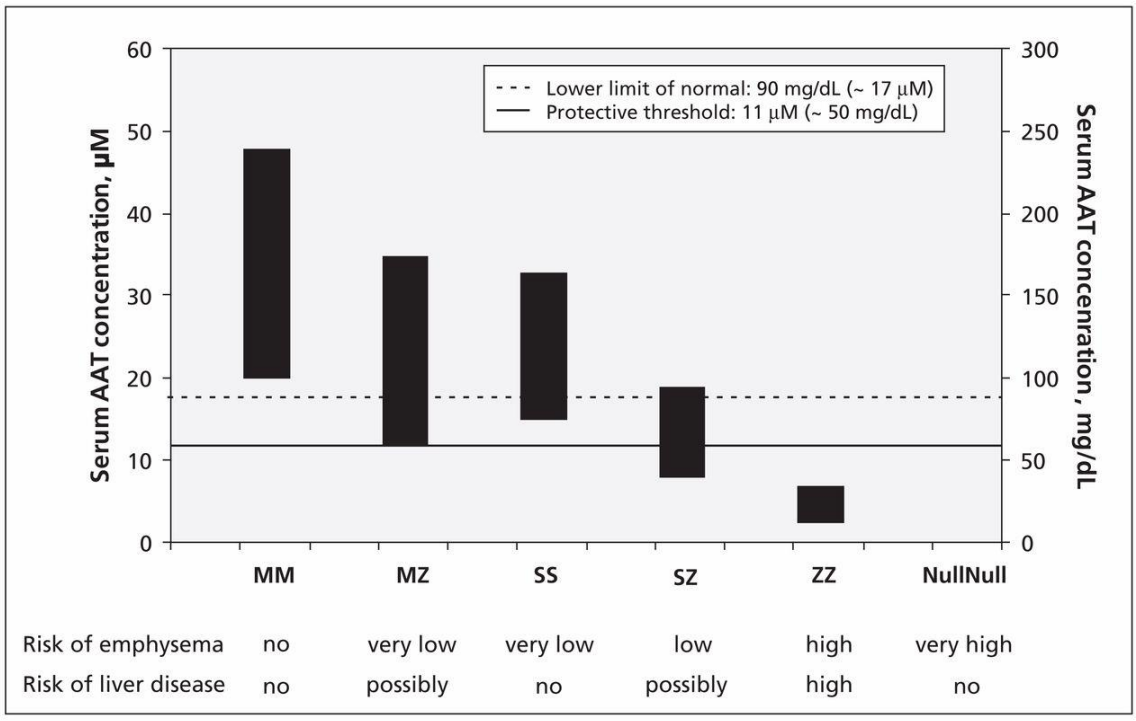


Figure 2. A1AT Allele Serum Levels.

Plot (Brode et al., 2012) showing the mg/dl and uM concentration of A1AT in serum for MM, MZ, SS, SZ, and ZZ genotypes. The solid line is the minimum clinically relevant target, the dotted line is the lower limit of healthy serum levels. Corresponding disease phenotypes listed at bottom.

Figure 3. A1AT target site.

The wildtype allele is depicted here. In the PiZ allele, the E(GAG) codon (blue) becomes a K(AAG) codon due to a G>A mutation (light grey).

The E342K G to A SNP is a strong candidate for the development of genome editing therapeutics due to the prevalence of the genotype among patients and the single base which needs to be changed. Additionally, due to the dual nature of the disease as a simultaneous loss-of-function and gain-of-function, simply eliminating loss of function by raising serum A1AT is insufficient, as it is necessary to not only increase the amount of circulating A1AT but also lower the amount of toxic misfolded A1AT. Correcting the PiZ allele into a wildtype PiM allele via genome editing is the ideal strategy to treat both mechanisms of the disease.

Current Limitations of Genome Editing Approaches Pertaining to A1AD

Nuclease genome editing, whether accomplished by homology-directed repair or nonhomologous end-joining, does not work with desirable frequencies in many primary cell types due to the dependency of HDR on cellular DNA repair processes that are only active in dividing cells (Nami et al., 2018). As A1AD is a disease affecting primary human hepatocytes which do not divide in culture, investigation of A1AD HDR faces limitations for cell-type relevant reagent development.

Despite these challenges, a number of research groups are exploring nuclease induced HDR as a potential therapeutic strategy for A1AD. One published study reports ~5% precise correction of the E342K PiZ mutation in a transgenic mouse model. To achieve this feat they employed dual adeno-associated viral (AAV) vectors to overcome transgene size limitations inherent to AAV. One AAV contained the donor template as well as a hU6 promoter driving expression of an sgRNA, and the second contained a staphylococcus aureus Cas9 (saCas9) (Shen et al., 2018). Several issues are readily apparent, with a high indel rate of 10-20% varying by formulation, and the introduction

of improper insertions, which themselves could introduce pathogenic misfolding consequences. The use of a dual adeno-associated viral vector requires codelivery to the same cell of both viral vectors to accomplish editing, and may result in challenging PK/PD properties. On the positive side, AAVs are clinically validated gene therapy vectors exhibiting highly efficient delivery to hepatocytes.

Lipid nanoparticle (LNP) mediated hepatic delivery of genome editing reagents offer numerous advantages over AAV. Unlike AAV, LNPs are not subject to pre-existing immunity and have the potential to be redosed. LNPs are typically used to deliver RNA reagents, resulting in transient expression of genome editing proteins. This is a particularly desirable feature to avoid accumulation of off-target edits over the course of a sustained expression which is often seen with AAVs. Unfortunately, LNPs have not been shown to be able to also deliver donor templates required to induce HDR.

Base editing however does not require a donor template and would be amenable to LNP-mediated delivery. Furthermore, the PiZ A1AT mutation is a class of transition mutation amenable to adenine base editors. However, gRNA options at this locus are limited as there is no NGG PAM in the vicinity that would position the target adenine in range of base editing (Figure 4). Additionally there are several bystander adenine bases immediately adjacent to the target SNP that if edited would yield non-synonymous changes in the A1AT gene with unknown phenotypic consequences. There is ongoing research to identify editors with altered PAM specificity and editing properties that can precisely correct PiZ A1AT.

Figure 4. Position of NGG protospacers relative to the E342K SNP.

Both are not in the optimal prime editing range. Pathogenic SNP represented by the blue triangle, with the pathogenic codon in the pink arrow.

Base editing additionally carries the risk of off-target deamination. Both cytosine base editors (CBEs) and adenine base editors (ABEs) contain deaminase domains, which retain a natural tendency to edit RNA substrates. This may cause issues for cell health due to erratic transcripts in the short term while base editors are active but the effect is expected to be transient. Base editors are additionally difficult to control, with their deaminase enzyme in an always-on configuration that often edits other nucleotides other than the intended target within the protospacer. Overexpression of CBEs, but not ABEs, has also been shown in some literature to generate genome-wide off target SNPs, (Zuo et al., 2019). Off-target analysis of all base editors is an ongoing area of research and there remains concern that off-targets could be harmful.

After work presented here had begun, several other groups have attempted to achieve prime editing of A1AT E342K. Initially, prime editing showed no detectable correction of the pathogenic SNP in liver organoids created from patients carrying the allele (Schene et al., 2020). Subsequently, a PE2 plasmid strategy with an RTT of 27bp

and PBS of 13bp and NGG PAM was shown to achieve 1.9% insertion of the pathogenic SNP as PE2 which increased to 9.9% with the addition of a nicking guide, and nuclear localization signal optimization to create an editor called PE* increased insertion to 6.4% for PE2*, and 15.8% for PE3*, with lower efficiency for SaKKHPE2* at 1.1-4.4%. The same paper demonstrated split-intein prime editors delivered via AAV achieved 2.1% editing as PE2 and 6.7% as PE2* (Liu et al., 2020). Indel rates were however very high for PE2 and PE2* at 0-2.2 to 0.1-3.8% respectively in Hek293T cells and 0.4-2.7% in vivo, and the authors comment on the lack of a nearby PAM within the optimal bounds for prime editing. Another paper attempting to correct patient-derived human iPSCs with an NGG PE3 strategy found no detectable editing (Habib et al., 2021).

All of the aforementioned challenges with existing genome editing techniques suggest that a newly emergent approach such as prime editing might be well suited to addressing PiZ Alpha-1 Antitrypsin Deficiency, and that existing prime editing agents for A1AT E342K could be greatly improved with alternative engineering methods.

Flap Excision Mechanisms Pertinent to Prime Editing

Published data thus far has not demonstrated exactly which DNA repair pathways are essential for making prime edits permanent. However the proposed mechanism in the seminal publication on prime editing does implicate a number of known factors that have been studied in the context of cellular replication. Large flaps hanging from DNA are common during the formation of Okazaki fragments as part of DNA replication. Flaps of 30 or more single stranded nucleotides are recognized and bound by Replication Protein A, which melts the secondary structure of the DNA flap, recruiting Dna2. Dna2 then cleaves the flap to be shorter and under the 30nt length needed for RPA recognition, at

which point FEN1 is recruited and then cleaves the flap (Bartos et al., 2018, Figure 41). FEN1 has been indicated in the literature to recognize the 5’ end of the downstream flap and the first 3’ nucleotide at the end of the upstream flap, (Finger et al., 2012) and is the surmised repair mechanism proposed by Anzalone et al in the prime editing paper. Any associated DNA repair enzymes would likely be of strong interest. Similar genes could include exonuclease 1 (EXO1), DNA repair protein complementing XP-G cells (XPG), Holliday Junction 5’ Flap Endonuclease (GEN1), 5’-3’ exoribonuclease (Xrn1), 5’-3’ exoribonuclease 2 (Rat1), and point to the need for a comprehensive screen.

RISC and siRNA

The RNA-induced silencing complex (RISC) is a ribonucleoprotein complex containing many RISC-associated proteins which is capable of incorporating single-stranded microRNA or double-stranded small interfering RNA (siRNA). Normally, the RISC complex exists in order to silence viral mRNA as part of the intracellular immune system, with the Dicer enzyme cleaving viral RNA which is then recognized by an Argonaute enzyme that uses it to hybridize to target transcripts. The RISC complex can accomplish its silencing through mRNA degradation, which is possible when the match between the target and siRNA is highly complementary, and causes the Argonaut enzyme to cleave the mRNA. This process is localized to cytoplasmic P-bodies (Dana, 2017). RISC complexes can also prevent localization of eukaryotic translation initiation factor (eIF) to the 5’cap of mRNA and can accelerate deadenylation of the 3’ poly(A) tail. RISC can also compete and prevent the binding of the 60S ribosomal unit. It is also able to block translation that has already begun by causing detachment of ribosomes and slowing the elongation by attaching downstream of the start. Finally, RISC also recruits histone

methyltransferases, which silences the gene through heterochromatin formation. The biology of RISC and RNA silencing is invaluable for use as a diagnostic tool for the effects of gene knockdown.

Pre-existing work in the field of siRNA silencing has shown that siRNA screens are a useful tool for determining genetic dependencies in specific biological processes. As with all genetic screens a high-throughput functional read out is paired with a genetic perturbation, in this case knockdown of gene expression with a suite of siRNAs. This process allows for functional characterization even without complete knowledge of the behavior of the gene (Sharma and Rao, 2009). Screens have been used to discover various pathways in disease, which have eventually led to targeted therapies. A druggable target in the pyrimidine host synthesis pathway was identified in an siRNA screen of Ebola virus replication (Martin et al., 2018). 26 upregulators and 13 downregulators of TNF-α were identified in another siRNA screen to understand LPS-induced response in macrophages (Sun et al., 2017). Some siRNA screens have also been used concurrently with CRISPR in order to double down on knockdown effects, achieving a combinatorially boost to genetic suppression (Zhao et al., 2017). An siRNA screen interrogating dependencies in genome editing and more specifically prime editing could illuminate limitations of the method as well as highlight avenues for future engineering and optimization.

Chapter II.

Materials and Methods

The design and construction of prime editor materials and development of experimental protocols are described in this chapter. Design was performed either manually in benchling or via the PrimeDesign algorithm. Molecular biology techniques were used to clone all plasmids, and DNA and RNA were synthesized by a CMO. All experiments were executed in compliance with BSL2 requirements in the labs at Beam Therapeutics.

PhusionU PCR

Cloning PCRs were all performed using deoxyuridine-tolerant PhusionU polymerase to allow for uridine excision ligation with the USER enzyme (NEB M5505S) (Geu-Flores et al., 2007). All primers were designed with a target melting temperature approximating 60°C and checked for spurious priming with a minimum of no off-targets with at least 3 mismatches, length of 16-30bp, and G/C content ~50% and G end clamps where possible, with a USER junction of 6-10bp where the primer internal sequence begins with a 5’ A and ends with a 3’-U followed by trailing bases. Some primers were designed with site directed mutagenesis regions.

Phusion U PCR Protocol:

- 25uL of PhusionU Green Mastermix (Thermofisher F564L)
- 1uL (100ng) template DNA
- 0.5uL 100uM primer 1
- 0.5uL 100uM primer 1
- 23uL H2O
- 95°C 2min, 30cycles(95°C 30sec, 62°C 30sec, 72°C 30sec/kb), 72C° 2min, 12°C hold.

USER Cloning

Fragments produced from Phusion U PCR were gel extracted (2% agarose gel with SYBR safe (Thermofisher S33102) running in 1X TAE at 140V, 300mA, 30min) with the Zymo DNA gel extraction kits and cloned using USER enzyme (New England Biolabs M5505S) to generate a gap at the uridine position within the primer junctions and DpnI (NEB R0176S) to digest parental methylated plasmid, and then transformed into NEB 5-alpha (C2987H), NEB stable (C3040H), or Mach1 (Thermofisher C862003) cells depending on manufacturer availability to allow for native bacterial ligation to repair the plasmid.

USER Cloning Protocol:

- 1uL Cutsmart 10X Buffer (New England Biolabs B7204S)
- 1uL 20000units/mL DpnI
- 1uL 1000units/mL USER enzyme
- 2uL H2O
- To 5uL total of equal volume for each DNA fragment
- 37°C 15min, 12°C hold for two-piece cloning
- 37C° 45min, 0.1°C/sec ramp down to 12°C hold for >2 piece cloning

Rolling Circle Amplification and Sanger Sequencing

Rolling circle amplification (RCA) from Templiphi kits (Cytiva, 25640080) was used to generate DNA for sequence verification. 5uL sample buffer was mixed with a picked colony and heated at 95°C for 3min. 5uL reaction buffer and 0.2uL enzyme were added and the solution was incubated at 30°C for 6 hours to overnight, and supplied with sequencing primers to Eton.

Bacterial Culture Methods

Bacterial transformation was used to incorporate plasmids (10min thaw, 5min incubation with 2uL USER reaction on ice, 30sec heat shock in 42°C water bath, 5min ice recovery, 15min(carb)-45min(kan) recovery in 200uL SOC media at 37°C in a bacterial shaker at 180RPM), and then transformation suspensions were plated on 100mm antibiotic resistance LB agar plates (100ug/mL carbenicillin or 50ug/mL kanamycin) and incubated overnight at 37°C. Bacterial colonies obtained from 100mM plates were picked

and seeded into 5mL or 30mL overnight cultures in LB with Carb100/Kan50 at 37°C and 230RPM in shakers. Glycerol stocks were created with 500uL of overnight culture and 500uL of 50% glycerol. Remaining culture volume was centrifuged for 10min at max (4700) RCF in benchtop centrifuges (eppendorf 2231010061), and lysis and DNA harvest was performed with Zymo plasmid prep kits according to manufacturer's protocol (Zymo Research D4208T/D4200). For multiplex cloning, 96-well format competent cells were used for transformation, and deep 2mL 96-well plates were used for mini/midiprep overnight antibiotic culture with airpore plate seals.

Design and Construction of Prime Editors

Gene blocks were designed and then synthesized by Thermofisher for PE2 (Figures 5, 42, 43). Three fragments were assembled into a pUTR backbone (Figure 44) using PhusionU PCR with custom primers (Table 24) followed by USER cloning to generate pUTR-PE2. Cloning of pCMV-PE2 was performed by a team at Beam investigating prime editing for CD34+ cells using these gene blocks and graciously provided (Figure 45). BbsI-cleavage sites were cloned out using site-directed mutagenesis with PhusionU PCR and USER cloning to enable compatibility of pUTR-PE2 with 120A restriction cloning (pUTR-PE2-BbsI--120A) in order to generate cDNA template for mRNA IVT (Table 2). Using preexisting plasmids in the Beam Therapeutics library generated in earlier work for optimized base editors which contained PAM-Interacting Domain (PID) mutations for NGC and NGA PAMS as well as pCMV/pUTR-PE2 as templates, pCMV-PE2-VRQR, pCMV-PE2-MQKFRAER, pCMV-PE2-spCas9(I322V,S409I,E427G,R654L, R753G)-VRQR, pCMV-PE2-spCas9(I322V,S409I,E427G,R654L,

R753G, R1114G)-MQKFRAER, and pUTR-PE2-VRQR were also assembled via PhusionU PCR and USER cloning (Figures 6, 7, 46, 47, 48, 49, 50, 51) (Table 1, 3). mRNA of PE2 and PE2-VRQR were generated from the associated pUTR constructs via restriction digest linearization followed by IVT conducted by Beam's mRNA production team using a proprietary in-house method. All plasmid production and selection was performed using Mach1 or NEB 5-alpha/stable cells depending on availability during the pandemic and Carb100/Kan50 medium and plates, with sequencing confirmation performed with rolling circle amplification and sanger sequencing at Eton Biosciences with custom primers (Table 25) and plasmid extraction executed with Zymo mini/midi/gigaprep kits (Zymo Research D4208T, D4200, D4204). Sequence verification was performed by Eton Biosciences using sanger sequencing.

```
>PE2Frag3
ACTGATAGCCGTTATGCTTTTGCTACTGCCCATATCCATGGAGAAATATACAGAAGGCG
TGGGTGGCTCACATCAGAAGGCAAAGAGATCAAAAATAAAGACGAGATCTTGGCCCTAC
TAAAAGCCCTCTTTCTGCCCAAAAGACTTAGCATAATCCATTGTCCAGGACATCAAAAG
GGACACAGCGCCGAGGCTAGAGGCAACCGGATGGCTGACCAAGCGGCCCGAAAGGCAGC
CATCACAGAGACTCCAGACACCTCTACCCTCCTCATAGAAAATTCATCACCCTCTGGCG
GCTCAAAAAGAACCGCCGACGGCAGCGAATTCGAGCCCAAGAAGAAGAGGAAAGTCTAA
```

Figure 5. Geneblock fragment 3.

Sequence of a section of PE2.

Table 1. Primer list for cloning pCMV/UTR-PE2-VRQR/MQKFRAER.

GL561	ACCGAGG/ideoxyU/GCAGACAG	Connect VRQR or MQKFRAER PID of pBTX460/pGL68 to pGL155 backbone F
GL562	ACCTCGG/ideoxyU/CTTTTTCACGATAT	Connect pGL155 backbone to VRQR or MQKFRAER PID of pBTX460/pGL68 R
GL563	AGGATC/ideoxyU/AGCGGAGGATCC	Connect pGL155 backbone to VRQR or MQKFRAER PID of pBTX460/pGL68 F
GL564	AGATCC/ideoxyU/CCAGAGTCGCCTCCGAGTTGAG	Connect VRQR or MQKFRAER PID of pBTX460/pGL68 to pGL155 backbone R

Table 2. Primer pairs for cloning out BbsI sites.

Site	Primer 1	Primer 1 Sequence	Primer 2	Primer 2 Sequence
BbsI site 1	HR311	AGCGGGTGGAAGA/ideoxyU/ATCCACC	HR312	ATCTTCCACCCGC/ideoxyU/TGTTGACTTC
BbsI site 2	HR313	ACTCCTAAGACCCC/ideoxyU/CGACAACTAAG	HR314	AGGGGTCTTAGGAG/ideoxyU/AGGCTGC

Table 3. USER primers for cloning Miller variant VRQR/MQKFRAER prime editors.

GL577	AAGAAGTAT/ideoxyU/CTATCGGACTGGC	USER cloning primer F to connect Miller spCas9 5' to VRQR/MQKFRAER backbone
GL578	AATACTTCT/ideoxyU/GTCGACTTTCCG	USER cloning primer R to connect VRQR/MQKFRAER backbone to Miller spCas9 5'
GL579	ACCGAGG/ideoxyU/GCAGACAGG	USER cloning primer F to connect VRQR backbone to Miller spCas9 5' without PID
GL580	ACCTCGG/ideoxyU/TTTCTTCACGATAT	USER cloning primer R to connect Miller spCas9 3' without PID to VRQR backbone
GL581	ACCTGTC/ideoxyU/CAACTCGGAGG	USER cloning primer F to connect MQKFRAER backbone to Miller spCas9 5'
GL582	AGACAGG/ideoxyU/CGATCCGTGT	USER cloning primer R to connect Miller spCas9 3' to MQKFRAER backbone

```
>PE2-VRQR PAM-Interacting Domain
GTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCTAAGCGGAACTCCGACAAGCT
GATCGCCAGAAAGAAGGACTGGGACCCCAAGAAATACGGCGGCTTCgtgTCTCCTACCG
TGGCCTATTCTGTTCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTCAAG
AGCGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCC
GATCGATTTCCTCGAGGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGC
TCCCCAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGCGGATGCTGGCCTCTGCT
aggGAACTGCAGAAGGGAAACGAACTGGCCCTGCCTAGCAAATATGTGAACTTCCTGTA
CCTGGCCAGCCACTATGAGAAGCTGAAGGGCAGCCCCGAGGACAATGAGCAAAAGCAGC
TGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTT
AGCAAGAGAGTGATTCTGGCCGACGCCAATCTGGACAAAGTGCTGTCCGCCTACAACAA
GCACCGGGACAAGCCTATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGA
CCAACCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGcag
TACcggTCCACCAAAGAGGTGCTGGACGCCACTCTGATCCACCAGTCTATCACCGGCCT
GTACGAGACACGGATCGACCTGTCTCAACTCGGAGGCGAC
```

Figure 6. PAM-Interacting Domain sequence for NGA PAM recognition.

For VRQR base editor plasmid used for cloning.

```
>PE2-MQKFRAER PAM-Interacting Domain
GTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCTAAGCGGAACTCCGACAAGCT
GATCGCCAGAAAGAAGGACTGGGACCCCAAGAAaTACGGCGGCTTTATGCAGCCCACCG
TGGCCTATtctGTtCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTCAAG
AGCGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCC
GATCGATTTCCTCGAGGCCAAGGGTTACAAAGAAGTGAAAAAGGACCTGATCATCAAGC
TCCCCAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCc
AAGttcCTGCAGAAGGGAAACGAACTGGCCCTGCCTAGCAAATATGTGAACTTCCTGTA
CCTGGCCAGCCACTATGAGAAGCTGAAGGGCAGCCCCGAGGACAATGAGCAAAAGCAGC
TGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTT
AGCAAGAGAGTGATTCTGGCCGACGCCAATCTGGACAAAGTGCTGTCCGCCTACAACAA
GCACCGGGACAAGCCTATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGA
CCAACCTGGGAGCCCCTagaGCCTTCAAGTACTTTGACACCACCATCgccCGGAAGGAG
TACcggTCCACCAAAGAGGTGCTGGACGCCACTCTGATCCACCAGTCTATCACCGGCCT
GTACGAGACACGGATCGACCTGTCTCAACTCGGAGGCGAC
```

Figure 7. PAM-Interacting Domain sequence for NGC PAM recognition.

From MQKFRAER base editor plasmid used for cloning.

Design and Construction of pegRNAs and nickRNAs

A manual design method was used for the initial batch of pegRNAs for A1AT, with minimum (truncation directly after the target SNP) and recommended maximum (34bp) (Anzalone et al., 2019) lengths for the RTT and a long or short PBS at 17bp or 13bp, for a total of 2 NGG, 2 NGA, and 6NGC protospacers (Table 4). A second design method utilizing the PrimeDesign software was conducted, for either correction with the lentivirus cassette or insertion into WT genomic DNA. pegRNAs GL556/545/600/544/549 were designed for correction while GL548/601/542 were designed for insertion. The following command was run with the target sequences for different PAMs:

docker run -v ${PWD}/:/DATA -w /DATA pinellolab/primedesign primedesign_cli --file [FILENAME].csv --pe_format NNNNNNNNNNNNNNNNN/NNN[PAM] --pbs_length_list 8 10 13 14 15 --rtt_length_list 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 --silent_mutation --number_of_pegrnas 6

For the PrimeDesign method, a total of 6 PBS lengths (8/10/12/13/14/15) and 3 RTT lengths spanning at least +2 past the pathogenic SNP and the 20bp maximum range where possible were chosen for a total of 6 novel protospacers (Figure 8). Due to limitations in the program for alternative PAMs, silent PAM disruption was added where possible manually to the RTT sequences. The protospacer from Liu et al. 2014 also had pegRNAs designed at 5 PBS lengths (8/10/13/14/15) and 4 RTT lengths (25/26/27/28) at, designated GL542. 8 nickRNAs were also returned by PrimeDesign for GL542, with 27 nickRNAs returned for NGA PAMs. Output sequences were ordered from IDT.

The pegRNAs were cloned into a BsaI-digested pU6-pegRNA-GG-Vector plasmid (Figure 54) via the method described in the original Anzalone paper (Figures 9, 52, 53), whereby top and bottom oligonucleotides with overhangs for the spacer, scaffold, and PBS/RTT components were annealed and cooled, then ligated with the vector (1uL T4 ligase (New England Biolabs M0202S), 1uL DpnI, 1uL Cutsmart 10X Buffer, to 10uL total reaction volume with H2O) (Table 5, 6, 25). Verification and production of guide plasmids was conducted as previously described in this work, with sequencing performed using LKO.1 ‘5 Eton universal primer.

Table 4. Protospacers for the first set of manually designed pegRNAs for E342K correction/insertion.

Spacer	Sequence
GL542	TCCCCTCCAGGCCGTGCATA
GL543	GGGTATGGCCTCTAAAAACA
GL544	GGCCGTGCATAAGGCTGTGC
GL545	CATAAGGCTGTGCTGACCAT
GL546	CTTCTCTCCCCTCCAGGCCG
GL547	CCCCTCCAGGCCGTGCATAA
GL548	CCAGGCCGTGCATAAGGCTG
GL549	TCTAAAAACATGGCCCCAGC
GL550	GCCTCTAAAAACATGGCCCC
GL551	GGTATGGCCTCTAAAAACAT
‘GL600	GGCTGTGCTGACCATCGACa

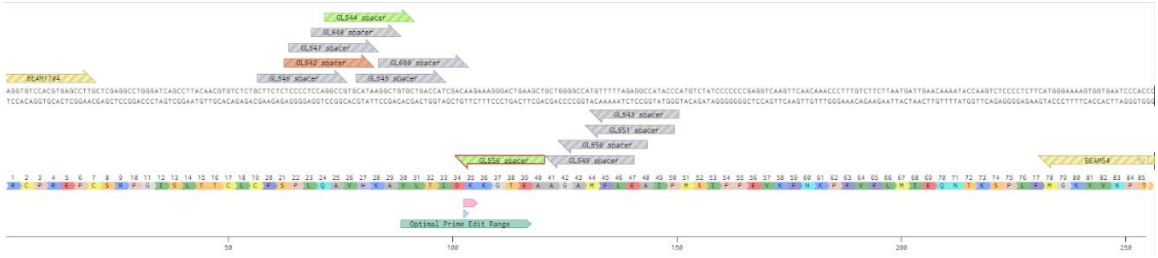


Figure 8. Positions of pegRNA spacers in the A1AT E342K genomic locus.

PAMs are downstream of arrows. E342K in pink, pathogenic G to A SNP in blue. NGA PAMs exhibiting editing in green. NGG PAM from Liu et al. in Orange. Amplicon PCR primers in yellow.

Step 4: pegRNA assembly

pegRNA Golden Gate assembly reaction

Digested pU6-pegRNA-GG plasmid-vector - Pre-cut, isolated 2.2-kb fragment	1.00 µL @ 30 ng/µL
Annealed protospacer oligonucleotides (component 2)	1.00 µL @ 1 µM
Annealed pegRNA 3'-extension oligonucleotides (component 4)	1.00 µL @ 1 µM
sgRNA scaffold annealed oligonucleotides (component 3) - Oligonucleotides ***must be phosphorylated***	1.00 µL @ 1 µM
*Bsa*I-HFv2 (NEB)	0.25 µL
T4 DNA ligase (NEB)	0.50 µL
10x T4 DNA ligase buffer (NEB)	1.00 µL
H_2O	4.25 µL
Total reaction volume:	10.0 µL

Incubate at room temperature for 10 min
Following incubation, perform the following program in a thermocycler:
5 min at 37 °C, then 5 min at 80 °C, then hold at 12 °C

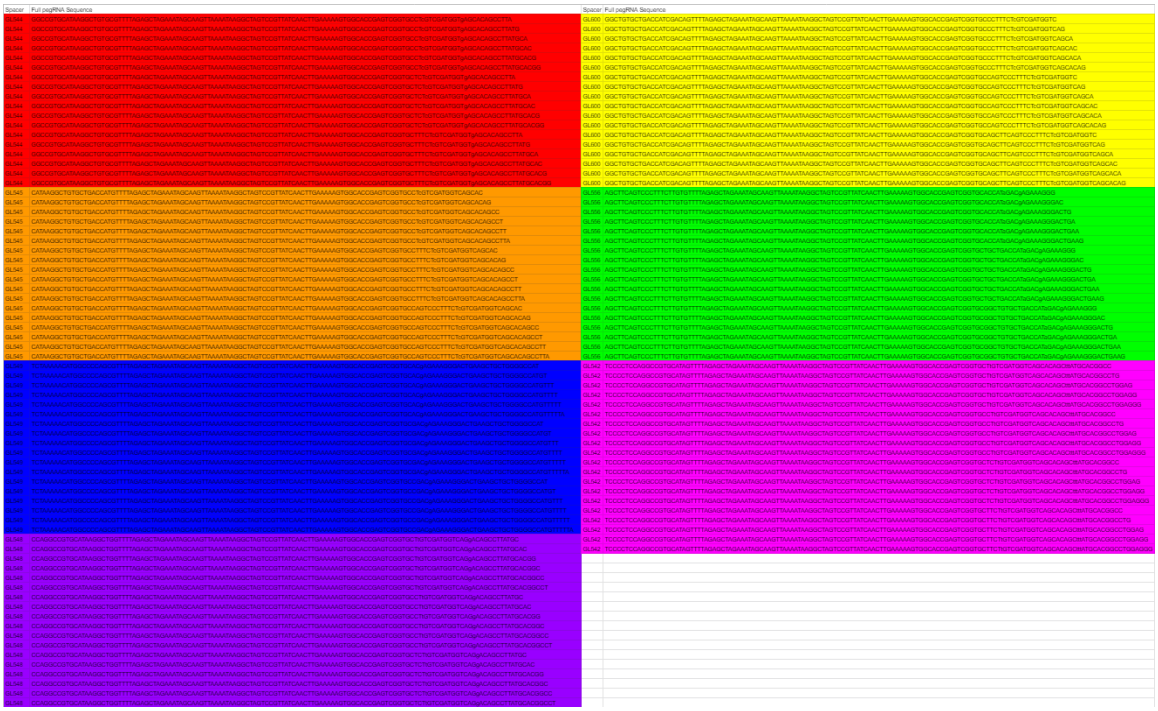
Figure 9. Supplementary note 3 (cont.) from Anzalone et al. 2019.

Protocol adapted in this work for cloning pegRNAs.

Table 5. pegRNA scaffold sequences.

pegRNA-scaffold-phos-F	/5Phos/AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG
pegRNA-scaffold-phos-R	/5Phos/GCACCGACTCGGTCCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAG

Table 6 Final full sequences of tested A1AT E342K insertion/correction pegRNAs.



The nickRNAs were cloned with NEB T4 ligase and SapI-digested (New England Biolabs R0569S) pU6-spdummy plasmid (Figure 55) (Tables 27, 28). For cloning nickRNA pU6 plasmids, a simpler version of the pegRNA protocol cloning was used where only step 4 of the original Anzalone protocol was used with 0.5uL of each 100uM guide oligonucleotide and 6.25uL H2O instead, with no pre-annealing. Not all of the nickRNAs produced by PrimeDesign were compatible with their given spacers for the

E342K lentivirus cassette, although all were cloned, with incompatible nickRNAs saved for potential use with genomic editing. GL544n10-16 and GL556n4-11 were found compatible and used in subsequent PE3 editing experiments.

Synthetic pegRNAs and nickRNAs for the HBB locus were graciously provided by another team at Beam that was working on the HBB sickle cell mutation and who had previously validated them after production by Agilent (Table 29). These guides were reordered as needed from Agilent. All other pegRNA and nickRNA sequences were manually converted from PrimeDesign sequences and ordered from Agilent with stabilizing modifications including a 4U tail and 2'-O-methyl and 3'-phosphorothionate linkages on the first three 5' bases and the fourth/third/second to last 3' bases.

siRNA Knockdown

Thermofisher carries a validated DNA Damage Response siRNA library known as Silencer. Similarly, Horizon Discovery retains a Dharmacon siGENOME siRNA SMARTPool DNA Damage Response library. However, the associations of these commercially available siRNA libraries with DNA damage effects was performed bioinformatically when the function of many of the genes is unknown. With the assistance of Beam's bioinformatics team, RNAseq data from the Protein Atlas was analyzed for Hek293T cells and human hepatocytes, and genes relating to DNA damage response that had divergent expression levels were added to the list (Table 30). The Thermofisher siRNA library contains 575 known DNA damage genes and 1725 siRNAs, and the Dharmacon siRNA library contains 242 genes and 968 siRNAs. The merged libraries were trimmed down to 235 genes (Table 31) to be manageable, and all were converted to manufacturer-guaranteed prevalidated SMARTPool siRNAs where possible.

Knockdown was performed according to manufacturer's specifications one day after seeding the cells at 7.5k or 26k cells/well in 48-well Poly-D-Lysine treated plates and after changing the media to OptiMEM (Fisher Scientific 31-985-062), using 0.5, 1, or 1.25uL 5uM siRNA and 0.2, 0.4, or 0.5uL Dharmafect 1 reagent respectively, incubated separately with 19uL OptiMEM each per well in a mastermix for all replicate wells with 10% extra volume to account for dead pipetting volume after mixing gently twice for 5min at room temperature, then combined and mixed gently twice and incubated at room temperature for 20 minutes prior to transfection of each well dropwise. The final condition chosen for the screen was 1uL 5uM siRNA and 0.4uL Dharmafect1.

RT-qPCR

Total RNA was harvested initially from samples during assay development using Zymo Direct-Zol RNA (Zymo Research R2051) kits according to manufacturer specifications for early experiments during assay development, but a switch was made to Zymo Quick-RNA (Zymo Research R1054) kits later to prevent toxicity from evaporated trizol for validating experiments. All RNA concentrations were standardized by dilution with H2O prior to RT-qPCR. RT-qPCR primer sequences were designed to target UNG1, UNG2, and the UNG1/2 shared transcript (Table 7, 8). These were converted into TaqMan probes with either VIC (for UNG1 only) or FAM channels and Nonfluorescent Quenchers (NFQ) (Thermofisher 4448508, 4331348). RT-qPCR was performed according to manufacturer's specifications using TaqPath™ 1-Step RT-qPCR Master Mix, CG (Thermofisher A15299), using a Thermofisher Quantstudio 7 Flex device

(Thermofisher 4485701) with 5uL of TaqPath mastermix, 13uL H2O, 1uL probe mix, and 1uL 20ng/uL total RNA sample.

Table 7. List of RT-qPCR primers designed for UNG1 and UNG2 single and combined amplicons.

GL555	ccgcccagtttggagaac	Reverse Transcriptase RT-qPCR primer F for UNG1&2 shared large amplicon coverage F
GL556	ccaaagaaccctctatacactgac	Reverse Transcriptase RT-qPCR primer R for UNG1&2 shared large amplicon coverage R
GL557	ggccaagcaaggtgttctc	Reverse Transcriptase RT-qPCR primer F for UNG1&2 shared small amplicon coverage F
GL558	ggtgccgcttcctatcaa	Reverse Transcriptase RT-qPCR primer R for UNG1&2 shared small amplicon coverage R
GL559	gatcggccagaagacgct	Reverse Transcriptase RT-qPCR primer F for UNG2 amplicon coverage F
GL560	atggccgccgcatctc	Reverse Transcriptase RT-qPCR primer R for UNG2 amplicon coverage R
GL561	accacttgcaggccatcc	Reverse Transcriptase RT-qPCR primer F for UNG1 amplicon coverage F
GL562	atcccattagcttgataaaatacgg	Reverse Transcriptase RT-qPCR primer R for UNG1 amplicon coverage R

Table 8. Amplicon sequences for UNG1 and UNG2 RT-qPCR.

UNG1&2 Shared RT-qPCR large amplicon (341bp):	ccgcccagtttggagaacatttataaagagttgtctacagacatagaggattttgttca tcctggccatggagatttatctgggtgggccaagcaaggtgttctccttctcaacgct gtcctcacggttcgtgcccatcaagccaactctcataaggagcgaggctgggagc agttcactgatgcagttgtgtcctggctaaatcagaactcgaatggccttgttttcttgc tctggggctcttatgctcagaagaagggcagtgccattgataggaagcggcaccat gtactacagacggctcatccctccccttgtcagtgtatagagggttctttgg
UNG1&2 Shared RT-qPCR small amplicon (200bp):	ggccaagcaaggtgttctccttctcaacgctgtcctcacggttcgtgcccatcaagc caactctcataaggagcgaggctgggagcagttcactgatgcagttgtgtcctggct aaatcagaactcgaatggccttgttttcttgctctggggctcttatgctcagaagaagg gcagtgccattgataggaagcggcacc
UNG2 RT-qPCR amplicon (135bp):	gatcggccagaagacgctctactccttttctcccccagccccgccaggaagcgac acgcccccagccccgagccggccgtccaggggaccggcgtggctggggtgcct gaggaaagcggagatgcggcggccat
UNG1 RT-qPCR amplicon (228):	accacttgcaggccatcccagccaagaaggccccggctgggcaggaggagcct gggacgccgccctcctcgccgctgagtgccgagcagttggaccggatccagagg aacaaggccgcggccctgctcagactcgcggcccgcaacgtgcccgtgggcttt ggagagagctggaagaagcacctcagcggggagttcgggaaaccgtattttatca agctaatgggat

Used in MiSeq sequencing for alignment of clusters.

Mammalian Cell Culture Methods

All cell culture was performed in 37°C 5% CO2-controlled incubators (Thermofisher 51030401) and in tissue culture hoods, with passaging and growth in tissue-culture treated T75 flasks (Corning 430641U) and experiments conducted in 48-well Poly-D-Lysine plates (Corning 354509) to encourage cell attachment. All cell detachment was performed using Tryp-LE (Thermofisher 12605010) detachment with 3-5min incubation in the incubators. Cell concentration was measured using Via2-Cassettes with a ChemoMetec NucleoCounter NC-200. Hek293T cells (ATCC CRL-3216) were grown in DMEM (Gibco 11965092) 10% FBS (Gibco A3160401) with P/S (Gibco 15140122). Hep G2 cells (ATCC HB-8065) were grown in E-MEM (ATCC 30-2003) 10% FBS with P/S. A lentivirus-transfected stable Hek293T cell line produced from prior

work at Beam containing a randomly integrated cassette (Table 9) with a fragment of the A1AT gene with the pathogenic E342K G>A SNP was utilized for various prime editing correction experiments. Human primary fibroblasts were grown in E-MEM with 15% FBS, 1X NEAA supplement (Gibco 11140050), and 1mM sodium pyruvate (Gibco11360070).

A long-lived coculture model with a feeder layer of mouse fibroblasts (Harvard University Howard Lab, 3T3-J2) alongside healthy primary human hepatocytes (BioIVT, donor RSE) that was previously developed and validated for liver identity expression profile by another team at Beam was utilized, with a 24-well plating on tissue culture treated plates at 300kcells/well of HPH cells and 20kcells/well of mouse fibroblasts. Cells were kept in 500uL CP media (BioIVT Z99029) with Torpedo antibiotic (BioIVT Z990007). A lentivirus with the A1AT E342K cassette designated pLV(Exp)-EGFP-TBG>(Serp1) was provided by another team at Beam (VectorBuilder LVLP(VB190903-1054hpz)-C), and 10uL corresponding to a MOI of 40 was transfected per well to insert the pathogenic SNP via random integration (Figure 10) (Table 10). Cassette integration was confirmed visually through a GFP reporter prior to editing experiments

Table 9. A1AT E342K lentivirus cassette sequence integrated into Hek293T cell line.

A1AT E342K pLenti Cassette	GTCACAGAGGAGGCACCCCTGAAGCTCTCCAAG GCCGTGCATAAGGCTGTGCTGACCATCGACAAG AAAGGGACTGAAGCTGCTGGGGCCATGTTTTTA GAGGCCATACCCATGTCTATCCCCCCCGAGGTCA AGTTCAACAAACCCTTTGTCTTCTTAATGATTGA ACAAAATACCAAGTCTCCCCTCTTCATGGGAAA AGTGGTGAATCCCACCCAAAAAGACCCAGCTTT CTTGTACAAAGTGGTTGATATCCAGCACAGTGG CGGCCGCTCGAGTCTAGAGGGCCCGCGGTTCGA AGGTAAGCCTATCCCTAACCCTCTCCTCGGTCT

```
>pLV(Exp)-EGFP-TBG>(Serp1)
CCAAGGCCGTGCATAAGGCTGTGCTGACCATCGACAAGAAAGGGACTGAAGCTGCTGGG
GCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGAGGTCAAGTTCAACAAACC
CTTTGTCTTCTTAATGATTGAACAAAATACCAAGTCTCCCCTCTTCATGGGAAAAGTGG
TGAATCCCACCCAAAAAGATCCTGCATTTTTGTATAAGGTCGTTGATATCCAGCACAGT
GGCG
```

Figure 10. Amplicon sequence of the lentivirus insert transfected into the HPH-fibroblast coculture model.

Used for alignment of MiSeq clusters.

Table 10. Primer sequences for amplicon PCR for the A1AT E342K lentivirus cassette inserted into HPH cells.

TFYA86	TGGAGTTCAGACGTGTGCTCTTCCGATCTCGCCACTGTGCTGGATAT
TFYA87	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCAAGGCCGTGCATAAG

Plasmid Transfection

Plasmid transfection was carried out for prime editors using 750ng or 150ng editor plasmid, 250ng or 50ng pegRNA plasmid, and 83ng nickRNA plasmid (for PE3 only for the 750/250 editor/pegRNA combination), with 1.1 or 0.3uL Lipofectamine 2000 (Thermofisher 11668019) in 10uL OptiMEM per half mix. For base editors used as controls, 150ng of editor and 50ng of gRNA were paired with 0.2uL Lipofectamine and 5uL OptiMEM per half mix. The DNA and transfection agent half mixes were combined separately in OptiMEM as mastermixes with volumes suitable for the number of wells to be transfected with 10% extra dead volume, then combined and incubated for 5min at room temperature. Transfections were performed dropwise for the 1083ng PE3, 1000ng PE2, 200ng PE3, and 200ng base editor plasmid conditions. For the overexpression

assays, 200ng of expression plasmid was used with 1uL Lipofectamine 2000 and 10uL OptiMEM per half mix.

RNA Transfection

In Hek293T and Hep G2 cells, prime editor RNA (3000ng editor mRNA, 400ng synthetic pegRNA, 150ng synthetic nickRNA) and OptiMEM (19uL) and Lipofectamine MessengerMax (Thermofisher LMRNA015) (1.1uL) and OptiMEM (19uL) per well were combined separately as mastermixes with 10% extra dead volume and incubated for 10 minutes and then incubated for 5 minutes together prior to use on cells. For base editor RNA, 150ng editor mRNA and 50ng gRNA with Lipofectamine MessengerMax (0.2uL) and OptiMEM (5uL per half mix) were combined and incubated as before. DNA was harvested 3 days after transfection.

For a control in the HPH A1AT E342K cassette, a previously published adenine base editor and guideRNA from prior work, ngcABEvar9/sgRNA025 (Packer, Chowdhary, Lung, et al., 2020), was transfected in experiments at 750ng mRNA and 250 gRNA. Given that previous internal work at Beam demonstrated that 1000ng of RNA was the optimal dose to preserve cell health in HPH cells, the prime editors were transfected in a ratio to preserve the 3000/400/150ng schema from Hek293T cells for PE3, with 845/112/42ng of mRNA/pegRNA/nickRNA, respectively. PE2 was transfected with a 3000:400 ratio of mRNA to pegRNA, at 882/117ng. Media was changed the day after and harvest of gDNA was carried out 3 days after transfection.

For primary human E342K fibroblasts, transfection was performed with the same 1000ng of RNA and preserved ratios for PE2/PE3 as the human primary hepatocytes. The optimal condition from prior work for base editors was used (100ng ngcABEvar9

mRNA, 50ng sgRNA025). In this case, cells were first trypsinized for 7 minutes in 4mL TrypLE in the flask, then the trypsin was neutralized with an equal volume of media. The needed amount of cells was aliquoted, and spun at 120RCF for 5min. The supernatant was aspirated and the cells were resuspended and washed with 25mL DPBS (Gibco 14190144), then centrifuged and aspirated as before again. The cells were then resuspended at 7.78*10^{6}cells/mL in Neon Transfection Buffer R (Invitrogen MPK1025) for a total of 70,000 cells per transfection well in 9uL, with an additional 1uL of RNA solution, and electroporated with the Neon system (Invitrogen MPK5000) using a protocol of 1000V, 40ms, 1 pulse before seeding into 500uL of media in 24-well tissue-culture treated plates (Corning 3337).

Production of mRNA

The production of mRNA was handled by Beam's mRNA platform team. Plasmids were digested by BbsI after cloning out extra BbsI cut sites, after which a 120A tail was cloned in by ligation. These plasmids were then transformed into bacteria and grown in large media batches at low temperature to prevent replication error in the polyA tail, after which the plasmid was extracted using Zymo Maxiprep or Gigaprep kits, and linearized by digestion. The mRNA was generated by in vitro transcription, purified with LiCl precipitation, and examined for full length product yield and purity by Fragment Analyzer (Agilent M5312AA).

Western Blots

Samples for western blots were generated via cell culture and transfection protocols as previously listed using either TREX2 siRNA or plasmid. Samples were

harvested in RIPA buffer (Pierce 89901) with benzonase nuclease (Sigma E1014-5KU) and Halt protease and phosphatase inhibitor cocktail (Thermo 78441) at 1uL/mL working volume each. Western blot analysis was performed by another team at Beam. Proteins were quantified using the Pierce BCA Assay kit (Thermofisher 23227) using 10uL of protein lysate or standard with the addition of 200uL of BCA working solution, then incubated at 37°C for 30min, then analyzed on a TECAN instrument (TECAN M200) with the default BCA protocol. Samples were then standardized to 50ng/uL. The Jess Simple Western system (ProteinSimple Jess) was used to run the western blot with Jess/Wess Separation 12-230kDa ladder (ProteinSimple SM-Woo4), Chemi Detection module (ProteinSimple DM-002), and anti-rabbit secondary antibody (ProteinSimple 042-206), using two different polyclonal TREX2-specific antibodies (Promocell PK-AB718-4971, Novus NBP1-76978-0.025mg). Image analysis was performed in the Compass for Simple Western software.

MiSEQ Library Preparation, Sequencing, and Analysis

MiSEQ libraries were prepared by aspirating media and treating cells with 200uL lysis buffer (10mM Tris pH8.0, 0.05% SDS, 2.5U/uL proteinase K (Invitrogen 25530049)) followed by a 45min 37°C incubation in a benchtop incubator with occasional agitation. 100uL of the lysis supernatant were transferred to 96-well PCR plates and treated with a 15min 95°C neutralization on a heat block. Amplicon generation was performed using 25uL Q5 Hotstart 2x Mastermix (New England Biosciences M0493S) and 0.25uL of each 100uM primer (Tables 10, 32), 1uL lysis supernatant, and 23uL H2O. The PCR protocol used was 95°C 2min, 30cycles(95°C 30sec, 62°C 30sec, 72°C 30sec), 72°C 2min. Quality control was conducted with gel electrophoresis using

5uL of completed PCR solution and 1uL NEB 6X Purple Dye (B7024S). and using Thermofisher E-Gels (Thermofisher G401004, G8100, G720802, EBM03) and SYBR-safe ChemiDoc imaging (Bio-Rad 12003154) and 6x Purple Dye. Barcoding PCR was performed with 25uL Q5 Hotstart 2x Mastermix, 1uL of 10uM combined barcode primer mastermix, 1uL of amplicon PCR reaction solution, and 23uL H2O, with the same PCR program but with 10 cycles instead of 30. QC was performed as before, and then the PCR reactions of amplicons of different sizes were pooled separately, mixed by vortexing, and 166uL or 333uL of the combined libraries were mixed with 33uL or 66uL 6X Purple Dye, respectively, and run on 2% agarose poured SYBR-safe gels, and gel extracted with Zymo DNA gel extraction kits, followed by another round of purification using Zymo DNA Clean and Concentrator (Zymo Research D4029) kits. The resultant product was nanodropped twice, taking the average value, and molar calculations were performed. 4nM suspensions of each library were created in 10mM Tris pH8.0 and a combined 4nM library with proportional volume of by number of samples per amplicon was subsequently created from those suspensions. 5uL of the combined 4nM library was combined with 5uL 0.2M NaOH freshly made from chilled 10M NaOH stock (to prevent loss of NaOH fidelity) and incubated to denature the libraries for 5min at room temperature. 990uL of HT1 MiSEQ buffer was added to quench the denaturation process, after which the libraries were at 20pM. 20% PhiX control was spiked into each library, using 120uL of 20pM denatured PhiX control that was generated using the same denaturation process and 480uL of the 20pM denatured sample library. All 600uL were loaded into MiSEQ cartridges and sequenced with V2 flow cells (Illumina MS-102-2003) for FASTQ data generation on an Illumina MiSEQ (Illumina SY-410-1003). Basecalls

were made by aligning reads from clusters against an amplicon sequence and searching for a target site sequence (Table 32, 33), followed by sorting and verbose output using a custom program created by Beam's bioinformatics team.

Chapter III

A1AT E342K Prime Editing Results

PE2/PE2_PAM and PE3/PE3b/PE3_PAM/PE3b_PAM strategy editor, pegRNA, and nickRNA PrimeDesign libraries were successfully designed for A1AT E342K. The experiments described herein identified NGA PAM prime editor combinations capable of precisely correcting the E342K mutation in SERPINA1.

NGA PAM PE2 Edits the A1AT E342K Locus

While several attempts in the literature have been made to prime edit the A1AT E342K locus, these attempts have focused on the usage of the canonical NGG PAM, which is problematic since the minimum RTT length for the closest NGG PAM is 25bp away from the pathogenic SNP, which is several bases out of the optimal range of 10-16bp recommended in the original Anzalone et al. paper. Therefore, in addition to designing NGG pegRNAs and cloning PE2 into a pCMV vector, VRQR and MQKFRAER variant prime editors with altered PAM-interacting domains were cloned into pCMV vectors and NGA and NGC PAM pegRNAs were cloned. Given that pegRNA PBS and RTT lengths can greatly influence the frequency of prime editing due to various causes such as internal homology of the PBS to a spacer, degradation of the RTT flap from exonucleases, etc., we sought to interrogate various pegRNA subcomponent lengths using PE2 strategy to identify candidate pegRNAs before proceeding to a PE3 strategy. The first round of pegRNAs were fixed at either 13bp or 17bp for the PBS length and with either the maximum or minimum RTT lengths as noted in the Anzalone paper. A total of 2 NGG (GL542/543), 2 NGA (GL544/545), and 6 NGC

(GL546/547/548/549/550/551) pegRNA spacers were included under this criteria for a total of 40 pegRNAs. Another NGA pegRNA (GL600) that was closest to the target and overlapped the pathogenic SNP was cloned with PBS lengths of 12/13/14/17bp and RTT lengths of 10/12/14/16bp. However, no detectable editing was observed with any of these pegRNAs in Hek293T cells.

In light of this negative data, a more comprehensively optimized screen was performed using the PrimeDesign software for the 6 closest protospacers on a pLenti cassette containing the E342K allele for correction. Where possible, manual design of PAM disruption was included (GL544/556/548) due to limitations of the software that prevented automated disruption of non-canonical PAMs. PBS lengths of 8/10/12/13/14/15bp and 3 RTT lengths at least +2 past the target SNP up to near the maximum optimal length (20bp) (GL544: 17/18/20, GL545: 10/14/20, GL556: 10/15/20, GL549: 23/24/25, GL548: 19/20/21). After work had begun, the GL542 spacer was shown in the literature to have activity, and a PrimeDesign was also conducted, resulting in PBS lengths of 8/10/13/14/15bp and RTT lengths of 25/26/27/28bp with PAM disruption. Unfortunately, GL542 lies outside of the bounds of the Lenti cassette, so the pegRNAs were designed against WT genomic DNA for insertion of the pathogenic SNP. Transfection with Lipofectamine 2000 in Hek293T cells produced significant detectable editing above background noise with two of the NGA protospacers (GL544 (~2-3%), GL556 (~0.5-1%)) and with the GL542 spacer (~1%) (Figure 11) (Table 34). Notably, the GL544 and GL556 NGA pegRNA variants which showed editing further demonstrated the incorporation of their respective PAM disruption silent mutations in the sequenced alleles (Figures 12, 13) (Tables 35, 36). Additionally, GL544 exhibited ~2-

fold higher editing than GL542 as PE2, and notably the leading GL542 pegRNA variant outperformed the 13bp PBS and 27bp RTT version that exhibited promising editing in the Liu et al. paper (Table 37) (Figure 14).

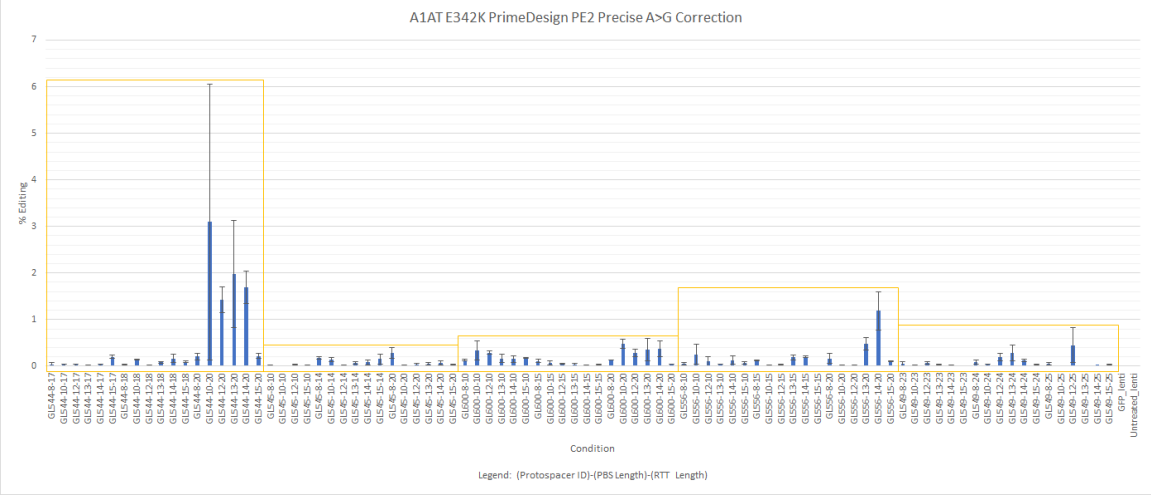


Figure 11. A1AT E342K PE2 PrimeDesign editing.

Percentage of alleles exhibiting precise correction of A1AT E342K following treatment by plasmid library of PrimeDesign pegRNAs and their corresponding plasmid PE2 editor variants. Each set of pegRNAs for a unique protospacer is indicated in separate boxes.

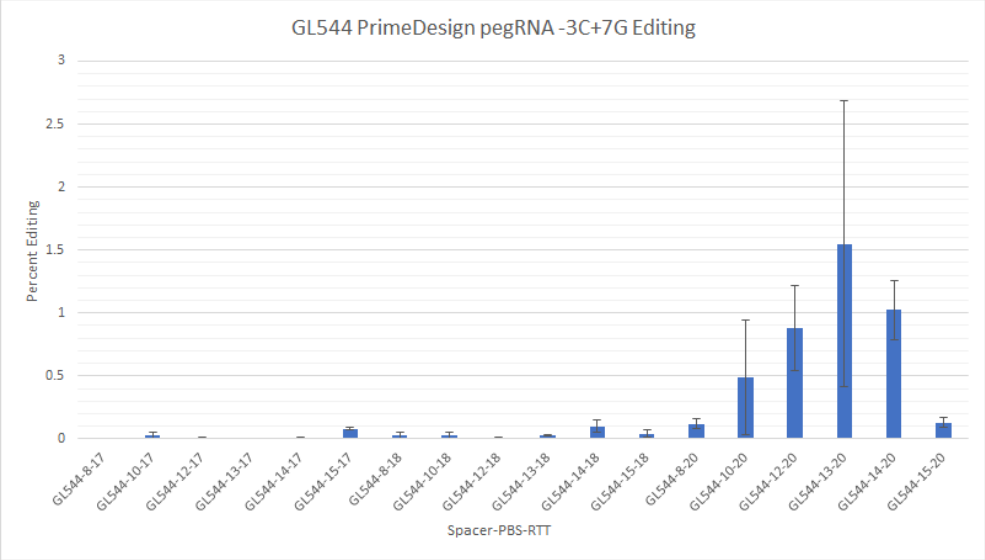


Figure 12. GL544 NGA A1AT E342K -3C+7G editing.

-3C+7G PAM disruption + precise correction allele is exhibited by the GL544 NGA pegRNA at A1AT E342K, proving specific correction.

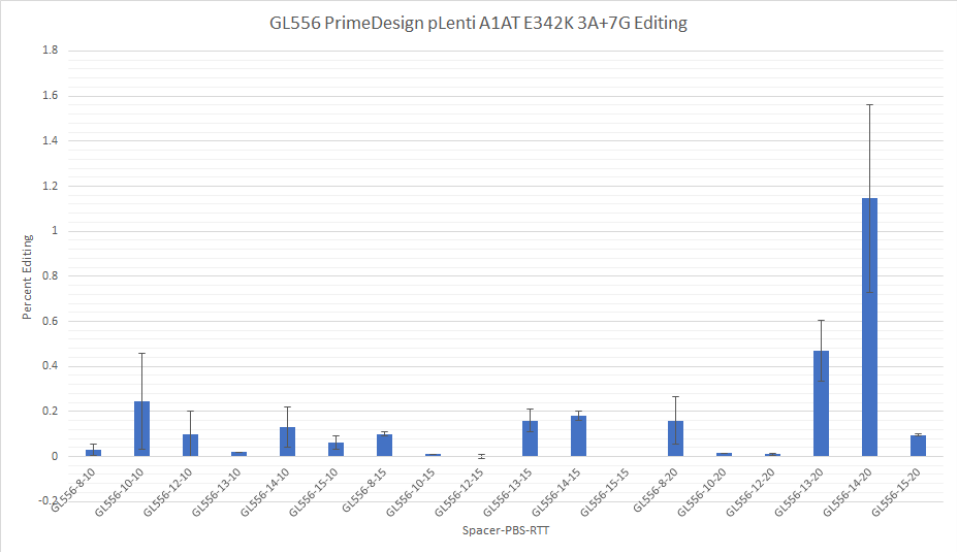


Figure 13. GL556 A1AT E342K 3A+7G editing.

3A+7G PAM disruption+precise correction allele is exhibited by the GL556 NGA pegRNA at A1AT E342K, proving specific correction.

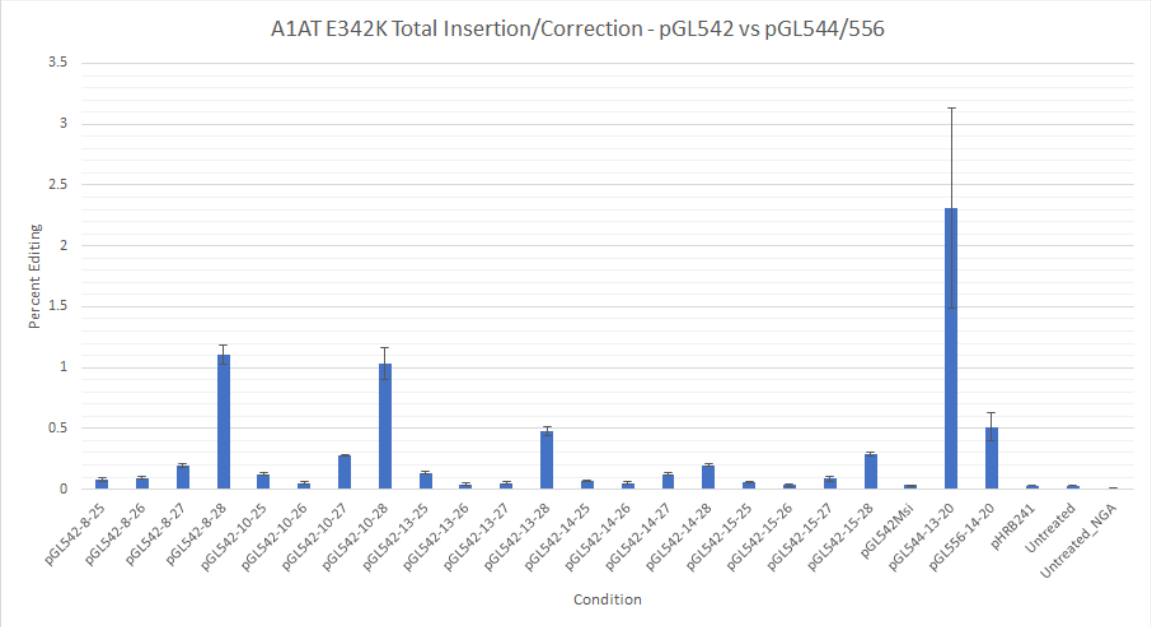


Figure 14. GL542 vs 556/544 A1AT E342K PE2 editing.

GL542 exhibited higher editing with the leading pegRNA candidates compared to GL556, but significantly lower (~2-fold) compared to GL544.

NGA PAM PE3 Competitively Edits the A1AT E342K Locus

Following this, nickRNAs from the PrimeDesign output were then cloned for all three pegRNAs, with a total of 7 for GL544, 8 for GL556, and 8 for GL542. Additionally, modified VRQR PIDs for NGA PAMs for base editors that exhibited greater affinity for NGA PAMs as a result of select mutations (I322V,S409I,E427G,R654L,R753G) (Miller et al. 2020) were also tested as PE2 against the VRQR (pGL191) condition. Moving from PE2 to PE3 appeared to double correction for GL556 and GL542, with an uncertain effect for GL544 that exhibited inconsistent editing but which did not fall below PE2 editing (Figure 15) (Table 38) the behavior of which is tentatively surmised to potentially be due to the cutting of multiple insertion

sites of the E342K lentivirus cassette. A direct comparison between the three PAMs revealed that the GL544 NGA PAM exceeded the editing of the GL542 NGG PAM for all conditions (Figure 16) (Table 39). However, the caveat is that the comparison is between genomic versus lentiviral loci. Unexpectedly, the heavily optimized mutant VRQR PID exhibited no detectable editing, suggesting that the mechanics of prime editing and base editing are sufficiently different that what works for one may not translate to the other. 3-5% editing was observed for GL544, 1-1.5% for GL556, and 2-3% for GL542.

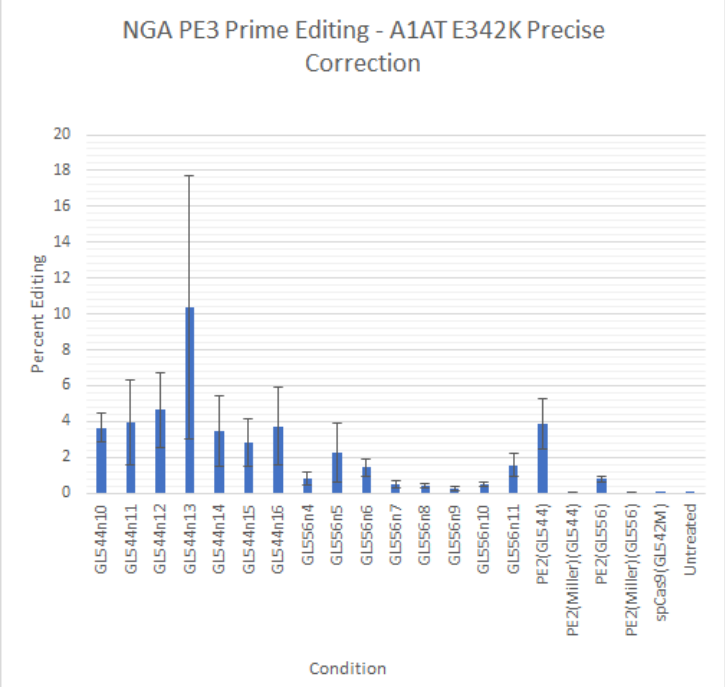


Figure 15. GL544/556 A1AT E342K PE3 editing.

NGA prime editing with PE3 strategy at A1AT E342K lentivirus cassette in Hek293T, compared against PE2 editing with or without the Miller et al. PID mutations and Cas9 control. All "GLXXXn##" nickRNAs were transfected with PE2-VRQR and either GL544-13-20 or GL556-14-20, according to their respective designations.

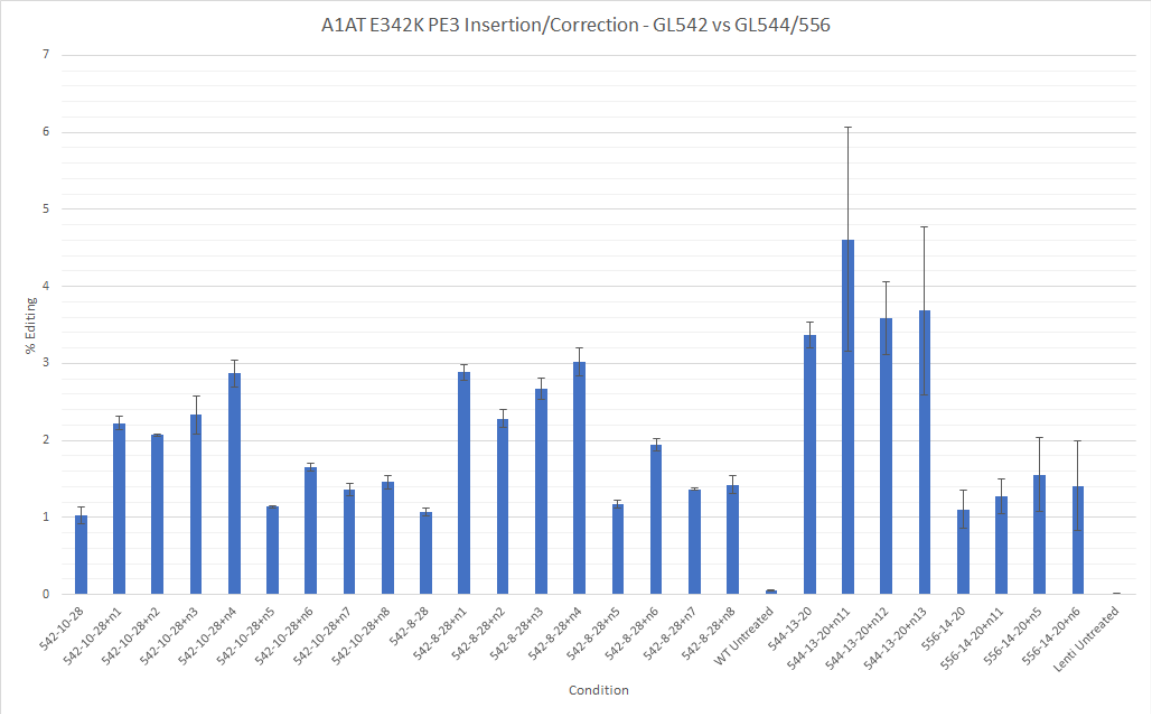


Figure 16. GL542 vs 544/556 A1AT E342K PE3 editing.

Comparison of GL542 genomic insertion versus GL544/556 correction of lentiviral cassette for A1AT E342K in Hek293T, with both PE2 and PE3 strategies.

Plasmid to mRNA conversion of NGA PE2/PE3 Maintains Prime Editing Frequency

As plasmid is toxic in vivo and to primary cells, testing of mRNA versions of prime editors was necessary. Hek293T cells containing the E342K lentivirus cassette were transfected with either plasmid or mRNA NGA prime editors. Conversion to mRNA sometimes increases editing frequencies, however, this was not clearly observed here. For all conditions tested, t-tests suggested plasmid and mRNA generated similar editing (Table 40) (Figure 17). We suspect that there may be a small increase in editing frequency for GL556, but the low levels of editing may have caused low assay sensitivity.

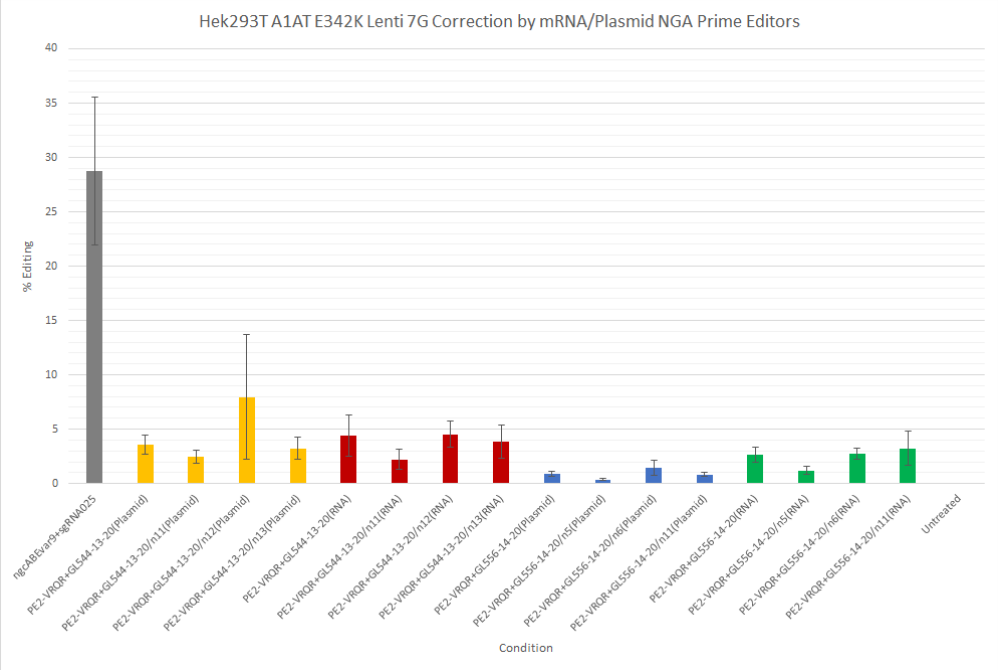


Figure 17. Comparison of mRNA versus plasmid correction at A1AT E342K in Hek cells.

GL544 plasmid (yellow) and RNA (red) show comparable levels of editing, and GL556 plasmid (blue) and RNA (green) show comparable levels of editing.

NGA PAM PE3 Detectably Edits A1AT E342K in Human Primary Fibroblasts.

Finally, PE2-VRQR mRNA was tested with synthetic peg/nickRNA GL544-13-20+GL544n11/12/13 and GL556-14-20+GL556n5/6/11 for correction in primary fibroblasts containing the E342K pathogenic allele. The cells were treated with either a prevalidated base editor (ngcBE4var9) and gRNA (sgRNA025) (M. Packer, V. Chowdhary, G. Lung et al., 2020) or the NGA prime editors and respective guides for the E324K site using electroporation. Detectable levels of editing were observed for the PE2 GL544-13-20 condition in primary fibroblasts, with one replicate generating 1.99%

editing, but with average editing at 0.836% and with a standard error of the mean of 0.58, whereas base editor generated 20% editing (Table 41) (Figure 18).

An additional attempt was made to attempt to develop an assay for human primary hepatocytes. However, these HPH cells were from healthy donors and lacked the pathogenic SNP, which needed to be inserted. Additionally, integration of a cassette with the SNP would take time longer than the lifetime of the HPH cells, which is roughly four days. Therefore a long-lived coculture model with a feeder layer of mouse fibroblasts alongside HPH cells was used, and was transfected with a lentivirus cassette containing A1AT E342K. Thereafter integration was confirmed by reporter fluorescence prior to transfection. However, control editor conditions failed to produce an acceptable baseline positive control rate of editing, likely due to the integration efficiency of the cassette, and so the assay was unable to be completed in time and was abandoned.

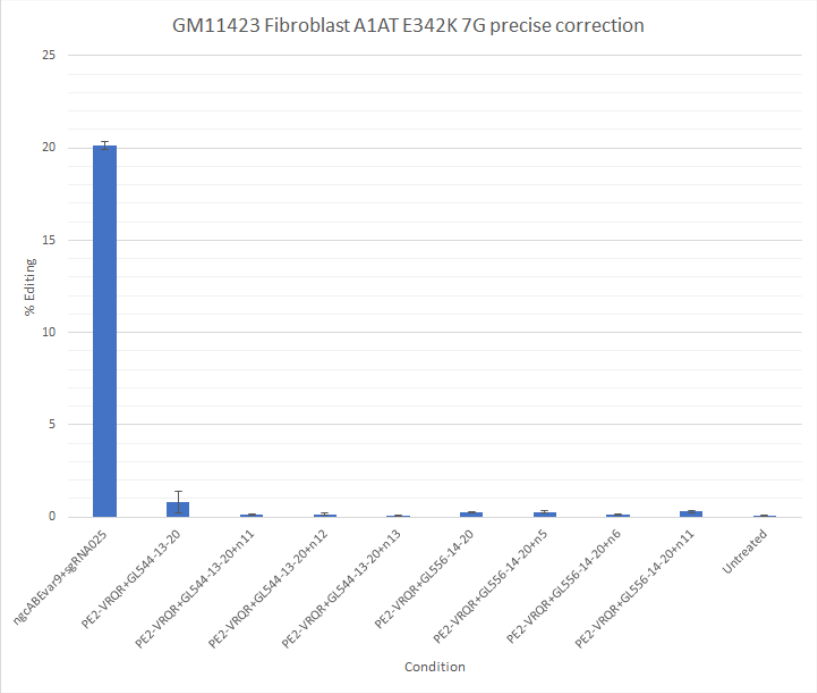


Figure 18. Precise correction of A1AT E342K by NGA prime editors in primary fibroblasts.

RNA was transfected into GM11423 primary fibroblasts containing the E342K SNP. GL544-13-20 PE2 strategy generated low (<1%) but detectable editing, although this is not statistically significant over noise.

Chapter IV

siRNA Knockdown Prime Editing Results

An siRNA library and lead HBB editor/pegRNA/nickRNA combinations were successfully chosen. Plasmid overexpression vectors to orthogonally validate siRNA hits were also tested. Several genes exhibited significant influence on prime editing outcomes, including TREX2.

Validation of siRNA Knockdown Timing and Influence on Gene Editing

Hek293T cells were transfected with siGENOME SMARTPool UNG siRNA and total RNA was harvested across five days, after which one-pot RT-qPCR was performed for UNG1, UNG2, and their combined transcript, which were chosen as controls due to their properties as uracil base excision enzymes which should theoretically have no effect on prime editors but a strong effect on cytidine base editors. Knockdown was measured via cycle count surrogate and was confirmed for all three transcripts. A difference of cycle number of between ~2-3 was observed across all five days, corresponding to at least the 75% knockdown guarantee from siGENOME (Figure 19) (Table 42). However, relative knockdown decreased daily. This informed future experiments, indicating that the prime editor ought to be transfected as soon as possible after knockdown. Accordingly, the assay for sequential transfection was designed with a day 0 seeding, day 1 knockdown, and day 2 prime editor transfection. The RT-qPCR experiment was repeated as during optimization for the sequential transfection, the format was changed from 96-well to 48-well as cell density was thought to play a role, and the knockdown effect was once again confirmed (Table 43) (Figure 20).

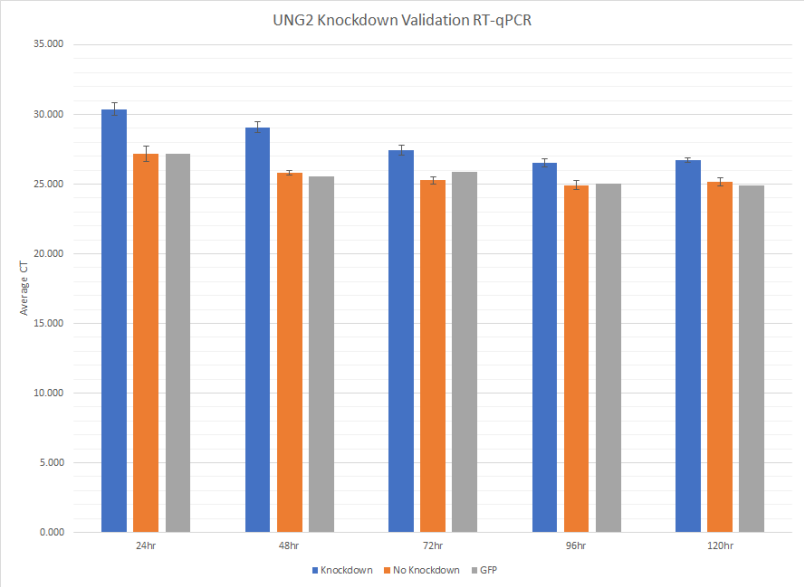


Figure 19. RT-qPCR results for UNG2 knockdown.

RT-qPCR CT scores for UNG2 transcript for 24/48/72/96/120hr, indicating a marked increase in CT count for the knockdown condition, corresponding to a decrease in transcript.

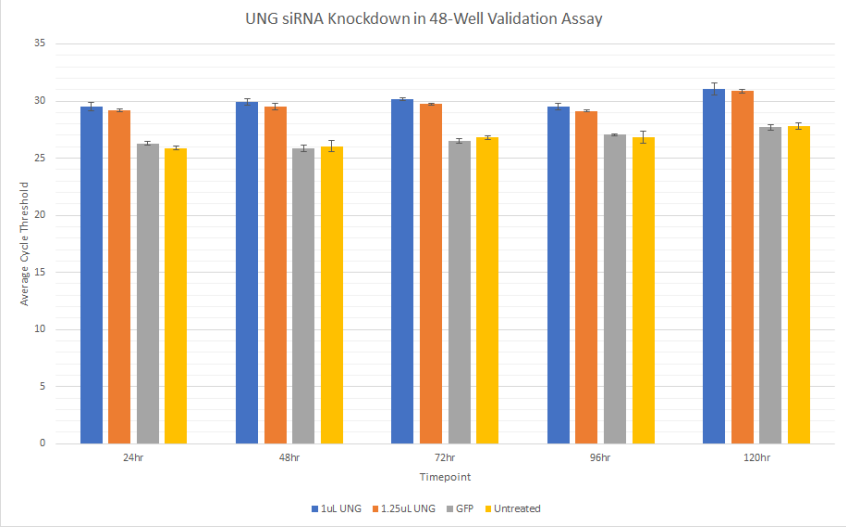


Figure 20. Ct scores for UNG2 in 48-well format.

Confirmed knockdown by 1uL 5uM UNG siRNA.

To validate the protein-level effects of siRNA knockdown, UNG knockdown followed by PE3 or base editor with (BE4) or without (BE4-noUGI, aka BE3B) uracil glycosylase inhibitor (Komor et al., 2017) transfection was performed. Since UNG excises uracil, it removes uridines that cytidine base editors generate, interfering with the C>U>T conversion process. As expected, knockdown of UNG greatly increased the proper conversion of C to T by ~20% and decreased C to G by ~12%, whereas for BE4 without UGI, knockdown increased C to T conversion by ~15% and decreased C to G editing by ~2% (Figure 21) (Table 11), thus validating the functionality of the assay format for knockdown followed by gene editing. For this set of experiments, the same amount of base editor and prime editor were used (150ng editor, 50ng total guide RNA (split 100:67 molar ratio between peg/nick for PE3), and PE3 mRNA did not show a significant difference in editing outcomes between the UNG and untreated control, validating that RNA transfection alone does not alter prime editing (Figure 22) (Table 12).

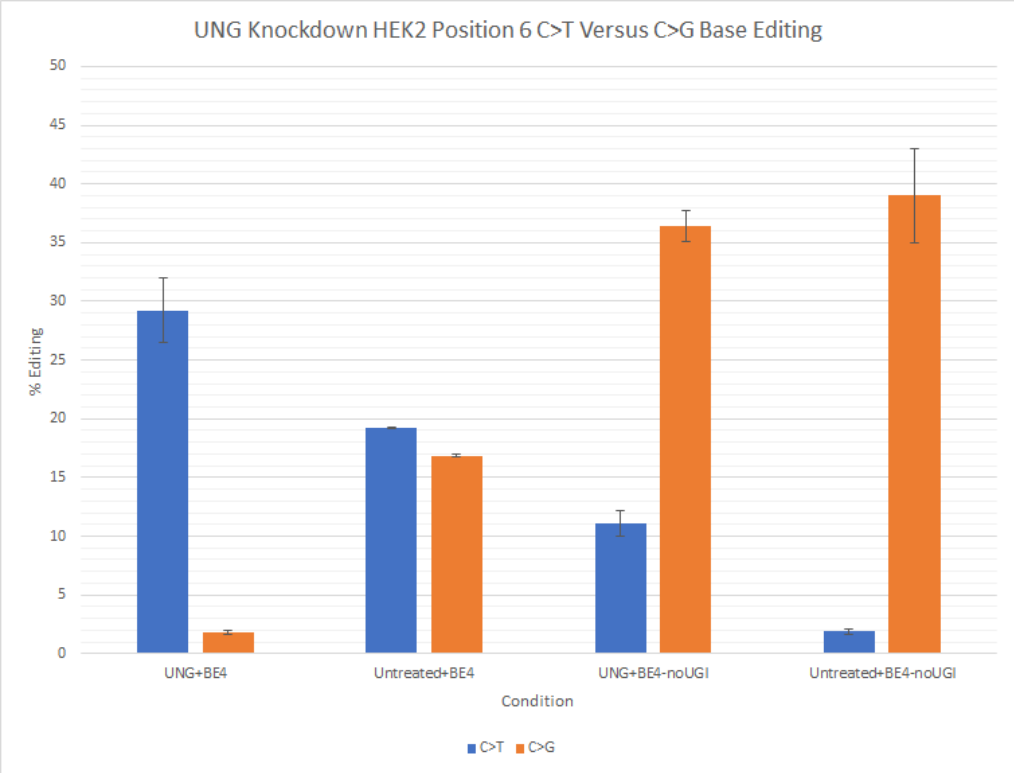


Figure 21. Hek2 target site C>T versus C>G editing.

UNG siRNA knockdown followed by base editor transfection.

Table 11. Percent editing of BE4 and BE4-noUGI with or without UNG siRNA knockdown.

Condition	Average C>T	Average C>G	C>T Error	C>G Error
UNG+BE4	29.20666667	1.846666667	2.742908027	0.174578858
Untreated+BE4	19.25333333	16.85666667	0.088380491	0.139323764
UNG+BE4-noUGI	11.06	36.43	1.065879918	1.339788541
Untreated+BE4-noUGI	1.886666667	39.02333333	0.244767463	4.036732727

Table 12. Percent editing of PE3 under knockdown with UNG or with GFP control transfection.

Condition	22T Average	22T Error
Untreated	0.013333333	0.008819171
GFP	0.01	0
UNG	0.013333333	0.003333333
Untreated+PE3	12.16666667	1.482299265
GFP+PE3	9.67	2.057968254
UNG+PE3	9.72	1.081310933

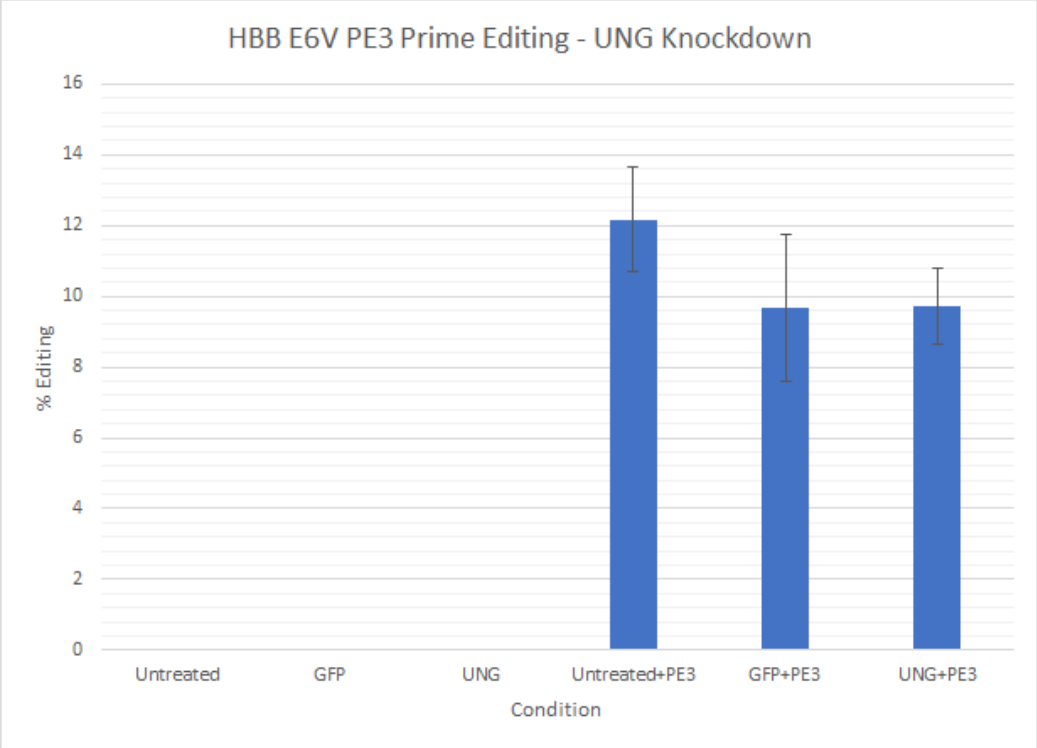


Figure 22. Prime editing with PE3 at HBB 22T with UNG siRNA.

Editing does not appear to be significantly affected.

siRNA Knockdown Suggests TREX2, FANCM, and LIG4 Effect on Prime Editing

Subsequently for the final screen, amounts of prime editor were increased to 3000ng of editor mRNA, 400ng of pegRNA, and 150ng of nickRNA and 1.1uL of Lipofectamine MessengerMax. A day 0 seeding at 7.5k Hek293T cells per 48-well followed by day 1 knockdown with 1uL 5uM SMARTPool siRNA and 0.4uL Dharmafect1 (Horizon Discovery T-2001-03), then a day 2 transfection with the prime editing reagents, and gDNA was harvested on day 5, after which amplicon PCR and barcoding PCR and MiSeq were performed. The screen was conducted as 4 staggered separate experiments for each 96-well of siRNAs in triplicate due to the sheer number of siRNAs involved, with 12 controls per experiment (Non-Targeting siRNA Pool #1/2, Non-Targeting siRNA #1/2/3/4/5, Lamin A/C, Cyclophilin B, GAPD Control siRNAs, Untreated+PE3, Untreated) (Figures 23, 24, 25, 26) (Tables 13, 14, 15, 16). siRNAs were selected based on an ontology sort of relevant functions to prioritize genes, with assistance from Beam's bioinformatics team to find differentially regulated genes between Hek293T/U2OS/HeLa cells, and combined with Thermofisher and Dharmacon libraries for DNA Damage siRNAs, before all were converted where available to Dharmacon siGENOME SMARTPool siRNAs.

Table 13. Percent editing of precise correction and undesired edits for siGENOME screen plate 1.

Condition	Average 22T	Average Undesired	22T Error	Undesired Error	Condition	Average 22T	Average Undesired	22T Error	Undesired Error	Condition	Average 22T	Average Undesired	22T Error	Undesired Error
GAPD Control siRNA	44.60333333	0.556666667	26.01520859	0.237299623	POLL	15.34333333	0.643333333	3.246805267	0.173237153	RPAIN	11.37333333	0.56	0.788972609	0.15
RAD17	44.25666667	1.513333333	23.97169048	0.52834122	XRCC5	15.15666667	0.676666667	2.196165851	0.161692438	ERCC3	11.18	0.433333333	1.498299036	0.076666667
MSH2	22.97333333	1.07	0.746688988	0.183575598	FANCA	15.09	0.693333333	1.066630208	0.158569999	OGG1	11.22	0.41	2.245818336	0.077674535
Control siRNA #5	22.58666667	0.79	0.639800316	0.055075705	MLH3	15.05666667	0.663333333	2.268790672	0.208832735	FANCL	11.12	0.436666667	0.612209115	0.018559215
ERCC2	22.79666667	1.373333333	0.382811239	0.18657736	POLB	14.87	0.536666667	1.985253972	0.133458275	NTHL1	11.47	0.763333333	1.72123018	0.156240555
EXO1	22.29	0.863333333	0.872314928	0.095974534	TP53BP1	14.61333333	0.626666667	0.884163132	0.020275875	XRCC6	10.92	0.386666667	0.551845389	0.216512766
BRCA2	21.86333333	1.013333333	1.437153823	0.247408794	POLG2	14.25333333	0.566666667	0.780220766	0.018559215	DCLRE1C	10.66	0.396666667	0.979047156	0.012018504
Untreated+PE3 Control	21.45	1.006666667	0.990420786	0.09171211	MLH1	14.51666667	0.946666667	0.935063872	0.239606158	CHEK2	10.55	0.496666667	0.798394222	0.049103066
FANCD2	21.33	0.95	0.975346092	0.096436508	Control siRNA Pool #2	14.20666667	0.563333333	1.475243859	0.092074849	POLE2	10.45	0.383333333	2.410276609	0.129142471
LIG1	20.95333333	0.756666667	0.338247115	0.089876458	MSH3	14.51666667	1.063333333	1.369359136	0.506304695	MSH6	10.54666667	0.546666667	0.964578204	0.032829526
LIG3	20.70333333	0.93	0.603995953	0.179257729	ERCC4	13.99333333	0.526666667	1.502412874	0.065659052	FANCF	10.33	0.353333333	1.226499083	0.068879928
RAD9A	20.07666667	0.873333333	1.114904082	0.143333333	POLQ	14.01333333	0.593333333	1.118158208	0.043333333	NEIL1	10.40666667	0.436666667	1.189430302	0.12251984
Lamin A/C Control siRNA	19.93666667	0.723333333	0.813333333	0.053644923	APEX2	13.9	0.586666667	3.096487257	0.127322862	NUDT1	10.19	0.323333333	0.759736797	0.060644685
POLE	19.82666667	0.806666667	1.412389622	0.141696075	RPA4	13.92333333	0.666666667	1.09138037	0.136788564	FANCE	10.29666667	0.493333333	1.44916451	0.086666667
MPG	19.16333333	0.753333333	3.067150324	0.143797697	ERCC8	13.52	0.406666667	0.23515952	0.052387445	CHEK1	10.3	0.5	0.816823114	0.09504385
RAD1	18.86	0.86	1.271154331	0.104403065	APEX1	13.69333333	0.623333333	1.123867331	0.090246576	MGMT	10.03333333	0.356666667	0.132455443	0.033333333
XRCC1	18.51	0.866666667	0.811007604	0.056960025	XRCC4	13.59333333	0.613333333	0.881709199	0.118649812	ATM	9.78	0.366666667	0.535007788	0.109898943
BRCA1	18.21	0.773333333	1.16258333	0.139562809	RAD51	13.42	0.533333333	2.726579053	0.123468395	POLH	9.736666667	0.3	0.747224494	0.03
POLM	17.58666667	0.6	0.921128535	0.06350853	DDB2	13.44333333	0.596666667	0.949075574	0.012018504	MSH5	9.863333333	0.516666667	1.710568457	0.163435342
RAD51C	17.55333333	0.68	3.155045342	0.126622799	NEIL2	13.39666667	0.573333333	0.193591781	0.074461026	RPA1	9.613333333	0.303333333	0.392612334	0.118930418
ERCC1	17.28	0.896666667	2.440409802	0.191688405	MUTYH	13.13	0.613333333	0.460470774	0.003333333	Control siRNA #2	9.503333333	0.403333333	0.980226731	0.026034166
FANCB	17.25333333	0.863333333	2.321639746	0.158569999	XRCC3	12.99	0.503333333	2.403060549	0.034801022	POLA1	9.343333333	0.37	0.58167383	0.081853528
POLI	16.94	0.64	0.441701257	0.073711148	Control siRNA #4	12.86333333	0.523333333	1.086298506	0.040960686	PNKP	9.183333333	0.256666667	0.422111096	0.08412953
Control siRNA #3	16.87333333	0.67	0.712327952	0.094516313	UNG	12.84333333	0.736666667	0.417306176	0.073333333	ERCC6	9.123333333	0.293333333	1.029924485	0.075129518
Cyclophilin B Control siRNA	16.46666667	0.656666667	0.925964962	0.053644923	ERCC5	12.58	0.52	0.516268664	0.160416126	DCLRE1A	8.613333333	0.32	1.2806422	0.047258156
FANCC	16.17	0.58	0.37072002	0.090184995	ATRX	12.47666667	0.74	1.127420852	0.136503968	MSH4	8.566666667	0.406666667	1.633527199	0.177043623
POLD1	16.25666667	0.696666667	2.134957715	0.175720738	FEN1	12.30333333	0.546666667	0.696762354	0.106822802	DDB1	8.403333333	0.396666667	1.291256924	0.124409717
TP53	16.26333333	0.716666667	0.487590447	0.118930418	DCLRE1B	12	0.566666667	0.369504172	0.117236703	LIG4	7.323333333	0.293333333	0.997084639	0.060092521
ATRIP	15.96	0.683333333	1.549913976	0.116237305	POLK	11.94	0.51	0.481283008	0.126622799	FANCM	7.2	0.3	0.731186251	0.11
Control siRNA #1	15.86	0.6	0.313209195	0.045092498	RPA2	11.79666667	0.42	0.279065902	0.089628864	Untreated Control	0.083333333	0.05	0.068394282	0.05
NEIL3	15.70666667	0.613333333	0.681428239	0.069602043	POLG	11.77666667	0.493333333	1.523836335	0.141106737					
XRCC2	15.81666667	0.686666667	1.793621414	0.109290642	ATR	11.75	0.56	0.962496753	0.115325626					
DNA2	15.42333333	0.63	2.498595161	0.213775583	Control siRNA Pool #1	11.41666667	0.406666667	1.222052008	0.091347931					

Table 14. Percent editing of precise correction and undesired edits for siGENOME screen plate 2.

Designation	Average 22T	Average Undesir	22T Error	Undesired Error	Designation	Average 22T	Average Undesir	22T Error	Undesired Error	Designation	Average 22T	Average Undesir	22T Error	Undesired Error
PAXIP1	16.23333	2.22	0.179475	0.072342	Control_Pool2	15.82333	2.13	0.223333	0.032146	RAD23A	15.67333	2.216667	0.112891	0.143798
SPRTN	16.16	2.043333	0.449778	0.124677	TWIST1	15.82	2.05	0.270555	0.04	BIVM	15.66667	2.13	0.184782	0.055076
APBB1	16.14	2.103333	0.310215	0.065659	RASSF7	15.81667	2.146667	0.23877	0.065659	CyclophilinB_Co	15.66333	2.11	0.28881	0.034641
Control_4	16.14	2.22	0.164418	0.089629	PML	15.81333	2.08	0.03383	0.068069	HMGA2	15.66	2.186667	0.3646	0.060645
USP16	16.10333	2.126667	0.234473	0.031798	MCM8	15.81	2.193333	0.088882	0.069602	POLD2	15.65	2.06	0.166533	0.015275
SIRT4	16.09333	2.21	0.14746	0.041633	CARM1	15.80667	2.063333	0.172466	0.052387	STK11	15.64667	1.986667	0.153442	0.036667
BCCIP	16.09	2.07	0.07	0.07	POLR2F	15.80667	2.2	0.297452	0.046188	SOX4	15.64333	2.076667	0.092616	0.040961
MDM2	16.05333	2.076667	0.121289	0.082932	POLR2H	15.80667	2.12	0.023333	0.011547	EEPD1	15.63	2.236667	0.037859	0.038442
BABAM1	16.01333	2.08	0.385155	0.170978	Control_Pool1	15.79667	2.116667	0.15836	0.008819	WRAP53	15.62667	2.123333	0.278109	0.094927
RNASEH2B	15.99333	2.246667	0.098376	0.006667	Control_2	15.79333	2.253333	0.087623	0.089505	BRSK1	15.61667	2.223333	0.235961	0.098376
PRMT1	15.97667	2.21	0.173429	0.121655	FNIP2	15.79333	2.196667	0.186577	0.112151	RAD52	15.61667	2.11	0.105251	0.100167
CXCL12	15.97667	2.21	0.057831	0.037859	MYC	15.79333	2.24	0.102035	0.015275	POLD4	15.61333	2.133333	0.082932	0.060093
SMARCB1	15.96	2.14	0.133167	0.028868	E2F8	15.79	2.086667	0.122882	0.061192	FOXO1	15.59333	2.056667	0.233548	0.083533
Control_1	15.95333	2.086667	0.058405	0.057831	FBXO18	15.79	2.226667	0.165227	0.023333	BCL2L12	15.57667	2.08	0.371678	0.055076
FMN2	15.95333	2.19	0.243607	0.081445	MCM7	15.78	2.1	0.06245	0.073711	KMT5C	15.57667	2.113333	0.248283	0.038442
RAD50	15.93667	2.14	0.065659	0.035119	GNL1	15.77333	2.156667	0.094927	0.012019	PMAIP1	15.57667	2.233333	0.131191	0.086859
NDRG1	15.92667	2.11	0.058119	0.005774	Lamin_Control	15.76667	2.096667	0.043716	0.064893	ALKBH8	15.57	2.176667	0.106927	0.044096
RAD9B	15.92667	2.043333	0.086859	0.040552	WRNIP1	15.76333	2.183333	0.226593	0.08413	RAD23B	15.54333	2.23	0.139801	0.035119
RAD54B	15.92	2.113333	0.040415	0.036667	MDM4	15.75667	2.013333	0.077531	0.033333	ELL3	15.52667	2.166667	0.144952	0.006667
CHD2	15.91667	2.19	0.221008	0.041633	RAD21	15.74667	2.083333	0.211371	0.026034	INTS3	15.52333	2.196667	0.136789	0.018559
MCM9	15.91333	2.106667	0.364067	0.057831	VAV3	15.74	2.13	0.162583	0.036056	BARD1	15.52	2.16	0.207445	0.08544
UBQLN4	15.90333	2.056667	0.273821	0.092796	RNF138	15.73667	2.203333	0.233833	0.026034	PRKCD	15.5	2.126667	0.175024	0.088192
CDC45	15.89333	2.16	0.192209	0.080829	Control_3	15.72333	2.163333	0.097011	0.069362	Untreated_PE3	15.5	2.086667	0.025166	0.01453
FANCG	15.88667	2.086667	0.104137	0.049777	CDC14B	15.72	2.06	0.242143	0.036056	AUNIP	15.49333	2.146667	0.247813	0.112891
FOXO3	15.88333	2.106667	0.108372	0.051747	DMAP1	15.71333	2.196667	0.080898	0.063333	CHD1L	15.48667	2.173333	0.143798	0.023333
EEF1E1	15.88	2.156667	0.151327	0.051747	Untreated	15.71333	2.213333	0.372439	0.056075	ANKLE1	15.48333	2.133333	0.153768	0.059255
JMY	15.88	2.18	0.170392	0.025166	GAPD_Control	15.71	2.233333	0.178979	0.078811	Control_5	15.47	2.166667	0.089629	0.046667
POLN	15.88	2.076667	0.208167	0.016667	TEX15	15.69667	2.14	0.259893	0.098658	EYA4	15.46333	2.086667	0.108372	0.038442
RAD54L	15.87	2.086667	0.225019	0.104137	PPP4C	15.69333	2.223333	0.112151	0.053645	NSMCE4A	15.45	2.23	0.196554	0.081445
TERF2IP	15.85333	2.173333	0.08838	0.023333	NUDT16L1	15.68667	2.043333	0.08838	0.026667	WDR48	15.31	2.143333	0.102144	0.044845
RAD18	15.85	2.143333	0.151327	0.078599	GINS2	15.68	2.113333	0.135769	0.0857					
ASCC2	15.84667	2.06	0.131698	0.005774	POLD3	15.67667	2.093333	0.034801	0.035277					
SESN2	15.83333	2.046667	0.188886	0.141696	COPS7B	15.67333	2.12	0.018559	0.065064					

Table 15. Percent editing of precise correction and undesired edits for siGENOME screen plate 3.

Gene Name	Average 22T	Average Undesired	22T Error	Undesired Error	Gene Name	Average 22T	Average Undesir	22T Error	Undesired Error	Gene Name	Average 22T	Average Undesir	22T Error	Undesired Error
siRNA #5 Contro	19.71333333	0.933333333	0.029627	0.708386	SOD1	16.24666667	0.516666667	1.221697	0.131191	RUVBL2	13.65	1.53	0.969347	0.183576
CTC1	19.61666667	0.32	0.576233	0.005774	siRNA Pool #2 C	16.22	0.186666667	0.626844	0.012019	REV3L	13.43666667	1.373333333	1.826512	0.10729
ZBTB18	19.36	1.146666667	1.084866	0.524479	CSNK1E	15.93	1.31	1.372309	0.273922	RFC3	13.42	0.84	1.697773	0.208407
MCRS1	19.3	1.003333333	0.520481	0.298682	siRNA Pool #1 C	15.92333333	0.43	2.081829	0.160104	SPIRE2	13.33	0.643333333	1.085188	0.097011
AEN	19.22333333	0.82	1.432554	0.385011	siRNA #1 Contro	15.87	0.55	0.856757	0.108167	GTF2H5	13.01666667	0.84	1.089118	0.135277
ZSWIM7	19.07666667	0.4	0.365529	0.017321	KLHL15	15.84333333	0.67	0.641413	0.150111	MAD2L2	12.74666667	0.27	0.299128	0.005774
SETMAR	18.81	0.783333333	2.728999	0.31866	MDC1	15.69666667	0.33	0.736599	0.02	PARP2	12.66333333	1.003333333	1.524569	0.262064
PAGR1	18.70666667	1.076666667	1.436017	0.497203	UBE2B	15.62333333	0.376666667	0.496969	0.024037	BRD4	12.46666667	0.61	1.005258	0.101489
PRKCG	18.18	0.583333333	0.668905	0.269588	TDG	15.61666667	0.993333333	0.624642	0.304923	BRCC3	12.46333333	0.956666667	1.277567	0.496801
HMGB2	18.16666667	0.556666667	0.295654	0.206989	UVRAG	15.47	0.323333333	0.420159	0.023333	PER1	12.25	0.713333333	0.168622	0.05696
PMS2	18.07666667	0.953333333	3.136167	0.688485	SMC3	15.35666667	0.543333333	0.884163	0.139324	PMS1	12.14333333	0.8	0.383159	0.111355
ATF2	18.07	0.52	1.154426	0.245017	SLFN11	15.30666667	0.556666667	0.466238	0.184421	MUS81	12.07	0.846666667	1.369319	0.443784
RASSF1	17.9	0.5	1.336076	0.072111	RPS27L	15.22666667	0.546666667	0.796709	0.034801	UBA52	12.02	1.993333333	0.441928	0.566755
GADD45A	17.87333333	0.62	1.466428	0.41	GTF2H2	15.22	0.656666667	1.248319	0.123873	GTF2H4	11.76333333	0.613333333	0.165563	0.093333
AATF	17.63666667	0.913333333	1.020087	0.233833	KMT5A	14.95	1.233333333	0.395011	0.223632	TYMS	11.75333333	0.71	0.606089	0.025166
TOPBP1	17.59333333	0.54	0.621673	0.023094	DDX11	14.93	1.476666667	0.464579	0.439179	XPC	11.73	0.496666667	0.240278	0.108985
MTA1	17.56666667	0.716666667	0.047022	0.17901	Untreated+PE3	14.88666667	1.046666667	0.490725	0.786751	siRNA #2 Contro	11.66	0.55	0.735006	0.16773
GTF2H1	17.49666667	0.833333333	2.312123	0.498743	NSMCE3	14.65666667	0.4	1.547227	0.056862	WRN	11.35	0.833333333	1.598166	0.145297
siRNA #3 Contro	17.40333333	0.616666667	1.357232	0.145297	CENPX	14.59333333	0.916666667	0.849987	0.145182	RFC4	11.23333333	1.18	0.382376	0.178979
siRNA #4 Contro	17.39666667	0.32	0.874115	0.043589	TRIM28	14.43333333	0.683333333	1.600566	0.208513	RTEL1	11.00666667	0.54	1.591072	0.157162
IRF7	17.32333333	0.466666667	1.514841	0.056667	APTX	14.4	0.856666667	1.034231	0.081921	CSNK1D	10.90333333	0.433333333	0.81736	0.07513
RECQL4	17.20333333	1.096666667	1.829611	0.450568	SIRT1	14.31666667	0.476666667	1.785258	0.172272	Cyclophilin B Co	10.62	0.77	1.658805	0.050332
HUS1	17.14666667	1.076666667	0.531706	0.299017	TRIP13	14.30666667	0.52	1.460415	0.140119	RFC2	10.09666667	0.82	1.361572	0.234379
DDIT4	17.12	0.41	0.876717	0.095394	GAPD Control	14.28333333	0.91	1.357649	0.490306	ALKBH2	9.96	0.566666667	0.457639	0.201025
RECQL5	17.12	0.37	3.117761	0.037859	KDM4D	14.25666667	1.14	0.61931	0.249466	POLR2A	9.666666667	2.83	0.750341	0.278388
RECQL	17.05666667	0.44	0.628234	0.01	IGHMBP2	14.23333333	0.986666667	3.38046	0.568106	UIMC1	9.363333333	0.796666667	0.85793	0.432139
RFC1	16.92	0.72	1.591446	0.200333	Lamin A/C Contr	14.20333333	0.796666667	1.978824	0.36177	PRKDC	9.07	0.67	0.975722	0.36074
YBX1	16.84666667	0.306666667	0.793501	0.016667	LYN	14.08333333	1.076666667	1.133966	0.580153	VCP	8.37	3.776666667	0.29023	0.737119
REV1	16.78333333	0.963333333	1.214857	0.408588	FOXO4	13.99333333	0.453333333	0.413333	0.095975	XAB2	7.613333333	2.023333333	0.658846	0.295541
BIVM-ERCC5	16.55333333	0.83	1.239009	0.215484	BRIP1	13.95	0.803333333	1.667723	0.45042	Untreated	0.083333333	0.086666667	0.012019	0.008819
TREX1	16.53666667	0.843333333	2.572666	0.123468	TCEA1	13.84	0.37	1.56513	0.185023					
GPS1	16.45	0.656666667	0.554286	0.163741	TOP2A	13.8	1.053333333	1.437115	0.384895					
PARP3	16.34333333	0.8	0.791672	0.081854	CCNO	13.79666667	0.536666667	0.537907	0.058973					

Table 16. Percent editing of precise correction and undesired edits of siGENOME screen plate 4.

Condition	Average 22T	Average Undesir	22T Error	Undesired Error	Condition	Average 22T	Average Undesir	22T Error	Undesired Error
MEN1	30.13667	0.65	2.912057	0.147309	RRM2B	22.17	0.43	1.02528	0.055678
TREX2	29.62667	0.58	0.020276	0.181475	ALKBH1	22.15333	0.413333	2.577275	0.124141
USP1	28.18	0.72	0.756241	0.058595	EYA1	21.96667	0.48	1.498158	0.035119
ASF1A	27.50333	0.746667	0.837383	0.046308	CDKN2D	21.93667	0.386667	1.938439	0.106823
MMS19	27.13667	0.71	1.286252	0.255343	PRMT6	21.74667	0.696667	1.15779	0.118369
UBE2A	26.04667	0.666667	0.9934	0.10729	RBBP8	21.72667	0.413333	0.126667	0.063857
SMC6	25.89333	0.683333	0.325849	0.082125	DNMT1	20.54667	0.493333	1.768628	0.060645
BAZ1B	25.78	0.456667	5.127586	0.098206	HMGB1	20.40333	0.266667	1.06054	0.07513
APLF	25.59	0.536667	2.005293	0.054569	DUT	20.39	0.416667	3.838129	0.087433
RNF168	25.56	0.413333	1.337273	0.029059	CNOT7	19.99667	0.3	1.484872	0.090185
TDP1	25.34333	0.74	1.28154	0.14	MNAT1	19.53333	0.356667	2.391682	0.101379
DMC1	25.20667	0.41	0.860278	0.133167	CHAF1A	19.27667	0.486667	1.004894	0.054569
EYA3	25.13667	0.513333	1.361241	0.020276	NBN	19.20667	0.396667	1.036827	0.1424
GEN1	25.07333	0.63	0.588907	0.060277	SMC1A	19.14667	0.32	1.025643	0.066583
GTF2H3	25.07333	0.59	0.369925	0.244404	CLK2	19.04667	0.256667	1.525389	0.07796
SMUG1	24.86333	0.47	1.530537	0.065574	UBE2V1	19.03333	0.36	0.239537	0.131149
XPA	24.72	0.526667	1.958937	0.156241	PARG	18.89	0.373333	0.274287	0.060093
GADD45G	24.43333	0.906667	1.248283	0.168259	SETX	18.85333	0.31	1.65288	0.055678
CIB1	24.4	0.38	2.668539	0.109697	PRPF19	18.71667	0.753333	1.406939	0.186577
ATXN3	24.35667	1.08	1.641639	0.680098	CDK7	18.55667	0.343333	1.642522	0.163333
PCNA	23.9	0.633333	0.459674	0.173813	H2AFX	18.51	0.27	2.354082	0.05
EME2	23.89667	0.566667	0.458597	0.078599	SMARCAL1	18.17333	0.313333	2.310024	0.136423
BTG2	23.73667	0.356667	1.125231	0.101708	TNP1	18.1	0.336667	0.753945	0.08838
TP73	23.55333	0.663333	1.729724	0.129142	ALKBH3	17.86333	0.616667	1.599858	0.056075
MBD4	23.52333	0.32	0.450383	0.087178	NPM1	17.27667	0.26	1.063709	0.066583
CETN2	23.45333	0.66	1.452542	0.138564	BLM	17.15333	0.266667	1.691669	0.054569
RDM1	23.41333	0.543333	2.687206	0.16231	SPO11	16.69333	0.41	1.041671	0.030551
EME1	23.09667	0.41	1.250418	0.061101	UBE2N	16.40667	0.306667	2.443101	0.069841
CCNH	23.05667	0.593333	0.80269	0.070553	SMARCAD1	16.27667	0.356667	1.053349	0.05696
RNF8	22.83333	0.29	1.7021	0.078102	PARP1	15.88	0.303333	2.522327	0.094927
ABL1	22.23667	0.543333	2.016601	0.096148	UPF1	14.93667	0.266667	0.690515	0.063596
PALB2	22.23	0.53	1.563724	0.046188	RRM2	12.25333	0.516667	1.831906	0.274489
Untreated+PE3_	22.2065	0.454333	1.472037	0.443333	Untreated	0.646667	0.026667	0.551674	0.006667
UBE2V2	22.2	0.48	0.475079	0.191398					

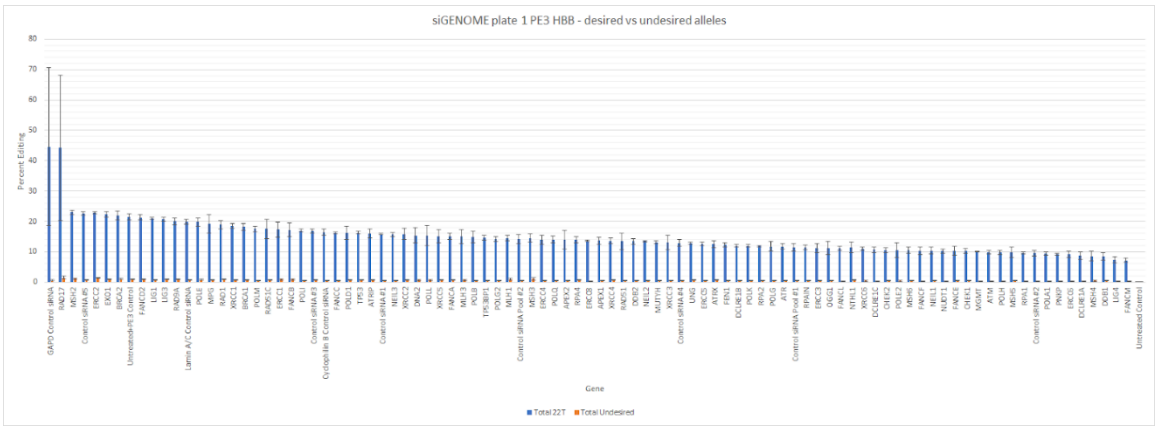


Figure 23. siGENOME plate 1 results.

Knockdown-accompanied prime editing results from the first plate of siGENOME siRNAs. 22T allele represents precise insertion of HBB E6V pathogenic SNP.

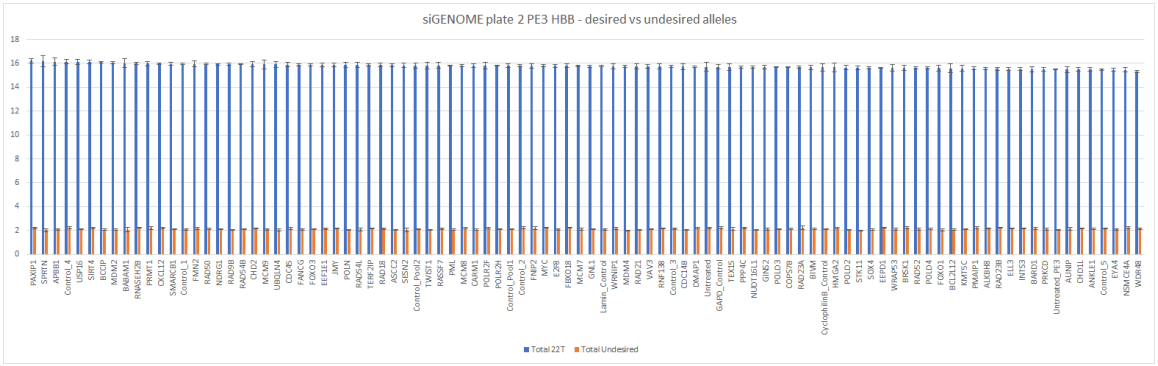


Figure 24. siGENOME plate 2 results.

Knockdown-accompanied prime editing results from the second plate of siGENOME siRNAs.22T allele represents precise insertion of HBB E6V pathogenic SNP.

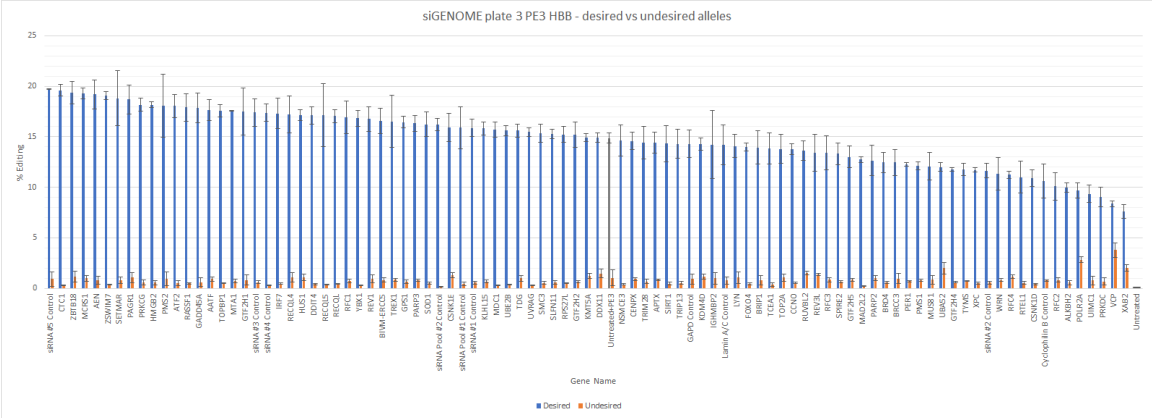


Figure 25. siGENOME plate 3 results.

Knockdown-accompanied prime editing results from the third plate of siGENOME siRNAs. 22T allele represents precise insertion of HBB E6V pathogenic SNP.

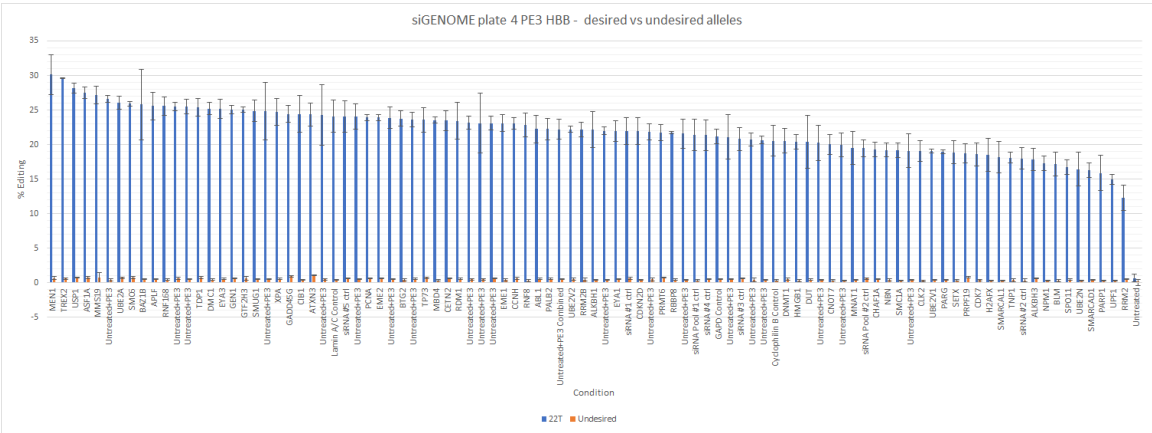


Figure 26. siGENOME plate 4 results.

Knockdown-accompanied prime editing results from the fourth plate of siGENOME siRNAs. 22T allele represents precise insertion of HBB E6V pathogenic SNP.

Statistical analysis showed that of the tested genes, 107/317 displayed statistically significant behavior deviating from the untreated+PE3 control at the individual t-test level (Table 17). However, once corrections were made for multiple hypotheses (Bonferroni/Holm's), only three gene hits remained statistically significant: FANCM, LIG4, and TREX2 (Table 18).

Table 17. T-Tests for siGENOME plates 1-4.

Plate1				Plate2		Plate3		Plate4	
Gene	TTEST	Gene	TTEST	Gene	TTEST	Gene Name	TTEST	Gene	TTEST
FANCM	0.000411165	Control siRNA #4	0.004613011	POLR2H	0.000890271	VCP	0.000987079	TREX2	1.11278E-20
LIG4	0.000513595	MLH1	0.004781848	RAD54B	0.001973541	XAB2	0.001278514	USP1	0.003827252
POLH	0.000834002	POLG2	0.005005589	PML	0.002379142	ALKBH2	0.001866937	ASF1A	0.010568053
POLA1	0.000839267	RPA2	0.005255841	COPS7B	0.006602022	ZSWIM7	0.00315779	SMC6	4.04975E-05
CHEK1	0.000955716	UNG	0.005415127	CXCL12	0.006602365	MCRS1	0.003544769	GEN1	0.024244068
NUDT1	0.000972757	RPA4	0.005684168	BCCIP	0.007731449	CTC1	0.003632358	GTF2H3	0.00193818
ERCC6	0.000985104	DCLRE1B	0.005884969	Control_1	0.007818926	RFC4	0.005026583	UBE2V1	8.39137E-08
CHEK2	0.001046924	FANCE	0.005980077	NDRG1	0.00899123	POLR2A	0.006793315	PARG	2.80679E-07
Control siRNA #2	0.001069614	MGMT	0.006171755	Lamin_Control	0.011245296	HMGB2	0.00823306	TNP1	0.006751263
DDB1	0.001377941	APEX1	0.006546322	RAD50	0.012988836	UIMC1	0.009651863	NPM1	0.019770617
RPAIN	0.00148424	MSH5	0.0072649	POLD3	0.01776067	siRNA #5 Control	0.009973558	SPO11	0.013882041
MSH6	0.001502943	TP53BP1	0.00738237	RAD9B	0.03095771	XPC	0.011240661	SMARCAD1	0.012102285
PNKP	0.001558343	POLQ	0.007447655	RNASEH2B	0.031054355	UBA52	0.012538893	UPF1	0.000462821
ATM	0.001563025	ERCC3	0.007654685	MCM7	0.032462388	PMS1	0.013211758	RRM2	0.022863961
DCLRE1C	0.001566104	MSH3	0.00806081	MDM2	0.039979437	PRKDC	0.013487834		
RPA1	0.00172683	FANCA	0.009715258	TERF2IP	0.048217036	GTF2H4	0.015929551		
FANCL	0.00178797	NTHL1	0.01032317			TYMS	0.017279397		
Untreated Control	0.001860808	ERCC8	0.010623237			PRKCG	0.019549625		
ATR	0.001985713	NEIL2	0.01087751			CSNK1D	0.020977467		
NEIL1	0.002037998	POLG	0.01107417			PER1	0.023301318		
FEN1	0.002096661	NEIL3	0.011187553			siRNA #2 Control	0.027528848		
DCLRE1A	0.002189974	TP53	0.018776186			MAD2L2	0.028538114		
FANCF	0.00223083	Control siRNA #1	0.01946952			TOPBP1	0.029145088		
XRCC6	0.002991448	Cyclophilin B Control siRNA	0.020564867			MTA1	0.030949558		
ATRX	0.003502681	Control siRNA Pool #2	0.021207866			HUS1	0.03571639		
Control siRNA Pool #1	0.003592372	ERCC4	0.021297045			ZBTB18	0.037413085		
XRCC4	0.003690038	FANCC	0.022701741						
POLK	0.003945705	Control siRNA #3	0.023100314						
DDB2	0.003993063	POLI	0.026615603						
ERCC5	0.004503704	POLE2	0.031376016						
MUTYH	0.00450567	OGG1	0.032179321						
MSH4	0.004509104								

Tests at the individual level prior to multiple hypothesis correction for genes with individual significance.

Table 18. siGENOME plates 1-4 Bonferroni and Holm's hits.

Gene	22T AVERAGE	22T ERROR	TTEST	Bonferroni Cutoff	Bonferroni P-Value	k	Holm's P-Value
FANCM	6.97	0.692411246	0.000411165	0.000520833	0.039471853	1	0.000411165
LIG4	7.073333333	0.945380582	0.000513595	0.000520833	0.049305127	2	0.000513595
TREX2	29.62666667	0.020275875	1.11278E-20	0.000520833	1.06827E-18	1	1.11278E-20

Hits from the siRNA knockdown prime editing screen for all four plates passing Bonferroni and Holm's cutoff scores.

To test the replicability of these results and further investigate the experiment, several top and bottom performing siGENOME plate 1 siRNAs were tested again regardless of significance in two repeat experiments (Figures 27, 28) (Tables 44, 45), along with a followup experiment containing only the individually significant hits from all four plates (Figure 29) (Table 19). Several genes reproduced their relative positions in the siGENOME plate 1 experiment, but of the original samples, only LIG4 and RAD9A remained statistically significant by t-test at the individual level in the first repeat, and no samples were statistically significant in the second repeat. In the plate1-4 repeat, no samples were again statistically significant although TREX2 and FANCM/LIG4 maintained their relative positions. These three genes exhibited slight upward, slight downward, and strong downward regulation, respectively.

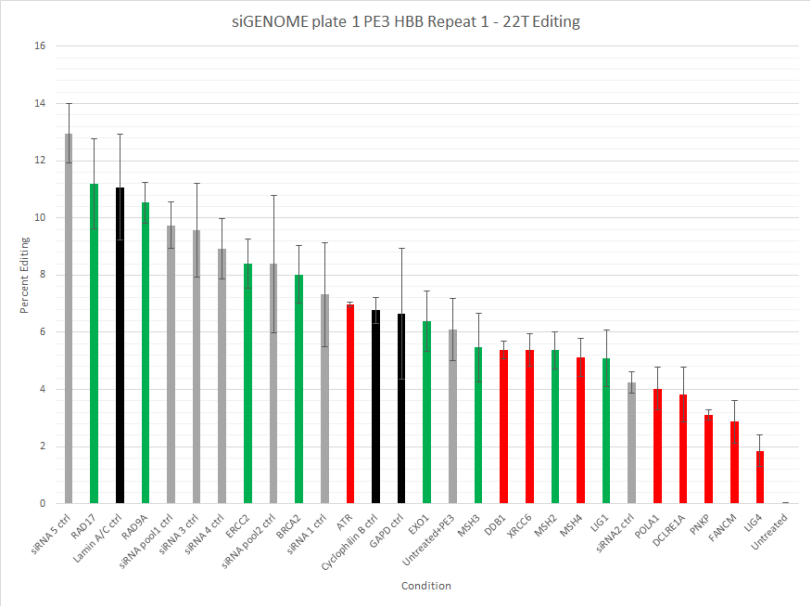


Figure 27. siGENOME plate 1 repeat knockdown prime editing.

HBB 22T precise correction. (green - high editing in original screen, red - low editing in original screen, grey - nontargeting control, black - targeting control).

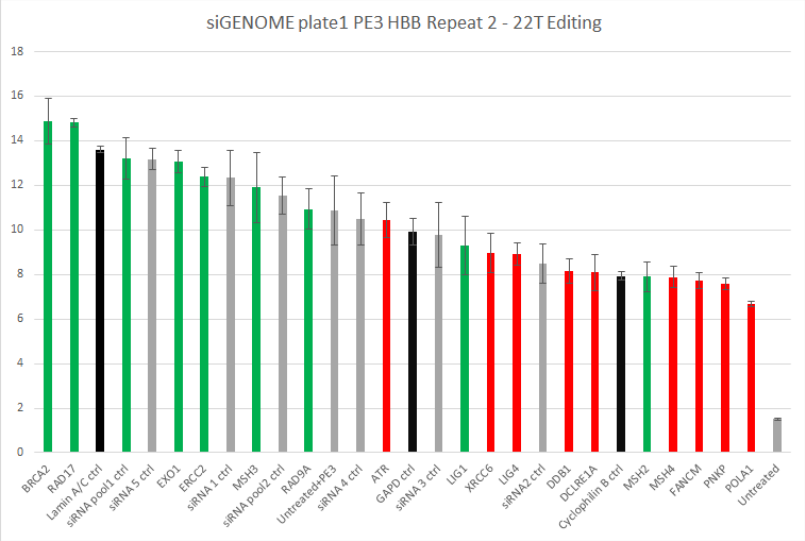


Figure 28. siGENOME plate 1 repeat 2 knockdown prime editing.

HBB 22T precise correction. Replicability of the relative positions of the genes is poor.

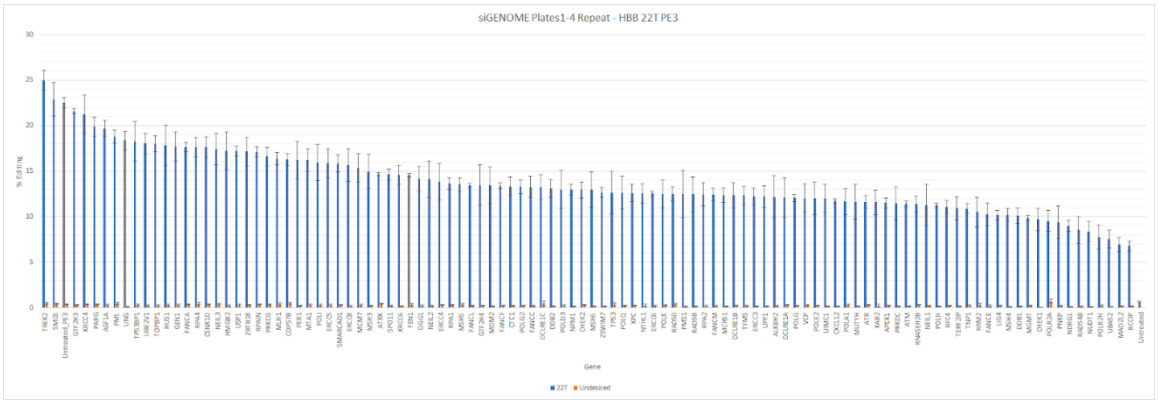


Figure 29. siGENOME plate 1-4 repeat knockdown prime editing.

HBB 22T precise correction results. TREX2 reproduces its location.

Table 19. Percent HBB PE3 editing of siGENOME plates 1-4 repeat screen.

Condition	Average 22T	Average Undesir	22T Error	Undesired Error	Condition	Average 22T	Average Undesir	22T Error	Undesired Error	Condition	Average 22T	Average Undesir	22T Error	Undesired Error
TREX2	24.97667	0.513333	1.12446	0.116237	OGG1	14.16667	0.243333	1.364323	0.120324	POLG	12.06	0.223333	0.42548	0.021858
SMC6	22.86333	0.44	1.833088	0.017321	NEIL2	14.13333	0.213333	1.995999	0.08838	VCP	12.04333	0.3	1.541183	0.06245
Untreated_PE3	22.52667	0.386667	0.561733	0.053645	ERCC4	13.85333	0.316667	1.973544	0.058119	POLE2	12.00667	0.246667	1.781734	0.063333
GTF2H3	21.59	0.3	0.285715	0.052915	RPA1	13.64667	0.33	0.621727	0.01	UIMC1	11.99	0.183333	1.563682	0.01453
XRCC4	21.26333	0.403333	2.107228	0.040552	MSH5	13.56	0.35	0.757914	0.160935	CXCL12	11.72	0.183333	0.292973	0.05696
PARG	19.85667	0.353333	1.054108	0.027285	FANCL	13.47333	0.28	0.235396	0.026458	POLA1	11.66	0.316667	1.431829	0.06888
ASF1A	19.7	0.323333	0.86031	0.066667	GTF2H4	13.46	0.246667	2.241837	0.063333	MUTYH	11.63667	0.22	1.888927	0.005774
PML	18.82	0.49	0.715285	0.111505	MDM2	13.44333	0.18	2.016353	0.030551	ATR	11.61333	0.303333	0.700484	0.049777
UNG	18.37	0.09	1.014741	0.036056	FANCF	13.41	0.273333	0.342394	0.01453	XAB2	11.59333	0.246667	1.353592	0.148474
TP53BP1	18.28333	0.323333	2.198184	0.092436	CTC1	13.33	0.246667	1.007786	0.046667	APEX1	11.52333	0.226667	0.548402	0.039299
UBE2V1	18.02667	0.226667	1.142517	0.091348	POLG2	13.28333	0.25	0.752381	0.036056	PRKDC	11.47333	0.183333	1.815804	0.049103
TOPBP1	18.01333	0.263333	0.908521	0.054874	FANCC	13.27333	0.22	1.145794	0.063509	ATM	11.41667	0.163333	0.306721	0.020276
HUS1	17.85333	0.31	2.219131	0.088882	DCLRE1C	13.25333	0.49	1.403096	0.251064	RNASEH2B	11.38333	0.283333	0.846174	0.02848
GEN1	17.72333	0.34	1.604061	0.089629	DDB2	13.1	0.156667	0.977088	0.038442	NEIL1	11.25667	0.216667	2.274953	0.057831
FANCA	17.65333	0.373333	0.497672	0.052387	POLD3	13	0.266667	2.096163	0.063857	POLH	11.24667	0.18	0.278109	0.051316
RPA4	17.63667	0.48	1.076202	0.128582	NPM1	12.96333	0.2	0.597365	0.041633	RFC4	11.05667	0.216667	0.717271	0.037118
CSNK1D	17.63333	0.363333	1.124357	0.043333	CHEK2	12.95333	0.296667	0.876553	0.052387	TERF2IP	10.93333	0.153333	1.245958	0.031798
NEIL3	17.46333	0.393333	1.729367	0.076884	MSH6	12.95	0.18	1.92424	0.034641	TNP1	10.89333	0.263333	0.536449	0.052387
HMGB2	17.24	0.25	2.043094	0.089629	ZSWIM7	12.66333	0.173333	0.542412	0.016667	RRM2	10.51667	0.31	1.622029	0.147986
USP1	17.21667	0.333333	0.549313	0.060093	TP53	12.63667	0.373333	2.313967	0.149034	FANCE	10.26	0.21	1.252837	0.145029
ZBTB18	17.14	0.313333	1.567237	0.038442	POLQ	12.62667	0.243333	1.791427	0.043716	LIG4	10.19667	0.236667	0.515957	0.024037
RPAIN	17.12667	0.366667	0.543885	0.040552	XPC	12.59	0.246667	0.956312	0.008819	MSH4	10.18333	0.126667	0.725496	0.012019
PRKCG	16.60667	0.36	1.020136	0.081854	NTHL1	12.57333	0.273333	1.037085	0.131191	DDB1	10.11667	0.203333	0.857619	0.018559
MLH1	16.37	0.406667	0.671441	0.112002	ERCC6	12.54333	0.223333	0.254122	0.083533	MGMT	9.876667	0.233333	0.313546	0.037118
COPS7B	16.29667	0.493333	0.643748	0.107134	POLK	12.51333	0.28	1.510346	0.050332	CHEK1	9.716667	0.056667	1.219868	0.047022
PER1	16.23667	0.253333	2.064593	0.050442	RAD50	12.51	0.393333	0.797517	0.0857	POLR2A	9.546667	0.723333	1.169919	0.208673
MTA1	16.23333	0.333333	1.236667	0.069841	PMS1	12.49	0.166667	2.617677	0.063333	PNKP	9.393333	0.216667	1.792339	0.043716
POLI	15.97	0.313333	1.989983	0.098206	RAD9B	12.47333	0.213333	1.886676	0.013333	NDRG1	9.003333	0.106667	0.634043	0.037118
ERCC5	15.9	0.336667	1.605034	0.064377	RPA2	12.45	0.2	1.281614	0.091652	RAD54B	8.553333	0.146667	1.496934	0.052387
SMARCAD1	15.84	0.323333	0.929157	0.044845	FANCM	12.42	0.22	0.734325	0.045092	NUDT1	8.393333	0.103333	1.114291	0.03383
ERCC8	15.66	0.203333	1.790503	0.051747	MCRS1	12.39	0.193333	0.778481	0.024037	POLR2H	7.756667	0.233333	1.340427	0.097011
MCM7	15.36333	0.3	1.558165	0.094516	DCLRE1B	12.37	0.213333	1.385364	0.036667	UBA52	7.543333	0.08	0.970057	0.040415
MSH3	14.96667	0.223333	1.915379	0.036667	TYMS	12.34333	0.226667	1.002486	0.052387	MAD2L2	6.943333	0.173333	0.770332	0.026034
ATRX	14.66333	0.446667	0.15836	0.055478	ERCC3	12.22	0.206667	0.939521	0.027285	BCCIP	6.803333	0.163333	0.548159	0.049777
SPO11	14.64333	0.23	0.597671	0.08	UPF1	12.21333	0.2	1.165795	0.02	Untreated	0.68	0.026667	0.090185	0.013333
XRCC6	14.59667	0.153333	1.048369	0.037118	ALKBH2	12.18333	0.21	2.305243	0.052915					
FEN1	14.57333	0.316667	0.186577	0.13691	DCLRE1A	12.06667	0.306667	2.200548	0.026667					

Overexpression of TREX2 Decreases Prime Editing

While some genes demonstrated replicability of their relative positions, the lack of consistent statistical significance and confounding possible factors complicated interpretation of results. Therefore to address this, orthogonal validation methods were required. As it is reasonable to expect that if knockdown increases prime editing, then overexpression may have a chance to decrease prime editing or vice versa. Thus, an overexpression assay was developed. Plasmids from OriGene were purchased (Table 46), containing constitutive expression cassettes for some of the top genes from plate1 and TREX2. An assay was developed to verify the functional effect of overexpression. A day 0 seeding at 26k cells per well, day 1 transfection with 200ng expression plasmid using 1uL Lipofectamine 2000, and a day 2 transfection with either 150ng BE4 mRNA and 50ng Hek2 gRNA or 3000ng PE2 mRNA and 400ng HBB E6V pegRNA and 150ng HBB E6V nickRNA was performed. pGFPMax expression was verified through FITC imaging as a double confirmation of protein-level expression (Figure 32), and held for all days up until harvest.

Overexpression of UNG decreased the relative C>T editing of BE4 and increased C>NonT relative to GFP control (Figure 30) (Table 20), but transfection with plasmid decreased overall prime editing compared to untreated controls by about 2-fold (Figure 31) (Table 21). No notable effect was seen with BE4-noUGI. Prime editing was decreased by about 5-fold with just the control GFP DNA, which is likely due to the toxicity of DNA transfection which interrupts cell expression and DNA repair. However, these results demonstrate that the overexpression screen was valid for enabling gene cofactors to influence editing outcomes. In the full screen, the untreated+PE3 condition

was again the highest editing condition, while UNG and GFP control plasmids decreased prime editing significantly. Within the tested samples, while FANCM exhibited the highest editing compared to other tested genes as the inverse of its knockdown effect, it was not statistically significant against UNG. Only TREX2 exhibited statistically significant deviation of behavior against all three controls, with a marked decrease in prime editing (Figure 33) (Table 47). This is in line with expectations and the inverse of its knockdown effect.

Table 20. Percent Hek2 editing in overexpression assay development for cytidine base editors and UNG.

Designation	Average C>T	Average C>NonT	C>T Error	C>NonT Error
Untreated+BE4	21.98667	23.89667	0.427798	0.666041
GFP+BE4	9.256667	12.91333	1.206653	1.253134
UNG+BE4	3.59	16.91667	0.246644	0.436438
Untreated+BE4-noUGI	2.406667	47.13	0.070553	1.08
GFP+BE4-noUGI	1.713333	28.07667	0.082932	1.085546
UNG+BE4-noUGI	2.503333	25.81667	0.072648	1.414983

Table 21. Percent editing in overexpression assay development for HBB PE3+plasmid.

Condition	22T AVERAGE	22T ERROR
Untreated+PE3	25.18166667	0.245444042
GFP+PE3	6.103333333	0.615476329
UNG+PE3	8.426666667	0.395362
Untreated	0.036666667	0.008819171
UNG	0.05	0.015275252
GFP	0.063333333	0.003333333

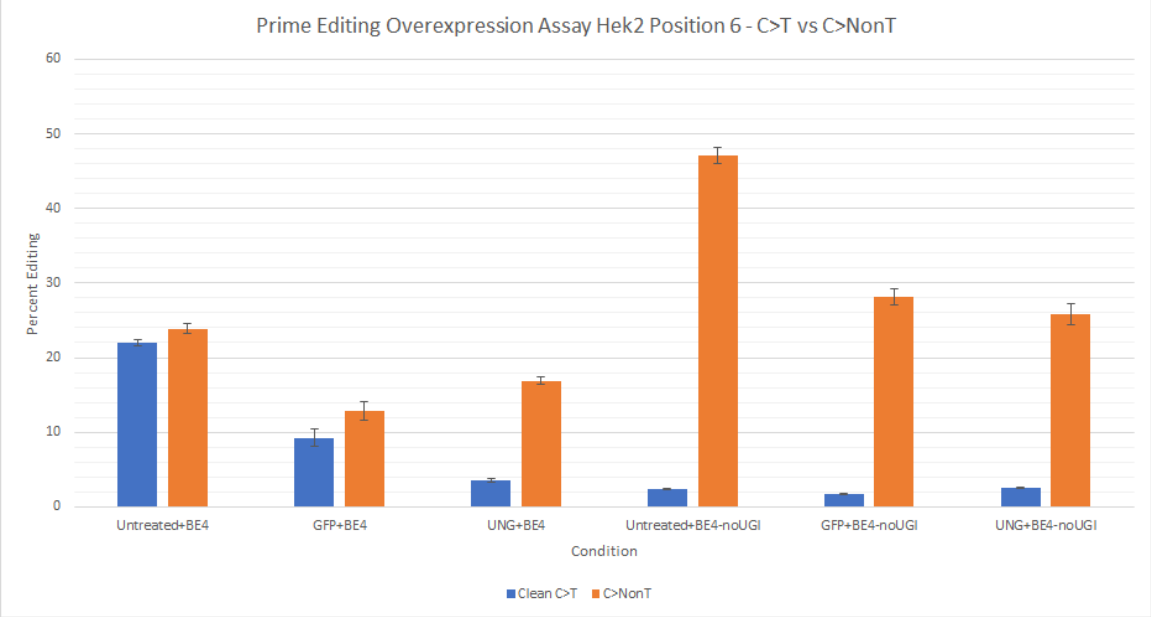


Figure 30. UNG expression effect on Hek2 BE4 editing.

Overexpression assay development experiment confirming that UNG expression inhibits C>T editing of BE4 relative to C>NonT compared to GFP.

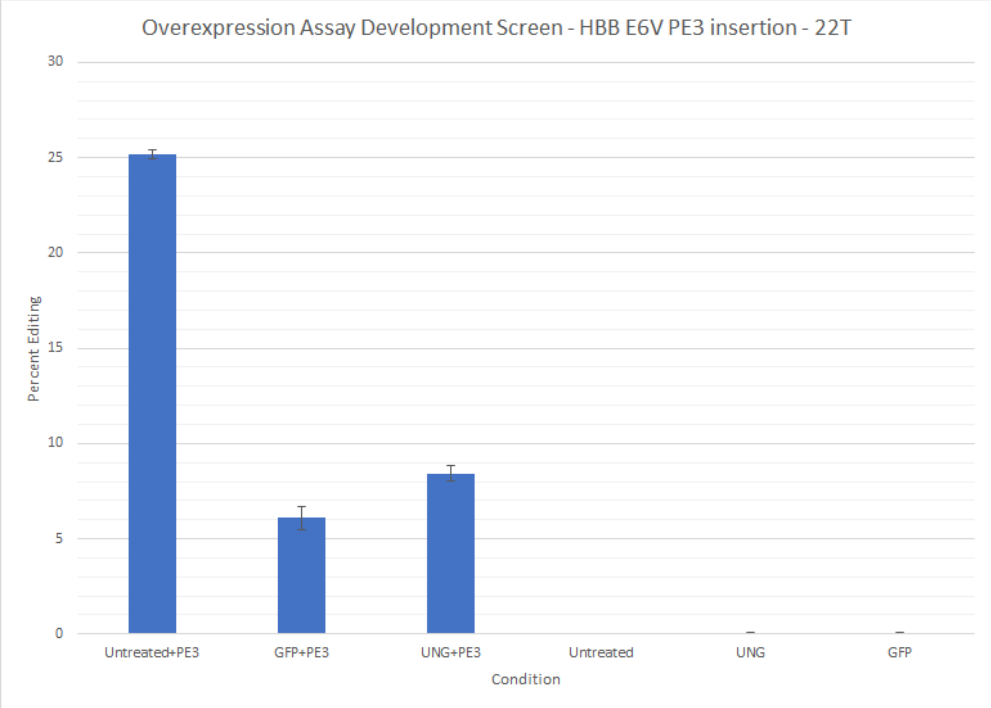


Figure 31. Control plasmid effect on prime editing.

Transfection of 200ng plasmid has a strong effect on decreasing prime editing, even with a control GFP plasmid.

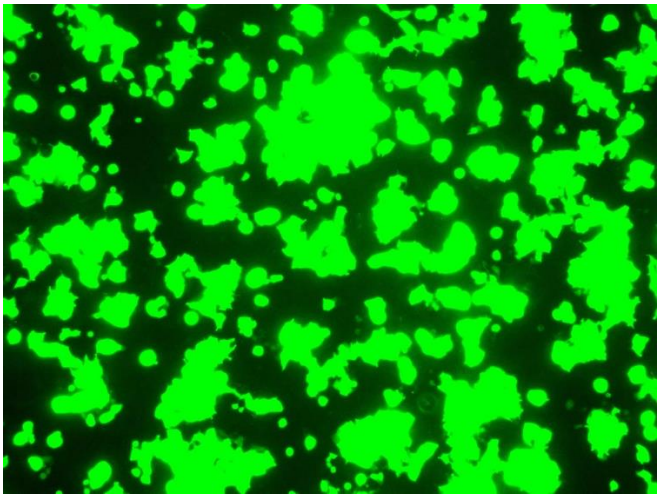


Figure 32. 24hr GFP FITC image.

Exemplary image of expression, which held for 5 days.

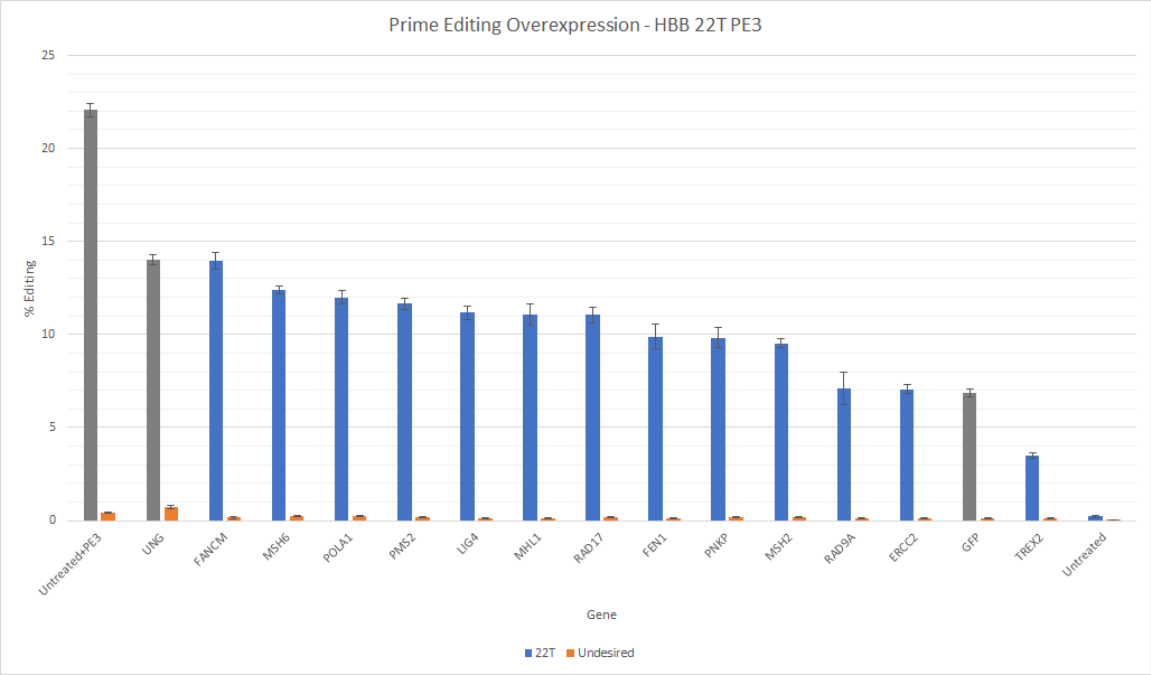


Figure 33. Plasmid overexpression prime editing.

TREX2 and FANCM recapitulate the inverse of their relative positions from the siRNA screen, with TREX2 exhibiting statistically significant downregulation of prime editing.

TREX2 Plasmid increases TREX2 Protein

Protein samples for various cell types were transfected with either TREX2 siRNA or plasmid, and analyzed via western blot and microplate reader using two different antibodies. TREX2 runs at 33kDa according to the antibody manufacturer despite having a molecular weight of 26kDa, and bands were observed in the 33kDA range. While the sensitivity of the assay does not allow us to determine the baseline expression of TREX2 in the tested cell types (Hek293T/HepG2/GM11423/RSE HPH) with confidence, overexpression clearly produces an increase in protein in each of the plasmid conditions,

confirming that presence of TREX2 that may be consequential to editing outcomes (Figure 34). Some off-target binding was observed, but did not interfere with the target band interpretation. Hek293T cells, GM11423 fibroblasts, HepG2 cells, and RSE HPH cells all appear to have some baseline expression of TREX2 at 33kDA, although it is much lower than the strong overexpression seen with plasmid in Hek293T cells, and it is uncertain whether knockdown by siRNA would be consistently consequential to editing outcomes in the desired primary cell types. This additionally explains why knockdown by TREX2 only sometimes generates significant results in Hek293T cells.

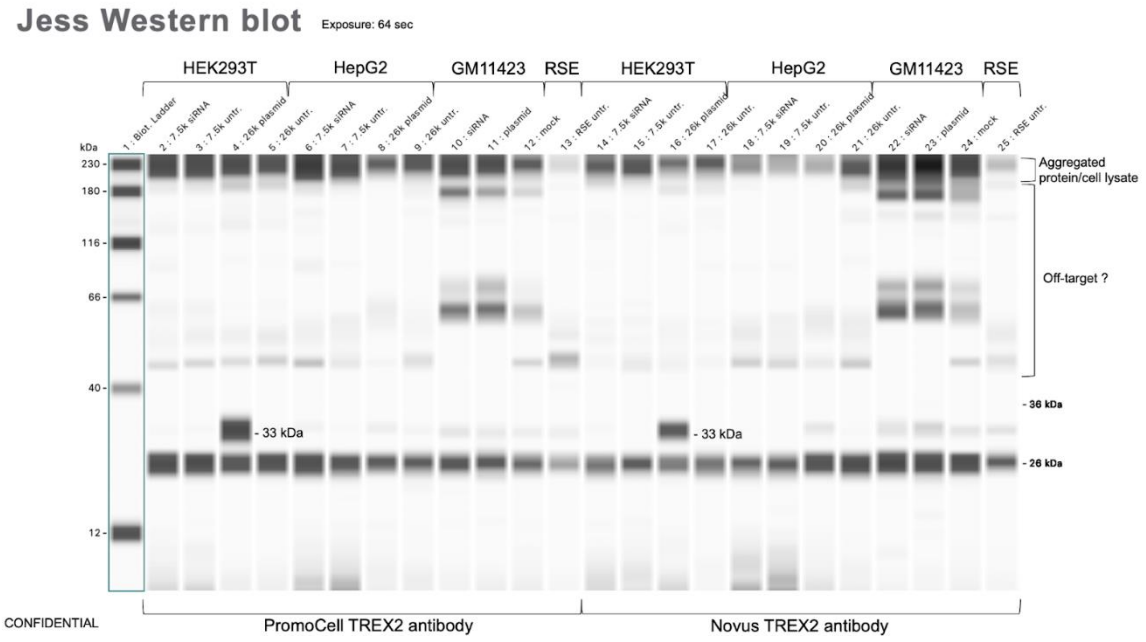


Figure 34. Western blot of TREX2 protein for various cell types and transfection conditions.

Exposure is with high dynamic range at 64sec. Hek293T, HepG2, GM11423 fibroblasts, RSE HPH cells were tested against two antibodies which both confirmed the presence of TREX2 at 33kDa

BRCA2 Knockout does not Influence Prime Editing

As knockout cell lines are a legitimate orthogonal validation method, attempts were made to procure them for hits from the siRNA screen. However, due to prohibitive timelines and cost, only a BRCA2 Hek293T knockout cell line was obtained. Seeding density conditions of 7.5/26k cells per well were tested with Hek293T parental WT and Beam's internal Hek293T WT cell lines as controls. Unfortunately, no significant difference was observed with knockout for the 7.5k cells per well condition against either control, and for 26k cells per well the BRCA2 knockout was not significant against the internal Beam WT cell line (Figure 35) (Table 48).

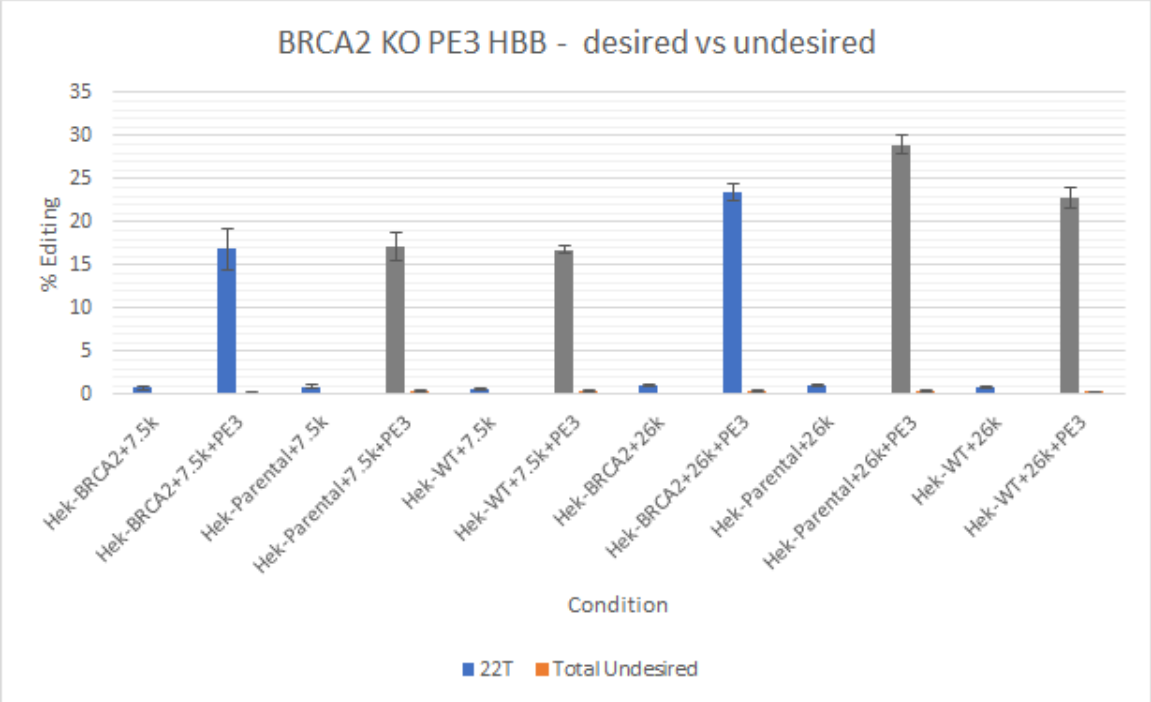


Figure 35. Results of the BRCA2 knockout cell line prime editing experiment.

TREX2 Overexpression Decreases Prime Editing at A Variety of Target Sites

To test if the effect of TREX2 was broadly applicable to prime editing rather than an artifact of a single target site, overexpression was performed as before with other target sites from the original Anzalone et al. paper. PE2 was also tested to determine whether the mechanism of TREX2 was specific to PE3. HBB, FANCF, and HEK3 mRNA/pegRNA/nickRNA PE2/3 strategy or A1AT E342K PE2/PE3 precise correction/insertion using pCMV-PE2-VRQR/GL544-13-20/GL544n11 were performed in Hek293T cells (Figures 36, 37) (Table 49). T-tests were significant for A1AT E342K NGG prime editing with TREX2 overexpression versus GFP for both PE2 (p-value 0.027092679) and PE3 (p-value 0.005452562). T-tests were significant for FANCF prime editing with TREX2 overexpression versus GFP for both PE2 (p-value 0.029748114) and PE3 (p-value 0.011872344). T-tests were not significant for HBB PE2, or either PE2/PE3 for HEK3 or the A1AT NGA sites, although similar relative positioning was observed. This may be due to the low levels of overall editing reducing assay sensitivity. No significant change in behavior was observed for the TREX2 siRNA knockdown prime editing positions at any of the other target sites relative to the respective UNG/GFP controls (Figure 38).

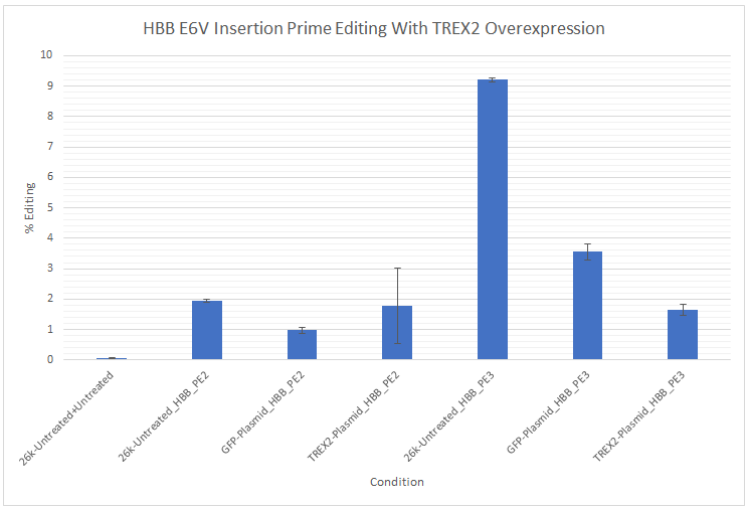


Figure 36. TREX2 overexpression effect on PE2/PE3 HBB prime editing.

TREX2 lowers editing for PE3 compared to GFP control. Error for PE2 made the results inconclusive.

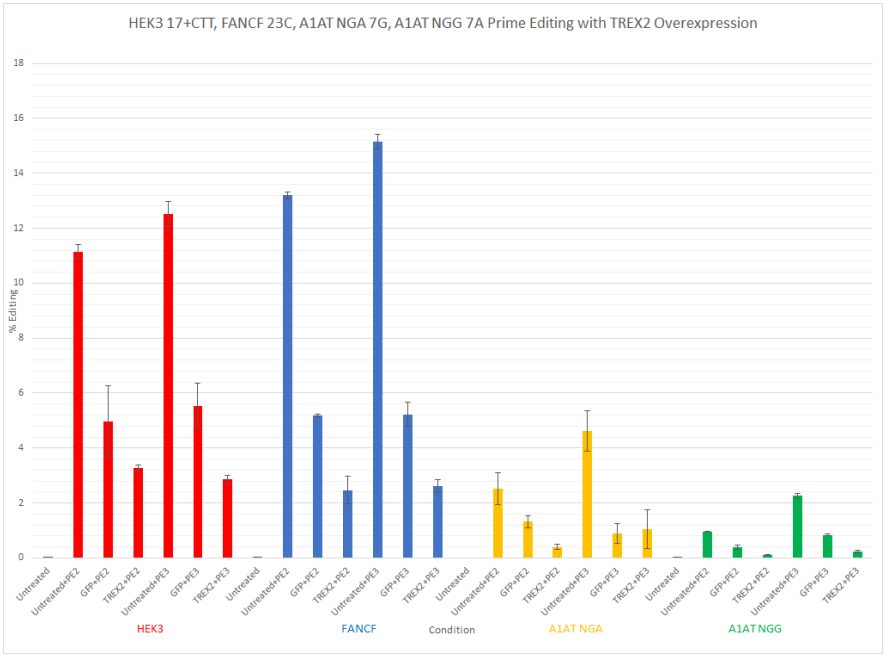


Figure 37. TREX2 overexpression effect on PE2/PE3 editing at various target sites.

Overexpression of TREX2 demonstrated statistically significant decreased prime editing for both PE2 and PE3 at FANCF and A1AT NGG target sites, with similar absolute behavior at HEK3 and A1AT NGA but the variability at those sites prevented significance.

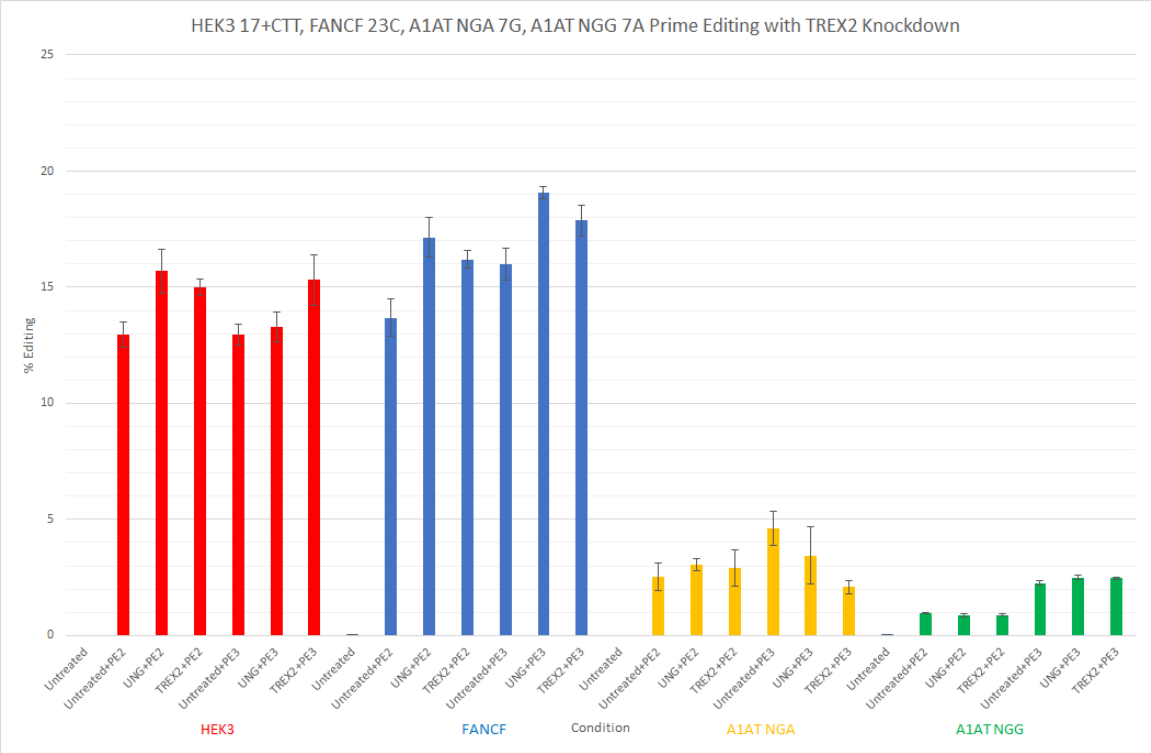


Figure 38. TREX2 knockdown effect on PE2/PE3 prime editing at various target sites.

No significant change in effect was detected for HEK3/FANCF/A1AT NGA/A1AT NGG.

TREX2 Influences Prime Editing but not Base Editing

To check whether TREX2 action was specific to normal prime editing, another experiment was conducted where the assay format was adapted for testing in Hek293T cells with a BE4 Hek2 gRNA (as used previously in this work) condition. Again, transfection of plasmid in general decreased prime editing. However, TREX2 overexpression did not demonstrate significantly different behavior from the GFP plasmid control. TREX2 knockdown also did not generate significant results against the untreated condition, while the UNG siRNA caused a large shift towards C>T as expected (Table 22) (Figure 39).

Table 22. Percent editing of HEK2 by BE4 with or without TREX2.

	AVERAGE		ERROR	
Condition	Total C>T	Total C>NonT	Total C>T	Total C>NonT
Untreated	0.01	0.013333333	0.005774	0.003333
Untreated+BE4 (26kcells)	31.95	29.18666667	1.504671	0.309641
GFP-Plasmid+BE4	8.39	10.91	0.522143	0.453689
TREX2-Plasmid+BE4	10.12667	14.1	0.301183	0.459021
Untreated	0.01	0.016666667	0	0.003333
Untreated+BE4 (7.5kcells)	26.48667	26.91333333	0.227327	0.167564
UNG-siRNA+BE4	49.7	2.04	6.370377	0.395095
TREX2-siRNA+BE4	28.63	30.86666667	1.278841	0.604161

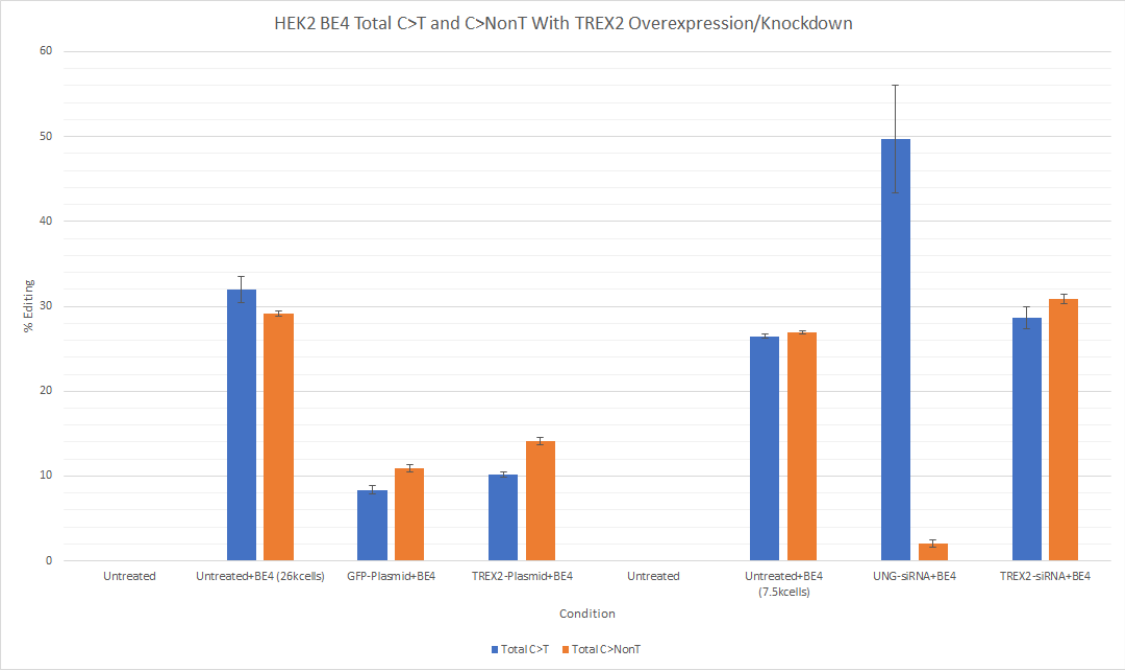


Figure 39. TREX2 overexpression/knockdown effect on BE4 HEK2 editing.

No notable changes to editing rates were observed for either knockdown or overexpression of TREX2.

TREX2 Overexpression Decreases Prime Editing In Liver Context

The assay format from the overexpression assay was adapted for Hep G2 cells, which are relevant to liver physiology as a liver cancer derived cell line in order to check for TREX2 relevance in target cell lines for A1AT E342K. TREX2 overexpression caused a strong reduction in prime editing compared to the GFP control, from ~16% to ~6% (p-value 0.000140992) (Table 23) (Figure 40). In Hep G2 cells, GFP transfection did not seem to decrease prime editing. siRNA knockdown of TREX2 did not seem to have an effect. Literature appears to indicate that TREX2 protein of the expected size is not found in HepG2 cells (Wang et al., 2005), and data from the Protein Atlas seems to indicate that while TREX2 is expressed as RNA more strongly than in kidney cells, no

protein expression has been found, while protein is found in kidney cells. Therefore while TREX2 can produce activity in HepG2 cells, knockdown is unlikely to improve prime editing rates as it might in Hek293T cells.

Table 23. Editing frequencies for the Hep G2 TREX2 experiment.

	AVERAGE		ERROR	
Designation	22T	Undesired	22T	Undesired
Untreated+PE3 (26K)	16.13333	2.036667	0.205291	0.097696
GFP-Plasmid+PE3	15.69	2.036667	0.450444	0.057831
TREX2-Plasmid+PE3	6.786667	0.623333	0.358345	0.219266
Untreated+PE3 (7.5K)	15.9	2.18	0.161658	0.15308
UNG-siRNA+PE3	16.03667	2.076667	0.226593	0.187735
TREX2-siRNA+PE3	15.68	2.276667	0.297041	0.170717

HBB E6V 22T PE3 editing is decreased by TREX2 overexpression.

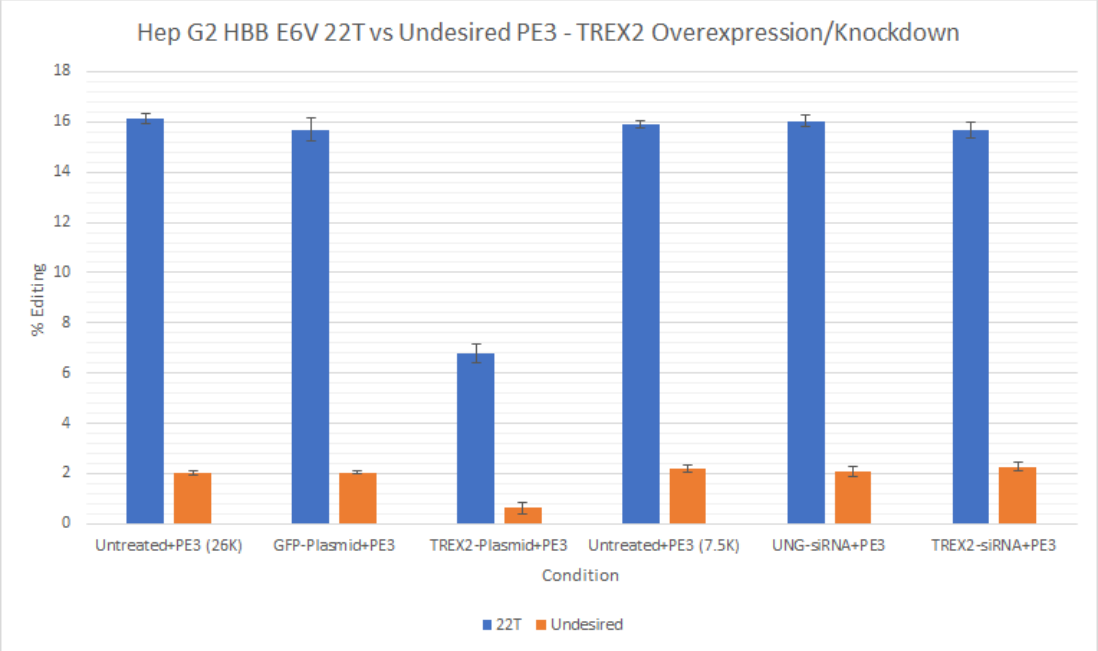


Figure 40. Effect of TREX2 overexpression/knockdown on HBB prime editing in HepG2 cells.

Knockdown has no effect, which overexpression strongly decreases prime editing by more than 2-fold.

Chapter V.

Discussion

NGA PAM Prime Editing Represents a Viable Alternative for A1AD

This work has demonstrated the value of exploring direct modifications to prime editors when faced with the challenge of inaccessible target sites. The use of VRQR prime editor and NGA pegRNAs at target sites closer to the pathogenic SNP has proven that such engineering may be advantageous, even if the prime editing levels reported in this thesis are quite low at 2-6%. These reagents have also demonstrated editing ability after conversion to RNA and detectable, if low editing in primary cells. As other literature demonstrates, engineering strategies such as NLS swapping (Liu et al., 2020) or linker flexibility or inlaid enzymes (Chu et al., 2021) can be leveraged to greatly increase editing frequency for one of the pegRNAs tested here, and other such related possibilities may in the future boost the relevance of what are currently low levels of editing at alternative target sites. Here, it is hopeful that alternative prime editors may someday represent a potential cure for A1AD disease as notable baseline activity has been confirmed, although at this time a multitude of improvements would need to be realized before any sort of clinical relevance could be considered reasonable.

These results further emphasize the precision and breadth with which screens for prime editing candidate reagents must be handled, given the high levels of variability in editing with a single protospacer when the PBS and RTT lengths are modified. Somewhat counterintuitively, occasionally very long RTTs far past the pathogenic SNP provide much more robust results. This may be due to the fact that sites with low activity may

benefit from having a longer reverse transcribed 3’ flap to protect the edit from exonuclease degradation and ensure incorporation into the repaired genome.

TREX2 Action on Prime Editing is a Novel Mechanism

The work here demonstrates that TREX2 is relevant to prime editing outcomes, and is a novel gene of interest in the context of genome editing. Specifically, the presence of TREX2 may decrease prime editing, and its absence may potentially aid prime editing. However, while TREX2 overexpression has a consistent and clear effect, TREX2 knockdown’s prime editing impacts are sensitive to variation in conditions and its mechanics are difficult to ascertain. The identity of TREX2 has hitherto been characterized as a nuclear protein with 3’ to 5’ exonuclease activity with interactions on double-stranded breaks. It may interact with both DNA Polymerase Delta and have a preference for mismatched 3’ termini (Mazur and Perrino, 2001) (Figure 56). Previous literature has indicated TREX2 to be relevant for interaction with double-stranded breaks. More specifically, TREX2 appears to have strong interactions with Cas9, suppressing insertions and increasing the frequency of deletions (Chari et al., 2015) (Bothmer et al., 2017).

Currently no literature seems to fully characterize the action of TREX2 on long 3’ flaps without double-stranded breaks. As TREX2 overexpression appears to strongly retard even PE2 editing which does not generate double-stranded breaks with high frequency, TREX2 may have additional properties as a flap exonuclease that have previously gone uncharacterized. Additionally, since knockdown and overexpression of FEN1 appeared to have no effect on prime editing, the specific type of flap generated by prime editing may have unique interactions with TREX2 and not FEN1. These results are

consequential in the broader context of understanding this emerging technology, as FEN1 was the originally surmised potential interacting enzyme with prime editing mechanics in the landmark paper detailing its invention. Literature also suggests that TREX2 knockdown reduces cell proliferation (Bothmer et al., 2017), which may account for the results seen here due to the fact that DNA repair processes are most active during cell proliferation, and thus if TREX2 knockdown has off-target connectome effects that reduce repair in general, even if the specific action of prime editing increases, overall edit incorporation may decrease as well, leading to inconsistent outcomes depending on the stages of cell culture confluence.

The translatability of these outcomes have various leverageable genome editing impacts. Optimized cotransfection of TREX2 siRNA may allow for improved prime editing frequencies, or may allow for decrease of unwanted insertions to bias for deletions if overexpressed, as shown with the reduction in editing with the HEK3 17+CTT pegRNA, in case there are target sites whose context appear to generate unwanted insertions. Additionally it can be tentatively surmised that inhibition of the specific exonuclease action of TREX2 without diminishing cell proliferation may theoretically achieve increased prime editing levels, that have been difficult to render consistently otherwise via knockdown methods presented here. Perhaps further engineering of prime editors with some TREX2-inhibiting enzyme could potentially represent a future endeavor for iterative improvement of the technology. It is currently already known that Cas9-TREX2 fusion enzymes exhibit markedly different editing outcomes, and so it is reasonable that concurrent usage of TREX2-inhibiting enzymes may allow colocalization effects to more specifically abrogate TREX2 interference than

generalized knockdown.Unfortunately, the protein expression profile and editing attempts in primary human cells appear to indicate that while editing reagents targeting A1AT E34K may be sensitive to the actions of TREX2, it is debatable whether TREX2 knockdown will be helpful in increasing prime editing for that goal due to low levels of sufficient protein presence in the primary cell type.

Limitations

The research methods are limited in terms of bandwidth by the number of samples able to be tested. Since each target site must be individually tested and screened for guides which may number in the dozens, and then for a candidate guide the length of components would be varied to ten or more subconfigurations multiplied against another ten, screening many disease target sites as would be done in an ideal case to assess the robustness of prime editing is not logistically feasible. As a result, generalization of the efficacy of prime editing to other sites is not necessarily reliably extensible, as guide efficacy in CRISPR-based technologies is heavily variable between sequence to sequence and thus between target site to target site.

Since prime editing is also not optimized, and there is no way to logistically perform a full optimization pass off all potential guides at once, it may be true that initial guide designs that were iterated off of may have skewed the data. It is possible to imagine that a full length RT/PBS guide at one spacer may not have performed well but that a shortened version may work very well and have the hidden potential to exceed the performance of other guides that appear to work better initially. By selecting candidates competitively against each other, some better guide designs may have been overlooked.

Additionally, due to the logistics involved in screening siRNAs for all of the genes involved in human DNA damage repair, it was not possible to perform a holistic all-encompassing screen of all genes. Validating over a thousand siRNAs individually and close to a thousand genes without a pre-established pipeline in all cell lines could potentially take years, so only a subset of pre-validated siRNAs were screened. This could potentially have allowed useful siRNAs to be left out and prevented understanding of the entire picture, although the data should still be valuable regardless.

It is also important to caution that all in vitro results, even those from human primary cells, do not necessarily translate into editing efficiency in vivo. There are many barriers to translating idealized lipofection into medicine, such as delivery to the correct organ, ensuring that the behavior of the cell type in vitro is representative of the behavior in vivo, and ensuring immunocompatibility. It is known that primary human hepatocytes exhibit very poor survival in vitro, and the current model only maintains cell viability about four days unless they are supplied in a coculture model to sustain them. Since other genome editing tools with well-validated expression timelines can take several days for their expression and editing followed by repair to occur, such as with 2 days for C base editors and three days for A base editors, it may be difficult to see the entire editing potential of prime editors without in situ conformation and a longer timescale. This problem could be compounded if performing a sequential transfection with siRNA followed by prime editors. Additionally, there remains the question of whether TREX2 is expressed in these cells at a transcript and protein level. The coculture model used with the primary human hepatocytes results in some degree of transfection of the mouse fibroblasts, thus allowing for the lentiviral cassette to be amplified from mouse gDNA,

resulting in some confounding of the prime editing data for A1AT E342K. Due to the difficulty in obtaining A1AT E342K pathogenic human primary hepatocytes this was a compromise that was made, but data interpretation must be taken with caution, and future work ought to be performed with E342K-carrying patient-derived hepatocytes.

No identification of off-targets beyond the alleles identified in the detectable amplicon range for MiSeq was performed. Large deletions and translocations as well as mismatched off-target sites may be difficult to detect, requiring the development of complex whole genome screening methods and were thus beyond the scope of this endeavor. This work merely assesses the feasibility of prime editing as it pertains to Alpha-1 in terms of corrective editing and other potential outcomes in the immediate locus. This limits the extensibility of conclusions due to other possible unknown negative effects.

Additionally, despite many attempts to develop a more robust assay, the siRNA cell culture knockdown sequential transfection method used in this work exhibited inconsistent editing outcomes with low overall editing that led to difficulty in extracting useful information. Such attempts constituted the bulk of the work, and is not presented here, but speaks to the challenges of reliable assay development for emergent technologies that are not yet fully understood.

Due to the limitations of the scope of this work, it was not possible to fully screen all direct comparisons of editor combinations at the A1AT E342 locus. The comparison between the correction of the E342K lenti insert by NGA pegRNAs may not be a perfect comparison against the genomic E342E insertion by NGG pegRNAs, which can lead to debate on their relative activity.

The overexpression assay utilized in this work as an orthogonal confirmation to the siRNA knockdown also presumes an opposite effect to occur with overexpression in order to generate a positive confirmation, which may not be the case mechanistically at the molecular level for various genes. It is easy to imagine a unidirectional effect where knockdown achieves an impact to prime editing whereas overexpression above baseline achieves nothing due to saturation of a genes effect even at normal expression levels. Due to logistics limitations with supply and cost, followups on other hits from the siRNA screen were not possible in more detail, which were made more unfortunate by the known deficiencies in overexpression as a confirmation method against siRNA knockdown. Constitutive knockdown by shRNA-integrated cell lines or knockout cell lines for TREX2 would have served as strong confirmation for the knockdown-specific directional effect, but were not procurable in time due to prohibitive cost and production turnover by CMOs. Additionally, the western blots performed on untreated cells indicate that TREX2 expression is possibly quite low, and the antibodies have not been fully validated beyond the manufacturer's claims for false positive results.

Work which remains to be done include the inclusion of a more robust pegRNA guide to raise the sensitivity of the knockdown assay, examination of a LIG4 or FANCM knockout cell line, protein assays of WT cell behavior for overexpression/knockdown and baseline comparisons, engineering of prime editors by rational design or directed evolution, testing of expanded pegRNA designs, and optimization of a primary human hepatocyte prime editing screen for A1AT E342K.

Appendix 1.

Additional Tables

Table 24. Primer list for USER cloning of pCMV-PE2 and pUTR-PE2.

GL496	AACCGG/ideoxyU/CATCATCACCATCA	USER cloning primer F to connect pCMV to PE2frag3
GL497	ACCGGT/ideoxyU/AGACTTTCCTCTTCTTCTTGGG	USER cloning primer R to connect PE2frag3 to pCMV
GL498	ATACTGA/ideoxyU/AGCCGTTATGCTTTTG	USER cloning primer F to connect PE2frag3 to PE2frag2
GL499	ATCAGTA/ideoxyU/AAACATTTAGCTTCTTACCTTCTGC	USER cloning primer R to connect PE2frag2 to PE2frag3
GL500	AACAAC/ideoxyU/ACCACCACGCCCACGACGC	USER cloning primer F to connect PE2frag2 to PE2frag1
GL501	AGTTGT/ideoxyU/GATCTCGCGCA	USER cloning primer R to connect PE2frag1 to PE2frag2
GL502	ATGAAACG/ideoxyU/ACAGCCGAC	USER cloning primer F to connect PE2frag1 to pCMV
GL503	ACGTTTCA/ideoxyU/GGTGGCGGCTCTCC	USER cloning primer R to connect pCMV to PE2frag1
GL504	AGCCACCA/ideoxyU/GAAACGGACAGCCGA	USER cloning primer F to connect PE2frag1 to pUTR
GL505	ATGGTGGC/ideoxyU/CTTATATTTCTTCTTACT	USER cloning primer R to connect pUTR to PE2frag1
GL506	ATTAAT/ideoxyU/AAGCTGCCTTCTGCG	USER cloning primer F to connect pUTR to PE2frag3
GL507	AATTAA/ideoxyU/TAGACTTTCCTCTTCTTCTTGGG	USER cloning primer R to connect PE2frag3 to pUTR

Table 25. Sanger sequencing primers for verification of pUTR plasmids.

GL508	TCGCGCGTTTCGGTGAT	Sequencing Primer for pUTR-PE2 constructions F1
GL509	ACCAACTCTGTGGGCTGG	Sequencing Primer for pUTR-PE2 constructions F2
GL510	ACGCCAAGGCCATCCTG	Sequencing Primer for pUTR-PE2 constructions F3
GL511	GCAGCGGACCTTCGACA	Sequencing Primer for pUTR-PE2 constructions F4
GL512	ATACCACGATCTGCTGAAAATTAT	Sequencing Primer for pUTR-PE2 constructions F5
GL513	GCATCAAAGAGCTGGGCAG	Sequencing Primer for pUTR-PE2 constructions F6
GL514	ACAACTACCACCACGCCC	Sequencing Primer for pUTR-PE2 constructions F7
GL515	CAGCTTCGAGAAGAATCCCA	Sequencing Primer for pUTR-PE2 constructions F8
GL516	GGAGGTGACTCTGGAGGATC	Sequencing Primer for pUTR-PE2 constructions F9
GL517	GACTCCACCCCACCAGT	Sequencing Primer for pUTR-PE2 constructions F10
GL518	TCTAACTGCCCCAGCCCT	Sequencing Primer for pUTR-PE2 constructions F11
GL519	CGAGGTAATCTGGGCTAAAGC	Sequencing Primer for pUTR-PE2 constructions F12
GL520	GCCTGAGTAGGAAGTCTAGAGT	Sequencing Primer for pUTR-PE2 constructions F13
GL521	TTTCCCCCTGGAAGCTCC	Sequencing Primer for pUTR-PE2 constructions F14
GL522	ATCAAAAAGGATCTTCACCTAGATC	Sequencing Primer for pUTR-PE2 constructions F15
GL523	ACCTGGAATGCTGTTTTCCC	Sequencing Primer for pUTR-PE2 constructions F16

Table 26. PrimeDesign output for A1AT E342K pegRNA library.

Designation	Annotation	PBS_length	RTT_length	Spacer_sequence_order_TOP	Spacer_sequence_order_BOTTOM	pegRNA_extension_sequence_order_TOP	pegRNA_extension_sequence_order_BOTTOM
GL544	PAM_disrupted	8	17	caccGGCCGTGCATAAGGCTGTGCgtttt	ctctaaaacGCACAGCCTTATGCACGGCC	gtgcCTcGTCGATGGTgAGCACAGCCTTA	aaaaTAAGGCTGTGCTcACCATCGACgAG
GL544	PAM_disrupted	10	17	caccGGCCGTGCATAAGGCTGTGCgtttt	ctctaaaacGCACAGCCTTATGCACGGCC	gtgcCTcGTCGATGGTgAGCACAGCCTTATG	aaaaCATAAGGCTGTGCTcACCATCGACgAG
GL544	PAM_disrupted	12	17	caccGGCCGTGCATAAGGCTGTGCgtttt	ctctaaaacGCACAGCCTTATGCACGGCC	gtgcCTcGTCGATGGTgAGCACAGCCTTATGCA	aaaaTGCATAAGGCTGTGCTcACCATCGACgAG
GL544	PAM_disrupted	13	17	caccGGCCGTGCATAAGGCTGTGCgtttt	ctctaaaacGCACAGCCTTATGCACGGCC	gtgcCTcGTCGATGGTgAGCACAGCCTTATGCAC	aaaaGTGCATAAGGCTGTGCTcACCATCGACgAG
GL544	PAM_disrupted	14	17	caccGGCCGTGCATAAGGCTGTGCgtttt	ctctaaaacGCACAGCCTTATGCACGGCC	gtgcCTcGTCGATGGTgAGCACAGCCTTATGCACG	aaaaCGTGCATAAGGCTGTGCTcACCATCGACgAG
GL544	PAM_disrupted	15	17	caccGGCCGTGCATAAGGCTGTGCgtttt	ctctaaaacGCACAGCCTTATGCACGGCC	gtgcCTcGTCGATGGTgAGCACAGCCTTATGCACGG	aaaaCCGTGCATAAGGCTGTGCTcACCATCGACgAG
GL544	PAM_disrupted	8	18	caccGGCCGTGCATAAGGCTGTGCgtttt	ctctaaaacGCACAGCCTTATGCACGGCC	gtgcTCTcGTCGATGGTgAGCACAGCCTTA	aaaaTAAGGCTGTGCTcACCATCGACgAGA
GL544	PAM_disrupted	10	18	caccGGCCGTGCATAAGGCTGTGCgtttt	ctctaaaacGCACAGCCTTATGCACGGCC	gtgcTCTcGTCGATGGTgAGCACAGCCTTATG	aaaaCATAAGGCTGTGCTcACCATCGACgAGA
GL544	PAM_disrupted	12	18	caccGGCCGTGCATAAGGCTGTGCgtttt	ctctaaaacGCACAGCCTTATGCACGGCC	gtgcTCTcGTCGATGGTgAGCACAGCCTTATGCA	aaaaTGCATAAGGCTGTGCTcACCATCGACgAGA
GL544	PAM_disrupted	13	18	caccGGCCGTGCATAAGGCTGTGCgtttt	ctctaaaacGCACAGCCTTATGCACGGCC	gtgcTCTcGTCGATGGTgAGCACAGCCTTATGCAC	aaaaGTGCATAAGGCTGTGCTcACCATCGACgAGA
GL544	PAM_disrupted	14	18	caccGGCCGTGCATAAGGCTGTGCgtttt	ctctaaaacGCACAGCCTTATGCACGGCC	gtgcTCTcGTCGATGGTgAGCACAGCCTTATGCACG	aaaaCGTGCATAAGGCTGTGCTcACCATCGACgAGA
GL544	PAM_disrupted	15	18	caccGGCCGTGCATAAGGCTGTGCgtttt	ctctaaaacGCACAGCCTTATGCACGGCC	gtgcTCTcGTCGATGGTgAGCACAGCCTTATGCACGG	aaaaCCGTGCATAAGGCTGTGCTcACCATCGACgAGA
GL544	PAM_disrupted	8	20	caccGGCCGTGCATAAGGCTGTGCgtttt	ctctaaaacGCACAGCCTTATGCACGGCC	gtgcTTTCTcGTCGATGGTgAGCACAGCCTTA	aaaaTAAGGCTGTGCTcACCATCGACgAGAAA
GL544	PAM_disrupted	10	20	caccGGCCGTGCATAAGGCTGTGCgtttt	ctctaaaacGCACAGCCTTATGCACGGCC	gtgcTTTCTcGTCGATGGTgAGCACAGCCTTATG	aaaaCATAAGGCTGTGCTcACCATCGACgAGAAA
GL544	PAM_disrupted	12	20	caccGGCCGTGCATAAGGCTGTGCgtttt	ctctaaaacGCACAGCCTTATGCACGGCC	gtgcTTTCTcGTCGATGGTgAGCACAGCCTTATGCA	aaaaTGCATAAGGCTGTGCTcACCATCGACgAGAAA
GL544	PAM_disrupted	13	20	caccGGCCGTGCATAAGGCTGTGCgtttt	ctctaaaacGCACAGCCTTATGCACGGCC	gtgcTTTCTcGTCGATGGTgAGCACAGCCTTATGCAC	aaaaGTGCATAAGGCTGTGCTcACCATCGACgAGAAA
GL544	PAM_disrupted	14	20	caccGGCCGTGCATAAGGCTGTGCgtttt	ctctaaaacGCACAGCCTTATGCACGGCC	gtgcTTTCTcGTCGATGGTgAGCACAGCCTTATGCACG	aaaaCGTGCATAAGGCTGTGCTcACCATCGACgAGAAA
GL544	PAM_disrupted	15	20	caccGGCCGTGCATAAGGCTGTGCgtttt	ctctaaaacGCACAGCCTTATGCACGGCC	gtgcTTTCTcGTCGATGGTgAGCACAGCCTTATGCACGG	aaaaCCGTGCATAAGGCTGTGCTcACCATCGACgAGAAA
GL545	PAM_intact	8	10	caccGCATAAGGCTGTGCTGACCATgtttt	ctctaaaacATGGTCAGCACAGCCTTATGC	gtgcCTcGTCGATGGTCAGCAC	aaaaGTGCTGACCATCGACgAG
GL545	PAM_intact	10	10	caccGCATAAGGCTGTGCTGACCATgtttt	ctctaaaacATGGTCAGCACAGCCTTATGC	gtgcCTcGTCGATGGTCAGCACAG	aaaaCTGTGCTGACCATCGACgAG
GL545	PAM_intact	12	10	caccGCATAAGGCTGTGCTGACCATgtttt	ctctaaaacATGGTCAGCACAGCCTTATGC	gtgcCTcGTCGATGGTCAGCACAGCC	aaaaGGCTGTGCTGACCATCGACgAG
GL545	PAM_intact	13	10	caccGCATAAGGCTGTGCTGACCATgtttt	ctctaaaacATGGTCAGCACAGCCTTATGC	gtgcCTcGTCGATGGTCAGCACAGCCT	aaaaAGGCTGTGCTGACCATCGACgAG
GL545	PAM_intact	14	10	caccGCATAAGGCTGTGCTGACCATgtttt	ctctaaaacATGGTCAGCACAGCCTTATGC	gtgcCTcGTCGATGGTCAGCACAGCCTT	aaaaAAGGCTGTGCTGACCATCGACgAG
GL545	PAM_intact	15	10	caccGCATAAGGCTGTGCTGACCATgtttt	ctctaaaacATGGTCAGCACAGCCTTATGC	gtgcCTcGTCGATGGTCAGCACAGCCTTA	aaaaTAAGGCTGTGCTGACCATCGACgAG
GL545	PAM_intact	8	14	caccGCATAAGGCTGTGCTGACCATgtttt	ctctaaaacATGGTCAGCACAGCCTTATGC	gtgcCTTTCTcGTCGATGGTCAGCAC	aaaaGTGCTGACCATCGACgAGAAAG
GL545	PAM_intact	10	14	caccGCATAAGGCTGTGCTGACCATgtttt	ctctaaaacATGGTCAGCACAGCCTTATGC	gtgcCTTTCTcGTCGATGGTCAGCACAG	aaaaCTGTGCTGACCATCGACgAGAAAG
GL545	PAM_intact	12	14	caccGCATAAGGCTGTGCTGACCATgtttt	ctctaaaacATGGTCAGCACAGCCTTATGC	gtgcCTTTCTcGTCGATGGTCAGCACAGCC	aaaaGGCTGTGCTGACCATCGACgAGAAAG
GL545	PAM_intact	13	14	caccGCATAAGGCTGTGCTGACCATgtttt	ctctaaaacATGGTCAGCACAGCCTTATGC	gtgcCTTTCTcGTCGATGGTCAGCACAGCCT	aaaaAGGCTGTGCTGACCATCGACgAGAAAG
GL545	PAM_intact	14	14	caccGCATAAGGCTGTGCTGACCATgtttt	ctctaaaacATGGTCAGCACAGCCTTATGC	gtgcCTTTCTcGTCGATGGTCAGCACAGCCTT	aaaaAAGGCTGTGCTGACCATCGACgAGAAAG
GL545	PAM_intact	15	14	caccGCATAAGGCTGTGCTGACCATgtttt	ctctaaaacATGGTCAGCACAGCCTTATGC	gtgcCTTTCTcGTCGATGGTCAGCACAGCCTTA	aaaaTAAGGCTGTGCTGACCATCGACgAGAAAG
GL545	PAM_intact	8	20	caccGCATAAGGCTGTGCTGACCATgtttt	ctctaaaacATGGTCAGCACAGCCTTATGC	gtgcCAGTCCCTTTCTcGTCGATGGTCAGCAC	aaaaGTGCTGACCATCGACgAGAAAGGGACTG
GL545	PAM_intact	10	20	caccGCATAAGGCTGTGCTGACCATgtttt	ctctaaaacATGGTCAGCACAGCCTTATGC	gtgcCAGTCCCTTTCTcGTCGATGGTCAGCACAG	aaaaCTGTGCTGACCATCGACgAGAAAGGGACTG
GL545	PAM_intact	12	20	caccGCATAAGGCTGTGCTGACCATgtttt	ctctaaaacATGGTCAGCACAGCCTTATGC	gtgcCAGTCCCTTTCTcGTCGATGGTCAGCACAGCC	aaaaGGCTGTGCTGACCATCGACgAGAAAGGGACTG
GL545	PAM_intact	13	20	caccGCATAAGGCTGTGCTGACCATgtttt	ctctaaaacATGGTCAGCACAGCCTTATGC	gtgcCAGTCCCTTTCTcGTCGATGGTCAGCACAGCCT	aaaaAGGCTGTGCTGACCATCGACgAGAAAGGGACTG
GL545	PAM_intact	14	20	caccGCATAAGGCTGTGCTGACCATgtttt	ctctaaaacATGGTCAGCACAGCCTTATGC	gtgcCAGTCCCTTTCTcGTCGATGGTCAGCACAGCCTT	aaaaAAGGCTGTGCTGACCATCGACgAGAAAGGGACTG
GL545	PAM_intact	15	20	caccGCATAAGGCTGTGCTGACCATgtttt	ctctaaaacATGGTCAGCACAGCCTTATGC	gtgcCAGTCCCTTTCTcGTCGATGGTCAGCACAGCCTTA	aaaaTAAGGCTGTGCTGACCATCGACgAGAAAGGGACTG
GL600	PAM_intact	8	10	caccGGCTGTGCTGACCATCGACAgtttt	ctctaaaacTGTCGATGGTCAGCACAGCC	gtgcCCTTTCTcGTCGATGGTC	aaaaGACCATCGACgAGAAAGG
GL600	PAM_intact	10	10	caccGGCTGTGCTGACCATCGACAgtttt	ctctaaaacTGTCGATGGTCAGCACAGCC	gtgcCCTTTCTcGTCGATGGTCAG	aaaaCTGACCATCGACgAGAAAGG
GL600	PAM_intact	12	10	caccGGCTGTGCTGACCATCGACAgtttt	ctctaaaacTGTCGATGGTCAGCACAGCC	gtgcCCTTTCTcGTCGATGGTCAGCA	aaaaTGCTGACCATCGACgAGAAAGG
GL600	PAM_intact	13	10	caccGGCTGTGCTGACCATCGACAgtttt	ctctaaaacTGTCGATGGTCAGCACAGCC	gtgcCCTTTCTcGTCGATGGTCAGCAC	aaaaGTGCTGACCATCGACgAGAAAGG
GL600	PAM_intact	14	10	caccGGCTGTGCTGACCATCGACAgtttt	ctctaaaacTGTCGATGGTCAGCACAGCC	gtgcCCTTTCTcGTCGATGGTCAGCACA	aaaaTGTGCTGACCATCGACgAGAAAGG
GL600	PAM_intact	15	10	caccGGCTGTGCTGACCATCGACAgtttt	ctctaaaacTGTCGATGGTCAGCACAGCC	gtgcCCTTTCTcGTCGATGGTCAGCACAG	aaaaCTGTGCTGACCATCGACgAGAAAGG
GL600	PAM_intact	8	15	caccGGCTGTGCTGACCATCGACAgtttt	ctctaaaacTGTCGATGGTCAGCACAGCC	gtgcCAGTCCCTTTCTcGTCGATGGTC	aaaaGACCATCGACgAGAAAGGGACTG
GL600	PAM_intact	10	15	caccGGCTGTGCTGACCATCGACAgtttt	ctctaaaacTGTCGATGGTCAGCACAGCC	gtgcCAGTCCCTTTCTcGTCGATGGTCAG	aaaaCTGACCATCGACgAGAAAGGGACTG
GL600	PAM_intact	12	15	caccGGCTGTGCTGACCATCGACAgtttt	ctctaaaacTGTCGATGGTCAGCACAGCC	gtgcCAGTCCCTTTCTcGTCGATGGTCAGCA	aaaaTGCTGACCATCGACgAGAAAGGGACTG
GL600	PAM_intact	13	15	caccGGCTGTGCTGACCATCGACAgtttt	ctctaaaacTGTCGATGGTCAGCACAGCC	gtgcCAGTCCCTTTCTcGTCGATGGTCAGCAC	aaaaGTGCTGACCATCGACgAGAAAGGGACTG
GL600	PAM_intact	14	15	caccGGCTGTGCTGACCATCGACAgtttt	ctctaaaacTGTCGATGGTCAGCACAGCC	gtgcCAGTCCCTTTCTcGTCGATGGTCAGCACA	aaaaTGTGCTGACCATCGACgAGAAAGGGACTG
GL600	PAM_intact	15	15	caccGGCTGTGCTGACCATCGACAgtttt	ctctaaaacTGTCGATGGTCAGCACAGCC	gtgcCAGTCCCTTTCTcGTCGATGGTCAGCACAG	aaaaCTGTGCTGACCATCGACgAGAAAGGGACTG
GL600	PAM_intact	8	20	caccGGCTGTGCTGACCATCGACAgtttt	ctctaaaacTGTCGATGGTCAGCACAGCC	gtgcAGCTTCAGTCCCTTTCTcGTCGATGGTC	aaaaGACCATCGACgAGAAAGGGACTGAAGCT
GL600	PAM_intact	10	20	caccGGCTGTGCTGACCATCGACAgtttt	ctctaaaacTGTCGATGGTCAGCACAGCC	gtgcAGCTTCAGTCCCTTTCTcGTCGATGGTCAG	aaaaCTGACCATCGACgAGAAAGGGACTGAAGCT
GL600	PAM_intact	12	20	caccGGCTGTGCTGACCATCGACAgtttt	ctctaaaacTGTCGATGGTCAGCACAGCC	gtgcAGCTTCAGTCCCTTTCTcGTCGATGGTCAGCA	aaaaTGCTGACCATCGACgAGAAAGGGACTGAAGCT
GL600	PAM_intact	13	20	caccGGCTGTGCTGACCATCGACAgtttt	ctctaaaacTGTCGATGGTCAGCACAGCC	gtgcAGCTTCAGTCCCTTTCTcGTCGATGGTCAGCAC	aaaaGTGCTGACCATCGACgAGAAAGGGACTGAAGCT
GL600	PAM_intact	14	20	caccGGCTGTGCTGACCATCGACAgtttt	ctctaaaacTGTCGATGGTCAGCACAGCC	gtgcAGCTTCAGTCCCTTTCTcGTCGATGGTCAGCACA	aaaaTGTGCTGACCATCGACgAGAAAGGGACTGAAGCT
GL600	PAM_intact	15	20	caccGGCTGTGCTGACCATCGACAgtttt	ctctaaaacTGTCGATGGTCAGCACAGCC	gtgcAGCTTCAGTCCCTTTCTcGTCGATGGTCAGCACAG	aaaaCTGTGCTGACCATCGACgAGAAAGGGACTGAAGCT
GL556	PAM_disrupted	8	10	caccGAGCTTCAGTCCCTTTCTTGTgtttt	ctctaaaacACAAGAAAGGGACTGAAGCTC	gtgcACCATaGACgAGAAAGGG	aaaaCCCTTTCTcGTCtATGGT
GL556	PAM_disrupted	10	10	caccGAGCTTCAGTCCCTTTCTTGTgtttt	ctctaaaacACAAGAAAGGGACTGAAGCTC	gtgcACCATaGACgAGAAAGGGAC	aaaaGTCCCTTTCTcGTCtATGGT
GL556	PAM_disrupted	12	10	caccGAGCTTCAGTCCCTTTCTTGTgtttt	ctctaaaacACAAGAAAGGGACTGAAGCTC	gtgcACCATaGACgAGAAAGGGACTG	aaaaCAGTCCCTTTCTcGTCtATGGT
GL556	PAM_disrupted	13	10	caccGAGCTTCAGTCCCTTTCTTGTgtttt	ctctaaaacACAAGAAAGGGACTGAAGCTC	gtgcACCATaGACgAGAAAGGGACTGA	aaaaTCAGTCCCTTTCTcGTCtATGGT
GL556	PAM_disrupted	14	10	caccGAGCTTCAGTCCCTTTCTTGTgtttt	ctctaaaacACAAGAAAGGGACTGAAGCTC	gtgcACCATaGACgAGAAAGGGACTGAA	aaaaTTCAGTCCCTTTCTcGTCtATGGT
GL556	PAM_disrupted	15	10	caccGAGCTTCAGTCCCTTTCTTGTgtttt	ctctaaaacACAAGAAAGGGACTGAAGCTC	gtgcACCATaGACgAGAAAGGGACTGAAG	aaaaCTTCAGTCCCTTTCTcGTCtATGGT
GL556	PAM_disrupted	8	15	caccGAGCTTCAGTCCCTTTCTTGTgtttt	ctctaaaacACAAGAAAGGGACTGAAGCTC	gtgcTGCTGACCATaGACgAGAAAGGG	aaaaCCCTTTCTcGTCtATGGTCAGCA
GL556	PAM_disrupted	10	15	caccGAGCTTCAGTCCCTTTCTTGTgtttt	ctctaaaacACAAGAAAGGGACTGAAGCTC	gtgcTGCTGACCATaGACgAGAAAGGGAC	aaaaGTCCCTTTCTcGTCtATGGTCAGCA
GL556	PAM_disrupted	12	15	caccGAGCTTCAGTCCCTTTCTTGTgtttt	ctctaaaacACAAGAAAGGGACTGAAGCTC	gtgcTGCTGACCATaGACgAGAAAGGGACTG	aaaaCAGTCCCTTTCTcGTCtATGGTCAGCA
GL556	PAM_disrupted	13	15	caccGAGCTTCAGTCCCTTTCTTGTgtttt	ctctaaaacACAAGAAAGGGACTGAAGCTC	gtgcTGCTGACCATaGACgAGAAAGGGACTGA	aaaaTCAGTCCCTTTCTcGTCtATGGTCAGCA
GL556	PAM_disrupted	14	15	caccGAGCTTCAGTCCCTTTCTTGTgtttt	ctctaaaacACAAGAAAGGGACTGAAGCTC	gtgcTGCTGACCATaGACgAGAAAGGGACTGAA	aaaaTTCAGTCCCTTTCTcGTCtATGGTCAGCA
GL556	PAM_disrupted	15	15	caccGAGCTTCAGTCCCTTTCTTGTgtttt	ctctaaaacACAAGAAAGGGACTGAAGCTC	gtgcTGCTGACCATaGACgAGAAAGGGACTGAAG	aaaaCTTCAGTCCCTTTCTcGTCtATGGTCAGCA
GL556	PAM_disrupted	8	20	caccGAGCTTCAGTCCCTTTCTTGTgtttt	ctctaaaacACAAGAAAGGGACTGAAGCTC	gtgcGGCTGTGCTGACCATaGACgAGAAAGGG	aaaaCCCTTTCTcGTCtATGGTCAGCACAGCC
GL556	PAM_disrupted	10	20	caccGAGCTTCAGTCCCTTTCTTGTgtttt	ctctaaaacACAAGAAAGGGACTGAAGCTC	gtgcGGCTGTGCTGACCATaGACgAGAAAGGGAC	aaaaGTCCCTTTCTcGTCtATGGTCAGCACAGCC
GL556	PAM_disrupted	12	20	caccGAGCTTCAGTCCCTTTCTTGTgtttt	ctctaaaacACAAGAAAGGGACTGAAGCTC	gtgcGGCTGTGCTGACCATaGACgAGAAAGGGACTG	aaaaCAGTCCCTTTCTcGTCtATGGTCAGCACAGCC
GL556	PAM_disrupted	13	20	caccGAGCTTCAGTCCCTTTCTTGTgtttt	ctctaaaacACAAGAAAGGGACTGAAGCTC	gtgcGGCTGTGCTGACCATaGACgAGAAAGGGACTGA	aaaaTCAGTCCCTTTCTcGTCtATGGTCAGCACAGCC
GL556	PAM_disrupted	14	20	caccGAGCTTCAGTCCCTTTCTTGTgtttt	ctctaaaacACAAGAAAGGGACTGAAGCTC	gtgcGGCTGTGCTGACCATaGACgAGAAAGGGACTGAA	aaaaTTCAGTCCCTTTCTcGTCtATGGTCAGCACAGCC
GL556	PAM_disrupted	15	20	caccGAGCTTCAGTCCCTTTCTTGTgtttt	ctctaaaacACAAGAAAGGGACTGAAGCTC	gtgcGGCTGTGCTGACCATaGACgAGAAAGGGACTGAAG	aaaaCTTCAGTCCCTTTCTcGTCtATGGTCAGCACAGCC
GL549	PAM_intact	8	23	caccGTCTAAAAACATGGCCCCAGCgtttt	ctctaaaacGCTGGGGCCATGTTTTTAGAC	gtgcACgAGAAAGGGACTGAAGCTGCTGGGGCCAT	aaaaATGGCCCCAGCAGCTTCAGTCCCTTTCTcGT
GL549	PAM_intact	10	23	caccGTCTAAAAACATGGCCCCAGCgtttt	ctctaaaacGCTGGGGCCATGTTTTTAGAC	gtgcACgAGAAAGGGACTGAAGCTGCTGGGGCCATGT	aaaaACATGGCCCCAGCAGCTTCAGTCCCTTTCTcGT
GL549	PAM_intact	12	23	caccGTCTAAAAACATGGCCCCAGCgtttt	ctctaaaacGCTGGGGCCATGTTTTTAGAC	gtgcACgAGAAAGGGACTGAAGCTGCTGGGGCCATGTTT	aaaaAAACATGGCCCCAGCAGCTTCAGTCCCTTTCTcGT
GL549	PAM_intact	13	23	caccGTCTAAAAACATGGCCCCAGCgtttt	ctctaaaacGCTGGGGCCATGTTTTTAGAC	gtgcACgAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTT	aaaaAAAACATGGCCCCAGCAGCTTCAGTCCCTTTCTcGT
GL549	PAM_intact	14	23	caccGTCTAAAAACATGGCCCCAGCgtttt	ctctaaaacGCTGGGGCCATGTTTTTAGAC	gtgcACgAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTT	aaaaAAAAACATGGCCCCAGCAGCTTCAGTCCCTTTCTcGT
GL549	PAM_intact	15	23	caccGTCTAAAAACATGGCCCCAGCgtttt	ctctaaaacGCTGGGGCCATGTTTTTAGAC	gtgcACgAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTA	aaaaTAAAAACATGGCCCCAGCAGCTTCAGTCCCTTTCTcGT
GL549	PAM_intact	8	24	caccGTCTAAAAACATGGCCCCAGCgtttt	ctctaaaacGCTGGGGCCATGTTTTTAGAC	gtgcGACgAGAAAGGGACTGAAGCTGCTGGGGCCAT	aaaaATGGCCCCAGCAGCTTCAGTCCCTTTCTcGTC
GL549	PAM_intact	10	24	caccGTCTAAAAACATGGCCCCAGCgtttt	ctctaaaacGCTGGGGCCATGTTTTTAGAC	gtgcGACgAGAAAGGGACTGAAGCTGCTGGGGCCATGT	aaaaACATGGCCCCAGCAGCTTCAGTCCCTTTCTcGTC
GL549	PAM_intact	12	24	caccGTCTAAAAACATGGCCCCAGCgtttt	ctctaaaacGCTGGGGCCATGTTTTTAGAC	gtgcGACgAGAAAGGGACTGAAGCTGCTGGGGCCATGTTT	aaaaAAACATGGCCCCAGCAGCTTCAGTCCCTTTCTcGTC
GL549	PAM_intact	13	24	caccGTCTAAAAACATGGCCCCAGCgtttt	ctctaaaacGCTGGGGCCATGTTTTTAGAC	gtgcGACgAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTT	aaaaAAAACATGGCCCCAGCAGCTTCAGTCCCTTTCTcGTC
GL549	PAM_intact	14	24	caccGTCTAAAAACATGGCCCCAGCgtttt	ctctaaaacGCTGGGGCCATGTTTTTAGAC	gtgcGACgAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTT	aaaaAAAAACATGGCCCCAGCAGCTTCAGTCCCTTTCTcGTC
GL549	PAM_intact	15	24	caccGTCTAAAAACATGGCCCCAGCgtttt	ctctaaaacGCTGGGGCCATGTTTTTAGAC	gtgcGACgAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTA	aaaaTAAAAACATGGCCCCAGCAGCTTCAGTCCCTTTCTcGTC
GL549	PAM_intact	8	25	caccGTCTAAAAACATGGCCCCAGCgtttt	ctctaaaacGCTGGGGCCATGTTTTTAGAC	gtgcCGACgAGAAAGGGACTGAAGCTGCTGGGGCCAT	aaaaATGGCCCCAGCAGCTTCAGTCCCTTTCTcGTCG
GL549	PAM_intact	10	25	caccGTCTAAAAACATGGCCCCAGCgtttt	ctctaaaacGCTGGGGCCATGTTTTTAGAC	gtgcCGACgAGAAAGGGACTGAAGCTGCTGGGGCCATGT	aaaaACATGGCCCCAGCAGCTTCAGTCCCTTTCTcGTCG
GL549	PAM_intact	12	25	caccGTCTAAAAACATGGCCCCAGCgtttt	ctctaaaacGCTGGGGCCATGTTTTTAGAC	gtgcCGACgAGAAAGGGACTGAAGCTGCTGGGGCCATGTTT	aaaaAAACATGGCCCCAGCAGCTTCAGTCCCTTTCTcGTCG
GL549	PAM_intact	13	25	caccGTCTAAAAACATGGCCCCAGCgtttt	ctctaaaacGCTGGGGCCATGTTTTTAGAC	gtgcCGACgAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTT	aaaaAAAACATGGCCCCAGCAGCTTCAGTCCCTTTCTcGTCG
GL549	PAM_intact	14	25	caccGTCTAAAAACATGGCCCCAGCgtttt	ctctaaaacGCTGGGGCCATGTTTTTAGAC	gtgcCGACgAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTT	aaaaAAAAACATGGCCCCAGCAGCTTCAGTCCCTTTCTcGTCG
GL549	PAM_intact	15	25	caccGTCTAAAAACATGGCCCCAGCgtttt	ctctaaaacGCTGGGGCCATGTTTTTAGAC	gtgcCGACgAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTA	aaaaTAAAAACATGGCCCCAGCAGCTTCAGTCCCTTTCTcGTCG
GL548	PAM_disrupted	8	19	caccGCCAGGCCGTGCATAAGGCTGgtttt	ctctaaaacCAGCCTTATGCACGGCCTGGC	gtgcTtGTCGATGGTCAGgACAGCCTTATGC	aaaaGCATAAGGCTGTcCTGACCATCGACaA
GL548	PAM_disrupted	10	19	caccGCCAGGCCGTGCATAAGGCTGgtttt	ctctaaaacCAGCCTTATGCACGGCCTGGC	gtgcTtGTCGATGGTCAGgACAGCCTTATGCAC	aaaaGTGCATAAGGCTGTcCTGACCATCGACaA
GL548	PAM_disrupted	12	19	caccGCCAGGCCGTGCATAAGGCTGgtttt	ctctaaaacCAGCCTTATGCACGGCCTGGC	gtgcTtGTCGATGGTCAGgACAGCCTTATGCACGG	aaaaCCGTGCATAAGGCTGTcCTGACCATCGACaA
GL548	PAM_disrupted	13	19	caccGCCAGGCCGTGCATAAGGCTGgtttt	ctctaaaacCAGCCTTATGCACGGCCTGGC	gtgcTtGTCGATGGTCAGgACAGCCTTATGCACGGC	aaaaGCCGTGCATAAGGCTGTcCTGACCATCGACaA
GL548	PAM_disrupted	14	19	caccGCCAGGCCGTGCATAAGGCTGgtttt	ctctaaaacCAGCCTTATGCACGGCCTGGC	gtgcTtGTCGATGGTCAGgACAGCCTTATGCACGGCC	aaaaGGCCGTGCATAAGGCTGTcCTGACCATCGACaA
GL548	PAM_disrupted	15	19	caccGCCAGGCCGTGCATAAGGCTGgtttt	ctctaaaacCAGCCTTATGCACGGCCTGGC	gtgcTtGTCGATGGTCAGgACAGCCTTATGCACGGCCT	aaaaAGGCCGTGCATAAGGCTGTcCTGACCATCGACaA
GL548	PAM_disrupted	8	20	caccGCCAGGCCGTGCATAAGGCTGgtttt	ctctaaaacCAGCCTTATGCACGGCCTGGC	gtgcCTtGTCGATGGTCAGgACAGCCTTATGC	aaaaGCATAAGGCTGTcCTGACCATCGACaAG
GL548	PAM_disrupted	10	20	caccGCCAGGCCGTGCATAAGGCTGgtttt	ctctaaaacCAGCCTTATGCACGGCCTGGC	gtgcCTtGTCGATGGTCAGgACAGCCTTATGCAC	aaaaGTGCATAAGGCTGTcCTGACCATCGACaAG
GL548	PAM_disrupted	12	20	caccGCCAGGCCGTGCATAAGGCTGgtttt	ctctaaaacCAGCCTTATGCACGGCCTGGC	gtgcCTtGTCGATGGTCAGgACAGCCTTATGCACGG	aaaaCCGTGCATAAGGCTGTcCTGACCATCGACaAG
GL548	PAM_disrupted	13	20	caccGCCAGGCCGTGCATAAGGCTGgtttt	ctctaaaacCAGCCTTATGCACGGCCTGGC	gtgcCTtGTCGATGGTCAGgACAGCCTTATGCACGGC	aaaaGCCGTGCATAAGGCTGTcCTGACCATCGACaAG
GL548	PAM_disrupted	14	20	caccGCCAGGCCGTGCATAAGGCTGgtttt	ctctaaaacCAGCCTTATGCACGGCCTGGC	gtgcCTtGTCGATGGTCAGgACAGCCTTATGCACGGCC	aaaaGGCCGTGCATAAGGCTGTcCTGACCATCGACaAG
GL548	PAM_disrupted	15	20	caccGCCAGGCCGTGCATAAGGCTGgtttt	ctctaaaacCAGCCTTATGCACGGCCTGGC	gtgcCTtGTCGATGGTCAGgACAGCCTTATGCACGGCCT	aaaaAGGCCGTGCATAAGGCTGTcCTGACCATCGACaAG
GL548	PAM_disrupted	8	21	caccGCCAGGCCGTGCATAAGGCTGgtttt	ctctaaaacCAGCCTTATGCACGGCCTGGC	gtgcTCTtGTCGATGGTCAGgACAGCCTTATGC	aaaaGCATAAGGCTGTcCTGACCATCGACaAGA
GL548	PAM_disrupted	10	21	caccGCCAGGCCGTGCATAAGGCTGgtttt	ctctaaaacCAGCCTTATGCACGGCCTGGC	gtgcTCTtGTCGATGGTCAGgACAGCCTTATGCAC	aaaaGTGCATAAGGCTGTcCTGACCATCGACaAGA
GL548	PAM_disrupted	12	21	caccGCCAGGCCGTGCATAAGGCTGgtttt	ctctaaaacCAGCCTTATGCACGGCCTGGC	gtgcTCTtGTCGATGGTCAGgACAGCCTTATGCACGG	aaaaCCGTGCATAAGGCTGTcCTGACCATCGACaAGA
GL548	PAM_disrupted	13	21	caccGCCAGGCCGTGCATAAGGCTGgtttt	ctctaaaacCAGCCTTATGCACGGCCTGGC	gtgcTCTtGTCGATGGTCAGgACAGCCTTATGCACGGC	aaaaGCCGTGCATAAGGCTGTcCTGACCATCGACaAGA
GL548	PAM_disrupted	14	21	caccGCCAGGCCGTGCATAAGGCTGgtttt	ctctaaaacCAGCCTTATGCACGGCCTGGC	gtgcTCTtGTCGATGGTCAGgACAGCCTTATGCACGGCC	aaaaGGCCGTGCATAAGGCTGTcCTGACCATCGACaAGA
GL548	PAM_disrupted	15	21	caccGCCAGGCCGTGCATAAGGCTGgtttt	ctctaaaacCAGCCTTATGCACGGCCTGGC	gtgcTCTtGTCGATGGTCAGgACAGCCTTATGCACGGCCT	aaaaAGGCCGTGCATAAGGCTGTcCTGACCATCGACaAGA
GL542	PAM_disrupted_silent_mutation	8	25	caccGTCCCCTCCAGGCCGTGCATAgtttt	ctctaaaacTATGCACGGCCTGGAGGGGAC	gtgcTtGTCGATGGTCAGCACAGCtttATGCACGGCC	aaaaGGCCGTGCATaaaGCTGTGCTGACCATCGACaA
GL542	PAM_disrupted_silent_mutation	10	25	caccGTCCCCTCCAGGCCGTGCATAgtttt	ctctaaaacTATGCACGGCCTGGAGGGGAC	gtgcTtGTCGATGGTCAGCACAGCtttATGCACGGCCTG	aaaaCAGGCCGTGCATaaaGCTGTGCTGACCATCGACaA
GL542	PAM_disrupted_silent_mutation	13	25	caccGTCCCCTCCAGGCCGTGCATAgtttt	ctctaaaacTATGCACGGCCTGGAGGGGAC	gtgcTtGTCGATGGTCAGCACAGCtttATGCACGGCCTGGAG	aaaaCTCCAGGCCGTGCATaaaGCTGTGCTGACCATCGACaA
GL542	PAM_disrupted_silent_mutation	14	25	caccGTCCCCTCCAGGCCGTGCATAgtttt	ctctaaaacTATGCACGGCCTGGAGGGGAC	gtgcTtGTCGATGGTCAGCACAGCtttATGCACGGCCTGGAGG	aaaaCCTCCAGGCCGTGCATaaaGCTGTGCTGACCATCGACaA
GL542	PAM_disrupted_silent_mutation	15	25	caccGTCCCCTCCAGGCCGTGCATAgtttt	ctctaaaacTATGCACGGCCTGGAGGGGAC	gtgcTtGTCGATGGTCAGCACAGCtttATGCACGGCCTGGAGGG	aaaaCCCTCCAGGCCGTGCATaaaGCTGTGCTGACCATCGACaA
GL542	PAM_disrupted_silent_mutation	8	26	caccGTCCCCTCCAGGCCGTGCATAgtttt	ctctaaaacTATGCACGGCCTGGAGGGGAC	gtgcCTtGTCGATGGTCAGCACAGCtttATGCACGGCC	aaaaGGCCGTGCATaaaGCTGTGCTGACCATCGACaAG
GL542	PAM_disrupted_silent_mutation	10	26	caccGTCCCCTCCAGGCCGTGCATAgtttt	ctctaaaacTATGCACGGCCTGGAGGGGAC	gtgcCTtGTCGATGGTCAGCACAGCtttATGCACGGCCTG	aaaaCAGGCCGTGCATaaaGCTGTGCTGACCATCGACaAG
GL542	PAM_disrupted_silent_mutation	13	26	caccGTCCCCTCCAGGCCGTGCATAgtttt	ctctaaaacTATGCACGGCCTGGAGGGGAC	gtgcCTtGTCGATGGTCAGCACAGCtttATGCACGGCCTGGAG	aaaaCTCCAGGCCGTGCATaaaGCTGTGCTGACCATCGACaAG
GL542	PAM_disrupted_silent_mutation	14	26	caccGTCCCCTCCAGGCCGTGCATAgtttt	ctctaaaacTATGCACGGCCTGGAGGGGAC	gtgcCTtGTCGATGGTCAGCACAGCtttATGCACGGCCTGGAGG	aaaaCCTCCAGGCCGTGCATaaaGCTGTGCTGACCATCGACaAG
GL542	PAM_disrupted_silent_mutation	15	26	caccGTCCCCTCCAGGCCGTGCATAgtttt	ctctaaaacTATGCACGGCCTGGAGGGGAC	gtgcCTtGTCGATGGTCAGCACAGCtttATGCACGGCCTGGAGGG	aaaaCCCTCCAGGCCGTGCATaaaGCTGTGCTGACCATCGACaAG
GL542	PAM_disrupted_silent_mutation	8	27	caccGTCCCCTCCAGGCCGTGCATAgtttt	ctctaaaacTATGCACGGCCTGGAGGGGAC	gtgcTCTtGTCGATGGTCAGCACAGCtttATGCACGGCC	aaaaGGCCGTGCATaaaGCTGTGCTGACCATCGACaAGA
GL542	PAM_disrupted_silent_mutation	10	27	caccGTCCCCTCCAGGCCGTGCATAgtttt	ctctaaaacTATGCACGGCCTGGAGGGGAC	gtgcTCTtGTCGATGGTCAGCACAGCtttATGCACGGCCTG	aaaaCAGGCCGTGCATaaaGCTGTGCTGACCATCGACaAGA
GL542	PAM_disrupted_silent_mutation	13	27	caccGTCCCCTCCAGGCCGTGCATAgtttt	ctctaaaacTATGCACGGCCTGGAGGGGAC	gtgcTCTtGTCGATGGTCAGCACAGCtttATGCACGGCCTGGAG	aaaaCTCCAGGCCGTGCATaaaGCTGTGCTGACCATCGACaAGA
GL542	PAM_disrupted_silent_mutation	14	27	caccGTCCCCTCCAGGCCGTGCATAgtttt	ctctaaaacTATGCACGGCCTGGAGGGGAC	gtgcTCTtGTCGATGGTCAGCACAGCtttATGCACGGCCTGGAGG	aaaaCCTCCAGGCCGTGCATaaaGCTGTGCTGACCATCGACaAGA
GL542	PAM_disrupted_silent_mutation	15	27	caccGTCCCCTCCAGGCCGTGCATAgtttt	ctctaaaacTATGCACGGCCTGGAGGGGAC	gtgcTCTtGTCGATGGTCAGCACAGCtttATGCACGGCCTGGAGGG	aaaaCCCTCCAGGCCGTGCATaaaGCTGTGCTGACCATCGACaAGA
GL542	PAM_disrupted_silent_mutation	8	28	caccGTCCCCTCCAGGCCGTGCATAgtttt	ctctaaaacTATGCACGGCCTGGAGGGGAC	gtgcTTCTtGTCGATGGTCAGCACAGCtttATGCACGGCC	aaaaGGCCGTGCATaaaGCTGTGCTGACCATCGACaAGAA
GL542	PAM_disrupted_silent_mutation	10	28	caccGTCCCCTCCAGGCCGTGCATAgtttt	ctctaaaacTATGCACGGCCTGGAGGGGAC	gtgcTTCTtGTCGATGGTCAGCACAGCtttATGCACGGCCTG	aaaaCAGGCCGTGCATaaaGCTGTGCTGACCATCGACaAGAA
GL542	PAM_disrupted_silent_mutation	13	28	caccGTCCCCTCCAGGCCGTGCATAgtttt	ctctaaaacTATGCACGGCCTGGAGGGGAC	gtgcTTCTtGTCGATGGTCAGCACAGCtttATGCACGGCCTGGAG	aaaaCTCCAGGCCGTGCATaaaGCTGTGCTGACCATCGACaAGAA
GL542	PAM_disrupted_silent_mutation	14	28	caccGTCCCCTCCAGGCCGTGCATAgtttt	ctctaaaacTATGCACGGCCTGGAGGGGAC	gtgcTTCTtGTCGATGGTCAGCACAGCtttATGCACGGCCTGGAGG	aaaaCCTCCAGGCCGTGCATaaaGCTGTGCTGACCATCGACaAGAA
GL542	PAM_disrupted_silent_mutation	15	28	caccGTCCCCTCCAGGCCGTGCATAgtttt	ctctaaaacTATGCACGGCCTGGAGGGGAC	gtgcTTCTtGTCGATGGTCAGCACAGCtttATGCACGGCCTGGAGGG	aaaaCCCTCCAGGCCGTGCATaaaGCTGTGCTGACCATCGACaAGAA

Oligonucleotides used for cloning.

Table 27. Primedesign output for A1AT E342K nickRNAs.

pegRNA	nickRNA	Spacer_sequence	PAM_sequence	Annotation	ngRNA to pegRNA_distance	Spacer_sequence_order_TOP	Spacer_sequence_order_BOTTOM
GL544	n1	GCCTCGAGCAAGGCTCACGT	GGA	PE3	77	caccGCCTCGAGCAAGGCTCACGT	aaacACGTGAGCCTTGCTCGAGGC
GL544	n2	GTAAGGCTGATCCCAGGCCT	CGA	PE3	61	caccGTAAGGCTGATCCCAGGCCT	aaacAGGCCTGGGATCAGCCTTAC
GL544	n3	GCAGAGACACGTTGTAAGGC	TGA	PE3	48	caccGCAGAGACACGTTGTAAGGC	aaacGCCTTACAACGTGTCTCTGC
GL544	n4	CCTGGAGGGGAGAGAAGCAG	AGA	PE3	32	caccGCCTGGAGGGGAGAGAAGCAG	aaacCTGCTTCTCTCCCCTCCAGGC
GL544	n5	GGCCTGGAGGGGAGAGAAGC	AGA	PE3	30	caccGGCCTGGAGGGGAGAGAAGC	aaacGCTTCTCTCCCCTCCAGGCC
GL544	n6	ATGCACGGCCTGGAGGGGAG	AGA	PE3	24	caccGATGCACGGCCTGGAGGGGAG	aaacCTCCCCTCCAGGCCGTGCATC
GL544	n7	TTATGCACGGCCTGGAGGGG	AGA	PE3	22	caccGTTATGCACGGCCTGGAGGGG	aaacCCCCTCCAGGCCGTGCATAAC
GL544	n8	CCTTATGCACGGCCTGGAGG	GGA	PE3	20	caccGCCTTATGCACGGCCTGGAGG	aaacCTCCAGGCCGTGCATAAGGC
GL544	n9	CACAGCCTTATGCACGGCCT	GGA	PE3	15	caccGCACAGCCTTATGCACGGCCT	aaacAGGCCGTGCATAAGGCTGTGC
GL544	n10	AGCTTCAGTCCCTTTCTcGT	CGA	PE3b seed	15	caccGAGCTTCAGTCCCTTTCTcGT	aaacACgAGAAAGGGACTGAAGCTC
GL544	n11	AACTTGACCTCGGGGGGGAT	AGA	PE3	71	caccGAACTTGACCTCGGGGGGGAT	aaacATCCCCCCCGAGGTCAAGTTC
GL544	n12	GTTGAACTTGACCTCGGGGG	GGA	PE3	75	caccGTTGAACTTGACCTCGGGGG	aaacCCCCCGAGGTCAAGTTCAAC
GL544	n13	GACAAAGGGTTTGTTGAACT	TGA	PE3	87	caccGACAAAGGGTTTGTTGAACT	aaacAGTTCAACAAACCCTTTGTC
GL544	n14	TAAGAAGACAAAGGGTTTGT	TGA	PE3	93	caccGTAAGAAGACAAAGGGTTTGT	aaacACAAACCCTTTGTCTTCTTAC
GL544	n15	ATTTTGTTCAATCATTAAGA	AGA	PE3	108	caccGATTTTGTTCAATCATTAAGA	aaacTCTTAATGATTGAACAAAATC
GL544	n16	GGTATTTTGTTCAATCATTA	AGA	PE3	111	caccGGTATTTTGTTCAATCATTA	aaacTAATGATTGAACAAAATACC
GL556	n1	GTGTCCACGTGAGCCTTGCT	CGA	PE3	84	caccGTGTCCACGTGAGCCTTGCT	aaacAGCAAGGCTCACGTGGACAC
GL556	n2	TGAGCCTTGCTCGAGGCCTG	GGA	PE3	75	caccGTGAGCCTTGCTCGAGGCCTG	aaacCAGGCCTCGAGCAAGGCTCAC
GL556	n3	GGCCGTGCATAAGGCTGTGC	TGA	PE3	15	caccGGCCGTGCATAAGGCTGTGC	aaacGCACAGCCTTATGCACGGCC
GL556	n4	CATAAGGCTGTGCTGACCAT	CGA	PE3	8	caccGCATAAGGCTGTGCTGACCAT	aaacATGGTCAGCACAGCCTTATGC
GL556	n5	GGCTGTGCTGACCATCGACg	AGA	PE3b seed	3	caccGGCTGTGCTGACCATCGACg	aaacCGTCGATGGTCAGCACAGCC
GL556	n6	GCTGACCATCGACgAGAAAG	GGA	PE3b seed	3	caccGCTGACCATCGACgAGAAAG	aaacCTTTCTcGTCGATGGTCAGC
GL556	n7	ACCATCGACgAGAAAGGGAC	TGA	PE3b nonseed	7	caccGACCATCGACgAGAAAGGGAC	aaacGTCCCTTTCTcGTCGATGGTC
GL556	n8	GCTGCTGGGGCCATGTTTTT	AGA	PE3	31	caccGCTGCTGGGGCCATGTTTTT	aaacAAAAACATGGCCCCAGCAGC
GL556	n9	ATACCCATGTCTATCCCCCC	CGA	PE3	58	caccGATACCCATGTCTATCCCCCC	aaacGGGGGGATAGACATGGGTATC
GL556	n10	CAAACCCTTTGTCTTCTTAA	TGA	PE3	93	caccGCAAACCCTTTGTCTTCTTAA	aaacTTAAGAAGACAAAGGGTTTGC
GL556	n11	CCCTTTGTCTTCTTAATGAT	TGA	PE3	97	caccGCCCTTTGTCTTCTTAATGAT	aaacATCATTAAGAAGACAAAGGGC
GL542	n1	GGCCTCGAGCAAGGCTCACG	TGG	PE3	67	caccGGCCTCGAGCAAGGCTCACG	aaacCGTGAGCCTTGCTCGAGGCC
GL542	n2	CTGATCCCAGGCCTCGAGCA	AGG	PE3	58	caccGCTGATCCCAGGCCTCGAGCA	aaacTGCTCGAGGCCTGGGATCAGC
GL542	n3	CACGTTGTAAGGCTGATCCC	AGG	PE3	46	caccGCACGTTGTAAGGCTGATCCC	aaacGGGATCAGCCTTACAACGTGC
GL542	n4	AGAAGCAGAGACACGTTGTA	AGG	PE3	35	caccGAGAAGCAGAGACACGTTGTA	aaacTACAACGTGTCTCTGCTTCTC
GL542	n5	GCtttATGCACGGCCTGGAG	GGG	PE3b nonseed	10	caccGCtttATGCACGGCCTGGAG	aaacCTCCAGGCCGTGCATaaaGC
GL542	n6	AGCtttATGCACGGCCTGGA	GGG	PE3b nonseed	9	caccGAGCtttATGCACGGCCTGGA	aaacTCCAGGCCGTGCATaaaGCTC
GL542	n7	CAGCtttATGCACGGCCTGG	AGG	PE3b nonseed	8	caccGCAGCtttATGCACGGCCTGG	aaacCCAGGCCGTGCATaaaGCTGC
GL542	n8	GCACAGCtttATGCACGGCC	TGG	PE3b nonseed	5	caccGCACAGCtttATGCACGGCC	aaacGGCCGTGCATaaaGCTGTGC

Oligonucleotide sequences used for cloning nickRNAs into pU6 expression plasmids.

Table 28. Full sequences of A1AT E342K nickRNAs.

nickRNA	Spacer_sequence	Compatible with Lentivirus cassette?
GL544n1	GCCTCGAGCAAGGCTCACGT	No
GL544n2	GTAAGGCTGATCCCAGGCCT	No
GL544n3	GCAGAGACACGTTGTAAGGC	No
GL544n4	CCTGGAGGGGAGAGAAGCAG	No
GL544n5	GGCCTGGAGGGGAGAGAAGC	No
GL544n6	ATGCACGGCCTGGAGGGGAG	No
GL544n7	TTATGCACGGCCTGGAGGGG	No
GL544n8	CCTTATGCACGGCCTGGAGG	No
GL544n9	CACAGCCTTATGCACGGCCT	No
GL544n10	AGCTTCAGTCCCTTTCTcGT	Yes
GL544n11	AACTTGACCTCGGGGGGGAT	Yes
GL544n12	GTTGAACTTGACCTCGGGGG	Yes
GL544n13	GACAAAGGGTTTGTTGAACT	Yes
GL544n14	TAAGAAGACAAAGGGTTTGT	Yes
GL544n15	ATTTTGTTCAATCATTAAGA	Yes
GL544n16	GGTATTTTGTTCAATCATTA	Yes
GL556n1	GTGTCCACGTGAGCCTTGCT	No
GL556n2	TGAGCCTTGCTCGAGGCCTG	No
GL556n3	GGCCGTGCATAAGGCTGTGC	No
GL556n4	CATAAGGCTGTGCTGACCAT	Yes
GL556n5	GGCTGTGCTGACCATCGACg	Yes
GL556n6	GCTGACCATCGACgAGAAAG	Yes
GL556n7	ACCATCGACgAGAAAGGGAC	Yes
GL556n8	GCTGCTGGGGCCATGTTTTT	Yes
GL556n9	ATACCCATGTCTATCCCCCC	Yes
GL556n10	CAAACCCTTTGTCTTCTTAA	Yes
GL556n11	CCCTTTGTCTTCTTAATGAT	Yes
GL542n1	GGCCTCGAGCAAGGCTCACG	pegRNA not compatible
GL542n2	CTGATCCCAGGCCTCGAGCA	pegRNA not compatible
GL542n3	CACGTTGTAAGGCTGATCCC	pegRNA not compatible
GL542n4	AGAAGCAGAGACACGTTGTA	pegRNA not compatible
GL542n5	GCtttATGCACGGCCTGGAG	pegRNA not compatible
GL542n6	AGCtttATGCACGGCCTGGA	pegRNA not compatible
GL542n7	CAGCtttATGCACGGCCTGG	pegRNA not compatible
GL542n8	GCACAGCtttATGCACGGCC	pegRNA not compatible

Lists compatibility with the pLenti A1AT E342K cassette (Table 17) described in the mammalian cell culture methods section.

Table 29. Sequences and information for synthetic peg/nickRNAs.

Component	Designation	Target Edit	Sequence
pegRNA	BE-Ag-021	HBB E6V 22T	mC*mA*mU*GGUGCACCUGACUCCUGGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAG UGGCACCGAGUCGGUGCAGACUUCUCCACAGGAGUCAGGUGCA CmU*mU*mU*U
nickRNA	BE-Ag-010	HBB E6V 22T	mC*mC*mU*UGAUACCAACCUGCCCAGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAG UGGCACCGAGUCGGUGCmU*mU*mU*U
pegRNA	sgRNA_254	FANCF +6G>C	mG*mG*mA*AUCCCUUCUGCAGCACCGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAG UGGCACCGAGUCGGUGCGGAAAAGCGAUGCAGGUGCUGCAGA AGGGAmU*mU*mU*U
nickRNA	sgRNA_255	FANCF +6G>C	mG*mG*mC*GGGGUCCCAGGUGCUGACGUGUUUUAGAGCUAGA AAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAA AAGUGGCACCGAGUCGGUGCmU*mU*mU*U
pegRNA	sgRNA_256	HEK3 +1CTTin s	mG*mG*mC*CCAGACUGAGCACGUGAGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAG UGGCACCGAGUCGGUGCGCCAUCAAAGCGUGCUCAGUCUGmU* mU*mU*U
nickRNA	sgRNA_257	HEK3 +1CTTin s	mC*mU*mU*GUCAACCAGUAUCCCGGUGCGUUUUAGAGCUAGA AAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAA AAGUGGCACCGAGUCGGUGCmU*mU*mU*U
pegRNA	GL544-13-20	A1AT E342K 7A>G	mG*mG*mC*CGUGCAUAAGGCUGUGCGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAG UGGCACCGAGUCGGUGCUUUCUCGUCGAUGGUGAGCACAGCCU UAUGCACmU*mU*mU*U
nickRNA	GL544n11	A1AT E342K 7A>G	mG*mA*mA*CUUGACCUCGGGGGGGAUGUUUUAGAGCUAGAAA UAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAA GUGGCACCGAGUCGGUGCmU*mU*mU*U
nickRNA	GL544n12	A1AT E342K 7A>G	mG*mU*mU*GAACUUGACCUCGGGGGGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAG UGGCACCGAGUCGGUGCmU*mU*mU*U
nickRNA	GL544n13	A1AT E342K 7A>G	mG*mA*mC*AAAGGGUUUGUUGAACUGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAG UGGCACCGAGUCGGUGCmU*mU*mU*U
pegRNA	GL556-14-20	A1AT E342K 7A>G	mA*mG*mC*UUCAGUCCCUUUCUUGUGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAG UGGCACCGAGUCGGUGCGGCUGUGCUGACCAUAGACGAGAAA GGGACUGAAmU*mU*mU*U
nickRNA	GL556n5	A1AT E342K 7A>G	mG*mG*mC*UGUGCUGACCAUCGACaGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAG UGGCACCGAGUCGGUGCmU*mU*mU*U
nickRNA	GL556n6	A1AT E342K 7A>G	mG*mC*mU*GACCAUCGACAAGAAAGGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAG UGGCACCGAGUCGGUGCmU*mU*mU*U
nickRNA	GL556n11	A1AT E342K 7A>G	mG*mC*mC*CUUUGUCUUCUUAAUGAUGUUUUAGAGCUAGAAA UAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAA GUGGCACCGAGUCGGUGCmU*mU*mU*U
			NOTE: A preceding "m" represents a 2'-O-methyl modification to a base. A trailing "*" represents a 3'-phosphorothioate modification to a base.

Here, "m" preceding a base represents a 2'-O-methyl modification and "" trailing a base represents a 3'-phosphorothioate modification.*

Table 30. List of differentially regulated genes between Hek239T/U2OS/HeLa.

ensemble_id	gene_symbol	description	logFC_u	logCPM_u	F_u	PValue_u	FDR_u	logFC_h	logCPM_h	F_h	PValue_h	FDR_h	subcellular location	siRNA plate	siRNA well
ENSG00000005007.12	UPF1	UPF1 RNA helicase and ATPase [Source:HGNC Symbol;Acc:HGNC:9962]	1.31936893	7.07093352	22.1946069	0.00429584	0.03016624	1.31348631	7.07093352	36.1782615	0.0013769	0.0130503	Nucleoplasm	3	3
ENSG00000005156.12	LIG3	DNA ligase 3 [Source:HGNC Symbol;Acc:HGNC:6600]						1.21935498	5.58662456	28.8720554	0.00235041	0.01827605	Nucleoplasm	1	1
ENSG00000010072.16	SPRTN	SprT-like N-terminal domain [Source:HGNC Symbol;Acc:HGNC:25356]						1.44870564	4.19861783	27.2679648	0.00268489	0.02012461	Nucleoplasm		
ENSG00000015475.18	BID	BH3 interacting domain death agonist [Source:HGNC Symbol;Acc:HGNC:1050]						1.02479037	6.00540197	17.8342987	0.00695574	0.03404219	Cytosol		
ENSG00000030110.13	BAK1	BCL2 antagonist/killer 1 [Source:HGNC Symbol;Acc:HGNC:949]						1.50039024	4.79971504	29.923981	0.00216142	0.01729133	not indicated		
ENSG00000052795.13	FNIP2	folliculin interacting protein 2 [Source:HGNC Symbol;Acc:HGNC:29280]						1.93990881	2.94478666	36.7867717	0.00132277	0.01292619	Cytosol		
ENSG00000062822.14	POLD1	DNA polymerase delta 1, catalytic subunit [Source:HGNC Symbol;Acc:HGNC:9175]	1.33621129	8.1648933	17.4119134	0.00732419	0.03956966	1.51817044	8.1648933	36.6335628	0.00133612	0.01292619	Nucleoplasm	1	1
ENSG00000064545.15	TMEM161A	transmembrane protein 161A [Source:HGNC Symbol;Acc:HGNC:26020]	3.89058424	6.17114239	175.730751	2.46E-05	0.00331642	2.32104779	6.17114239	131.673653	5.25E-05	0.00244782	Cytosol		
ENSG00000064655.19	EYA2	EYA transcriptional coactivator and phosphatase 2 [Source:HGNC Symbol;Acc:HGNC:3520]						7.3471773	2.54781489	79.9389299	0.00019129	0.00427024	Cytosol		
ENSG00000070010.19	UFD1	ubiquitin recognition factor in ER associated degradation 1 [Source:HGNC Symbol;Acc:HGNC:12520]						0.74828938	6.32625182	15.1939443	0.00977476	0.04348982	Cytosol;Nucleoplasm		
ENSG00000072133.11	RPS6KA6	ribosomal protein S6 kinase A6 [Source:HGNC Symbol;Acc:HGNC:10435]						11.4131378	2.53969046	270.328419	3.94E-06	0.00223273	Mitochondria;Nucleoplasm		
ENSG00000072501.18	SMC1A	structural maintenance of chromosomes 1A [Source:HGNC Symbol;Acc:HGNC:11111]	1.38849715	8.43486458	21.1419633	0.00478908	0.03112903						Nucleoplasm	3	3
ENSG00000076248.10	UNG	uracil DNA glycosylase [Source:HGNC Symbol;Acc:HGNC:12572]	1.57696717	6.65187028	15.2875838	0.00965011	0.04667962	1.79179876	6.65187028	32.508672	0.00177751	0.01540505	Cytosol	1	1
ENSG00000078900.15	TP73	tumor protein p73 [Source:HGNC Symbol;Acc:HGNC:12003]						1.54493258	4.90397688	43.2908965	0.00089093	0.00950326	Nucleoplasm	3	3
ENSG00000081377.17	CDC14B	cell division cycle 14B [Source:HGNC Symbol;Acc:HGNC:1719]						1.81173275	4.54697933	56.0271232	0.00047052	0.00711775	Nucleoplasm		
ENSG00000087088.20	BAX	BCL2 associated X, apoptosis regulator [Source:HGNC Symbol;Acc:HGNC:959]						0.90405713	6.16599128	14.5976214	0.01062281	0.04555762	not indicated		
ENSG00000088038.19	CNOT3	CCR4-NOT transcription complex subunit 3 [Source:HGNC Symbol;Acc:HGNC:7879]	2.07566909	7.69429425	39.5339433	0.0011115	0.01716871	2.2408965	7.69429425	77.9476397	0.00020404	0.00427024	Cytosol		
ENSG00000089163.4	SIRT4	sirtuin 4 [Source:HGNC Symbol;Acc:HGNC:14932]						3.25103995	-1.1617573	16.7960114	0.00791112	0.03739031	not indicated		
ENSG00000093009.10	CDC45	cell division cycle 45 [Source:HGNC Symbol;Acc:HGNC:1739]						1.45303588	7.07756622	57.6830244	0.00043742	0.0068667	Nucleoplasm		
ENSG00000096401.8	CDC5L	cell division cycle 5 like [Source:HGNC Symbol;Acc:HGNC:1743]						1.46118597	4.92861927	13.7549634	0.01200303	0.0491947	Nucleoplasm		
ENSG00000099956.19	SMARCB1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1 [Source:HGNC Symbol;Acc:HGNC:11103]						2.15335678	7.57424549	98.8526045	0.00011074	0.0031771	Nucleoplasm		
ENSG00000100109.17	TFIP11	tuftelin interacting protein 11 [Source:HGNC Symbol;Acc:HGNC:17165]						1.18265018	6.48195677	16.9919378	0.00771762	0.03697292	Nucleoplasm		
ENSG00000100142.15	POLR2F	RNA polymerase II subunit F [Source:HGNC Symbol;Acc:HGNC:9193]						1.31675159	6.52250097	43.79096	0.00086622	0.00935971	Nucleoli fibrillar center;Nucleoplasm		
ENSG00000100325.15	ASCC2	activating signal cointegrator 1 complex subunit 2 [Source:HGNC Symbol;Acc:HGNC:24103]	2.46407463	6.26181363	60.5731338	0.00038688	0.00919662	1.10726924	6.26181363	23.2761976	0.00385877	0.02447633	Cytosol;Focal adhesion sites		
ENSG00000102081.14	FMR1	FMRP translational regulator 1 [Source:HGNC Symbol;Acc:HGNC:3775]	1.08379253	7.59191092	16.625616	0.00808485	0.0413123						Cytosol		
ENSG00000102882.12	MAPK3	mitogen-activated protein kinase 3 [Source:HGNC Symbol;Acc:HGNC:6877]	1.80687338	5.96180938	15.5530326	0.00930842	0.04529009	1.36808308	5.96180938	15.4764499	0.00940527	0.04304147	Nucleoplasm		
ENSG00000103266.11	STUB1	STIP1 homology and U-box containing protein 1 [Source:HGNC Symbol;Acc:HGNC:11427]	1.72204066	7.32853282	27.7242159	0.00258345	0.02415089	1.18880018	7.32853282	22.8810096	0.00401108	0.0250142	Cytosol;Nucleoplasm		
ENSG00000104313.20	EYA1	EYA transcriptional coactivator and phosphatase 1 [Source:HGNC Symbol;Acc:HGNC:3519]	2.58715559	4.43351224	41.9005334	0.00096491	0.01681605	1.64056555	4.43351224	31.8160049	0.00187057	0.0155866	Nuclear bodies;Nucleoplasm	3	3
ENSG00000104419.15	NDRG1	N-myc downstream regulated 1 [Source:HGNC Symbol;Acc:HGNC:7679]	2.70489897	8.71893367	51.5137283	0.00058015	0.01237645	2.50607285	8.71893367	79.039344	0.00019691	0.00427024	Cytosol		
ENSG00000105327.17	BBC3	BCL2 binding component 3 [Source:HGNC Symbol;Acc:HGNC:17868]						2.12058522	3.00463571	21.8773753	0.00443684	0.02630512	Cytosol		
ENSG00000105373.19	NOP53	NOP53 ribosome biogenesis factor [Source:HGNC Symbol;Acc:HGNC:4333]						0.89334608	6.49279936	21.8978382	0.00442756	0.02630512	Nucleoli		
ENSG00000105393.16	BABAM1	BRISC and BRCA1 A complex member 1 [Source:HGNC Symbol;Acc:HGNC:25008]	1.32961899	6.65055817	23.6566193	0.00371955	0.02813334						Nuclear bodies		
ENSG00000107562.16	CXCL12	C-X-C motif chemokine ligand 12 [Source:HGNC Symbol;Acc:HGNC:10672]						8.39897787	-0.6305059	60.5260447	0.00097014	0.0101978	not indicated		
ENSG00000107672.15	NSMCE4A	NSE4 homolog A, SMC5-SMC6 complex component [Source:HGNC Symbol;Acc:HGNC:25935]	1.40450039	5.06414978	24.9177521	0.00330437	0.02749238	1.81092903	5.06414978	66.7569567	0.00030269	0.00535827	Nucleoplasm		
ENSG00000107949.17	BCCIP	BRCA2 and CDKN1A interacting protein [Source:HGNC Symbol;Acc:HGNC:978]						0.75056464	5.95864056	13.86356	0.01181182	0.04872206	Nucleoplasm		
ENSG00000108984.15	MAP2K6	mitogen-activated protein kinase kinase 6 [Source:HGNC Symbol;Acc:HGNC:6846]	2.77165126	3.41742606	19.9803888	0.00542815	0.03367095						Cytosol;Nucleoplasm		
ENSG00000111358.14	GTF2H3	general transcription factor IIH subunit 3 [Source:HGNC Symbol;Acc:HGNC:4657]						1.27903236	5.40922699	25.0965793	0.00325072	0.02235206	Nucleoplasm	3	3
ENSG00000111877.17	MCM9	minichromosome maintenance 9 homologous recombination repair factor [Source:HGNC Symbol;Acc:HGNC:21484]	1.30123106	3.75661921	14.3695235	0.01097395	0.0493631						Nucleoplasm		
ENSG00000112029.10	FBXO5	F-box protein 5 [Source:HGNC Symbol;Acc:HGNC:13584]						1.10184705	5.04684581	14.6244334	0.01058255	0.04555762	Nucleoplasm		
ENSG00000112130.17	RNF8	ring finger protein 8 [Source:HGNC Symbol;Acc:HGNC:10071]						0.94917735	5.20089221	18.3643593	0.00652841	0.03269395	Nucleoplasm	3	3
ENSG00000112319.19	EYA4	EYA transcriptional coactivator and phosphatase 4 [Source:HGNC Symbol;Acc:HGNC:3522]	13.8667285	6.54145561	376.379578	3.24E-06	0.00132221						Nucleoplasm		
ENSG00000114126.17	TFDP2	transcription factor Dp-2 [Source:HGNC Symbol;Acc:HGNC:11751]	1.07155829	5.64899987	16.1764961	0.0085686	0.04243499						Nucleoplasm		
ENSG00000114742.14	WDR48	WD repeat domain 48 [Source:HGNC Symbol;Acc:HGNC:30914]	1.33543188	5.9847078	21.9536815	0.00440235	0.03027073						Vesicles		
ENSG00000115274.15	INO80B	INO80 complex subunit B [Source:HGNC Symbol;Acc:HGNC:13324]						1.25823688	3.6245402	21.9988441	0.00438211	0.02630512	Nucleoli;Nucleoplasm		
ENSG00000116750.13	UCHL5	ubiquitin C-terminal hydrolase L5 [Source:HGNC Symbol;Acc:HGNC:19678]	1.53424885	6.14098229	35.7889829	0.00141314	0.01808815						Mitochondria;Nucleoplasm		
ENSG00000118046.16	STK11	serine/threonine kinase 11 [Source:HGNC Symbol;Acc:HGNC:11389]						5.60357791	4.81899736	49.2590236	0.00064827	0.00829784	Nucleoplasm		
ENSG00000118689.15	FOXO3	forkhead box O3 [Source:HGNC Symbol;Acc:HGNC:3821]						1.79782491	5.25774593	47.9567961	0.00069268	0.00856619	Nucleoplasm		
ENSG00000122547.11	EEPD1	endonuclease/exonuclease/phosphatase family domain containing 1 [Source:HGNC Symbol;Acc:HGNC:22223]						2.98299295	1.70181445	53.8003376	0.00052069	0.00721322	Nuclear speckles		
ENSG00000122691.13	TWIST1	twist family bHLH transcription factor 1 [Source:HGNC Symbol;Acc:HGNC:12428]	9.42926955	0.89451217	70.617419	0.00026245	0.00752972						not indicated		
ENSG00000124535.16	WRNIP1	WRN helicase interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:20876]	1.20573653	6.15066706	21.8635113	0.00444315	0.03030083	1.17683618	6.15066706	34.0629047	0.00159057	0.01422959	Nucleoplasm		
ENSG00000124766.7	SOX4	SRY-box transcription factor 4 [Source:HGNC Symbol;Acc:HGNC:11200]	2.21768953	4.60758063	43.1816541	0.00089646	0.01681605	2.75023587	4.60758063	109.375663	8.52E-05	0.00287894	Nucleoplasm		
ENSG00000124802.12	EEF1E1	eukaryotic translation elongation factor 1 epsilon 1 [Source:HGNC Symbol;Acc:HGNC:3212]						1.03042123	4.66207449	16.9768228	0.00773232	0.03697292	Cytosol;Nucleoplasm		
ENSG00000125901.6	MRPS26	mitochondrial ribosomal protein S26 [Source:HGNC Symbol;Acc:HGNC:14045]	1.7144326	3.4516353	19.9637134	0.00543815	0.03367095	1.46417443	3.4516353	24.9421294	0.00329699	0.02248438	Mitochondria		
ENSG00000126453.9	BCL2L12	BCL2 like 12 [Source:HGNC Symbol;Acc:HGNC:13787]						1.24271285	5.58764577	19.7631104	0.00556051	0.02913118	Vesicles		
ENSG00000126457.21	PRMT1	protein arginine methyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:5187]						2.13908931	8.66836375	53.1178928	0.00053752	0.00721322	Nucleoplasm		
ENSG00000127423.10	AUNIP	aurora kinase A and ninein interacting protein [Source:HGNC Symbol;Acc:HGNC:28363]						1.15790719	2.74206974	15.3843515	0.00952359	0.04304147	Centrosome		
ENSG00000128886.12	ELL3	elongation factor for RNA polymerase II 3 [Source:HGNC Symbol;Acc:HGNC:23113]	4.21695546	2.69519956	89.0053857	0.00014518	0.00603959	1.63992628	2.69519956	33.2974705	0.00167908	0.01478987	Nucleoli		
ENSG00000129173.13	E2F8	E2F transcription factor 8 [Source:HGNC Symbol;Acc:HGNC:24727]						1.72977674	3.88124834	45.9375328	0.00077019	0.00889994	Nucleoli		
ENSG00000129245.12	FXR2	FMR1 autosomal homolog 2 [Source:HGNC Symbol;Acc:HGNC:4024]	1.41151663	4.51237982	20.9866599	0.00486834	0.03115736						Cytosol		
ENSG00000130413.15	STK33	serine/threonine kinase 33 [Source:HGNC Symbol;Acc:HGNC:14568]						2.2700091	4.42353397	44.155536	0.0008488	0.00929211	Nucleoplasm		
ENSG00000130726.12	TRIM28	tripartite motif containing 28 [Source:HGNC Symbol;Acc:HGNC:16384]	3.20786368	10.2261475	51.7471532	0.00057366	0.01237645	3.13403589	10.2261475	88.9838709	0.00014527	0.00377709	Nucleoplasm	2	2
ENSG00000130766.5	SESN2	sestrin 2 [Source:HGNC Symbol;Acc:HGNC:20746]						1.27995669	2.70567253	19.8274619	0.00552084	0.02913118	Cytosol		
ENSG00000131153.9	GINS2	GINS complex subunit 2 [Source:HGNC Symbol;Acc:HGNC:24575]						1.42258575	4.34459852	15.3998206	0.00950357	0.04304147	Nucleoplasm		
ENSG00000131778.19	CHD1L	chromodomain helicase DNA binding protein 1 like [Source:HGNC Symbol;Acc:HGNC:1916]						0.89990702	8.93871028	15.0531499	0.00996636	0.04387309	Nucleoplasm	2	2
ENSG00000133247.14	KMT5C	lysine methyltransferase 5C [Source:HGNC Symbol;Acc:HGNC:28405]	3.04043892	4.99841645	16.3279625	0.00840112	0.04210683						Nucleoplasm		
ENSG00000133863.8	TEX15	testis expressed 15, meiosis and synapsis associated [Source:HGNC Symbol;Acc:HGNC:11738]	12.1880487	2.88723491	165.678769	0.00011392	0.00592386	12.1880487	2.88723491	324.713327	2.64E-05	0.00219566	Cytosol;Nucleoplasm		
ENSG00000133895.15	MEN1	menin 1 [Source:HGNC Symbol;Acc:HGNC:7010]	1.16769865	7.74777615	14.6071191	0.01060853	0.04930891	1.08436573	7.74777615	19.9450457	0.00544697	0.02913118	Nucleoplasm	3	3
ENSG00000134215.16	VAV3	vav guanine nucleotide exchange factor 3 [Source:HGNC Symbol;Acc:HGNC:12659]	10.8922072	2.99981023	91.5742283	0.00013492	0.00603959						not indicated		
ENSG00000134758.14	RNF138	ring finger protein 138 [Source:HGNC Symbol;Acc:HGNC:17765]						1.62605488	4.43260261	14.1007189	0.01140853	0.04789912	Mitochondria		
ENSG00000134897.14	BIVM	basic, immunoglobulin-like variable motif containing [Source:HGNC Symbol;Acc:HGNC:16034]	1.72744525	6.17713756	32.1310173	0.00182744	0.02159692						Nucleoplasm		
ENSG00000135972.9	MRPS9	mitochondrial ribosomal protein S9 [Source:HGNC Symbol;Acc:HGNC:14501]						0.93070564	4.2856312	15.2512962	0.00969816	0.043381	Mitochondria		
ENSG00000136104.21	RNASEH2B	ribonuclease H2 subunit B [Source:HGNC Symbol;Acc:HGNC:25671]						1.14925819	7.02288439	21.8310625	0.00445796	0.02630512	Nucleoplasm		
ENSG00000136997.19	MYC	MYC proto-oncogene, bHLH transcription factor [Source:HGNC Symbol;Acc:HGNC:7553]	1.15521536	8.2261393	16.901837	0.00780579	0.0413123						Nucleoplasm		
ENSG00000137337.15	MDC1	mediator of DNA damage checkpoint 1 [Source:HGNC Symbol;Acc:HGNC:21163]						1.46018566	6.0945322	23.2112242	0.00388326	0.02447633	Nuclear bodies	2	2
ENSG00000137760.14	ALKBH8	alkB homolog 8, tRNA methyltransferase [Source:HGNC Symbol;Acc:HGNC:25189]						1.51522478	3.73723632	22.4427387	0.00418979	0.02582152	Nuclear bodies		
ENSG00000139372.15	TDG	thymine DNA glycosylase [Source:HGNC Symbol;Acc:HGNC:11700]	1.43012273	5.7680573	28.3056226	0.00246164	0.0238149	1.00473114	5.7680573	23.6162718	0.00373399	0.02427093	Nucleoplasm	2	2
ENSG00000140464.20	PML	promyelocytic leukemia [Source:HGNC Symbol;Acc:HGNC:9113]	1.29085691	6.30091014	23.7517678	0.00368581	0.02813334						Nuclear bodies		
ENSG00000141499.16	WRAP53	WD repeat containing antisense to TP53 [Source:HGNC Symbol;Acc:HGNC:25522]	1.57444193	5.5667427	32.0684651	0.0018359	0.02159692	0.96474668	5.5667427	20.7355275	0.00500035	0.02849514	Nuclear bodies		
ENSG00000141510.17	TP53	tumor protein p53 [Source:HGNC Symbol;Acc:HGNC:11998]	1.70537109	8.35064633	14.7000229	0.01047018	0.04921578						Nucleoplasm	1	1
ENSG00000141682.11	PMAIP1	phorbol-12-myristate-13-acetate-induced protein 1 [Source:HGNC Symbol;Acc:HGNC:9108]	1.38028799	5.38874399	24.1704372	0.00354222	0.02793433	2.71736134	5.38874399	138.744417	4.58E-05	0.00238131	Vesicles		
ENSG00000142453.11	CARM1	coactivator associated arginine methyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:23393]						1.05853203	5.96275955	16.0547497	0.00870656	0.04030402	Nucleoplasm		
ENSG00000143624.14	INTS3	integrator complex subunit 3 [Source:HGNC Symbol;Acc:HGNC:26153]	1.18660103	7.36858319	17.0396122	0.00767152	0.041068						Nucleoli		
ENSG00000143799.13	PARP1	poly(ADP-ribose) polymerase 1 [Source:HGNC Symbol;Acc:HGNC:270]						1.72649853	7.93837323	35.0342572	0.0014872	0.01359724	Nucleoplasm	3	3
ENSG00000144524.17	COPS7B	COP9 signalosome subunit 7B [Source:HGNC Symbol;Acc:HGNC:16760]						1.28018333	6.81542436	45.3442833	0.00079521	0.00891063	Nucleoplasm		
ENSG00000145736.14	GTF2H2	general transcription factor IIH subunit 2 [Source:HGNC Symbol;Acc:HGNC:4656]						0.94226496	6.02574765	19.4535768	0.00575682	0.0295659	Nuclear speckles	2	2
ENSG00000147231.14	RADX	RPA1 related single stranded DNA binding protein, X-linked [Source:HGNC Symbol;Acc:HGNC:25486]	3.81550549	5.06316858	55.9250874	0.00047267	0.01081275	9.89585771	5.06316858	318.269237	5.08E-06	0.00140878	not indicated		
ENSG00000149269.9	PAK1	p21 (RAC1) activated kinase 1 [Source:HGNC Symbol;Acc:HGNC:8590]						1.33062708	5.77369179	19.7592082	0.00556292	0.02913118	Cytosol;Plasma membrane		
ENSG00000149923.14	PPP4C	protein phosphatase 4 catalytic subunit [Source:HGNC Symbol;Acc:HGNC:9319]	1.40203986	7.2236936	28.3421983	0.00245424	0.0238149	1.41625967	7.2236936	47.7457459	0.00070012	0.00856619	Nucleoplasm		
ENSG00000149948.13	HMGA2	high mobility group AT-hook 2 [Source:HGNC Symbol;Acc:HGNC:5009]						7.24918981	4.86255986	174.850283	2.49E-05	0.00219566	Nucleoli;Nucleoplasm		
ENSG00000150907.10	FOXO1	forkhead box O1 [Source:HGNC Symbol;Acc:HGNC:3819]	2.42445021	2.40998864	25.0373281	0.00326836	0.02746746	2.83047156	2.40998864	56.9829847	0.00045101	0.00694896	Cytosol;Nucleoplasm		
ENSG00000152409.9	JMY	junction mediating and regulatory protein, p53 cofactor [Source:HGNC Symbol;Acc:HGNC:28916]	1.86858916	3.38341547	23.4117701	0.00380835	0.02854549						Nucleoplasm		
ENSG00000153094.23	BCL2L11	BCL2 like 11 [Source:HGNC Symbol;Acc:HGNC:994]	1.62418566	6.10518195	38.0102609	0.00122243	0.0172383	1.49861397	6.10518195	54.5599778	0.00050279	0.00721322	not indicated		
ENSG00000155229.21	MMS19	MMS19 homolog, cytosolic iron-sulfur assembly component [Source:HGNC Symbol;Acc:HGNC:13824]	1.42528802	7.11369779	28.8831361	0.0023483	0.02363958	0.85272463	7.11369779	17.5486532	0.00720204	0.0350415	Nucleoplasm	3	3
ENSG00000155816.20	FMN2	formin 2 [Source:HGNC Symbol;Acc:HGNC:14074]	4.8418456	1.75967316	77.2772099	0.0002086	0.00694211	2.91916325	1.75967316	70.2296823	0.00026615	0.00481388	Actin filaments;Plasma membrane		
ENSG00000156256.15	USP16	ubiquitin specific peptidase 16 [Source:HGNC Symbol;Acc:HGNC:12614]						0.83780126	6.35159163	14.0430235	0.01150488	0.04789912	Cytosol;Nucleoplasm		
ENSG00000157212.19	PAXIP1	PAX interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:8624]	3.45926662	5.31641393	61.3407058	0.00037481	0.00919662						Nucleoplasm;Vesicles		
ENSG00000159685.11	CHCHD6	coiled-coil-helix-coiled-coil-helix domain containing 6 [Source:HGNC Symbol;Acc:HGNC:28184]						1.30441904	3.38862358	16.045188	0.00871752	0.04030402	Mitochondria		
ENSG00000160117.16	ANKLE1	ankyrin repeat and LEM domain containing 1 [Source:HGNC Symbol;Acc:HGNC:26812]	3.5921338	3.46098793	21.0304459	0.00484581	0.03115736						Cytosol;Nucleoplasm		
ENSG00000160469.17	BRSK1	BR serine/threonine kinase 1 [Source:HGNC Symbol;Acc:HGNC:18994]						3.76590798	3.59213241	80.4546142	0.00018817	0.00427024	Nucleoplasm		
ENSG00000160803.8	UBQLN4	ubiquilin 4 [Source:HGNC Symbol;Acc:HGNC:1237]						0.9299878	4.67264553	16.89794	0.00780963	0.03712923	Nucleoplasm		
ENSG00000163071.11	SPATA18	spermatogenesis associated 18 [Source:HGNC Symbol;Acc:HGNC:29579]	4.29124076	-0.0569771	23.7630661	0.00634396	0.03665399	9.09877144	-0.0569771	72.5662205	0.00066374	0.00836714	Mitochondria		
ENSG00000163882.9	POLR2H	RNA polymerase II subunit H [Source:HGNC Symbol;Acc:HGNC:9195]	1.25436304	6.03631026	21.6350406	0.00454887	0.03046294	1.36327344	6.03631026	41.6250245	0.00098056	0.0101978	Nucleoplasm		
ENSG00000163932.15	PRKCD	protein kinase C delta [Source:HGNC Symbol;Acc:HGNC:9399]	1.74355683	5.22786418	24.5922219	0.0034052	0.02793433						Cytosol		
ENSG00000164305.19	CASP3	caspase 3 [Source:HGNC Symbol;Acc:HGNC:1504]						1.36393497	4.83880572	29.6069727	0.00221611	0.01739913	Mitochondria;Nucleoplasm		
ENSG00000166199.13	ALKBH3	alkB homolog 3, alpha-ketoglutaratedependent dioxygenase [Source:HGNC Symbol;Acc:HGNC:30141]	1.72008008	4.47660655	27.7627945	0.00257511	0.02415089	1.84325282	4.47660655	53.204927	0.00053534	0.00721322	Nucleoplasm	3	3
ENSG00000166313.19	APBB1	amyloid beta precursor protein binding family B member 1 [Source:HGNC Symbol;Acc:HGNC:581]						2.23755531	6.48219215	107.103165	9.00E-05	0.00287894	Plasma membrane		
ENSG00000166848.7	TERF2IP	TERF2 interacting protein [Source:HGNC Symbol;Acc:HGNC:19246]						1.18186444	4.42834852	22.1753045	0.00430424	0.02630512	Nuclear bodies;Nucleoplasm		
ENSG00000168101.14	NUDT16L1	nudix hydrolase 16 like 1 [Source:HGNC Symbol;Acc:HGNC:28154]						1.01159228	4.24478455	14.1748311	0.01128639	0.04789912	Cytosol;Plasma membrane		
ENSG00000168209.5	DDIT4	DNA damage inducible transcript 4 [Source:HGNC Symbol;Acc:HGNC:24944]	1.65601494	5.15730257	14.4418235	0.01086094	0.0493531						Cytosol		
ENSG00000169689.15	CENPX	centromere protein X [Source:HGNC Symbol;Acc:HGNC:11422]	0.92568375	6.45681973	14.4163714	0.01090054	0.0493531	1.13646827	6.45681973	34.6522095	0.00153731	0.01390261	not indicated		
ENSG00000169727.12	GPS1	G protein pathway suppressor 1 [Source:HGNC Symbol;Acc:HGNC:4549]	1.20930723	8.54537596	14.704203	0.01046402	0.04921578	1.48209124	8.54537596	35.5758219	0.00143353	0.01325221	Cytosol;Nucleoplasm		
ENSG00000172716.16	SLFN11	schlafen family member 11 [Source:HGNC Symbol;Acc:HGNC:26633]	8.44872063	5.65654326	399.395481	2.77E-06	0.00132221	6.65369382	5.65654326	597.221474	9.41E-07	0.00066427	Nucleoplasm		
ENSG00000174010.9	KLHL15	kelch like family member 15 [Source:HGNC Symbol;Acc:HGNC:29347]	2.62890865	3.89682134	41.1460435	0.00100861	0.0171257						Cytosol		
ENSG00000174371.17	EXO1	exonuclease 1 [Source:HGNC Symbol;Acc:HGNC:3511]						0.88094694	5.78394336	18.3205162	0.00656237	0.03269395	Nucleoplasm	1	1
ENSG00000175104.15	TRAF6	TNF receptor associated factor 6 [Source:HGNC Symbol;Acc:HGNC:12036]						1.20295949	3.50465887	19.3597581	0.0058182	0.02969781	Mitochondria		
ENSG00000178971.16	CTC1	CST telomere replication complex component 1 [Source:HGNC Symbol;Acc:HGNC:26169]	2.06798969	4.81576301	17.54276	0.00720725	0.03945019						not indicated		
ENSG00000179262.10	RAD23A	RAD23 homolog A, nucleotide excision repair protein [Source:HGNC Symbol;Acc:HGNC:9812]	1.33368642	7.67869548	17.4471478	0.00729245	0.03956966						Nucleoplasm	2	2
ENSG00000179456.10	ZBTB18	zinc finger and BTB domain containing 18 [Source:HGNC Symbol;Acc:HGNC:13030]						2.56828782	3.70949646	27.4229619	0.00264983	0.02004234	Nuclear speckles		
ENSG00000179965.12	ZNF771	zinc finger protein 771 [Source:HGNC Symbol;Acc:HGNC:29653]	2.01974503	2.28243969	25.1489645	0.00323523	0.02746643	3.81065811	2.28243969	124.062733	6.14E-05	0.00244782	Nucleoli;Nucleoplasm		
ENSG00000180336.18	MEIOC	meiosis specific with coiled-coil domain [Source:HGNC Symbol;Acc:HGNC:26670]	2.84333487	0.9972489	29.6590784	0.00220699	0.02324322	3.22596935	0.9972489	63.9429923	0.00033752	0.00573098	Cell Junctions		
ENSG00000181026.15	AEN	apoptosis enhancing nuclease [Source:HGNC Symbol;Acc:HGNC:25722]						1.10689935	5.72202844	20.000811	0.00541593	0.02913118	Nucleoplasm		
ENSG00000181218.5	H2AW	H2A.W histone [Source:HGNC Symbol;Acc:HGNC:20507]	10.9360066	1.79658803	118.829321	6.87E-05	0.00508486	3.25324979	1.79658803	79.1570249	0.00019616	0.00427024	not indicated		
ENSG00000181222.17	POLR2A	RNA polymerase II subunit A [Source:HGNC Symbol;Acc:HGNC:9187]	0.99788548	6.8422883	15.9097186	0.00887491	0.04343487						Nucleoplasm		
ENSG00000182979.18	MTA1	metastasis associated 1 [Source:HGNC Symbol;Acc:HGNC:7410]	2.10230783	6.96743632	34.0665112	0.00159017	0.01974653						Nucleoplasm		
ENSG00000183765.22	CHEK2	checkpoint kinase 2 [Source:HGNC Symbol;Acc:HGNC:16627]	1.16152531	6.14195601	18.0369316	0.00678797	0.03868213	0.8339538	6.14195601	15.3482531	0.00957062	0.04304147	Nucleoplasm	1	1
ENSG00000183955.13	KMT5A	lysine methyltransferase 5A [Source:HGNC Symbol;Acc:HGNC:29489]						0.87434846	6.14593878	14.2183844	0.01121544	0.04785255	Nucleoplasm		
ENSG00000184481.17	FOXO4	forkhead box O4 [Source:HGNC Symbol;Acc:HGNC:7139]	2.98625884	3.77280572	62.9925964	0.00035052	0.00911362						Nuclear speckles		
ENSG00000185115.6	NSMCE3	NSE3 homolog, SMC5-SMC6 complex component [Source:HGNC Symbol;Acc:HGNC:7677]						1.64177647	2.35141368	25.4591475	0.00314555	0.02199239	not indicated		
ENSG00000185507.21	IRF7	interferon regulatory factor 7 [Source:HGNC Symbol;Acc:HGNC:6122]	9.21092161	3.52194422	46.7214325	0.00073875	0.01463432						Nucleoplasm		
ENSG00000186280.7	KDM4D	lysine demethylase 4D [Source:HGNC Symbol;Acc:HGNC:25498]						2.61373779	0.68907643	40.1054288	0.00107318	0.0110233	Nucleoplasm		
ENSG00000186868.15	MAPT	microtubule associated protein tau [Source:HGNC Symbol;Acc:HGNC:6893]	2.84875021	3.20827632	38.4677804	0.00118758	0.01716871	3.82425529	3.20827632	109.790722	8.44E-05	0.00287894	Plasma membrane		
ENSG00000187778.14	MCRS1	microspherule protein 1 [Source:HGNC Symbol;Acc:HGNC:6960]	1.14175503	7.04178993	19.6712915	0.00561777	0.03436753	1.10975401	7.04178993	30.2687137	0.00210386	0.01699425	Nucleoplasm		
ENSG00000188486.3	H2AX	H2A.X variant histone [Source:HGNC Symbol;Acc:HGNC:4739]						3.00558698	5.31815676	176.7393	2.42E-05	0.00219566	not indicated		
ENSG00000188766.12	SPRED3	sprouty related EVH1 domain containing 3 [Source:HGNC Symbol;Acc:HGNC:31041]						3.22530595	0.88709195	20.8999049	0.0049134	0.02838851	Nucleoplasm		
ENSG00000196586.16	MYO6	myosin VI [Source:HGNC Symbol;Acc:HGNC:7605]						3.07072554	7.75008457	72.8520422	0.00024246	0.00458462	Nucleoplasm		
ENSG00000197061.4	H4C3	H4 clustered histone 3 [Source:HGNC Symbol;Acc:HGNC:4787]						3.7143336	-0.5879416	22.0478017	0.00436031	0.02630512	not indicated		
ENSG00000198369.10	SPRED2	sprouty related EVH1 domain containing 2 [Source:HGNC Symbol;Acc:HGNC:17722]						1.27856196	4.04952798	24.3992987	0.00346695	0.02326214	Nucleoplasm		
ENSG00000204590.12	GNL1	G protein nucleolar 1 (putative) [Source:HGNC Symbol;Acc:HGNC:4413]						0.85811536	5.02720608	15.1511763	0.00983242	0.04351369	Vesicles	2	2
ENSG00000204991.11	SPIRE2	spire type actin nucleation factor 2 [Source:HGNC Symbol;Acc:HGNC:30623]	2.18283701	3.33852027	16.2587344	0.00847711	0.04223325						Nucleoplasm		
ENSG00000214941.8	ZSWIM7	zinc finger SWIM-type containing 7 [Source:HGNC Symbol;Acc:HGNC:26993]	1.66255917	4.04552915	17.8889302	0.00690995	0.03880927						Cytosol		
ENSG00000221983.8	UBA52	ubiquitin A-52 residue ribosomal protein fusion product 1 [Source:HGNC Symbol;Acc:HGNC:12458]	1.89108523	9.84570519	19.921754	0.00546344	0.03367095						Cytosol;Endoplasmic reticulum		
ENSG00000240972.2	MIF	macrophage migration inhibitory factor [Source:HGNC Symbol;Acc:HGNC:7097]	2.19476675	7.19011577	50.77409	0.00060136	0.01250838	3.27839221	7.19011577	178.094272	2.37E-05	0.00219566	Cytosol		
ENSG00000254087.8	LYN	LYN proto-oncogene, Src family tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:6735]						2.02394896	4.27466819	18.8996203	0.0061326	0.03111174	Golgi apparatus;Plasma membrane		
ENSG00000270181.3	BIVM-ERCC5	BIVM-ERCC5 readthrough [Source:HGNC Symbol;Acc:HGNC:43690]	1.37410826	5.2309897	14.8007988	0.0103229	0.04907803						Nucleoplasm		
ENSG00000275700.5	AATF	apoptosis antagonizing transcription factor [Source:HGNC Symbol;Acc:HGNC:19235]						0.99813332	5.38063518	16.1148621	0.00863807	0.04030402	Nucleoli		
ENSG00000280789.2	PAGR1	PAXIP1 associated glutamate rich protein 1 [Source:HGNC Symbol;Acc:HGNC:28707]	1.77284652	3.83647504	26.7780324	0.00280001	0.02514327						not indicated		
ENSG00000283782.2	AC116366.3	novel protein	1.04516713	7.12773178	17.6990795	0.00707087	0.03896001						Nucleoplasm		

(Produced with the BeamTX bioinformatics team)

Table 31. Final list of genes which were selected for knockdown experimentation as siGENOME SMARTPool siRNAs

UNG	FANCA	ERCC8	MSH3	POLQ	MCM8	NSMCE4A	TEX15	ANKLE1	NSMCE3	REV1	PMS1	SOD1	RRM2B	DNMT1	EME1
FEN1	FANCB	APEX1	MSH4	RAD1	MDM4	BCCIP	VAV3	BRSK1	IRF7	REV3L	HMGB2	CSNK1D	GEN1	USP1	ALKBH3
RPA1	FANCC	APEX2	MSH5	RAD9A	POLD2	MCM9	RNF138	UBQLN4	KDM4D	RUVBL2	GADD45A	XAB2	TDP1	EYA1	SMC1A
RPA2	FANCD2	DDB1	MSH6	RAD17	POLD3	EYA4	BIVM	POLR2H	MCRS1	PRKCG	IGHMBP2	CCNO	GADD45G	RDM1	CDK7
RPA4	FANCE	DDB2	MUTYH	RAD21	POLD4	WDR48	RNASEH2B	PRKCD	SPIRE2	TCEA1	PMS2	GTF2H2	EYA3	CIB1	SETX
RPAIN	FANCF	ATR	NUDT1	RAD23A	POLN	STK11	MYC	APBB1	ZSWIM7	RTEL1	CSNK1E	YBX1	XPA	BTG2	UBE2V2
DNA2	FANCL	ATRX	OGG1	RAD23B	RAD18	FOXO3	ALKBH8	TERF2IP	UBA52	APTX	BRIP1	GTF2H1	UBE2A	TNP1	SMUG1
BRCA1	FANCM	ATRIP	PNKP	RAD52	RAD50	EEPD1	PML	NUDT16L1	LYN	GTF2H5	SMC3	MUS81	PRMT6	PALB2	DUT
BRCA2	XRCC1	DCLRE1A	POLA1	RAD54B	RAD9B	TWIST1	WRAP53	DDIT4	BIVM-ERCC5	UBE2B	TRIP13	UIMC1	RNF8	CNOT7	BLM
EXO1	XRCC2	DCLRE1B	POLB	RAD54L	SPRTN	WRNIP1	PMAIP1	CENPX	AATF	MDC1	TYMS	TRIM28	MMS19	CDKN2D	CLK2
LIG1	XRCC3	DCLRE1C	POLD1	RASSF7	FNIP2	SOX4	CARM1	GPS1	PAGR1	SIRT1	XPC	BRCC3	EME2	PARP1	NPM1
LIG3	XRCC4	NEIL1	POLE	MDM2	CDC14B	EEF1E1	INTS3	SLFN11	RASSF1	TREX1	HUS1	UVRAG	APLF	UPF1	ASF1A
LIG4	XRCC5	NEIL2	POLE2	FANCG	SIRT4	BCL2L12	COPS7B	KLHL15	RFC1	WRN	RPS27L	TREX2	MNAT1	NBN	H2AFX
CHEK1	XRCC6	NEIL3	POLG	BARD1	CDC45	PRMT1	PPP4C	CTC1	RFC2	DDX11	MAD2L2	MBD4	SMC6	DMC1	UBE2N
CHEK2	ERCC1	NTHL1	POLG2	CHD1L	SMARCB1	AUNIP	HMGA2	ZBTB18	RFC3	TDG	SETMAR	RNF168	CCNH	PCNA	RRM2
ATM	ERCC2	MGMT	POLH	CHD2	POLR2F	ELL3	FOXO1	AEN	RFC4	TOPBP1	PRKDC	PRPF19	RBBP8	BAZ1B	SMARCAD1
TP53	ERCC3	MLH1	POLI	DMAP1	ASCC2	E2F8	JMY	POLR2A	BRD4	ATF2	PARP2	ATXN3	ABL1	ALKBH1	SMARCAL1
TP53BP1	ERCC4	MLH3	POLK	FBXO18	NDRG1	SESN2	FMN2	MTA1	RECQL	VCP	TOP2A	CETN2	HMGB1	PARG	
RAD51	ERCC5	MPG	POLL	GNL1	BABAM1	GINS2	USP16	KMT5A	RECQL5	ALKBH2	PER1	TP73	CHAF1A	UBE2V1	
RAD51C	ERCC6	MSH2	POLM	MCM7	CXCL12	KMT5C	PAXIP1	FOXO4	RECQL4	GTF2H4	PARP3	MEN1	SPO11	GTF2H3	

Pooled from a combination of Thermofisher Silencer, Dharmacon siGENOME, and Beam's differential expression screen and trimmed using ontology relevance.

Table 32. MiSEQ target sites and associated amplicon PCR primer sequences.

Target Site Name	Target Site Sequence	Amplicon Primer 1	Amplicon Primer 1 Sequence	Amplicon Primer 2	Amplicon Primer 2 Sequence
HBB Genomic	CCATGGTG CACCTGAC TCCTGAGG	HR323	ACACTCTTTCCCTA CACGACGCTCTTCC GATCTNNNNAGGG TTGGCCAATCTACT CCC	HR324	TGGAGTTCAGA CGTGTGCTCTTC CGATCTGTCTTC TCTGTCTCCACA TGCC
A1AT E342 Genomic	ATCGACGA GAAAGGGA CTGAAGC	BEAM54	TGGAGTTCAGACG TGTGCTCTTCCGAT CTGGGTGGGATTC ACCACTTTTCCCAT G	BEAM1704	ACACTCTTTCCC TACACGACGCT CTTCCGATCTNN NNAGGTGTCCA CGTGAGCCTTG
A1AT E342K Lenti Cassette	ATCGACAA GAAAGGGA CTGAAGC	BEAM800	TGGAGTTCAGACG TGTGCTCTTCCGAT CTAGACCGAGGAG AGGGTTAGGGATA G	BEAM801	ACACTCTTTCCC TACACGACGCT CTTCCGATCTNN NNGTCACAGAG GAGGCACCCCT G
HEK2 Genomic	GAACACAA AGCATAGA CTGCGGG	HR043	ACACTCTTTCCCTA CACGACGCTCTTCC GATCTNNNNTGAA TGGATTCCTTGGAA ACAATG	HR044	TGGAGTTCAGA CGTGTGCTCTTC CGATCTCCAGCC CCATCTGTCAAA CT
FANCF Genomic	GGAATCCC TTCTGCAG CACCTGG	ST023_Fwd	ACACTCTTTCCCTA CACGACGCTCTTCC GATCTNNNNATAG CATTGCAGAGAGG CGT	ST023_Rev	TGGAGTTCAGA CGTGTGCTCTTC CGATCTGACCA AAGCGCCGATG GAT
HEK3 Genomic	GGCCCAGA CTGAGCAC GTGATGGC AGAGGAAA GGAAGCCC	HR459	ACACTCTTTCCCTA CACGACGCTCTTCC GATCTNNNNATGT GGGCTGCCTAGAA AGG	HR460	TGGAGTTCAGA CGTGTGCTCTTC CGATCTCCCAGC CAAACTTGTCA ACC
GL544 pegRNA	GACCATCG ACAAGAAA GGGACTGA	Same as A1AT E342K Lenti Cassette	Same as A1AT E342K Lenti Cassette	Same as A1AT E342K Lenti Cassette	Same as A1AT E342K Lenti Cassette
GL556 pegRNA	AGCTTCAG TCCCTTTCT TGT	Same as A1AT E342K Lenti Cassette	Same as A1AT E342K Lenti Cassette	Same as A1AT E342K Lenti Cassette	Same as A1AT E342K Lenti Cassette

The spacer sequences define the numbering of the bases in the editing frequency tables.

Table 33. MiSEQ amplicon sequences.

Target Site Name	Amplicon Sequence
HBB Genomic	AGGGTTGGCCAATCTACTCCCAGGAGCAGGGAGGGCAGGAGC CAGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTT ACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAAC AGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGT TACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGG TGAGGCCCTGGGCAGGTTGGTATCAAGGTTACAAGACAGGTT TAAGGAGACCAATAGAAACTGGGCATGTGGAGACAGAGAAG A
A1AT E342 Genomic	AGGTGTCCACGTGAGCCTTGCTCGAGGCCTGGGATCAGCCTTA CAACGTGTCTCTGCTTCTCTCCCCTCCAGGCCGTGCATAAGGC TGTGCTGACCATCGACGAGAAAGGGACTGAAGCTGCTGGGGC CATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGAGGTC AAGTTCAACAAACCCTTTGTCTTCTTAATGATTGAACAAAATA CCAAGTCTCCCCTCTTCATGGGAAAAGTGGTGAATCCCACCC
A1AT E342K Lenti Cassette	GTCACAGAGGAGGCACCCCTGAAGCTCTCCAAGGCCGTGCAT AAGGCTGTGCTGACCATCGACAAGAAAGGGACTGAAGCTGCT GGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCG AGGTCAAGTTCAACAAACCCTTTGTCTTCTTAATGATTGAACA AAATACCAAGTCTCCCCTCTTCATGGGAAAAGTGGTGAATCCC ACCCAAAAAGACCCAGCTTTCTTGTACAAAGTGGTTGATATCC AGCACAGTGGCGGCCGCTCGAGTCTAGAGGGCCCGCGGTTCG AAGGTAAGCCTATCCCTAACCCTCTCCTCGGTCT
HEK2 Genomic	TGAATGGATTCCTTGGAAACAATGATAACAAGACCTGGCTGA GCTAACTGTGACAGCATGTGGTAATTTTCCAGCCCGCTGGCCC TGTAAAGGAAACTGGAACACAAAGCATAGACTGCGGGGCGG GCCAGCCTGAATAGCTGCAAACAAGTGCAGAATATCTGATGA TGTCATACGCACAGTTTGACAGATGGGGCTGG
FANCF Genomic	ATAGCATTGCAGAGAGGCGTATCATTTCGCGGATGTTCCAATC AGTACGCAGAGAGTCGCCGTCTCCAAGGTGAAAGCGGAAGTA GGGCCTTCGCGCACCTCATGGAATCCCTTCTGCAGCACCTGGA TCGCTTTTCCGAGCTTCTGGCGGTCTCAAGCACTACCTACGTC AGCACCTGGGACCCCGCCACCGTGCGCCGGGCCTTGCAGTGG GCGCGCTACCTGCGCCACATCCATCGGCGCTTTGGTC
HEK3 Genomic	CCCAGCCAAACTTGTCAACCAGTATCCCGGTGCAGGAGCTGC ACATACTAGCCCCTGTCTAGGAAAAGCTGTCCTGCGACGCCCT CTGGAGGAAGCAGGGCTTCCTTTCCTCTGCCATCACGTGCTCA GTCTGGGCCCCAAGGATTGACCCAGGCCAGGGCTGGAGAAGC AGAAAAAAAGCATCAAGCCTACAAATGCATGCTTACTTGCAG CAGAAATAGACTAATTGCATGGGCGTTTCCCTGGGATCCCTGT CTCCAGGCTTCTCTCATCCATGCCTTTCTAGGCAGCCCACAT

Table 34. Correction percentage scores for the A1AT E342K PE2 PrimeDesign screen.

Condition	Average 7G Correction	7G Error	Condition	Average 7G Correction	7G Error	Condition	Average 7G Correction	7G Error
GL545-8-10	0.03	0	GL600-12-15	0.055	0.005	GL556-12-20	0.015	0.005
GL545-10-10	0.01	N/A	GL600-13-15	0.035	0.025	GL556-13-20	0.485	0.135
GL545-12-10	0.03	0.01	GL600-14-15	0.015	0.005	GL556-14-20	1.185	0.415
GL545-15-10	0.025	0.005	GL600-15-15	0.035	0.005	GL556-15-20	0.105	0.005
GL545-8-14	0.18	0.03	GL600-8-20	0.13	0	GL549-8-23	0.045	0.045
GL545-10-14	0.135	0.045	GL600-10-20	0.48	0.09	GL549-10-23	0.02	0.01
GL545-12-14	0.025	0.005	GL600-12-20	0.29	0.08	GL549-12-23	0.065	0.025
GL545-13-14	0.07	0.03	GL600-13-20	0.355	0.235	GL549-13-23	0.045	0.005
GL545-14-14	0.08	0.05	GL600-14-20	0.38	0.17	GL549-14-23	0.02	0
GL545-15-14	0.15	0.11	GL600-15-20	0.05	0	GL549-15-23	0.02	N/A
GL545-8-20	0.285	0.115	GL556-8-10	0.045	0.025	GL549-8-24	0.09	0.05
GL545-10-20	0.015	0.005	GL556-10-10	0.255	0.215	GL549-10-24	0.03	0.02
GL545-12-20	0.035	0.025	GL556-12-10	0.11	0.1	GL549-12-24	0.2	0.08
GL545-13-20	0.05	0.03	GL556-13-10	0.04	0	GL549-13-24	0.28	0.17
GL545-14-20	0.075	0.045	GL556-14-10	0.13	0.09	GL549-14-24	0.12	0.03
GL545-15-20	0.03	0.01	GL556-15-10	0.07	0.03	GL549-15-24	0.035	0.015
GL600-8-10	0.13	0.02	GL556-8-15	0.12	0.01	GL549-8-25	0.045	0.025
GL600-10-10	0.34	0.2	GL556-10-15	0.02	0	GL549-10-25	0.01	N/A
GL600-12-10	0.29	0.04	GL556-12-15	0.03	0.01	GL549-12-25	0.445	0.375
GL600-13-10	0.165	0.085	GL556-13-15	0.19	0.05	GL549-13-25	0.01	N/A
GL600-14-10	0.15	0.07	GL556-14-15	0.2	0.02	GL549-14-25	0.03	N/A
GL600-15-10	0.17	0.01	GL556-15-15	0.02	N/A	GL549-15-25	0.035	0.005
GL600-8-15	0.11	0.04	GL556-8-20	0.165	0.105	GFP_lenti	0.02	N/A
GL600-10-15	0.07	0.05	GL556-10-20	0.03	0	Untreated_lenti	0.01	0

"N/A" means that only one sample had detectable reads so no error was calculable.

Table 35. Allele percentages for GL544 PAM disruption and precise correction combined edit.

ProtospacerID-PBSlength-RTTlength	Average -3C+7G	-3C+7G Error
GL544-8-17	0	0
GL544-10-17	0.025	0.025
GL544-12-17	0.005	0.005
GL544-13-17	0	0
GL544-14-17	0.005	0.005
GL544-15-17	0.08	0.01
GL544-8-18	0.025	0.025
GL544-10-18	0.025	0.025
GL544-12-18	0.005	0.005
GL544-13-18	0.025	0.005
GL544-14-18	0.1	0.05
GL544-15-18	0.04	0.03
GL544-8-20	0.12	0.04
GL544-10-20	0.485	0.455
GL544-12-20	0.88	0.34
GL544-13-20	1.55	1.14
GL544-14-20	1.025	0.235
GL544-15-20	0.13	0.04

Designations described via the legend in the upper left box.

Table 36. Allele percentages for GL556 PAM disruption and precise correction combined edit.

ProtospacerID-PBSlength-RTTlength	Average 3A+7G	3A+7G Error
GL556-8-10	0.03	0.02
GL556-10-10	0.245	0.215
GL556-12-10	0.1	0.09
GL556-13-10	0.02	0.01
GL556-14-10	0.13	0.09
GL556-15-10	0.06	0.04
GL556-8-15	0.1	0.02
GL556-10-15	0.01	0
GL556-12-15	0	0
GL556-13-15	0.16	0.02
GL556-14-15	0.18	0
GL556-15-15	0	N/A
GL556-8-20	0.16	0.1
GL556-10-20	0.015	0.005
GL556-12-20	0.01	0
GL556-13-20	0.47	0.14
GL556-14-20	1.145	0.385
GL556-15-20	0.095	0.005

"N/A" means that only one sample had detectable reads so no error was calculable.

Table 37. Percent correction for the GL542 PrimeDesign PE2 experiment against GL544/556 spacers.

ProtospacerID-PBSlength-RTTlength	Average Precise Correction/Insertion	Error
pGL542-8-25	0.08	0.017321
pGL542-8-26	0.093333	0.008819
pGL542-8-27	0.19	0.015275
pGL542-8-28	1.103333	0.079652
pGL542-10-25	0.123333	0.008819
pGL542-10-26	0.053333	0.008819
pGL542-10-27	0.283333	0.003333
pGL542-10-28	1.036667	0.131698
pGL542-13-25	0.136667	0.008819
pGL542-13-26	0.04	0.01
pGL542-13-27	0.053333	0.01453
pGL542-13-28	0.476667	0.037565
pGL542-14-25	0.066667	0.003333
pGL542-14-26	0.053333	0.008819
pGL542-14-27	0.123333	0.008819
pGL542-14-28	0.2	0.01
pGL542-15-25	0.06	0.005774
pGL542-15-26	0.033333	0.008819
pGL542-15-27	0.086667	0.018559
pGL542-15-28	0.286667	0.017638
pGL542Msi	0.033333	0.003333
pGL544-13-20	2.31	0.824197
pGL556-14-20	0.513333	0.11348
pHRB241	0.03	0
Untreated	0.03	0
Untreated_NGA	0.01	0

Table 38. Percent precise correction of A1AT E342K NGA PE2 vs PE3 vs PE2-Miller.

Condition	Average Precise Correction	Error	Condition	Average Precise Correction	Error
GL544n1	5.956666667	1.166223725	GL556n2	1.163333333	0.609927136
GL544n2	5.826666667	2.60834771	GL556n3	0.396666667	0.12251984
GL544n3	2.436666667	0.537969433	GL556n4	0.84	0.380175398
GL544n4	3.526666667	1.320509161	GL556n5	2.276666667	1.62670765
GL544n5	1.816666667	0.210739439	GL556n6	1.436666667	0.504160468
GL544n6	5.276666667	2.389437405	GL556n7	0.493333333	0.204640606
GL544n7	4.386666667	1.382779005	GL556n8	0.416666667	0.11780398
GL544n8	1.81	0.427200187	GL556n9	0.253333333	0.125476868
GL544n9	3.376666667	0.853020776	GL556n10	0.486666667	0.110201835
GL544n10	3.646666667	0.809739327	GL556n11	1.566666667	0.661118581
GL544n11	3.963333333	2.351923563	PE2(GL544)	3.873333333	1.383730867
GL544n12	4.64	2.110505469	PE2(Miller)(GL544)	0.043333333	0.014529663
GL544n13	10.37333333	7.366619155	PE2(GL556)	0.783333333	0.16596519
GL544n14	3.476666667	1.977846753	PE2(Miller)(GL556)	0.02	0
GL544n15	2.83	1.332153645	spCas9(GL542M)	0.01	0
GL544n16	3.746666667	2.168396745	Untreated	0.03	
GL556n1	0.49	0.015275252			

Table 39. Percentage precise correction/insertion of GL542 PE2/PE3 vs GL544/556.

Condition	Precise Correction/Insertion	Error
542-10-28	1.026667	0.115662
542-10-28+n1	2.226667	0.093512
542-10-28+n2	2.07	0.011547
542-10-28+n3	2.33	0.243379
542-10-28+n4	2.873333	0.173813
542-10-28+n5	1.14	0.015275
542-10-28+n6	1.656667	0.052387
542-10-28+n7	1.356667	0.082125
542-10-28+n8	1.456667	0.093868
542-8-28	1.076667	0.049777
542-8-28+n1	2.886667	0.102035
542-8-28+n2	2.28	0.115326
542-8-28+n3	2.67	0.136137
542-8-28+n4	3.023333	0.17901
542-8-28+n5	1.17	0.052915
542-8-28+n6	1.95	0.079373
542-8-28+n7	1.366667	0.016667
542-8-28+n8	1.426667	0.115662
WT Untreated	0.05	0.005774
544-13-20	3.373333	0.170229
544-13-20+n11	4.61	1.454522
544-13-20+n12	3.586667	0.478063
544-13-20+n13	3.683333	1.095389
556-14-20	1.103333	0.248283
556-14-20+n11	1.276667	0.222586
556-14-20+n5	1.553333	0.48074
556-14-20+n6	1.41	0.579396
Lenti Untreated	0.013333	0.003333

Table 40. Editing percentages of mRNA vs. plasmid correction at A1AT E342K in Hek cells.

Condition	Average 7G	7G Error	TTEST P-Value
ngcABEvar9+sgRNA025	28.75	6.792896	
PE2-VRQR+GL544-13-20(Plasmid)	3.566666667	0.847945	
PE2-VRQR+GL544-13-20/n11(Plasmid)	2.47	0.570468	
PE2-VRQR+GL544-13-20/n12(Plasmid)	7.94	5.736282	
PE2-VRQR+GL544-13-20/n13(Plasmid)	3.243333333	1.036731	
PE2-VRQR+GL544-13-20(RNA)	4.4	1.876708	0.714795914
PE2-VRQR+GL544-13-20/n11(RNA)	2.23	0.960226	0.842627459
PE2-VRQR+GL544-13-20/n12(RNA)	4.506666667	1.206874	0.613094569
PE2-VRQR+GL544-13-20/n13(RNA)	3.886666667	1.533431	0.747980439
PE2-VRQR+GL556-14-20(Plasmid)	0.85	0.228108	
PE2-VRQR+GL556-14-20/n5(Plasmid)	0.37	0.083267	
PE2-VRQR+GL556-14-20/n6(Plasmid)	1.403333333	0.69451	
PE2-VRQR+GL556-14-20/n11(Plasmid)	0.833333333	0.156773	
PE2-VRQR+GL556-14-20(RNA)	2.626666667	0.715945	0.120514194
PE2-VRQR+GL556-14-20/n5(RNA)	1.183333333	0.368299	0.152181226
PE2-VRQR+GL556-14-20/n6(RNA)	2.733333333	0.471393	0.197792627
PE2-VRQR+GL556-14-20/n11(RNA)	3.216666667	1.578199	0.269452796
Untreated	0.023333333	0.003333	

T-test values for difference between RNA and plasmid paired conditions was not significant.

Table 41. Editing percentages of A1AT E342K by NGA prime editors in GM11423 fibroblasts.

Designation	7G AVERAGE	7G ERROR
ngcABEvar9+sgRNA025	20.14666667	0.21789396
PE2-VRQR+GL544-13-20	0.836666667	0.579549634
PE2-VRQR+GL544-13-20+n11	0.136666667	0.012018504
PE2-VRQR+GL544-13-20+n12	0.163333333	0.040960686
PE2-VRQR+GL544-13-20+n13	0.103333333	0.020275875
PE2-VRQR+GL556-14-20	0.253333333	0.052387445
PE2-VRQR+GL556-14-20+n5	0.286666667	0.068394282
PE2-VRQR+GL556-14-20+n6	0.15	0.005773503
PE2-VRQR+GL556-14-20+n11	0.296666667	0.080897741
Untreated	0.096666667	0.023333333

Table 42. RT-qPCR Ct scores for UNG knockdown of various transcripts.

Target	Sample	Average Ct	SD	Target	Sample	Average Ct	SD	Target	Sample	Average Ct	SD	Target	Sample	Average Ct	SD
UNG1	24hr+	31.40209	0.354229	UNG2	24hr+	30.37754	0.449261	UNG1_2_L	24hr+	35.68135	0.750496	UNG1_2_S	24hr+	29.37316	0.461284
UNG1	48hr+	30.50991	0.113681	UNG2	48hr+	29.08256	0.380299	UNG1_2_L	48hr+	33.59191	0.433163	UNG1_2_S	48hr+	27.64209	0.420182
UNG1	72hr+	29.09341	0.237939	UNG2	72hr+	27.41338	0.345403	UNG1_2_L	72hr+	31.78099	0.247661	UNG1_2_S	72hr+	25.83984	0.317827
UNG1	96hr+	28.07385	0.297139	UNG2	96hr+	26.52611	0.273933	UNG1_2_L	96hr+	30.62184	0.28502	UNG1_2_S	96hr+	24.7696	0.259469
UNG1	120hr+	27.59835	0.120746	UNG2	120hr+	26.71687	0.146243	UNG1_2_L	120hr+	31.02448	0.187379	UNG1_2_S	120hr+	24.93341	0.084557
UNG1	24hr-	30.0845	0.518052	UNG2	24hr-	27.18532	0.566783	UNG1_2_L	24hr-	31.31104	0.503946	UNG1_2_S	24hr-	25.45825	0.473891
UNG1	48hr-	28.73899	0.100962	UNG2	48hr-	25.81702	0.166214	UNG1_2_L	48hr-	29.95921	0.287532	UNG1_2_S	48hr-	24.21514	0.199509
UNG1	72hr-	27.23985	0.219548	UNG2	72hr-	25.26531	0.24586	UNG1_2_L	72hr-	29.32939	0.09247	UNG1_2_S	72hr-	23.48414	0.202226
UNG1	96hr-	26.53816	0.247153	UNG2	96hr-	24.91523	0.33168	UNG1_2_L	96hr-	28.86352	0.097148	UNG1_2_S	96hr-	22.91397	0.117838
UNG1	120hr-	26.28122	0.136696	UNG2	120hr-	25.15912	0.288487	UNG1_2_L	120hr-	29.20757	0.299146	UNG1_2_S	120hr-	23.20329	0.336568
UNG1	24hr GFP	29.93258	N/A	UNG2	24hr GFP	27.18782	N/A	UNG1_2_L	24hr GFP	31.49147	N/A	UNG1_2_S	24hr GFP	25.71425	N/A
UNG1	48hr GFP	28.52766	N/A	UNG2	48hr GFP	25.53117	N/A	UNG1_2_L	48hr GFP	29.5902	N/A	UNG1_2_S	48hr GFP	24.0011	N/A
UNG1	72hr GFP	27.70194	N/A	UNG2	72hr GFP	25.87248	N/A	UNG1_2_L	72hr GFP	29.78874	N/A	UNG1_2_S	72hr GFP	24.12414	N/A
UNG1	96hr GFP	26.68658	N/A	UNG2	96hr GFP	25.00686	N/A	UNG1_2_L	96hr GFP	28.97398	N/A	UNG1_2_S	96hr GFP	23.28731	N/A
UNG1	120hr GFP	26.28857	N/A	UNG2	120hr GFP	24.92086	N/A	UNG1_2_L	120hr GFP	28.81556	N/A	UNG1_2_S	120hr GFP	23.21648	N/A

Average UNG1/2/1&2 Ct scores and standard deviations.

Table 43. Ct values for UNG2 knockdown.

Name	Average CT	CT Error
24hr-1uL-UNG	29.49067942	0.379495
24hr-1.25uL-UNG	29.174263	0.103572
24hr-GFP	26.28395844	0.154252
24hr-Untreated	25.8890241	0.131328
48hr-1uL-UNG	29.91177877	0.282967
48hr-1.25uL-UNG	29.51558749	0.264475
48hr-GFP	25.87082227	0.286053
48hr-Untreated	26.05685488	0.501822
72hr-1uL-UNG	30.17690214	0.118841
72hr-1.25uL-UNG	29.71838506	0.118728
72hr-GFP	26.52925237	0.19621
72hr-Untreated	26.7971433	0.139003
96hr-1uL-UNG	29.50931803	0.310827
96hr-1.25uL-UNG	29.12909762	0.087181
96hr-GFP	27.04476992	0.101618
96hr-Untreated	26.82254219	0.535756
120hr-1uL-UNG	31.05862109	0.549356
120hr-1.25uL-UNG	30.87568283	0.156981
120hr-GFP	27.70372645	0.262739
120hr-Untreated	27.80135282	0.314483

This experiment was validated in 48-well format.

Table 44. Percent HBB PE3 editing of siGENOME plate 1 repeat screen 1

Condition	AVERAGE 22T	ERROR
siRNA 5 ctrl	12.95333	1.040662
RAD17	11.19667	1.566411
Lamin A/C ctrl	11.07667	1.856397
RAD9A	10.53	0.725282
siRNA pool1 ctrl	9.736667	0.811672
siRNA 3 ctrl	9.556667	1.641791
siRNA 4 ctrl	8.913333	1.06021
ERCC2	8.41	0.863095
siRNA pool2 ctrl	8.393333	2.403264
BRCA2	8.02	1.000267
siRNA 1 ctrl	7.313333	1.828718
ATR	6.97	0.068069
Cyclophilin B ctrl	6.766667	0.451085
GAPD ctrl	6.643333	2.291378
EXO1	6.386667	1.052843
Untreated+PE3	6.1	1.084866
MSH3	5.466667	1.211202
DDB1	5.386667	0.308455
XRCC6	5.383333	0.570273
MSH2	5.366667	0.660614
MSH4	5.126667	0.655396
LIG1	5.076667	0.986667
siRNA2 ctrl	4.26	0.373631
POLA1	4.03	0.751953
DCLRE1A	3.826667	0.96596
PNKP	3.1	0.1823
FANCM	2.873333	0.751805
LIG4	1.846667	0.546697
Untreated	0.03	0

Table 45. Percent HBB PE3 editing of siGENOME plate 1 repeat screen 2.

Gene	AVERAGE	ERROR
Untreated	1.503333333	0.044845413
BRCA2	14.89	1.02001634
POLA1	6.666666667	0.118368539
RAD17	14.81666667	0.193419521
PNKP	7.596666667	0.259122451
FANCM	7.713333333	0.37176755
MSH2	7.89	0.65184354
MSH4	7.87	0.475709996
Cyclophilin B ctrl	7.926666667	0.198522319
DCLRE1A	8.09	0.820812605
DDB1	8.136666667	0.544834328
Lamin A/C ctrl	13.60333333	0.132455443
siRNA2 ctrl	8.473333333	0.878357052
siRNA 5 ctrl	13.17666667	0.473298121
siRNA pool1 ctrl	13.19	0.931683065
EXO1	13.06	0.50083264
LIG4	8.923333333	0.494177206
XRCC6	8.956666667	0.898337972
ERCC2	12.37333333	0.43910262
LIG1	9.283333333	1.302117933
siRNA 1 ctrl	12.34333333	1.237326868
GAPD ctrl	9.91	0.583894968
siRNA 3 ctrl	9.78	1.454922678
MSH3	11.89666667	1.589384926
siRNA pool2 ctrl	11.52	0.831444526
ATR	10.45	0.780534005
siRNA 4 ctrl	10.49333333	1.15328998
RAD9A	10.94	0.89985184
Untreated+PE3	10.86	1.54282209

Table 46. List of OriGene plasmids used in the overexpression assay.

Gene	Vendor	Catalogue Number	Description
RAD17	OriGene	SC120501	pCMV6-XL5 RAD17 (NM_133338) Human Untagged Clone transcript variant 1
RAD9A	OriGene	SC321679	pCMV6-AC Rad9 (RAD9A) (NM_004584) Human Untagged Clone transcript variant 1
FEN1	OriGene	SC110892	pCMV6-XL5 FEN1 (NM_004111) Human Untagged Clone
UNG	OriGene	SC109875	pCMV6-XL5 UNG (NM_080911) Human Untagged Clone transcript variant 2
ERCC2	OriGene	SC106224	pCMV6-XL4 XPD (ERCC2) (NM_000400) Human Untagged Clone transcript variant 1
LIG4	OriGene	SC118710	pCMV6-XL5 DNA Ligase IV (LIG4) (NM_002312) Human Untagged Clone transcript variant 1
FANCM	OriGene	RC211791	pCMV6-Entry FANCM (NM_020937) Human Tagged ORF Clone
POLA1	OriGene	RC214011	pCMV6-AC-Myc-DDK DNA polymerase alpha (POLA1) (NM_016937) Human Tagged ORF Clone
PNKP	OriGene	SC128040	pCMV6-XL4 PNK (PNKP) (NM_007254) Human Untagged Clone
DCLRE1A	OriGene	SC333120	pCMV6 series SNM1A (DCLRE1A) (NM_001271816) Human Untagged Clone transcript variant 1
PMS2	OriGene	SC119825	pCMV6-XL5 PMS2 (NM_000535) Human Untagged Clone
MSH6	OriGene	SC120035	pCMV6-XL5 MSH6 (NM_000179) Human Untagged Clone
MLH1	OriGene	SC323792	pCMV6-AC MLH1 (NM_000249) Human Untagged Clone
MSH2	OriGene	SC119995	pCMV6-XL5 MSH2 (NM_000251) Human Untagged Clone
TREX2	OriGene	SC309626	pCMV6-XL5 TREX2 (NM_080701) Human Untagged Clone

Table 47. Overexpression assay HBB PE3 percent editing.

Condition	Average 22T	Average Undesired	22T Error	Undesired Error
Untreated+PE3	22.06333333	0.423333333	0.372394653	0.043716257
UNG	14.02666667	0.72	0.274246442	0.113578167
FANCM	13.94666667	0.16	0.447263283	0.045092498
MSH6	12.38	0.216666667	0.215483951	0.033333333
POLA1	11.98	0.243333333	0.367559519	0.043716257
PMS2	11.65333333	0.176666667	0.284155669	0.012018504
LIG4	11.16333333	0.126666667	0.377903574	0.016666667
MHL1	11.05333333	0.143333333	0.577042267	0.017638342
RAD17	11.04666667	0.18	0.423687516	0.011547005
FEN1	9.89	0.113333333	0.664630223	0.024037009
PNKP	9.833333333	0.183333333	0.540411366	0.020275875
MSH2	9.516666667	0.2	0.239188443	0.03
RAD9A	7.103333333	0.123333333	0.871403721	0.047022453
ERCC2	7.05	0.116666667	0.246644143	0.046666667
GFP	6.843333333	0.12	0.213098829	0.034641016
TREX2	3.503333333	0.113333333	0.149480582	0.026034166
Untreated	0.226666667	0.03	0.046308147	0.005773503

Table 48. Percent HBB PE3 editing in BRCA2 knockout cell line.

Condition	Average 22T	Average Undesired	22T Error	Undesired Error
Hek-BRCA2+7.5k	0.803333	0.026666667	0.211844	0.012018504
Hek-BRCA2+7.5k+PE3	16.90333	0.213333333	2.368968	0.046308147
Hek-Parental+7.5k	0.886667	0.03	0.196497	0.005773503
Hek-Parental+7.5k+PE3	17.17	0.406666667	1.613722	0.048419463
Hek-WT+7.5k	0.563333	0.026666667	0.108985	0.006666667
Hek-WT+7.5k+PE3	16.81333	0.296666667	0.462685	0.119210365
Hek-BRCA2+26k	1.08	0.06	0.068069	0.03
Hek-BRCA2+26k+PE3	23.44	0.376666667	0.972471	0.046666667
Hek-Parental+26k	1.103333	0.04	0.06888	0.011547005
Hek-Parental+26k+PE3	28.97	0.456666667	1.157684	0.027284509
Hek-WT+26k	0.91	0.02	0.066583	0.01
Hek-WT+26k+PE3	22.78	0.336666667	1.163286	0.038441875

Table 49. Percentages for precise target editing with TREX2 influence on at various target sites.

Condition	22T AVERAGE	22T ERROR
26k-Untreated+Untreated	0.07	0.015275
26k-Untreated_HBB_PE2	1.94	0.060828
GFP-Plasmid_HBB_PE2	0.97	0.085049
TREX2-Plasmid_HBB_PE2	1.773333	1.23382
26k-Untreated_HBB_PE3	9.21	0.07
GFP-Plasmid_HBB_PE3	3.563333	0.263902
TREX2-Plasmid_HBB_PE3	1.653333	0.186667
Condition	Average 7G	7G ERROR
Untreated+Untreated	0.006666667	0.003333333
Untreated+pGL544-13-20	2.16	0.315330514
GFP-Plasmid+pGL544-13-20	1.546666667	0.211765069
TREX2-Plasmid+pGL544-13-20	1.22	0.412431813
UNG-siRNA+pGL544-13-20	3.073333333	0.233547521
TREX2-siRNA+pGL544-13-20	2.686666667	0.652184364
Untreated+pGL544-13-20+n11	4.44	0.572916515
GFP-Plasmid+pGL544-13-20+n11	0.686666667	0.228789083
TREX2-Plasmid+pGL544-13-20+n11	0.48	0.134288247
UNG-siRNA+pGL544-13-20+n11	3.613333333	0.972733834
TREX2-siRNA+pGL544-13-20+n11	2.24	0.211266025
Condition	7A AVERAGE	7A ERROR
Untreated+Untreated	0.036666667	0.003333333
Untreated+pGL542-8-28	0.93	0.017320508
GFP-Plasmid+pGL542-8-28	0.406666667	0.066416196
TREX2-Plasmid+pGL542-8-28	0.113333333	0.043716257
UNG-siRNA+pGL542-8-28	0.913333333	0.106822802
TREX2-siRNA+pGL542-8-28	0.93	0.02081666
Untreated+pGL542-8-28+n11	2.246666667	0.054873592
GFP-Plasmid+pGL542-8-28+n11	0.773333333	0.063333333
TREX2-Plasmid+pGL542-8-28+n11	0.226666667	0.029059326
TREX2-siRNA+pGL542-8-28+n11	2.436666667	0.161795481
UNG-siRNA+pGL542-8-28+n11	2.426666667	0.127191894

Condition	23C AVERAGE	23C ERROR
26k-Untreated+Untreated	0.023333	0.006667
26k-Untreated+FANCF-PE2	13.19667	0.11921
GFP-Plasmid+FANCF-PE2	5.19	0.052915
TREX2-Plasmid+FANCF-PE2	2.473333	0.489535
26k-Untreated+FANCF-PE3	15.15	0.278747
GFP-Plasmid+FANCF-PE3	5.226667	0.434255
TREX2-Plasmid+FANCF-PE3	2.613333	0.245515
7.5k-Untreated+Untreated	0.03	0.005774
7.5k-Untreated+FANCF-PE2	13.69	0.786532
UNG-siRNA+FANCF-PE2	17.15	0.84309
TREX2-siRNA+FANCF-PE2	16.19	0.373631
7.5k-Untreated+FANCF-PE3	15.99333	0.683236
UNG-siRNA+FANCF-PE3	19.08667	0.261683
TREX2-siRNA+FANCF-PE3	17.88	0.669552
Condition	AVERAGE 17+CTT	17+CTT ERROR
26k-Untreated	0.016666667	0.003333333
26k-Untreated+HEK3-PE2	12.30666667	0.172851895
GFP-Plasmid+HEK3-PE2	5.606666667	0.421676548
TREX2-Plasmid+HEK3-PE2	2.933333333	0.383941546
26k-Untreated+HEK3-PE3	13.07	0.52848841
GFP-Plasmid+HEK3-PE3	6.14	0.551935987
TREX2-Plasmid+HEK3-PE3	3.486666667	0.19272894
7.5k-Untreated	0.03	0.005773503
7.5k-Untreated+HEK3-PE2	13.92	0.626498204
UNG-siRNA+HEK3-PE2	15.84666667	1.379194612
TREX2-siRNA+HEK3-PE2	15.15333333	0.410297995
7.5k-Untreated+HEK3-PE3	13.81333333	0.915283805
UNG-siRNA+HEK3-PE3	15.21333333	1.012016688
TREX2-siRNA+HEK3-PE3	15.50333333	0.852766739

Appendix 2.

Additional Figures

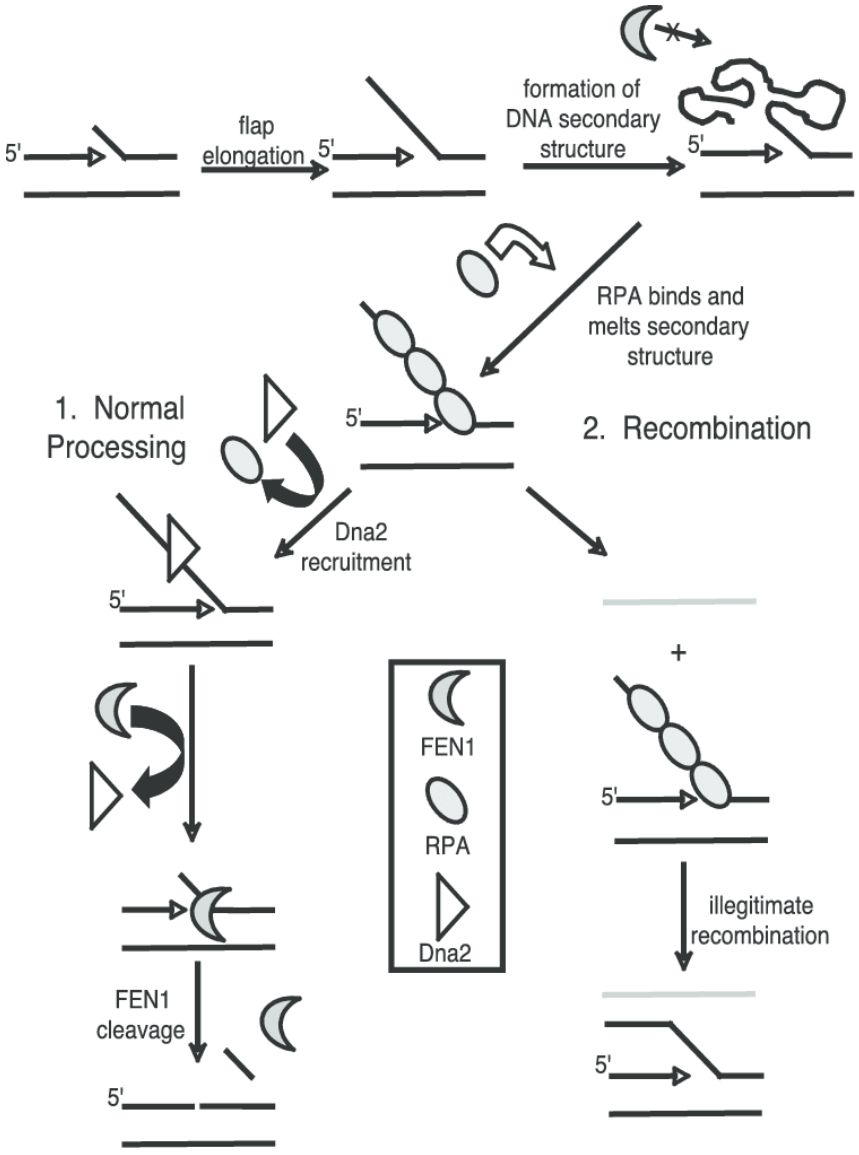


Figure 41. DNA Flap Repair Mechanism.

Diagram (Bartos et al 2008) showing the mechanism of Okazaki fragment flap recognition by RPA followed by Dna2 binding and subsequent shortening and cleavage, completed by FEN1 cleavage of the short downstream flap in human cells. There is also the possibility that a long flag can hybridize and recombine with another nearby ssDNA.

>PE2Frag1
ATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAAAGTCGACAAGAAGTACAG
CATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCA
AGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTC
GACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAA
CCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGAC
TGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGAC
GAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAA
GGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGG
GCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTG
TTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAG
CAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTG
CCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTG
AGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTT
TCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCA
AGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCT
CTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGG
CTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACG
GCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATT
CCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGG
CCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAG
GAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCC
CAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAG
TGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGAC
CTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTG
CTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGC
TGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTG
ACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGA
CAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACG
GCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAAC
TTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCA
GGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGA
CAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATG
GCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGG
CATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGC
TGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCC
GACTACGATGTGGACGCTATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGAC
CAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACT
ACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGA
GGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAA
GCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAG
TGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGC
GAGATCAACAACTACCAC

Figure 42. Geneblock fragment 1.

Sequence of a section of PE2.

>PE2Frag2
CACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAG
CGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCG
GCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCC
AACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGG
CCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGC
AGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAG
GACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAA
AGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAA
GCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATC
ATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGA
ACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATG
AGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTG
GACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGT
GCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTA
CCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACC
AGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGA
CCTGTCTCAGCTGGGAGGTGACTCTGGAGGATCTAGCGGAGGATCCTCTGGCAGCGAGACACCAGGAACAA
GCGAGTCAGCAACACCAGAGAGCAGTGGCGGCAGCAGCGGCGGCAGCAGCACCCTAAATATAGAAGATGAG
TATCGGCTACATGAGACCTCAAAAGAGCCAGATGTTTCTCTAGGGTCCACATGGCTGTCTGATTTTCCTCA
GGCCTGGGCGGAAACCGGGGGCATGGGACTGGCAGTTCGCCAAGCTCCTCTGATCATACCTCTGAAAGCAA
CCTCTACCCCCGTGTCCATAAAACAATACCCCATGTCACAAGAAGCCAGACTGGGGATCAAGCCCCACATA
CAGAGACTGTTGGACCAGGGAATACTGGTACCCTGCCAGTCCCCCTGGAACACGCCCCTGCTACCCGTTAA
GAAACCAGGGACTAATGATTATAGGCCTGTCCAGGATCTGAGAGAAGTCAACAAGCGGGTGGAAGACATCC
ACCCCACCGTGCCCAACCCTTACAACCTCTTGAGCGGGCTCCCACCGTCCCACCAGTGGTACACTGTGCTT
GATTTAAAGGATGCCTTTTTCTGCCTGAGACTCCACCCCACCAGTCAGCCTCTCTTCGCCTTTGAGTGGAG
AGATCCAGAGATGGGAATCTCAGGACAATTGACCTGGACCAGACTCCCACAGGGTTTCAAAAACAGTCCCA
CCCTGTTTAATGAGGCACTGCACAGAGACCTAGCAGACTTCCGGATCCAGCACCCAGACTTGATCCTGCTA
CAGTACGTGGATGACTTACTGCTGGCCGCCACTTCTGAGCTAGACTGCCAACAAGGTACTCGGGCCCTGTT
ACAAACCCTAGGGAACCTCGGGTATCGGGCCTCGGCCAAGAAAGCCCAAATTTGCCAGAAACAGGTCAAGT
ATCTGGGGTATCTTCTAAAAGAGGGTCAGAGATGGCTGACTGAGGCCAGAAAAGAGACTGTGATGGGGCAG
CCTACTCCGAAGACCCCTCGACAACTAAGGGAGTTCCTAGGGAAGGCAGGCTTCTGTCGCCTCTTCATCCC
TGGGTTTGCAGAAATGGCAGCCCCCCTGTACCCTCTCACCAAACCGGGGACTCTGTTTAATTGGGGCCCAG
ACCAACAAAAGGCCTATCAAGAAATCAAGCAAGCTCTTCTAACTGCCCCAGCCCTGGGGTTGCCAGATTTG
ACTAAGCCCTTTGAACTCTTTGTCGACGAGAAGCAGGGCTACGCCAAAGGTGTCCTAACGCAAAAACTGGG
ACCTTGGCGTCGGCCGGTGGCCTACCTGTCCAAAAAGCTAGACCCAGTAGCAGCTGGGTGGCCCCCTTGCC
TACGGATGGTAGCAGCCATTGCCGTACTGACAAAGGATGCAGGCAAGCTAACCATGGGACAGCCACTAGTC
ATTCTGGCCCCCCATGCAGTAGAGGCACTAGTCAAACAACCCCCCGACCGCTGGCTTTCCAACGCCCGGAT
GACTCACTATCAGGCCTTGCTTTTGGACACGGACCGGGTCCAGTTCGGACCGGTGGTAGCCCTGAACCCGG
CTACGCTGCTCCCACTGCCTGAGGAAGGGCTGCAACACAACTGCCTTGATATCCTGGCCGAAGCCCACGGA
ACCCGACCCGACCTAACGGACCAGCCGCTCCCAGACGCCGACCACACCTGGTACACGGATGGAAGCAGTCT
CTTACAAGAGGGACAGCGTAAGGCGGGAGCTGCGGTGACCACCGAGACCGAGGTAATCTGGGCTAAAGCCC
TGCCAGCCGGGACATCCGCTCAGCGGGCTGAACTGATAGCACTCACCCAGGCCCTAAAGATGGCAGAAGGT
AAGAAGCTAAATGTTTAT

Figure 43. Geneblock fragment 2.

Sequence of a section of PE2.

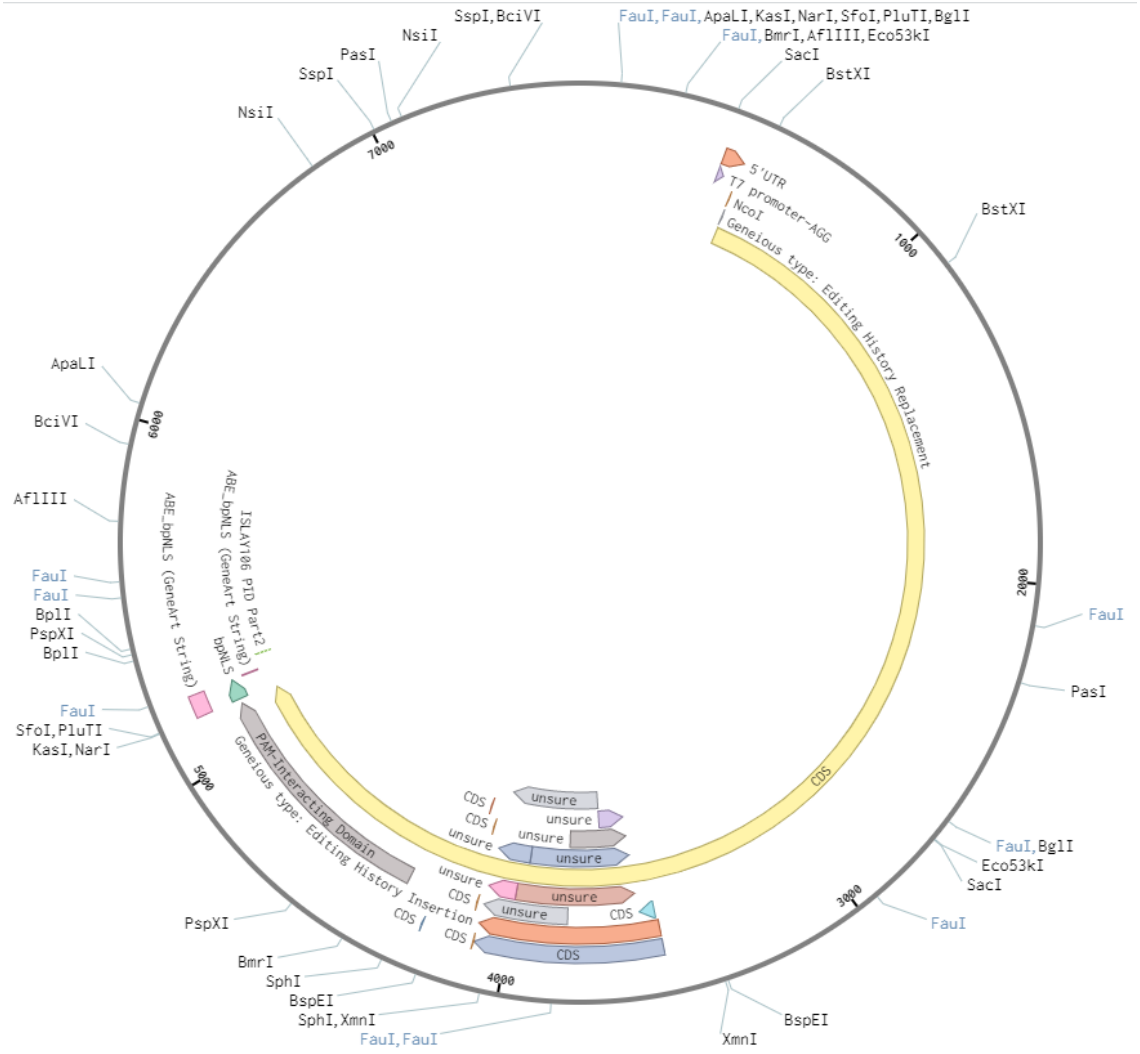


Figure 44. Plasmid map of pGL111(pUTR-Trilink-ISLAY1-monoTadA-ABE7.10(V82S)-MQKFRAER).

Source plasmid for pUTR backbone.

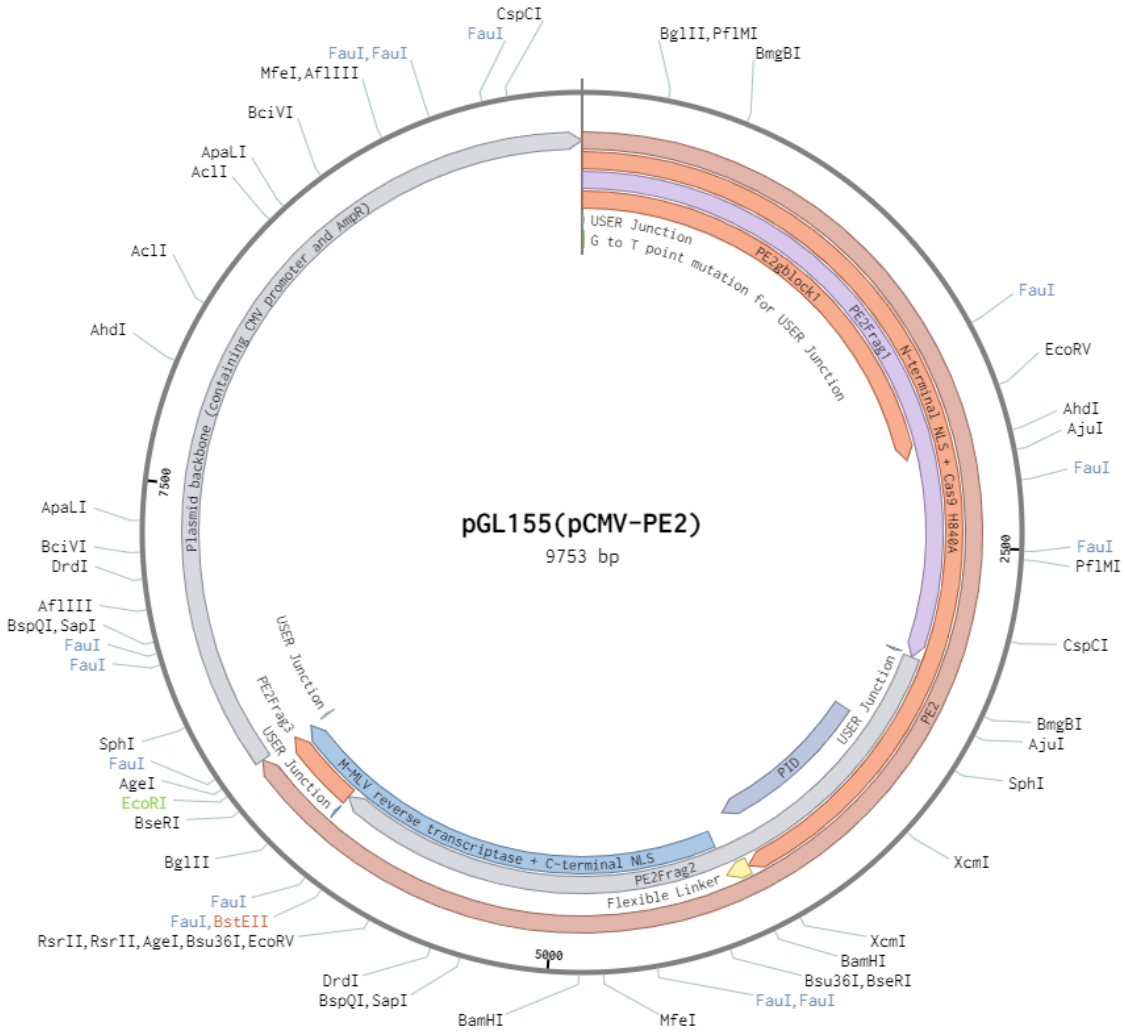


Figure 45. pCMV-PE2 plasmid map.

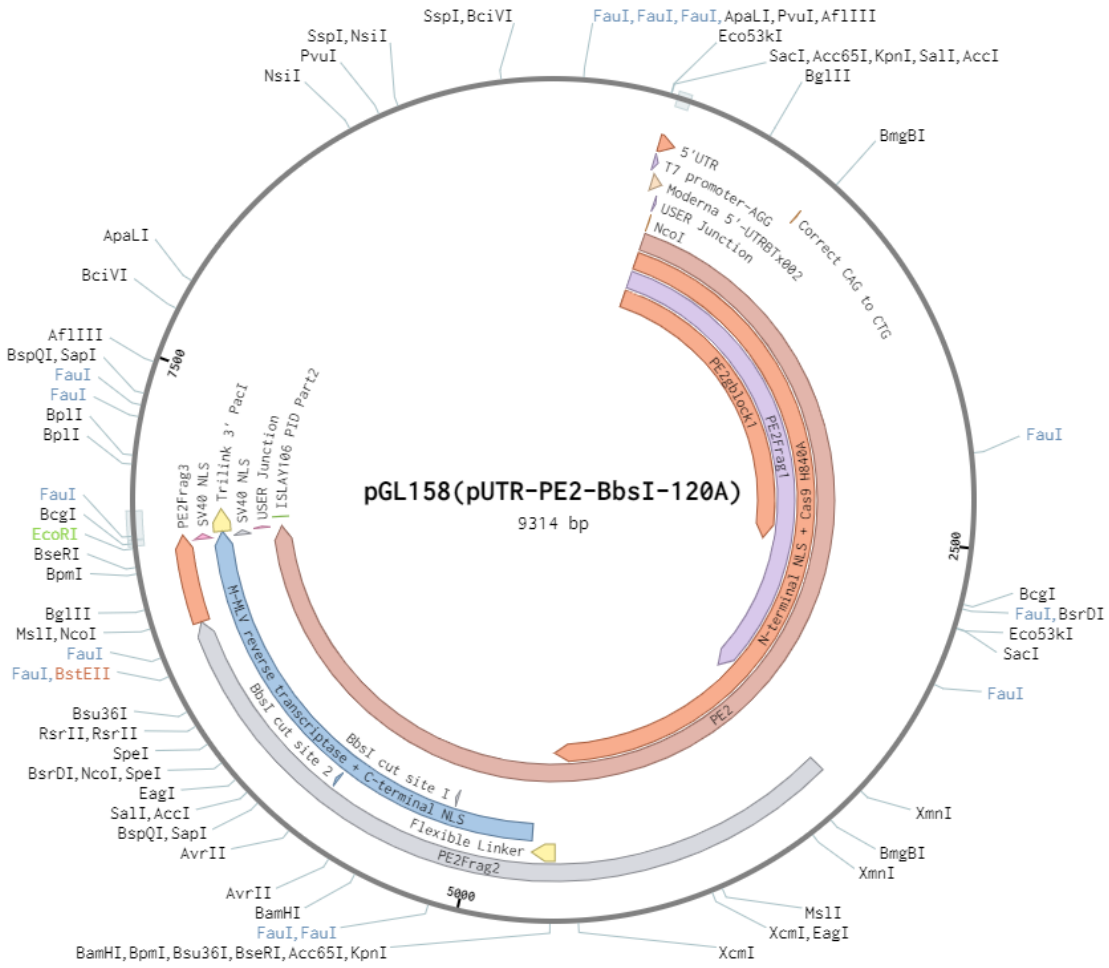


Figure 46. Plasmid map of pUTR-PE2-BbsI-120A.

>PE2
MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH
SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA
HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSK
SRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN
LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNR
EDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL
LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI
ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREM
IEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
EKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKS
DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQ
ITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH
DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE
LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQ
KGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTST
KEVLDATLIHQSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSG
GSSTLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATS
TPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDL
REVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRD
PEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSE
LDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQP
TPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQAL
LTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCL
RMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDR
VQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSL
LQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYA
FATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEA
RGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSEFEPKKKRKV*

Figure 47. Amino acid sequence for PE2.

>PE2-VRQR
MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH
SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA
HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSK
SRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN
LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNR
EDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL
LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI
ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREM
IEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
EKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKS
DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQ
ITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH
DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKE
LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQ
KGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRST
KEVLDATLIHQSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSG
GSSTLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATS
TPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDL
REVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRD
PEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSE
LDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQP
TPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQAL
LTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCL
RMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDR
VQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSL
LQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYA
FATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEA
RGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSEFEPKKKRKV*

Figure 48. Amino acid sequence for PE2-VRQR.

VRQR mutations enable targeting of NGA noncanonical PAMs.

>PE2-MQKFRAER
MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH
SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA
HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSK
SRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN
LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNR
EDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL
LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI
ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREM
IEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
EKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKS
DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQ
ITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH
DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVKE
LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKFLQ
KGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRST
KEVLDATLIHQSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSG
GSSTLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATS
TPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDL
REVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRD
PEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSE
LDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQP
TPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQAL
LTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCL
RMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDR
VQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSL
LQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYA
FATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEA
RGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSEFEPKKKRKV*

Figure 49. Amino acid sequence for PE2-MQKFRAER.

These mutations enable targeting of NGC noncanonical PAMs.

>PE2-(I322V,S409I,E427G,R654L,R753G)-VRQR
MKRTADGSEFESPKKKRKVDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH
SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA
HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSK
SRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN
LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQDLTLLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNR
EDLLRKQRTFDNGIIPHQIHLGELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL
LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI
ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREM
IEERLKTYAHLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKS
DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQ
ITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH
DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKE
LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQ
KGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRST
KEVLDATLIHQSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSG
GSSTLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATS
TPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDL
REVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRD
PEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSE
LDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQP
TPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQAL
LTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCL
RMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDR
VQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSL
LQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYA
FATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEA
RGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSEFEPKKKRKV*

Figure 50. Amino acid sequence for PE2-(I322V,S409I,E427G,R654L,R753G)-VRQR.

These additional mutations are theorized to increase residence time of Cas9 to increase editing rates.

>PE2-(I322V,S409I,E427G,R654L,R753G, R1114G)-MQKFRAER
MKRTADGSEFESPKKKRKVDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLI
ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILR
VNTEITKAPLSASMVKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQGDFYPFLKDNR
EKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNL
PNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED
YFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMI
EERLKTYAHLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI
HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGGHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE
LDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI
TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMI
AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKGNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKG
KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA
KFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATL
IHQSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTLNIEDEYRLH
ETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKP
HIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPP
SHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRD
LADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGY
LLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNW
GPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDP
VAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLD
TDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQE
GQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGE
IYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAI
TETPDTSTLLIENSSPSGGSKRTADGSEFEPKKKRKV*

Figure 51. Amino acid sequence for PE2-(I322V,S409I,E427G,R654L,R753G, R1114G)-MQKFRAER.

These additional mutations are theorized to increase residence time of Cas9 to increase editing rates.

Supplementary Note 3. Protocol for cloning 3'-extended pegRNAs into mammalian U6 expression vectors by Golden Gate assembly.

Cloning overview

1. **Digest pU6-pegRNA-GG-Vector plasmid (component 1)** with *Bsa*I and isolate the plasmid fragment (~2.2kb) containing the origin of replication, U6 promoter, U6 poly-T termination sequence, and Amp^R gene

2. **Order oligonucleotides for:**
    a. The desired spacer (target) sequence flanked by indicated overhangs (**component 2**)
        i. Use the desired target's 5'-3' sequence for the top strand oligonucleotide (incldluing the 5' CACC and 3' GTTTT overhangs) and use the reverse complement of the target sequence for the bottom strand oligonucleotide (including the 5' CTCTAAAAC overhang). Spacer sequences must begin with a G nucleotide for efficient transcription initiation.
    b. The desired pegRNA 3' extension template flanked by the indicated overhangs (**component 3**)
        i. Use the RNA sense sequence as the top strand oligonucleotide (featuring the 5' GTGC overhang) and use the reverse complement of this sequence for the bottom strand oligonucleotide (featuring the 5' AAAA overhang).
    c. SpCas9 sgRNA scaffold sequence featuring compatible golden gate overhangs (**component 4**)
        i. These oligonucleotides are not the complete scaffold sequence, as overhangs from the remaining components contribute several missing nucleotides
        ii. Note: these oligonucleotides must be 5' phosphorylated. Oligonucleotides can be 5' phosphorylated by the manufacturer or 5' phosphorylated enzymatically using T4 PNK (see protocol below)
3. **Anneal top and bottom oligonucleotides** for components 2, 3, and 4 in separate annealing reactions according to the protocol below. If the SpCas9 sgRNA scaffold sequence (component 4) was not phosphorylated, phosphorylate with T4 PNK.
4. **Golden Gate assembly of isolated 2.2-kb fragment from component 1 with components 2, 3, and 4**
5. **Transform the ligation product into *E. coli*.** The antibiotic resistance conferred by component 1 from the pU6-pegRNA-GG-vector plasmid is ampicillin and carbenicillin resistance.
6. **Isolate and sequence** plasmids from the resulting clonal transformants

pegRNA cloning protocol

Step 1: Digest pU6-pegRNA-GG-Vector plasmid (component 1)

Combine the following in a PCR tube:

2000 ng pU6-pegRNA-GG-Vector (component 1)	X µL
Bsa1-HFv2 (NEB)	1.0 µL
10x Cutsmart Buffer	3.0 µL
H_2O	to 30.0 µL
Total reaction volume	30.0 µL

Incubate at 37 °C for 4-16 hours
Isolate ~2.2-kb fragment from cut plasmid.

Figure 52. Supplementary Note 3 from Anzalone et al. 2019.

Protocol adapted in this work for cloning pegRNAs.

Steps 2 and 3: Order and anneal oligonucleotide parts (components 2, 3, and 4)

Materials

Annealing buffer: H_2O supplemented with 10 mM Tris-Cl pH 8.5 and 50 mM NaCl
Complementary oligonucleotide pairs

Protocol

Combine the following in a PCR tube:

Top oligonucleotide, 100 µM	1.0 µL
Bottom oligonucleotide, 100 µM	1.0 µL
Annealing buffer (components 2, 3, and 4)	23.0 µL
Total reaction volume	25.0 µL

In thermocycler, heat at 95 °C for 3 minutes, then cool gradually (0.1 °C/s) to 22 °C

Dilute annealed oligonucleotides 1:4 by adding 75 µL H_2O. The final concentration of each oligonucleotide will be 1 µM after this dilution. Do not dilute the sgRNA scaffold (component 4) if phosphorylating by PNK in step 2.5.

Step 2.b.ii.: sgRNA scaffold phosphorylation (unnecessary if oligonucleotides were purchased phosphorylated)

Protocol

Combine the following in a PCR tube:

4 µM oligonucleotide duplex from step 1	6.25 µL
10x T4 DNA ligase buffer (NEB)	2.50 µL
T4 PNK (NEB)	0.50 µL
H_2O	15.75 µL
Total reaction volume	20.0 µL

In thermocycler, incubate at 37 °C for 60 minutes
Following this phosphorylation, annealed scaffold oligonucleotides are now at a concentration of 1 µM. Proceed to step 3.

Figure 53. Supplementary Note 3 (cont.) from Anzalone et al. 2019.

Protocol adapted in this work for cloning pegRNAs.

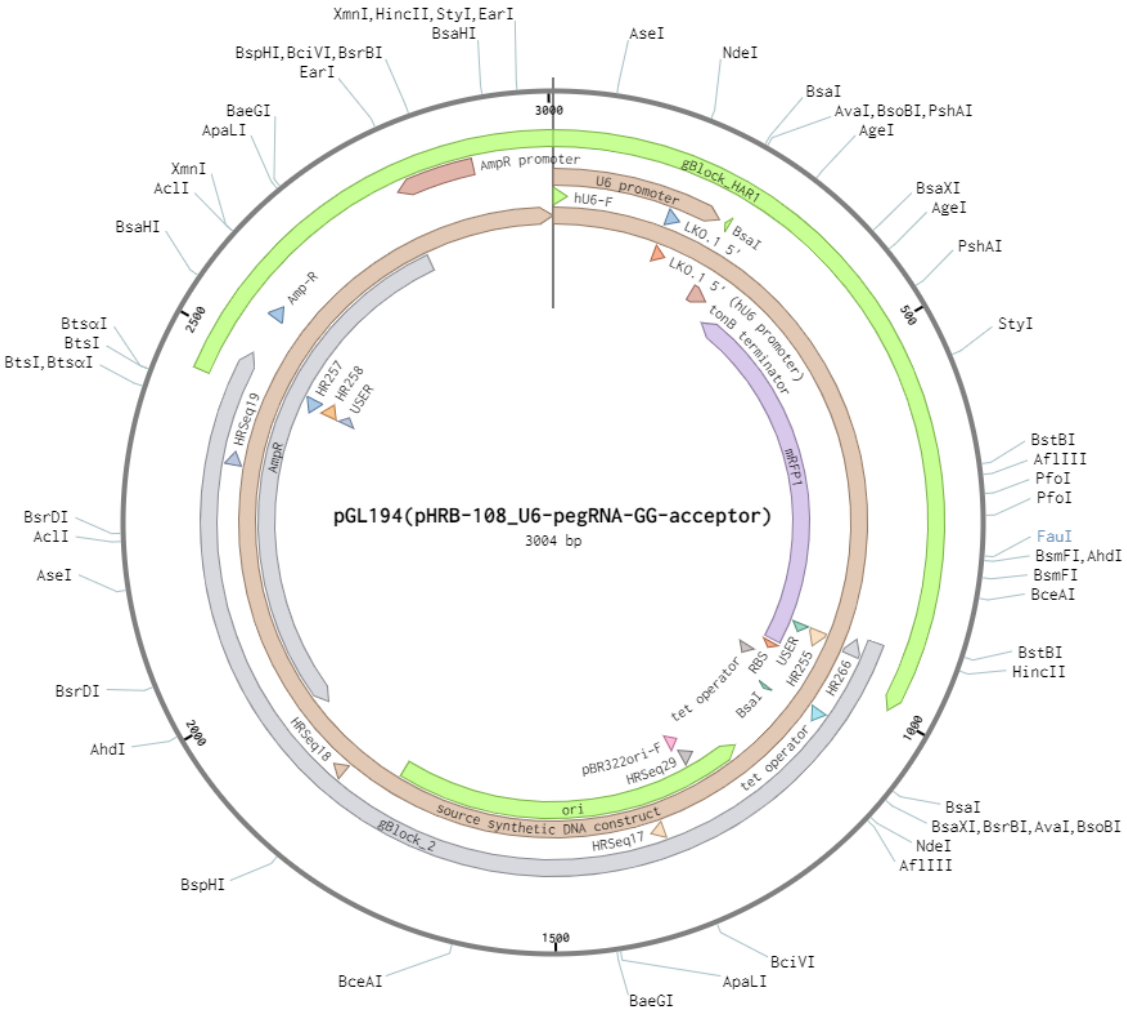


Figure 54. Plasmid map for pU6-pegRNA-GG-Acceptor vector.

Used for cloning mammalian expression of pegRNAs.

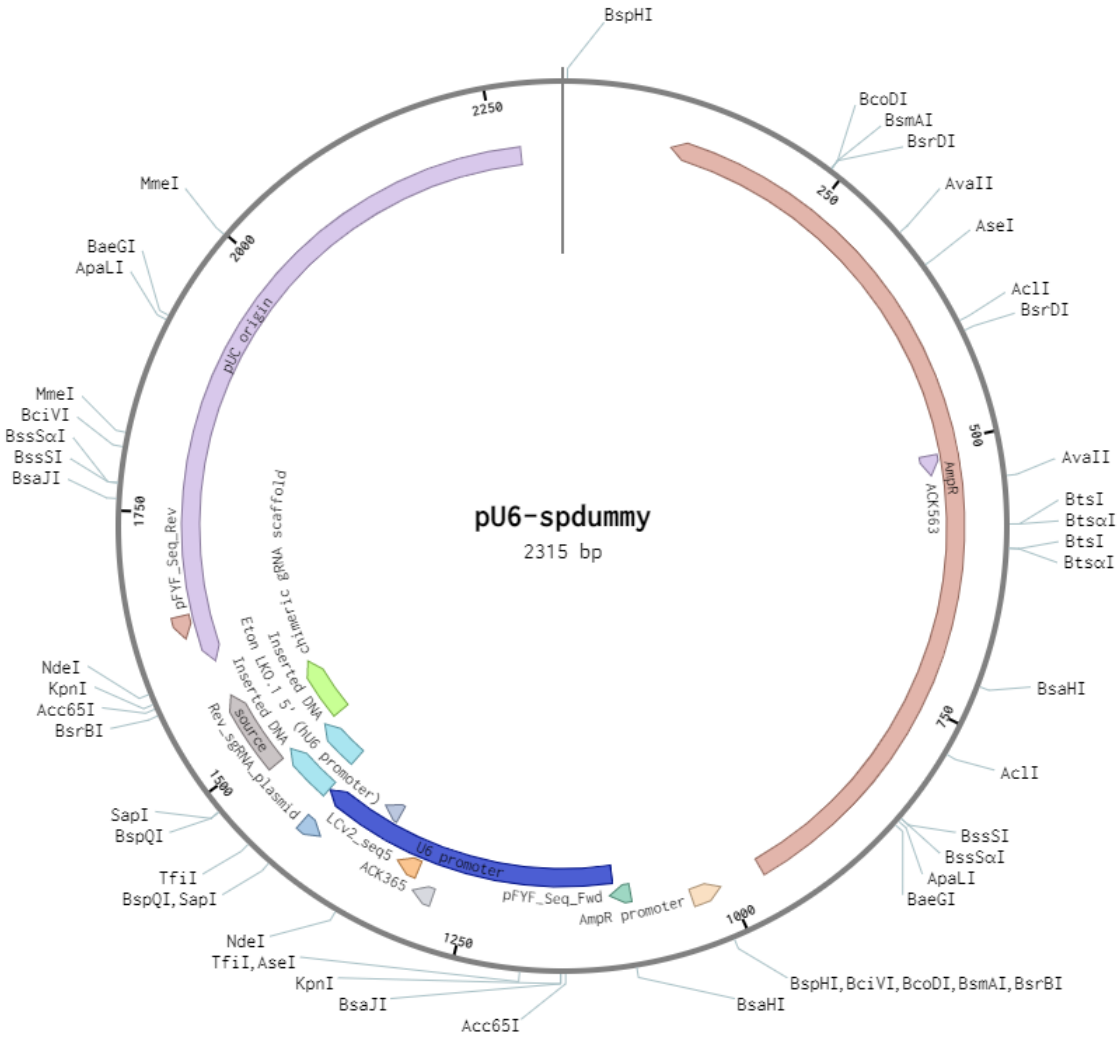


Figure 55. pU6-spdummy vector.

Used for cloning mammalian expression of nickRNAs.

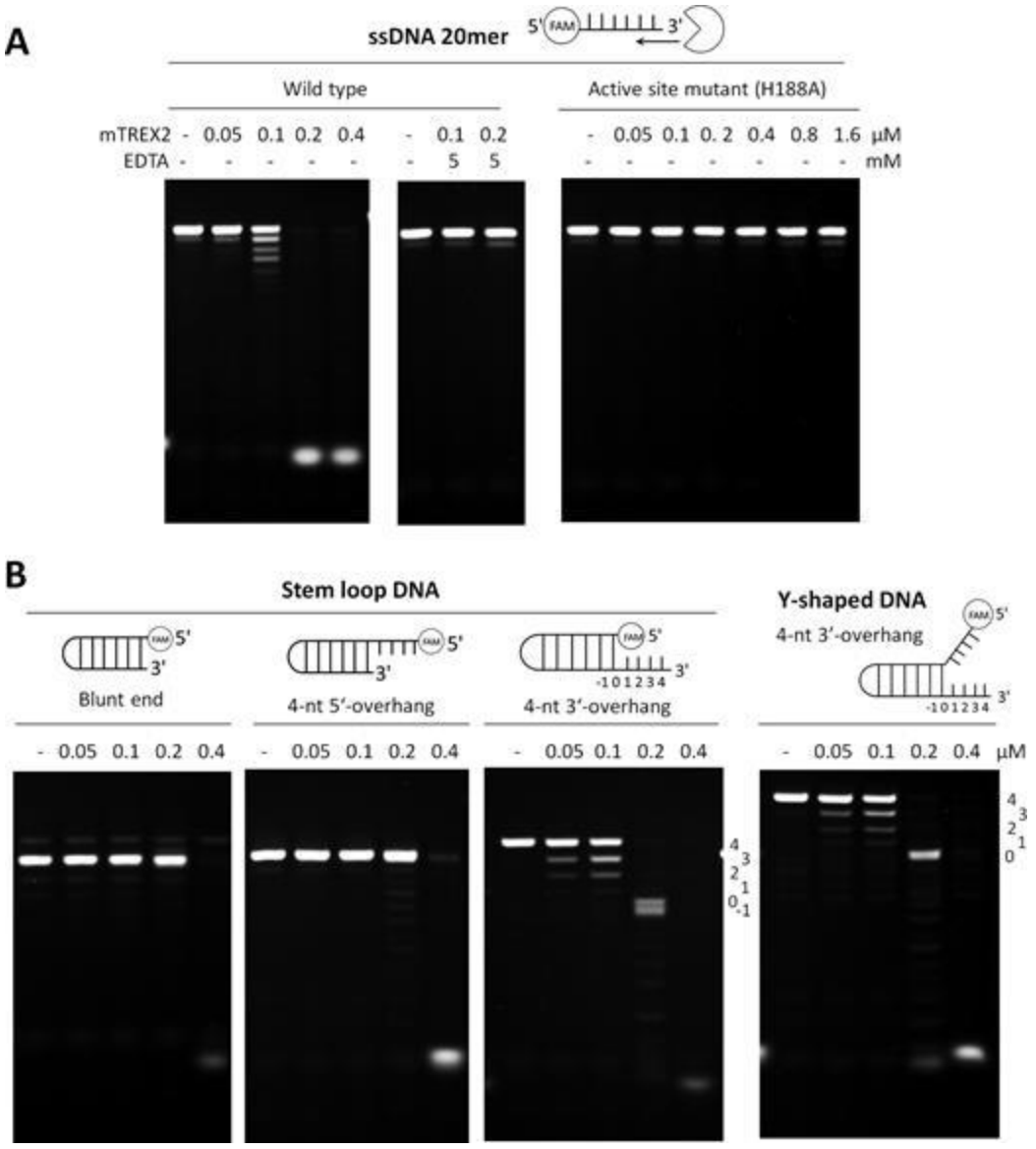


Figure 56. Figure 1 from Cheng et al., 2018.

Depicts the structure and cleavage analysis gels of various DNA duplexes by TREX2 action.

References

Takeuchi, Ryo et al. “Redesign of extensive protein-DNA interfaces of meganucleases using iterative cycles of in vitro compartmentalization.” *Proceedings of the National Academy of Sciences of the United States of America* vol. 111,11 (2014): 4061-6. doi:10.1073/pnas.1321030111

Carroll, Dana. “Genome engineering with zinc-finger nucleases.” *Genetics* vol. 188,4 (2011): 773-82. doi:10.1534/genetics.111.131433

Wang, Luyao et al. “In Vivo Delivery Systems for Therapeutic Genome Editing.” *International journal of molecular sciences* vol. 17,5 626. 27 Apr. 2016, doi:10.3390/ijms17050626

Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat Commun* 9, 1911 (2018). https://doi.org/10.1038/s41467-018-04252-2

Menon, Vijay, and Lawrence Povirk. “Involvement of p53 in the repair of DNA double strand breaks: multifaceted Roles of p53 in homologous recombination repair (HRR) and non-homologous end joining (NHEJ).” *Sub-cellular biochemistry* vol. 85 (2014): 321-36. doi:10.1007/978-94-017-9211-0_17

Brunet, Erika, and Maria Jasin. “Induction of Chromosomal Translocations with CRISPR-Cas9 and Other Nucleases: Understanding the Repair Mechanisms That Give Rise to Translocations.” *Advances in experimental medicine and biology* vol. 1044 (2018): 15-25. doi:10.1007/978-981-13-0593-1_2

Gaudelli, N., Komor, A., Rees, H. et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature 551, 464–471 (2017). https://doi.org/10.1038/nature24644

Komor, A., Kim, Y., Packer, M. *et al.* Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424 (2016). https://doi.org/10.1038/nature17946

Anzalone, A.V., Randolph, P.B., Davis, J.R. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157 (2019). https://doi.org/10.1038/s41586-019-1711-4

Stoller, James K., Aboussouan, Loutfi S.. “A Review of α1-Antitrypsin Deficiency.” *American Journal of Respiratory and Critical Care Medicine* (2011): https://doi.org/10.1164/rccm.201108-1428CI

Brode, Sarah K., Ling, Simon C., Chapman, Kenneth R.. “Alpha-1 antitrypsin deficiency: a commonly overlooked cause of lung disease” *CMAJ* 184 (12) 1365-1371 (2012). DOI: https://doi.org/10.1503/cmaj.111749

Shen, Shen., Sanchez, Minerva E.., Blomenkamp, Keith., Corcoran, Erik M.., Marco, Eugenio., Yudkoff, Clifford J.., Jiang, Haiyan., Teckman, Jeffrey H.., Bumcrot, David., Albright, Charles F.. "Amelioration of Alpha-1 Antitrypsin Deficiency Diseases with Genome Editing in Transgenic Mice" Human Gene Therapy (2018) 29:8, 861-873 https://doi.org/10.1089/hum.2017.227

Zuo, Erwei., Sun, Yidi., Wei, Wu., Yuan, Tanglong., Ying, Wenqin., Sun, Hao., Yuan, Liyun., Steinmetz, Lars M.., Li, Yixue., Yang, Hui. "Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos" *Science.* (2019) : 289-292 DOI: 10.1126/science.aav9973

Bartos, Jeremy D et al. "Catalysis of strand annealing by replication protein A derives from its strand melting properties." *The Journal of biological chemistry* vol. 283,31 (2008): 21758-68. doi:10.1074/jbc.M800856200

Finger, L. David et al. "The wonders of flap endonucleases: structure, function, mechanism and regulation." *Sub-cellular biochemistry* vol. 62 (2012): 301-26. doi:10.1007/978-94-007-4572-8_16

Dana, Hassan et al. "Molecular Mechanisms and Biological Functions of siRNA." *International journal of biomedical science : IJBS* vol. 13,2 (2017): 48-57.

Sharma, Sonia., Rao, Anjana. "RNAi screening: tips and techniques." *Nature immunology* vol. 10,8 (2009): 799-804. doi:10.1038/ni0809-799

Martin, S., Chiramel, A.I., Schmidt, M.L. et al. A genome-wide siRNA screen identifies a druggable host pathway essential for the Ebola virus life cycle. *Genome Med* 10, 58 (2018). https://doi.org/10.1186/s13073-018-0570-1

Sun, Jing et al. "Genome-wide siRNA screen of genes regulating the LPS-induced TNF-α response in human macrophages." *Scientific data* vol. 4 170007. 1 Mar. 2017, doi:10.1038/sdata.2017.7

Zhao N, Wang G, Das AT, Berkhout B. "Combinatorial CRISPR-Cas9 and RNA interference attack on HIV-1 DNA and RNA can lead to cross-resistance." *Antimicrob Agents Chemother.* 2017 61:e01486-17. https://doi.org/10.1128/AAC.01486-17.

Schene, I.F., Joore, I.P., Oka, R. et al. Prime editing for functional repair in patient-derived disease models. *Nat Commun* 11, 5352 (2020). https://doi.org/10.1038/s41467-020-19136-7

Liu, P., Liang, SQ., Zheng, C. et al. Improved prime editors enable pathogenic allele correction and cancer modelling in adult mice. *Nat Commun* 12, 2121 (2021). https://doi.org/10.1038/s41467-021-22295-w

Habib, O., Habib, G., Hwang, GH., Bae, S. Comprehensive analysis of prime editing outcomes in human embryonic stem cells. *bioRxiv* 439533, (2021). https://doi.org/10.1101/2021.04.12.439533

Miller, Shannon M et al. "Continuous evolution of SpCas9 variants compatible with non-G PAMs." *Nature biotechnology* vol. 38,4 (2020): 471-481. doi:10.1038/s41587-020-0412-8

Geu-Flores, Fernando et al. "USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products." *Nucleic acids research* vol. 35,7 (2007): e55. doi:10.1093/nar/gkm106

Komor, Alexis C et al. "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity." *Science advances* vol. 3,8 eaao4774. 30 Aug. 2017, doi:10.1126/sciadv.aao4774

Nami, Fatemeharefeh et al. "Strategies for In Vivo Genome Editing in Nondividing Cells." *Trends in biotechnology* vol. 36,8 (2018): 770-786. doi:10.1016/j.tibtech.2018.03.004

Chu, S Haihua et al. "Rationally Designed Base Editors for Precise Editing of the Sickle Cell Disease Mutation." The CRISPR journal vol. 4,2 (2021): 169-177. doi:10.1089/crispr.2020.0144

Mazur, D J, and F W Perrino. "Excision of 3' termini by the Trex1 and TREX2 3'-->5' exonucleases. Characterization of the recombinant proteins." *The Journal of biological chemistry* vol. 276,20 (2001): 17022-9. doi:10.1074/jbc.M100623200

Chari, Raj et al. "Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach." *Nature methods* vol. 12,9 (2015): 823-6. doi:10.1038/nmeth.3473

Bothmer, Anne et al. "Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus." *Nature communications* vol. 8 13905. 9 Jan. 2017, doi:10.1038/ncomms13905

Cheng, Hiu-Lo, et al. "Structural insights into the duplex DNA processing of TREX2". *Nucleic Acids Research*, Volume 46, Issue 22, 14 December 2018, Pages 12166–12176, https://doi.org/10.1093/nar/gky970

Wang, Chuan-Jen, et al. "Cellular expression and localization of DNA exonucleases, Trex, and their response to chemotherapeutic compounds in different tumor cell lines". *Proc Amer Assoc Cancer Res*, Volume 46, 2005. https://cancerres.aacrjournals.org/content/65/9_Supplement/418.5