

[*This is a protocol for inclusion in CPMB. There are 8 figures and 3 tables. A PMC file is available]

Base Editing in Human Cells to Produce Single Nucleotide Variant Clonal Cell Lines

Carlos A. Vasquez^{1,2}, Quinn T. Cowan^{1,2}, and Alexis C. Komor^{1*}

¹Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA, 92093, U.S.A.

²These authors contributed equally to this work.

*Correspondence should be addressed to Alexis C. Komor (akomor@ucsd.edu)

CONFLICT OF INTEREST STATEMENT:

A.C.K. is a member of the SAB of Pairwise Plants and is an equity holder for Pairwise Plants and Beam Therapeutics. A.C.K.'s interests have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies. The authors declare no other competing financial interests.

ABSTRACT:

Base editing technologies enable point mutation introduction at targeted genomic sites in mammalian cells, with higher efficiency and precision than traditional genome editing methods that employ DNA double-strand breaks, such as zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and the CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats-CRISPR-associated protein 9) system. This allows for the generation of single nucleotide variant isogenic cell lines (i.e. cell lines whose genomic sequences differ from each other only at a single, edited nucleotide) in a more time- and resource-effective manner. These single nucleotide variant clonal cell lines

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/cpmb.129](https://doi.org/10.1002/cpmb.129).

This article is protected by copyright. All rights reserved.

represent a powerful tool to assess the functional role of genetic variants in a native cellular context. Base editing can, therefore, facilitate genotype-to-phenotype studies in a controlled laboratory setting, with applications in both basic research and clinical applications. Here, we provide optimized protocols (including experimental design, methods, and analyses) to design base editing constructs, transfect adherent cells, quantify base editing efficiencies in bulk, and generate single nucleotide variant clonal cell lines.

Basic Protocol 1: Design and Production of Plasmids for Base Editing Experiments

Basic Protocol 2: Transfecting Adherent Cells and Harvesting Genomic DNA

Basic Protocol 3: Genotyping Harvested Cells using Sanger Sequencing

Alternate Protocol 1: Next-Generation Sequencing to Quantify Base Editing

Basic Protocol 4: Single Cell Isolation of Base Edited Cells using FACS

Alternate Protocol 2: Single Cell Isolation of Base Edited Cells Using Dilution Plating

Basic Protocol 5: Clonal Expansion to Generate Isogenic Cell Lines and Genotyping of Clones

KEYWORDS:

base editing, genome editing, single nucleotide variant, isogenic cell lines

INTRODUCTION:

Base editing is a genome editing technique capable of precisely introducing single nucleotide variants (SNVs) into the genome of living cells (Rees & Liu, 2018). Base editors (BEs) consist of a catalytically inactivated or impaired CRISPR-associated protein [dead Cas9 (dCas9), nickase Cas9 (nCas9), or dead Cas12a (dCas12a)] tethered to a single-stranded DNA (ssDNA)-modifying enzyme. In this system, the Cas protein will complex with a user-programmed guide RNA (gRNA), and bind to a genomic locus (termed the “protospacer”) that is complementary to the sequence of the gRNA and

harbors a protospacer adjacent motif (PAM, a short DNA sequence specific to the Cas protein used) (Figure 1; Jinek et al., 2012). The Cas protein locally denatures the target double-stranded DNA (dsDNA) region upon binding, to form an R-loop, exposing a small window (typically ~5 nucleotides long) of ssDNA (Jiang et al., 2016). Once bound, the ssDNA-modifying enzyme will catalyze the deamination of target nucleobases within this window. Subsequent DNA replication or repair of these modified base intermediates (uracil or inosine) results in permanent introduction of single-base substitutions. To date, two major classes of BEs have been described: cytosine base editors (CBEs), which change C•G base pairs to T•A (Komor et al., 2016), and adenosine base editors (ABEs), which convert A•T base pairs to G•C (Gaudelli et al., 2017). Because BEs do not rely on double-stranded DNA breaks (DSBs), the introduction of unwanted gene alterations and random insertions and deletions (indels) of nucleotides at the targeted site are minimized compared to traditional genome editing methods such as ZFNs, TALENs, and the CRISPR-Cas9 system (Hess et al., 2017; Ran et al., 2013). As a testament to their utility, BEs have been optimized and employed for a wide range of applications, including synthetic biology, therapeutics, and agriculture (Evanoff & Komor, 2019; Molla & Yang, 2019; Rees & Liu, 2018). Of particular interest to the field of functional genomics, base-editing technologies are uniquely positioned to facilitate studies investigating genotype-to-phenotype relationships by enabling researchers to generate isogenic cell lines with enhanced throughput.

[*Figure 1 near here]

Advances in next-generation sequencing are increasingly making the detection of SNVs routine, particularly in clinical settings (Lappalainen, et al., 2019). Our ability to interpret the functional consequences of these SNVs, however, has lagged far behind. While the identification of clinically actionable and pathogenic mutations has revolutionized and improved the field of precision

medicine (Ahmed et al., 2020; Malone et al., 2020), these only represent a minority of reported human genetic variants. Indeed, while there are currently over 685 million human SNVs identified in the dbSNP database (Sherry et al., 2001), less than 1% have a defined clinical interpretation in ClinVar (Landrum et al., 2018). This issue is particularly endemic to rare genetic variants and those discovered in minoritized populations and indigenous people (1000 Genomes Project Consortium et al., 2015; Fox et al., 2020; Popejoy & Fullerton, 2016), highlighting the need for a significant increase in studies that can functionally assess human genetic variants in laboratory settings. At the core of these studies, is the ability to introduce genetic variants in living cells with high efficiency and precision, something that can be accomplished using base editing technologies. Such approaches allow researchers to study variants of interest in their native cellular context, without confounding background genetic variation (which can occur in studies using patient-derived cell lines), and base editing facilitates the generation of these cellular models in a more time- and resource-effective manner than those that can be achieved with DSB-dependent editing strategies. Specifically, because SNV introduction efficiencies using DSB-mediated editing is so low (typically <1% in many cell types, compared to >20% for most targets using base editing), researchers must screen through hundreds of clones to obtain one with the desired genotype, or utilize DNA repair manipulation or donor template design “tricks” that can extend the timing of cell line generation by months or require the installation of additional, unwanted edit “scars” in the locus of interest (see the Commentary section for more details) (Coggins et al., 2017; Kwart et al., 2017; Ousterout et al., 2015; Paquet et al., 2016; Ran et al., 2013; Riesenber et al., 2019; Yeh et al., 2019). Using base editing, these SNV-containing isogenic cell lines can be generated in as little as ~2 weeks, with both heterozygous and homozygous genotypes generated with high efficiencies (~25% each of clones screened). This rapid and efficient generation of isogenic cell line models could subsequently

facilitate the clinical interpretation of variants with increased throughput compared to DSB-reliant methods, and the identification of new therapeutic strategies in the field of precision medicine.

The protocols outlined below describe how to introduce site-specific C•G to T•A or A•T to G•C mutations in the genome of mammalian cells, with the ultimate goal of generating SNV isogenic cell lines (cell lines whose genomic sequences differ from each other at a single nucleotide). The methodology is broken down into five basic protocols, starting with designing and generating BE:gRNA combinations to introduce an SNV of interest (Basic Protocol 1). We then describe the process for validating these BE:gRNA combinations in bulk mammalian cells by Sanger or next-generation sequencing (Basic Protocol 2, Basic Protocol 3, and Alternate Protocol 1). Lastly, we describe the procedure to isolate single cells (Basic Protocol 4 and Alternate Protocol 2), and then expand and genotype isogenic cell lines harboring the SNV of interest (Basic Protocol 5). We describe the whole process using four SNVs in the *MUTYH* gene [1: 45340220 (GRCh38); c.35G>A p.W12*, c.809T>G p.V270A; c.886C>T p.L296L; and c.909G>A p.E303E] as illustrative examples. Additionally, this whole process (outlined in Figure 2) is designed to be modular, such that the researcher can adapt the presented methodology according to their specific experimental goals.

[*Figure 2 near here]

STRATEGIC PLANNING

The optimal BE:gRNA combination will depend on the experimental goal and sequence surrounding the target base. While selecting a base editor, gRNA spacer, and suitable PAM, there are several considerations that should be reviewed before starting (discussed in the Critical Parameters section). These include sequence context surrounding the target base, PAM and Cas variant selection, and acceptable levels of indels or off-target editing.

A particular challenge that many researchers encounter is “bystander editing”. This can occur if other C or A nucleobases fall within the same exposed ssDNA window as the target, and are unintentionally modified; these edited bases are referred to as “bystander mutations” and can be problematic for phenotyping experiments. However, bystander editing can be minimized by using a particular deaminase variant (such as one with a shifted editing window or strict sequence preference) and careful PAM selection. The nucleotides flanking the target base —especially the 5’ nucleotide for CBEs— will also influence deaminase selection (see Arbab et al. 2020 for a detailed analysis of deaminase sequence preferences). Finally, while certain experiments require no detectable indels and off-target editing (where high-fidelity Cas and deaminase variants can be used), others will benefit from using BEs with the highest on-target efficiency (see Critical Parameters and Table 2). As base editing systems are modular, we recommend testing multiple BE:gRNA options initially (through Basic Protocol 3), and then selecting the combination that produces the best editing profile for isogenic cell line generation or downstream assays.

An additional matter to consider before beginning is the fact that immortalized cell lines may contain genetic variation or mutations in the target gene compared to the human reference genome; we thus recommend sequencing the locus of interest in your cell line before beginning. Within this protocol, we have included details on how to do this: see Basic Protocol 1 step 6A for primer design, Basic Protocol 2 for “Harvesting Genomic DNA”, and Basic Protocol 3 for sequencing the target site. Additionally, be sure to plan proper negative controls when generating SNV-containing isogenic cell lines (see Critical Parameters).

BASIC PROTOCOL 1: DESIGN AND PRODUCTION OF PLASMIDS FOR BASE EDITING EXPERIMENTS

Base editing experiments rely on the preparation of two high-quality plasmids: one expresses the gRNA that designates the target genomic location, and the other, expresses the BE protein

component (Cas9-deaminase fusion). Accurate and efficient base editing relies on careful design of the most appropriate BE:gRNA combination. Central to this, is the identification of a PAM that enables the Cas protein to bind in a manner that only positions the target C or A within the base editing window. We recommend the use of BEs that incorporate the engineered SpCas9-NG for minimal targeting restrictions with the smallest reduction in on-target activity in mammalian cells (Nishimasu et al., 2018). BEs that incorporate other engineered variants, Cas9 homologs, and Cas12a orthologs, such as SpRY, Sau/Sauri/Spy-mac, and Lb/As/Aa (Gaudelli et al., 2020; Hu et al., 2020; Y. B. Kim et al., 2017; Li et al., 2018; Z. Liu et al., 2019; Richter et al., 2020; Walton et al., 2020) offer additional PAM options but may have reduced on-target efficiencies at certain genomic loci or broadened editing windows, leading to bystander mutations (see the Critical Parameters section). The observed editing efficiency will depend on the context of the genomic site and base editor used. We have listed recommended CBE and ABE variants in Table 2 and discuss alternatives in Critical Parameters. Of note, optimized “BE-max” plasmids contain GenScript human codon optimization for maximal expression levels, bipartite NLSs on both termini for enhanced nuclear import, and an optional bicistronic enhanced green fluorescent protein (EGFP) for assessment of transfection efficiency and BE expression (Koblan et al., 2018; Zafra et al., 2018). The gRNA expression vector should contain a U6 promoter driving expression of an *S. pyogenes* Cas9 gRNA (e.g. Addgene #47511). The initial spacer sequence is irrelevant because site-directed mutagenesis will be used in this protocol to replace the gRNA spacer using a 5' overhang on the forward primer.

This basic protocol outlines the steps to properly design gRNAs for BE experiments, replace the spacer sequence in a gRNA expression vector, and prepare the appropriate plasmids for mammalian cell transfection. Additionally, Basic Protocol 1 details primer design for the amplification of a target genomic locus to quantify targeted SNV introduction by Sanger or next-generation sequencing.

NOTE: If multiple BE:gRNA options are available, we recommend experimentally validating all of them to identify the optimal combination that maximizes target C or A editing and minimizes bystander editing. Automated programs discussed in the Commentary section can also be used to predict potential off-target sites.

[*Table 1 near here]

[*Table 2 near here]

Materials:

- Plasmids expressing the base editing machinery (available at Addgene): *S. pyogenes* Cas9 gRNA vector (#47511), BE4max-NG-GFP (equivalent to #125616 with ancAPOBEC or #140001 with CMV promoter) for C•G to T•A editing, and/or ABE4max-NG-GFP (#140005) for A•T to G•C editing.
- Nuclease-free water (VWR, cat. no. 82007-328)
- T4 PNK (NEB, cat. no. M0201L)
- T4 ligase buffer [ATP source] (NEB, cat. no. B0202S), or leftover from ligase kits
- Primers (Integrated DNA Technologies (IDT) or similar)
- Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, cat. no. F534L)
- 10 mM dNTPs (NEB, cat. no. N0447S)
- Reagents for agarose gel electrophoresis
- 1 kb DNA ladder (NEB cat. no. N3232L)

- QIAquick PCR purification kit (Qiagen, cat. no. 28106)
- Nanodrop spectrophotometer
- Quick ligation kit (NEB, cat. no. M2200L)
- Competent cells (Mach1, DH5a, or 10B)
- Ampicillin or carbenicillin, 100mg/mL stock solution (see Reagents and Solutions)
- Agar plates (Elbing & Brent, 2019) with ampicillin/carbenicillin, final conc. 100 µg/mL
- Liquid LB or 2xYT media (Elbing & Brent, 2019) with ampicillin/carbenicillin, final conc. 100 µg/mL
- 0.2 mL PCR tubes (Fisher Scientific, cat. no. 07-200-259)
- 1.5 mL nuclease-free microcentrifuge tubes (VWR, cat. no. 87003-294)
- Cryogenic storage vials, 2mL (VWR, cat. no. 10018-760)
- Plasmid mini prep kit (QIAGEN, cat. no. 28106)
- Endotoxin-free plasmid midi prep kit (Fisher Scientific, cat. no. NC0919795)

Protocol Steps

Design gRNA

1. Import genomic locus of interest into sequence viewing software of choice. We frequently use Benchling (<https://www.benchling.com>) or SnapGene (<https://www.snapgene.com>).

Be sure to import in the sense orientation using the reference genome ('build') and isoform that matches the SNV annotation.

2. Identify the target nucleotide (C or A) and find an NG PAM exactly 12 to 16 nucleotides downstream (i.e. towards the 3' end of the strand containing the target nucleotide, see Figure 3A for an example). The target can be on either strand, as long as there is a suitable PAM on the same strand. Annotate the protospacer. Alternately, free software programs such as the Benchling wizard (Komor, 2016), BE-designer (Hwang et al., 2018), and BE-Hive (Arbab et al., 2020) can be used for automated gRNA design (see Critical Parameters). As most automated programs lack flexibility in deaminase and Cas variants, we recommend manually designing your protospacers (Figure 3A) and checking predicted editing efficiencies in BE-Hive (Arbab et al., 2020).

If bystander edits are unavoidable, choose the protospacer with bystander nucleotides farthest from the center of the window (position 6) or one that will incorporate silent mutations (as long as it is not a splice site).

[*Figure 3 near here]

3. If the protospacer does not start with guanine, add a 5'G to create a 21-nucleotide spacer in the gRNA. For protospacers that already start with a 5'G (such as the example V270 protospacer), move to the next step.
- Adding a guanine drastically increases gRNA transcription from the U6 promoter, yielding high editing efficiencies, despite a possible 5' mismatch.*
4. Design a custom forward primer with the sequence 5'-[N₂₀₋₂₁]GTTT TAGAGCTAGAAATAGCA-3', replacing the [N₂₀₋₂₁] portion with the protospacer sequence from the previous step (e.g. Table 1, #2 for MUTYH V270). This will be used with the universal reverse primer (Table 1, #1). Analyze this primer pair for homo- and hetero-dimers using the IDT oligo

analyzer (<https://www.idtdna.com/pages/tools/oligoanalyzer>). Note that amplification may be difficult using primers with a $\Delta G < -10$ kcal/mol.

The custom primer sequence should only include the protospacer sequence without the PAM. In base editing experiments, the spacer sequence will always be identical to the protospacer and contain the target A or C because direct nucleobase modification occurs on ssDNA of the canonical "non-target strand" that is not duplexed with gRNA. See Figure 3A for an example. It is also recommended to include a non-targeting gRNA as a control (see Critical Parameters and Table 1 primer #6).

5. Order the primers from IDT or another manufacturer of choice for custom gRNA construction via site-directed mutagenesis of the spacer or "around-the-horn" cloning.

Design primers for amplification of genomic locus

6. A. Use a primer design tool (e.g. Primer3 or other) to generate a pair of primers that amplify a ~1 kilobase (kb) region containing the protospacer, for Sanger sequencing. See Figure 4A and Table 1, primers #11-12 for V270 example using the Benchling Primer3 design wizard.

These primers should ideally be 18-24 base pairs (bp) long, and have 40-60% GC content and melting temperatures of 60-65°C. To ensure that at least one of these primers is appropriately positioned to yield maximum quality sequencing data of the protospacer region of the amplicon, the target sequencing region should be at least 50 bp away from the primer used for sequencing.

B. Alternatively, design primers for Illumina next-generation sequencing (NGS).

This requires specific adapter sequences be added to both the forward and reverse primers, and a much shorter amplicon. Design a primer pair that amplifies 200-250 bp of DNA (for a 300-cycle NGS run; if a 150-cycle NGS run will be used, the amplicon should be 100-125bp long) with the same parameters as above. Then, add the adapter sequences #13 and #14 from Table 1 onto the 5' end of the forward and reverse primers, respectively. For the MUTYH V270 example, the Primer3 Wizard on Benchling was used to identify 22-nt PCR primers that amplify a 249 bp region around the protospacer, and the Illumina NGS adapters were added to create primers #15 and 16 in Table 1. Order these as the round 1 PCR primers. A second round of amplification is required to provide the samples with P5/P7 tails and unique barcodes. Order round 2 PCR primers based on the Illumina barcoding system to be used during library preparation. See Alternate Protocol 1 for more information.

[*Figure 4 near here]

Clone new spacer into gRNA expression plasmid via site-directed mutagenesis

7. In 0.2 mL PCR tubes, 5' phosphorylate each primer (forward and reverse, from step #5 above) individually in separate 20 μ L reactions by combining the following (in the order stated):

- 15 μ L nuclease-free water
- 2 μ L primer (100 μ M stock)
- 2 μ L T4 DNA ligase buffer
- 1 μ L T4 polynucleotide kinase

Mix well and quickly spin to collect sample. The 20 μL of phosphorylated primer is enough for 8 PCR reactions—scale up if the primer is needed for more than 8 gRNAs (i.e. the universal reverse).

8. Place the reaction tube in a thermocycler and run the following program:

- 37°C for 20 min
- 95°C for 5 min
- Hold at 12°C

9. In a PCR tube on ice, combine the following reagents for a 50 μL reaction (in the order stated):

- X μL nuclease-free water (fill to a total volume of 50 μL)
- 10 μL Phusion HF buffer
- 1 μL dNTP mix
- 2.5 μL FWD phosphorylated primer (10 μM from previous step)
- 2.5 μL REV phosphorylated primer (10 μM from previous step)
- \sim 1 ng DNA template: gRNA expression plasmid with different spacer sequence, (Addgene # 47511, noted above)
- 0.5 μL Phusion polymerase

10. Mix, quickly spin, and run a thermocycler program with the following cycling conditions:

1 cycle:	30 sec	98°C	(initial denaturation)
----------	--------	------	------------------------

11. Run

~35 cycles:	10 sec	98°C	(denaturation)
	20 sec	65°C	(annealing)
	35-45 sec	72°C	(extension)
1 cycle:	5 min	72°C	(final extension)
	hold	12°C	(hold)

a 1 μ L aliquot

out on an agarose gel (1-1.5%) with a 1kb DNA ladder to check for a PCR product that is ~2.3 kb long (see Figure 3B for an example).

If there is no PCR product, use the GC buffer or additives such as DMSO or formamide (5%). If there are unwanted products, increasing the annealing temperature or decreasing the primer concentration can improve specificity. Gel extracting the band of interest is also possible during step 13.

12. Add 1 μ L DpnI to the PCR product, mix, and spin. Incubate at 37°C for 1 hr.

13. Purify PCR products using the Qiagen PCR clean-up protocol.

Using a vacuum manifold followed by thoroughly drying the column in a centrifuge is suggested.

14. Quantify the concentration of the PCR product using a Nanodrop. Then, prepare a 20 μ L ligation reaction to circularize the linear PCR product, which contains the new spacer sequence and 5' phosphate groups. In a PCR tube, combine:

~50 μ g PCR product after clean-up (from previous step)

X μ L nuclease-free water (fill to a total volume of 20 μ L)

10 μ L 2X Quick Ligase buffer

1 μ L Quick Ligase

15. Mix, quickly spin, and incubate at room temperature for 5-10 min.
16. Using sterile technique, transform \sim 5 μ L of the ligated product into competent bacterial cells of choice. We recommend Alternate Protocol 1 described by Seidman and coworkers (Seidman et al., 2001). Plate 10-50 μ L cells on an agar plate containing 100 μ g/mL ampicillin or carbenicillin in a manner that will yield single colonies, and grow at 37°C overnight.

The BE and gRNA plasmids we recommend confer ampicillin/carbenicillin resistance. Expect \sim 20-200 colonies, with the majority containing the desired spacer sequence replaced.

17. Pick 1-2 colonies per clone and grow in liquid LB or 2xYT media supplemented with 100 μ g/mL ampicillin/carbenicillin until saturated.

Using cells that double rapidly, such as Mach1, allows for short growth periods (\sim 6 hrs). Inoculating colonies in 1-3 mL of media allows for sequencing on the same day. Save the plates airtight at 4°C (using Parafilm) and mark the colonies screened.

18. Mini-prep saturated cultures.

We recommend saving the original culture tubes with un-lysed bacteria at 4°C to create starter cultures for correct clones

19. Quantify the prep using a Nanodrop, and then run an aliquot on an agarose gel (1.5%) to check the quality and size of the plasmid (see Figure 3C for an example).

The supercoiled gRNA plasmid runs around ~1.5_kb on a linearized ladder.

20. Send in preps for Sanger sequencing using gRNA sequencing primer #7 in Table 1 (U6), to confirm the presence of the new spacer sequence.

- *Sequence the entire backbone using primers #8-10 in Table 1 (at this step or after step 26, the midi-prep) as mutations may occur during PCR. Note that the universal U6 promoter primer that many sequencing companies use anneals very close to the spacer region and may not provide high-quality sequencing data of this region of the plasmid.*

Prepare High Quality, Endotoxin-free gRNA and BE plasmid DNA

21. Once the base editor plasmids are obtained from Addgene, streak the surface of the bacterial stab onto an LB or 2xYT plate containing 100 µg/mL ampicillin or carbenicillin. Use sterile technique and streak for single colonies. Incubate at 37°C overnight (see <https://www.addgene.org/recipient-instructions/myplasmid/>).
22. In the morning, grow 0.5 mL starter cultures of LB or 2xYT media with 100 µg/mL ampicillin or carbenicillin for each BE and gRNA plasmid required, at 37°C for ~6 hr or until saturated. For BEs, inoculate a single colony from the plate in the previous step. For gRNAs, add antibiotic-containing media to culture tube saved at step 18 or inoculate from the original plate to avoid re-transforming.
23. Dilute into 50 mL for midi-prep (or 200 mL for maxi-prep) cultures for overnight growth.
24. Once saturated, remove 0.5 mL of culture and combine with 0.5 mL 50% glycerol in a 2 mL cryogenic vial. Freeze stock at -80°C.

25. Use the remainder of the culture to prepare endotoxin-free midi- or maxi-preps based on the manufacturer's protocol.

If solution is cloudy after the syringe filter step of the midi- or maxi-prep, centrifuge the 50 mL tube for 10 min at 4,000 rcf and proceed with supernatant. Using a vacuum manifold is suggested. Drying the wash buffer from the column is critical (spin in microcentrifuge for 2-3 min at $\geq 12,000$ rcf). At the elution step, pre-incubating the buffer at 55°C for 3-5 min increases DNA yield.

26. Quantify the plasmid sample using a Nanodrop and run an aliquot out on an agarose gel to confirm the quality of midi- or maxi-prep (Figure 3C).

BEs plasmids are high-copy while the gRNA plasmid is medium- to low-copy.

BASIC PROTOCOL 2: TRANSFECTING ADHERENT CELLS AND HARVESTING GENOMIC DNA

Before proceeding to generate isogenic cell lines, it is highly recommended to validate the BE:gRNA combination(s) from Basic Protocol 1. The most time- and resource-effective way to accomplish this, is with an easy-to-transfect cell line, such as Human Embryonic Kidney 293 (HEK293) cells. While many different methods are available to deliver the BE and gRNA into cells (such as viral transduction, mRNA transfection or electroporation, and ribonucleoprotein transfection or electroporation), the protocol outlined here utilizes commercial cationic lipid reagents to transfect cells with plasmids encoding the BE and gRNA. This basic protocol details how to deliver transfection-quality plasmids into HEK293 cells and extract the genomic DNA (gDNA) for downstream sequencing.

We utilize HEK293T cells for base editing experiments. HEK293T cells contain an endogenous copy of the SV40 large T antigen. As such, plasmids containing an SV40 origin of replication will be replicated

in HEK293T cells, resulting in longer-term expression of recombinant proteins, which can cause higher levels of off-target editing with genome editing agents. Because of that, the plasmids used by the authors do not contain an SV40 origin of replication.

NOTE: All culture incubations should be performed in a humidified 37°C, 5% CO₂ (g) incubator. Use 10% FBS in DMEM with or without added pen/strep for all steps as indicated. All experimental work in this section must be done using sterile reagents and proper aseptic techniques (Phelan, 2006). Failure to do so, will result in possible contamination, which will disrupt downstream workflow and cause wasting of valuable laboratory resources.

Materials:

- HEK293T cells (ATCC CRL-3216)
- Nuclease-free water (VWR, cat. no. 82007-328)
- Endotoxin-free base editor and gRNA plasmids from Basic Protocol 1
- 48-well plates (VWR, cat. no. 10062-898)
- Lipofectamine 2000 (Life Technologies, cat. no. 11668-019)
- Opti-MEM Reduced Serum Medium, Gibco (Life Technologies cat. no. 31985070)
- Phosphate-buffered saline (PBS; Gibco Life, cat. no. 10010049)
- Dulbecco's Modified Eagle's Medium (DMEM; Thermo Fisher Scientific, cat. no. 10566024)
- Fetal Bovine Serum (FBS; Life Technologies, cat. no. 10437-028)
- 1.5 mL nuclease-free microcentrifuge tubes (VWR, cat. no. 87003-294)

- 10% (w/v) sodium dodecyl sulfate (SDS) stock solution (see Reagents and Solutions)
- 1M Tris HCl stock solution (pH 8.0; see Reagents and Solutions).
- Proteinase K enzyme (NEB, cat. no. P8107S)
- 0.2 mL PCR tubes (Fisher Scientific, cat. no. 07-200-259)
- 1.5 mL nuclease-free microcentrifuge tubes (VWR, cat. no. 87003-294)
- Fluorescence microscope
- 37°C incubator with humidified atmosphere of 5% CO₂ (g)

Protocol Steps

Transfecting HEK293T Cells with BE and gRNA

1. Plate HEK293T cells in medium without antibiotics such that they will be 70-85% confluent upon transfection (see Figure 5A for example).

For a 48-well plate, 100,000 cells/well (in a well consisting of 250 μL of 10% FBS in DMEM culture medium, or a cell solution of 4.0×10^5 cell/mL) is usually ready to transfect 4 hours after plating. 50,000 cells/well (in 250 μL of medium per well, or a cell solution of 2.0×10^5 cell/mL) is usually ready to transfect 16 hours after plating.

2. Warm Opti-MEM to room temperature prior to adding Lipofectamine 2000.
3. For each transfection, 200 ng of gRNA plasmid and 800 ng of BE plasmid is needed if done in a 48-well plate format. Also, include a negative control transfection sample that contains a non-targeting gRNA or no gRNA (see Commentary section, "Inclusion of proper controls"). For different transfection formats, the total amount of plasmid and

Lipofectamine will need to be appropriately scaled up or down. Prepare two different tubes per transfection:

- a. Tube A: add 200 ng of gRNA plasmid and 800 ng of BE plasmid. Dilute the DNA with Opti-MEM to a total volume of 12.5 μ L.
- b. Tube B: add 1.5 μ L of Lipofectamine 2000 and 11 μ L of Opti-MEM.

In some instances, you may be required to generate multiple edits in parallel. In this case, follow step 3a and 3b for each BE:gRNA combination.

4. Mix the contents of tube A and tube B and incubate the mixture for 15 minutes at room temperature prior to transfection.
5. Carefully add the 25 μ L mixture dropwise to the cells and place back into the incubator.

Be careful not to touch the bottom of the well with the pipet tip to avoid disturbing the cell monolayer.

6. Monitor the cells under a fluorescence microscope after 24 hr. EGFP-positive cells should be observed evenly across the surface of the plate (see Figure 5A). Transfection efficiency should be at least 70% (ideally over 90%). If not, please refer to the Troubleshooting section.

Lipofectamine can be toxic to the cells, so we recommend checking on the health status of the cells on the microscope every day. If cells appear healthy, changing medium after 48 hr (instead of 24 h) is also sufficient. If high cell toxicity is observed, as seen in Figure 5B, change to fresh DMEM medium before 24 hr.

7. After 24 hr, aspirate the old medium and gently rinse cells with 150 μ L of PBS
Be careful, as roughly adding the PBS will result in dislodging adherent cells.

8. Gently aspirate off the PBS and add 250 μL of fresh pre-warmed (37°C) culture medium (10% FBS in DMEM with 1% Pen/Strep).
9. Return to incubator for two additional days after media is replaced.

[*Figure 5 near here]

Harvesting Genomic DNA

10. Prepare cell lysis solution prior to taking cells from step 9 out of the incubator. 100 μL of cell lysis solution will be required per well when done in a 48-well plate format. As such, prepare a master mix solution containing the following: 10 mM Tris (pH 7.5), 0.05% SDS, 25 $\mu\text{g}/\text{mL}$ proteinase K.

For example, a 1-mL master mix cell lysis solution will contain the following: 983.75 μL of nuclease-free water, 10 μL of 1M Tris HCL stock solution (pH 8.0), 5 μL of 10% SDS stock solution, and 1.25 μL of Proteinase K enzyme. The Tris and SDS may be prepared in advance, but the Proteinase K must be added immediately prior to use.

11. Aspirate 250 μL of the old medium and gently rinse the cells with 150 μL of PBS. Carefully aspirate off the PBS.
12. Add 100 μL of cell lysis solution prepared in step 10 to each well and wait for 3-5 minutes, or until cells are completely dislodged from the bottom of the well.
13. Collect the cell lysate into PCR tubes.

Slightly tilt the 48-well plate. Prior to contacting the cell lysate with the pipette tip, displace 100 μL of air. Collect 100 μL of cell lysate on the first suction and transfer to PCR tubes. Pipetting the cell lysate up and down, especially vigorously, will result in

air bubbles that will reduce the amount of harvested genomic DNA. Be advised—the viscosity of genomic DNA increases the chance for cross-contamination of samples to occur.

14. Incubate the cell lysis reaction in the thermocycler:

- 37°C for 1 hr 30 min
- 80°C for 30 min
- Hold at 4°C

Keep cell lysate at 4°C and avoid freeze/thaw cycles. Amplification works best on fresh gDNA, and the lysate can be used directly or diluted (1:5-1:50) depending on the cell density before harvesting. After incubation in the thermocycler, cell lysate normally appears slightly cloudy and will be more viscous when higher amounts of gDNA are present. Optionally, a Nanodrop can be used to estimate gDNA concentration (blank with lysis solution), although impurities in the lysate reduce the accuracy of the readings.

BASIC PROTOCOL 3: GENOTYPING HARVESTED CELLS USING SANGER SEQUENCING

After transfection with the BE:gRNA combination(s), base editing efficiency needs to be assessed to determine the feasibility of generating an isogenic cell line harboring the target SNV. Generally, an editing efficiency >10% is sufficient to proceed (Figures 2C; 6A). Sanger sequencing is a cost-effective alternative to next-generation sequencing (Alternate Protocol 1) and is usually sufficient to determine the feasibility of generating cell lines with the SNV of interest. However, accurate quantification of editing efficiencies with Sanger sequencing is only possible when efficiencies are between 10-90% and when using high-quality sequencing traces (See Figure 4). Please refer to the “Using EditR to Quantify Base Editing Efficiencies” section to understand how to determine editing

efficiency using Sanger Sequencing (directly after step 4 of this protocol). If sequencing many different samples (>100), or if absolute quantification of base editing efficiencies less than 10% is required, we recommend using NGS (see Alternate Protocol 1).

Materials:

- QIAquick PCR purification kit (Qiagen, cat. no. 28106)
- Cell lysate from Basic Protocol 2
- Primers
- dNTP mix, 10mM each (NEB, cat. no. N0447L)
- Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, cat. no. F534L)
- DMSO
- 0.2 mL PCR tubes (Fisher Scientific, cat. no. 07-200-259)
- Thermocycler
- Reagents for agarose gel electrophoresis
- 100 bp DNA ladder (NEB, cat. no. N3231L)
- Nanodrop spectrophotometer
- Access to Sanger sequencing facility (e.g. GENEWIZ)
- Data analysis platform to analyze the sequencing results (e.g. Benchling)

Protocol Steps

PCR of genomic DNA

1. For each sample, combine the following reagents (in the order stated) for a 50 μ L PCR reaction to amplify the genomic locus of interest:

- X μ L Nuclease-free water (fill to a total volume of 50 μ L)
- 10 μ L GC Buffer
- 1 μ L dNTP mix (10 mM each)
- 0.5 μ L 10 μ M forward primer (in this example, #11 **Table 1**)
- 0.5 μ L 10 μ M reverse primer (in this example, #12 **Table 1**)
- 1.5 μ L 100% DMSO
- 0.5 μ L cell lysate (gDNA) from step 14 of Basic Protocol 2
- 0.5 μ L of Phusion DNA polymerase

It is imperative to include a negative control sample every time amplification from gDNA is performed. In this sample, 0.5 μ L of water is added instead of the cell lysis solution (gDNA) to control for reagent contamination. We suggest making a master mix, then adding the cell lysate and polymerase to individual aliquots.

2. Mix, quickly spin, and run in a thermocycler with the following cycling conditions:

1 cycle:	30 sec	98°C	(initial denaturation)
25 to 34 cycles*:	10 sec	98°C	(denaturation)
	25 sec	65°C**	(annealing)
	10 sec	72°C	(extension)
1 cycle:	5 min	72°C	(final extension)
1 cycle:	∞	12°C	(hold)

*See “Quantification of base editing efficiency in bulk cells” in the Critical Parameters section for information regarding determining the correct number of cycles to use during gDNA amplification

**The annealing temperature is specific to the primers chosen. In this protocol, it is recommended that researchers use an automated primer design tool (Step 6A of Basic Protocol 1). Using the Benchling Primer3 design wizard, the default optimal annealing temperature is 65°C. If the primer pair’s annealing temperature is not 65°C, please adjust accordingly.

- Run a 5 µL aliquot of each PCR reaction on a 2-3% agarose gel using a 100 bp ladder for size comparison (see Figure 4B for an example).

If you do not get the correct-size product bands, please refer to the Troubleshooting section.

4. After confirming the presence of a correctly-sized PCR product, prepare the samples for sequencing. Genewiz (<https://www.genewiz.com/en>) offers two Sanger sequencing options, but this may depend on the sequencing vendor of choice:

a. Sending purified PCR product: Purify the product with QIAquick PCR purification kit following the manufacturer's instructions. Measure the DNA concentration after PCR clean-up of each sample using a Nanodrop. Prepare the appropriate amount of DNA along with the sequencing primer (we recommend using either the forward or reverse primer from the PCR reaction that is positioned at least 50-75 bp away from the target base of interest) in the same tube according to company protocol.

b. Sending unpurified PCR product: unpurified PCR samples sent for Sanger sequencing will undergo an enzymatic PCR purification protocol. Provide the appropriate amount of the successful PCR reaction in one tube according to sequencing company guidelines.

Provide the appropriate amount of sequencing primer (we recommend using either the forward or reverse primer from the PCR reaction, whichever one is at least 75-bp away from the target base of interest) in a separate tube. We recommend submitting a picture of the labelled gel as well to the sequencing company as this can help with quantification and troubleshooting, if necessary.

Sending unpurified PCR product (step 4b.) usually results in better sequencing coverage quality and thus, an easier to interpret chromatogram, but costs extra.

Using EditR to Quantify Base Editing Efficiencies

Quantifying base editing efficiency of the target nucleotide is needed either to (a) determine the feasibility of generating a cell line with a given SNV as seen in Basic Protocol 3, or (b) confirm the

introduction of either a homozygous edit or heterozygous edit into mammalian cells as seen in Basic Protocol 5. In both instances, we recommend measuring the base editing efficiency of the total harvested cells in bulk using a simple and publicly available program, "EditR." (https://moriaritylab.shinyapps.io/editr_v10/; Kluesner et al., 2018). EditR is a free online tool or desktop application which requires an .ab1 Sanger sequencing file of the potentially edited region (~300-700bp) and the gRNA protospacer sequence (~20bp) to predict where a base edit occurred. Once the .ab1 file and DNA protospacer sequence are correctly uploaded, EditR generates a plot displaying editing efficiencies at each base within the protospacer (Figure 6D and 6E).

NOTE: If we state, "an editing efficiency of ~10%", this means that ~10% of the cells of the total harvested cell population have a successful edit at the target base. This is will be indicated by overlapping chromatogram peaks at the targeted nucleotide. In this hypothetical example, it would show ~90% unedited base and ~10% edited base in the sequencing chromatogram. Using MUTYH V270 as an example, Figure 6A and 6D shows a chromatogram where the base edit efficiency is reported as 38% (P = .01) by EditR. It is important to note that a poor-quality Sanger sequencing chromatograph (such as that shown in Figure 6C) may produce an EditR output file that falsely shows editing at the target base (see Figure 6E). It is, therefore, important to confirm that non-A/C bases in the protospacer display less than ~7% editing.

The following steps have been adapted from the online protocol found at https://moriaritylab.shinyapps.io/editr_v10/:

[*Figure 6 near here]

5. Upload your .ab1 file of the sequenced region.
6. Enter the gRNA sequence protospacer sequence.

- If your gRNA is antisense to the .ab1 file, check the “Guide sequence is reverse complement” box.
7. Click the “Predicted Editing” tab.
 8. Examine the gRNA protospacer chromatogram and underlying tile plot to determine if base editing occurred. Observing >10% editing efficiency at the target base (as indicated by a double peak at the targeted nucleotide showing <90% unedited base and >10% edited base in the sequencing chromatogram) is sufficient to proceed (Figure 6A).
 - All colored tiles represent base calls that are deemed significant, i.e. if there are multiple colored tiles under a single base call, base editing likely occurred.
 9. If you wish to download a report of the operations performed on your data, click the "Download Report" tab on the top of the page.

ALTERNATE PROTOCOL 1: NEXT-GENERATION SEQUENCING TO QUANTIFY BASE EDITING

Some base editing experiments require rigorous quantification to accurately determine absolute editing efficiencies or to deconvolute editing patterns in bulk. Next-generation sequencing (NGS) enables the researcher to quantify individual edited alleles, which can be helpful to establish the frequency of bystander mutations in bulk before isolating single cells. NGS is also desirable when genotyping edited cell lines at target sites with increased copy number (due to gene duplication) or if robust quantification of indels is needed. In these cases, the harvested genomic DNA samples from Basic Protocol 2 or Basic Protocol 5 should be amplified and prepared for Illumina NGS. Targeted amplicon sequencing for genome editing experiments has been previously described in depth by Veeranagouda and coworkers, as well as by Yang and coworkers (Veeranagouda, et al., 2018; Yang, et al., 2014). We recommend quantifying base editing using a 300-cycle, paired-end NGS run with a 200-250 bp amplicon. However, the amplicon length

can be easily altered for other types of sequencing runs. Primers for the initial amplification (round 1) should be designed with the proper adapter sequence and distance from the target site, as described in step 6B of Basic Protocol 1, and Figure 4A. Then, sample barcodes (for de-multiplexing) and Illumina-specific P5/P7 tail sequences will be added during the round 2 PCR. Once the data is acquired, we suggest performing data analysis with the free CRISPResso2 software using the batch mode and base editor output (Clement et al., 2019). See the Critical Parameters section for additional considerations.

Materials:

- Cell lysate from Basic Protocol 2
- Primers for round 1 PCR (IDT or any other DNA provider)
- Primers for round 2 PCR (IDT or any other DNA provider) from **Table 3**, or TG Nextera® XT Index Kit v2 (Illumina, cat. no. TG-131-2001)
- dNTP mix, 10mM each (NEB, cat. no. N0447L)
- Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, cat. no. F534L)
- DMSO
- Nuclease-free water (VWR, cat. no. 82007-328)
- 0.2 mL PCR tubes (Fisher Scientific, cat. no. 07-200-259)
- Thermocycler
- Reagents for agarose gel electrophoresis
- 100 bp DNA ladder (NEB, cat. no. N3231L)

- QIAquick PCR purification kit (Qiagen, cat. no. 28106)
- Qubit fluorometer with dsDNA HS assay kit (Q32854), or qPCR with NEBNext Library Quant Kit for Illumina (E7630L)

[*Table 3 near here]

Protocol Steps

Next-Generation Sequencing

1. To amplify the genomic locus of interest, create a 25 μ L round 1 (rd1) PCR reaction for each gDNA sample (BE:gRNA combination), and a negative control containing water. For each sample, combine the following reagents (in the order stated):

- X μ L Nuclease-free water (fill to a total volume of 25 μ L)
- 5 μ L GC Buffer
- 0.5 μ L 10 mM dNTPs
- 0.5 μ L 10 μ M forward NGS rd1 primer (for the MUTYH V270 example, primer #15
Table 1)
- 0.5 μ L 10 μ M reverse NGS rd1 primer (for the MUTYH V270 example, primer #16
Table 1)

- 0.75 μ L 100% DMSO
- 0.5 μ L cell lysate (gDNA)
- 0.25 μ L of Phusion DNA polymerase

It is imperative to include a negative control sample every time amplification from gDNA is performed. In this sample, 0.5 μ L of water is added instead of the cell lysis solution, to control for gDNA contamination. We suggest making a master mix, then adding the cell lysate and polymerase to individual aliquots.

- Mix, quickly spin, and run in a thermocycler with the following cycling conditions:

1 cycle:	60 sec	98°C	(initial denaturation)
22 to 28 cycles*:	10 sec	98°C	(denaturation)
	20 sec	65°C	(annealing)
	10 sec	72°C	(extension)
1 cycle:	5 min	72°C	(final extension)
1 cycle:	∞	12°C	(hold)

**To avoid PCR bias, use the minimum number of cycles that provides robust amplification. This may require optimization for each target locus (see also “Quantification of base editing efficiency in bulk cells” in the Critical Parameters section).*

- Run a 1-2 μ L aliquot of each PCR reaction on a 2% agarose gel using a 100 bp ladder (see Figure 4C for an example).

NGS round 1 adapter sequences add 66 bp to the length of your amplicon. If you do not get the correct-size product bands, please refer to the Troubleshooting section.

4. Set up the round 2 (rd2) PCR. This second round adds a unique barcode designated by an 8-nt sequence in each of the primers: A1 in this example (Fwd-A/Rev-1). For each rd1 sample, create a 25 μ L rd2 PCR reaction by combining the following reagents (in the order stated):

- X μ L Nuclease-free water (fill to a total volume of 25 μ L)
- 5 μ L GC Buffer
- 0.5 μ L 10 mM dNTPs
- 0.5 μ L 10 μ M forward NGS rd2 primer (custom barcode; #1 **Table 3**, A for example)
- 0.5 μ L 10 μ M reverse NGS rd2 primer (custom barcode; #13 **Table 3**, 1 for example)
- 0.75 μ L 100% DMSO
- 0.5 μ L rd1 PCR product
- 0.25 μ L of Phusion DNA polymerase

The negative control sample in this PCR should contain 0.5 μ L of the round 1 negative control PCR product. Barcodes can be custom-generated and input into the sequencer, such as those provided in Table 3, or ordered in a kit through Illumina (TG Nextera[®] XT Index Kit v2).

5. Mix, quickly spin, and run in a thermocycler with the following cycling conditions:

1 cycle:	60 sec	98°C	(initial denaturation)
8 to 16 cycles*:	10 sec	98°C	(denaturation)
	20 sec	65°C	(annealing)
	10 sec	72°C	(extension)
1 cycle:	5 min	72°C	(final extension)
1 cycle:	∞	12°C	(hold)

**Use the minimum number of cycles that provides robust amplification.*

- Run a 1-2 μ L aliquot of each PCR reaction on a 2% agarose gel using a 100 bp ladder (see Figure 4C for an example).

NGS round 2 adapter sequences add 74 bp to the length of your rd1 PCR product (or a total of 140 bp to your original amplicon).

- Pool rd2 PCR products from all samples together (e.g. no gRNA control and each BE:gRNA combination at the locus of interest). Perform a gel extraction on the pooled samples to eliminate lower molecular weight products that would decrease the quality of the NGS data.

We recommend doing a second PCR clean-up on the elution to remove any agarose (peak absorbance at 230nm) that remains after gel purification. A Nanodrop can be used to determine purity and roughly quantify concentration. Then, dilute into the quantification range needed for the next step.

8. Accurately quantify the DNA concentration of the libraries following the Qubit dsDNA HS assay kit protocol.

Alternatively, the NEBNext Library Quant Kit for Illumina protocol describes quantification by qPCR.

9. Set up the NGS run according to Illumina protocols or submit to a sequencing core facility.

We recommend analyzing data with CRISPResso2 (Clement et al., 2019), discussed in the Critical Parameters section.

BASIC PROTOCOL 4: SINGLE CELL ISOLATION OF BASE EDITED CELLS USING FACS

Following confirmation of >10% editing efficiency, the next step is to isolate single cells and clonally expand them, to obtain isogenic cell lines harboring the SNV of interest. In this article, we describe two different methods for doing this: one utilizes fluorescence-activated cell sorting (FACS) (this protocol), and the other, utilizes dilution plating (Alternate Protocol 2). Using FACS followed by clonal expansion is preferred over dilution plating for two key reasons. Firstly, FACS allows for the discrimination between single cells and multipllets, which eliminates the chances of obtaining doublet cell-derived colonies (a frequent observation seen in dilution plating, Figure 7E). Secondly, FACS allows for the discrimination between untransfected and transfected cells when using plasmids with fluorescent markers, which is particularly important when working with cell lines with low transfection efficiencies. Dilution plating, however, is more cost-effective, gentler on the cells, and does not require specialized instrumentation. Dilution plating, therefore, may be preferred over FACS when working with sensitive cells or if the researcher does not have access to FACS instrumentation (or simply when working with plasmids which do not possess fluorescent markers).

NOTE: We used ABE_{max}-NG-GFP (Addgene #140005) to generate the V270A MUTYH variant used in this example. In this plasmid, although EGFP and ABE are both transcribed in a single mRNA transcript, they are translated into unique and separate proteins via the “self-cleaving” P2A linker. Therefore, FACS can be used to sort individual EGFP-positive cells (Figure 2E), which selects cells that are actively expressing the BE. There are also methods to select for cells with high BE activity (rather than simply expression); see the Critical Parameters section for more details on this.

Materials:

- HEK293T cells (ATCC CRL-3216)
- Nuclease-free water (VWR, cat. no. 82007-328)
- Endotoxin-free base editor plasmids and cloned gRNA plasmids from **Basic Protocol 1**
- 48-well plates (VWR, cat. no. 10062-898)
- 96-well plates (VWR, cat. no. 10062-900)
- Lipofectamine 20000 (Thermo Fisher Scientific, cat. no. 11668-019)
- Opti-MEM Reduced Serum Medium, Gibco (Life Technologies cat. no. 31985070)
- Phosphate-buffered saline (PBS; Gibco, cat. no. 10010049)
- Dulbecco’s Modified Eagle’s Medium (DMEM; CoreBio, cat. no. 10566-024)
- Fetal Bovine Serum (FBS; Thermo Fisher Scientific, cat. no. 10437-028)
- Pen/Strep (Gibco, cat. no. 15070063)
- 1.5 mL nuclease-free microcentrifuge tubes (VWR, cat. no. 87003-294)

- 5 mL round bottom polystyrene test tube, with cell strainer snap cap (Thermo Fisher Scientific, cat. no. 352235)
- Sterile 15 mL conical tubes
- Accumax (Accutase, cat. no. AM105)
- Propidium Iodide (PI; Thermo Fisher Scientific, cat. no. P3566)
- Fluorescence microscope
- Centrifuge
- 37°C incubator with humidified atmosphere of 5% CO₂ (g)

Protocol Steps

Transfect for Single Cell Isolation

1. Transfect the HEK293T cells with the BE:gRNA combination(s) that show desired editing activity in bulk, as described in Basic Protocol 2 steps 1 to 8. Incubate the cells for three days post transfection.
It is important to include a proper negative control sample (see Critical Parameters), such as a non-targeting gRNA or no gRNA.
2. 1-2 days after the transfections, prepare 96-well plates containing 100 µL of culture medium in each well (50% FBS, 1% Pen/Strep). Generally, 2 plates per well in step 1 is sufficient to obtain cells containing the SNVs of interest, as referred to in the Understanding Results section. Place the plates into the tissue-culture incubator overnight.

Single transfected cells from step 1 will be sorted into individual wells of the 96-well plate. For generating single nucleotide variant clonal cell lines, two 96-well plates per BE:gRNA combination containing DMEM with 50% FBS, 1% Pen/Strep is used but other sorting culture medium is also commercially available. Preparing extra collection plates with medium is optional.

Preparation of the Cells for FACS

3. Three days post-transfection, aspirate 250 μ L of the old medium from the transfected cells and gently rinse cells with 150 μ L of PBS. Carefully aspirate off the PBS.
4. Add 300 μ L of Accumax to each well and incubate at 37° for 5 to 10 min, or until cells are completely dislodged from the bottom of the well.

The use of trypsin is not advised to prepare cells for sorting as it is too harsh and will result in fewer cells surviving the clonal expansion process.

5. Create a suspension by gently pipetting up and down several times.
6. Transfer the cell suspension to a 15-mL conical tube, then centrifuge for 5 min at 100 RCF at room temperature. Aspirate supernatant.
7. Resuspend each sample with 1 ml of PBS supplemented with 0.5 μ l of the viability dye PI.

Take an aliquot and count the cells using a hemocytometer.
3-7 million cells per mL is recommended. This target concentration should be met if the well surface is near confluency.

Although not a requirement, PI will improve efficiency of obtaining single cell clones when sorting.

8. Filter cells through a sterile 35 μ m cell strainer to collect a uniform suspension in a 5 mL polystyrene round-bottom tube. Place on ice until sort.
9. Using the FACSria II system (or equivalent), and under sterile conditions, sort single EGFP-positive cells into individual wells of the 96-well plates containing 100 μ L of culture media containing DMEM with 50% FBS, 1% Pen/Strep.
10. Place plates into tissue-culture incubator as soon as possible after sorting.

ALTERNATE PROTOCOL 2: SINGLE CELL ISOLATION OF BASE EDITED CELLS USING DILUTION PLATING

Dilution plating offers a cost-effective alternative when researchers do not have access to a sorting machine or when they are working with plasmids which do not possess fluorescent markers. Furthermore, dilution plating might be preferred when working with sensitive cell lines that are less likely to survive the sorting conditions. However, there are two major limitations to dilution plating. Firstly, the chances of obtaining doublet cell-derived colonies increases. We have found, however, that using a 35 μ m sterile cell strainer prior to dilution plating reduces the chances of obtaining doublet-derived colonies. Secondly, dilution plating does not allow for the discrimination between untransfected and transfected cells, something that is possible when using FACS. Therefore, the rates of successfully obtaining a cell line harboring the SNV of interest are reduced (please refer to the Understanding Results section). Therefore, screening more single cell-derived colonies per transfection may be necessary to obtain the cell line of interest when using this method. This can be accomplished by using more than two plates per transfection. This protocol uses an adaptation of methods previously described (Giuliano, et al., 2019).

Materials:

- 96-well plates (VWR, cat. no. 10062-900)
- Dulbecco's Modified Eagle's Medium (DMEM; CoreBio, cat. no. 10566-024)
- Fetal Bovine Serum (FBS; Thermo Fisher Scientific, cat. no. 10437-028)
- Pen/Strep (Gibco, cat. no. 15070063)
- 35µm sterile cell strainer snap cap (Life Sciences, cat. no. 352235)
- 8-channel 200µL micropipettor
- Reagent dispensing reservoir/tray (VWR, cat. no. 89094-662)
- TrypLE Express (Thermo Fisher Scientific, cat. no. 12605028)
- Transfected HEK293T cells from Basic Protocol 2
- 37°C incubator with humidified atmosphere of 5% CO₂ (g)

Protocol Steps

Single Cell Isolation Using Dilution Plating

1. Add 100 µL of pre-warmed (37°C) DMEM medium (10% FBS, 1% Pen/Strep) to all the wells in a 96-well plate, except to well A1.

In general, we recommend doing four plates per SNV when the editing efficiency from Basic Protocol 3 is > 20 %. More plates may be needed per SNV when editing efficiency is < 20%.

2. After the cells from Basic Protocol 2 step 6 have been transfected and incubated for three days, pipette off 250 μL of the old medium from the cells and gently rinse cells with 150 μL of PBS. Carefully aspirate off the PBS.
3. Add 50 μL of TrypLE to each well and incubate at 37°C for 5 to 10 min, or until cells are completely dislodged from the bottom of the well.
4. Resuspend the transfected cells with culture medium.
5. Add 200 μL of the cell suspension from step 4 to well A1 through a 35 μm sterile cell strainer.

The use of the strainer is optional, but when used, we have observed a higher frequency of isolating single cell-derived colonies as opposed to doublet cell-derived colonies.
4. Using a single channel pipette, make 1:2 dilutions by transferring 100 μL of the cell suspension from A1 down the first column (B1 to H1) using the same tip. Discard 100 μL of cells from the last well.

Mix gently before each transfer.
5. Add an additional 100 μL of medium to each well in column 1.
6. Using an 8-channel pipette, make 1:2 dilutions by transferring 100 μL of the cell suspension across each column of the plate starting from column 1 and ending at column 12. Discarding 100 μL of cell suspension from the last column is optional.
7. Place plates into tissue-culture incubator undisturbed after dilutions are made. Observe cells 4-7 days later (Figure 7).

[*Figure 7 near here]

BASIC PROTOCOL 5: CLONAL EXPANSION TO GENERATE ISOGENIC CELL LINES AND GENOTYPING OF CLONES

Regardless of which method the researcher uses to isolate single cell-derived colonies (FACS —Basic Protocol 4— or dilution plating —Alternate Protocol 2—), colony formation of HEK293T cells should be apparent after 4-7 days and they should be ready to subculture 10-14 days after the isolation protocol. Figures 7B and 7D provide examples of single cell-derived colonies. This Basic Protocol describes the process for clonally expanding the isolated cells from Basic Protocol 4 into isogenic cell lines, genotyping the resultant lines, and subculturing the appropriate colonies into larger flasks for storage or downstream experiments. For the SNV of interest, aim to obtain at least three different clones for each genotype (i.e. three harboring a wild-type genotype, three harboring a heterozygous genotype, and three harboring a homozygous genotype). All these cell lines can be used for comparison and in downstream experiments (Figure 6B).

Materials:

- 48-well plates (VWR, cat. no. 10062-898)
- Cells from Basic Protocol 4, step 10 or Alternate Protocol 2, step 7
- T-75 flask (Victorio, cat. no. 10062-860-KLM)
- Phosphate-buffered saline (PBS; Gibco, cat. no. 10010049)
- Dulbecco's Modified Eagle's Medium (DMEM; CoreBio, cat. no. 10566-024)
- Fetal Bovine Serum (FBS; Thermo Fisher Scientific, cat. no. 10437-028)

- Pen/Strep (Gibco, cat. no. 15070063)
- 37°C incubator with humidified atmosphere of 5% CO₂ (g)
- TrypLE Express (Thermo Fisher Scientific, cat. no. 12605028)
- Microscope
- 37°C incubator with humidified atmosphere of 5% CO₂ (g)

Protocol Steps

Subculturing and Genotyping Single Cells

1. Carefully inspect each well of the plates from either Basic Protocol 4 or Alternate Protocol 2 under the microscope and circle the wells that are harboring single cell-derived clones (usually, it takes around 3-4 days to notice distinguishable colony formation, but it can also be longer depending on clonal expansion rate, see Time Considerations). Be wary of potential doublet cell-derived colonies (see Figure 7 for examples).

We have found that each 96-well plate will yield around 8 colonies per plate when prepared according to the FACS method, and around 11 single cell-derived colonies per plate when prepared according to the dilution method.

2. Once the colony covers at least 30% of the well's surface area, pipette off 100 µL of the old medium and gently rinse the cells with 50 µL of PBS.

Since clonal expansion rate may differ from clone to clone, we recommend checking the colonies daily after they have been identified in step 1 of Basic Protocol 5. This step usually takes 1-2 weeks but can vary. Please see Time Considerations.

3. Add 30 μL of TrypLE Express. Wait 3-5 minutes, or until cells are completely dislodged from the bottom of the well.
4. While cells are being trypsinized, add 225 μL of pre-warmed (37°C) culture medium (10% FBS, 1% Pen/Strep) to each well of a separate 48-well plate. Allocate two wells for each single cell-derived colony.

Half of the cells will be clonally expanded while the other half will be used for genotyping.
5. Resuspend the trypsinized cells with 120 μL of pre-warmed (37°C) DMEM medium (10% FBS, 1% Pen/Strep).
6. Passage two individual 75 μL aliquots of the clonal cell suspension into each well from step 4.
7. Allow the cells to reach 80-90% confluency before proceeding to step 8. This will usually take 3 days, but it is recommended to check the cells under the microscope every day until they reach the appropriate confluency.
8. Two simultaneous steps must be followed. Thus, proper labelling of wells for identification of which clone it corresponds to is crucially important:
 - a. Once both wells have reached the appropriate confluency, take one of the wells, and harvest the genomic DNA and genotype, as previously described in Basic Protocols 2 and 3 or Alternate Protocol 1.
 - b. For the other well containing the same clone, continue passaging the cells into larger wells or flasks before storage and until the sequencing data from step 8a is obtained. We

recommend moving from a 48-well plate (250 μ L of DMEM medium with 10% FBS, 1% Pen/Strep), to a 6-well plate (1 mL of DMEM medium with 10% FBS, 1% Pen/Strep), then to a T25 flask (5 mL of DMEM medium with 10% FBS, 1% Pen/Strep), and finally to a T75 flask (10 mL of DMEM medium with 10% FBS, 1% Pen/Strep). Passage the cells when they are 85-90% confluent; it is, therefore, recommended to monitor the cells under the microscope every day. After obtaining the genotyping results, we recommend keeping at least three clones, three cell lines containing wild-type, three clones containing heterozygous, and three clones containing homozygous genotypes (Figure 6B). Please refer to the Understanding Results section.

9. Continue to clonally expand the sequenced-validated cell lines by passaging the cells into larger wells or flasks as indicated in step 8b. We additionally recommend preparing a fourth cell line per each genotype that is a mixture of each of the clones with that specific genotype.
10. Cryopreserve cell lines.

There are many protocols that show researchers how to cryopreserve mammalian cell lines (Stacey & Masters, 2008; Yokoyama et al., 2012). We also recommend using <https://www.abcam.com/protocols/cryopreservation-of-mammalian-cell-lines-video-protocol> as a resource.

REAGENTS AND SOLUTIONS

Ampicillin or carbenicillin 1000X stock solution (100 mg/mL; 5 mL)

Dissolve 500 mg of powdered ampicillin or carbenicillin sodium salt (such as GoldBio, cat. no. A-301 or C-103) in 5 mL of deionized water to create a 100 mg/mL stock solution. Store at -20°C. Note: carbenicillin is more stable but more expensive.

Agar plates and liquid media with ampicillin or carbenicillin, final concentration 100 µg/mL

Prepare LB or 2xYT agar media or LB or 2xYT liquid media as previously described (Elbing & Brent, 2019). Add the ampicillin/carbenicillin stock solution 1000X to a final concentration of 100 µg/mL before pouring plates or inoculating liquid cultures. For example, add 500 µL of a 100 mg/mL antibiotic solution to 500 mL of agar-containing media, to make ~20 plates with a final concentration of 100 µg/mL, with 25 mL of media each. When preparing plates, wait until the molten agar media has cooled down to ~55°C before adding the antibiotic.

10% (w/v) SDS stock solution (100 mL)

Dissolve 10 g of SDS (Thermo Fisher Scientific, cat. no. 15525017) in 80 mL of deionized water. Bring final volume to 100 mL with deionized water. This stock solution is stable for 6 months at room temperature.

Tris HCL (1M) stock solution (pH 8.0; 1 Liter)

Dissolve 121.14 g Tris (Thermo Fisher Scientific, cat. no. PI17926) in 800 ml of deionized water. Adjust pH to 8 with 1 M NaOH. Bring final volume to 1 liter with deionized water. Autoclave and store at room temperature.

COMMENTARY

BACKGROUND INFORMATION:

Base editing is a technique capable of precisely and efficiently introducing point mutations in mammalian cells by relying on nucleobase chemistry rather than on the introduction of double-stranded DNA breaks (DSBs) (Hess et al., 2017; Ran et al., 2013). The system utilizes the targeting ability of the Cas9:gRNA complex but avoids the pitfalls of DSB-mediated genome editing, as discussed below.

All targeted nuclease systems, such as ZFNs, TALENs, and Cas enzymes rely on phosphodiester backbone cleavage and subsequent DNA repair either to create random insertions and deletions (indels) through the non-homologous end-joining (NHEJ) pathway (Hsu, et al., 2014), or for the incorporation of a donor DNA sequence containing the mutation of interest via homology-directed repair (HDR) (M. Liu et al., 2019). While NHEJ is active in all phases of the cell cycle, HDR normally requires a sister chromatid as a template and is, therefore, only active during the late S and G2 phases, making HDR-mediated genome editing only possible in actively dividing cells. Thus, attempting to create targeted point mutations using traditional DSB-mediated genome editing inevitably leads to low levels of accurate SNV introduction (typically, less than 1%, unless cell cycle or DNA repair perturbations are employed, or silent “blocking mutations” are incorporated (Paquet et al., 2016; Riesenberger et al., 2019; Yeh et al., 2019). Further, DSBs can be toxic to cells and cause p53-dependent apoptosis (Haapaniemi et al., 2018) or large genomic rearrangements (Kosicki et al., 2018), which is exacerbated during multiplexed genome editing (i.e. the simultaneous targeting of distinct loci) (Shin et al., 2017; Webber et al., 2019). This results in decreased viability, the requirement for screening hundreds of clones, the introduction of genomic modifications other than the modification of interest, and the potential enrichment for cells with p53 mutations (the latter two of which can cause artifacts in phenotypic characterization). Recently, deleterious on-target effects, such as large deletions or loss-of-heterozygosity, were shown to

escape standard genotyping methods in up to 40% of edited clones (where hemizygous ones appeared homozygous) (Weisheit et al., 2020).

A common method to facilitate the enrichment and selection of genome edited cell lines rely on knock-in strategies of selectable markers through homologous recombination, such as antibiotic resistance or fluorescent markers (Martin et al., 2019; Norris et al., 2015; Ochiai et al., 2014; Sur et al., 2012). However, removing these integrated selection markers without permanently damaging or “scarring” the host genome is non-trivial and can sometimes require multiple clonal expansion events (H et al., 1993; Watkins et al., 2014). Another alternative method to DSB-dependent editing strategies is the use of adeno-associated-viral (AAV)-mediated insertion of genomic cassettes into human somatic cells (Di Nicolantonio et al., 2008; S et al., 2009). However, the packaging limit for AAVs is 4.7kb (smaller than some genes of interest), and as this involves ectopic expression of the gene of interest, it is less physiologically-relevant than editing of the endogenous locus (Gray et al., 2010; Kohlbrenner et al., 2005).

Base editors avoid DSBs by instead relying on the direct chemical modification of target nucleobases. When the edited base is used as a template for DNA replication or repair, a transition mutation is faithfully introduced with high efficiency (up to 80% in many cases). This addition to the genome engineering toolkit appeared in 2016 with the advent of CBEs (which introduce C•G to T•A point mutations using uracil intermediates), followed by ABEs (which introduce A•T to G•C point mutations using inosine intermediates) the following year. Because of their dependence on more ubiquitous and reliable DNA repair pathways, base editors have the added advantage of working in non-dividing cells.

However, like all current genome editing technologies, base editing has limitations. Most importantly, base editors cannot be used to introduce transversion mutations, insertions, or

deletions. Second, base editors have a strict requirement for a PAM to be positioned 12-16 nucleotides from the target base, making certain targets inaccessible. However, ongoing efforts to engineer Cas enzymes with PAMs complimentary to Cas9-NG have yielded successful base editors (Gaudelli et al., 2020; Hu et al., 2020; Y. B. Kim et al., 2017; Li et al., 2018; Z. Liu et al., 2019; Richter et al., 2020; Walton et al., 2020). Additionally, it is important to bear in mind that even with a suitable PAM, not all genomic sites are amenable to base editing. As with other forms of genome editing, the sequence or chromatin context of the target site may limit the Cas9:gRNA complex's ability to bind DNA. Some base editors additionally contain deaminase enzymes with an intrinsic sequence preference. Finally, CBEs can produce C-to-non-T edits, likely due to high efficiencies of excision of the uracil intermediate by endogenous DNA repair enzymes. A better understanding of the cellular mechanisms responsible for high BE efficiency and product purity (i.e. target conversion only into the desired base rather than unwanted bases) is needed to improve the system.

Recently described “prime” editors, a novel CRISPR-Cas9 gene editing technique, utilize a reverse transcriptase and a template sequence as a 3'-extension to the gRNA to precisely edit DNA (Anzalone et al., 2019). This tool is not limited to transition mutations, avoids any bystander nucleotides, and has more flexible PAM positioning. However, prime editing is less efficient than base editing when a target A or C can be optimally positioned within the protospacer, and requires more extensive gRNA optimization per target (Anzalone et al., 2019). Base editors are also more compact and produce fewer indels (Anzalone et al., 2019), which can be important factors with *in vivo* applications. It is important to consider the goal of your study while designing an experiment to select the most appropriate editing technique and set of tools.

CRITICAL PARAMETERS:

gRNA design considerations

Programs including the Benchling wizard (<https://benchling.com/pub/liu-base-editor>); limited to CBEs), BE-designer (<http://www.rgenome.net/be-designer/>), DeepHF (<http://www.deephf.com/>), and BE-Hive (<https://www.crisprbehave.design/>) can be used for automated gRNA design or scoring (Komor, 2016; Hwang et al., 2018; Wang et al., 2019; Arbab et al., 2020). BE-Hive incorporates ABEs and alternate PAMs, making it the most comprehensive software. This machine learning algorithm enables predictions of both editing outcome and efficiency for a given BE:gRNA combination. Other programs are available for designing and scoring gRNAs for traditional Cas9 genome editing, and have been previously reviewed (Tycko et al., 2016). These may be helpful for predicting potential off-targets or gauging how effectively Cas9 will bind to a given target site.

The human U6 promoter drastically increases expression of the gRNA if guanine is the first nucleotide transcribed. Any decrease in Cas9 binding due to a mismatch at the first position of the protospacer that this may cause is more than compensated for by the increased expression levels. The mouse U6 promoter exhibits high expression with A or G as the first nucleotide, and can be used as an alternative (Wang et al., 2019). Additionally, the *S. pyogenes* gRNA backbone contains an early transcription termination signal (UUUU) that may decrease expression levels, and can be mutated to increase cellular gRNA levels (Dang et al., 2015). While we have found that low base editing efficiency is usually caused by other factors, this may be an issue in certain cell types.

Bystander editing

Bystander edits occur when C's or A's beyond the desired target are also present in the editing window and become mutated concurrently with the target base. Bystander editing can be avoided by designing a protospacer that pushes potential bystander bases out of the editing window, as shown with the gRNA example targeting V270 with an adenine in position

9 (Figure 3A). Bystander edits may also be mitigated by using deaminase enzymes that are modified to be less processive (such as the YE1, YE2, and YEE CBEs) (Y. B. Kim et al., 2017; Z. Liu et al., 2020) or sequence-specific (such as the eA3A CBE) (Gehrke et al., 2018), or by alternative BE architectures, such as rigid linkers (Tan et al., 2019). In some cases, bystander edits are acceptable, such as in gene knockout experiments (Figure 8C) or if the bystander creates a silent mutation or intronic mutation (as long as it is not a splice site). When attempting to knock out genes via premature stop codon introduction or splice site disruption, disruptions targeted to the first and last few exons can be inefficient due to translation reinitiating and/or alternative splicing (Smits et al., 2019). We suggest base editing splice donor sites in the middle of the gene body to achieve the highest level of protein loss (Webber et al., 2019).

[*Figure 8 near here]

Base editor selection considerations

There is a large selection of base editor constructs from which to choose, and the most suitable choice depends on your experimental goal. The two editors that we recommend (BE4max-NG-P2A-EGFP and ABEmax-NG-P2A-EGFP) are good choices with which to start, and modifications can be made according to specific experimental requirements. In most cases, nCas9 (D10A) should be utilized, which will direct DNA repair machinery to use the modified base as a template, but dCas9 can be employed if indels must be avoided at all costs (we note that indel formation with CBEs is target site-dependent, and ABEs generally do not introduce indels even when using nCas9). If an NG PAM is not available, we suggest using SaCas9-KKH (which recognizes an NNNRRT PAM, but will result in a widened window which can cause bystander editing (Y. B. Kim et al., 2017) or the SpRY variant (Walton et al., 2020) instead of the SpCas9-NG. A major consideration is the ability of the deaminase enzyme portion of the BE to deaminate other free ssDNA or RNA in the cell

(Grünwald et al., 2019; Jin et al., 2019; Lee et al., 2020; Zhou et al., 2019; Zuo et al., 2019), causing off-target editing. Engineering efforts have rationally designed deaminase variants that substantially decrease off-target editing (Doman et al., 2020; Grünwald et al., 2019; Yu et al., 2020; Zuo et al., 2020) but many times do so at the expense of sequence specificity, causing decreased on-target activity at certain sites. Specifically, rAPOBEC1-YE1 and the next-generation CBEs listed in Table 2 are examples of CBEs with greatly reduced off-target DNA editing, and ABEs containing the mutations V106W, F148A, and V82G show almost no off-target RNA editing (ABEs inherently do not display off-target DNA editing) (Zhou et al., 2019). If small amounts of off-target editing are acceptable, utilizing the most efficient base editors will aid in isogenic cell line generation. ABE8s significantly increase editing efficiencies but also have a wider editing window, increasing the chance of bystander mutations.

Inclusion of proper controls

The process of clonal expansion represents an enormous genetic bottleneck and puts the cells under extreme selective pressure. This can result in genetic, epigenetic, and/or phenotypic variation of single cell-derived lines (Giuliano et al., 2019; Grav et al., 2018). Additionally, some CBEs can increase 2-fold the inherent mutation rate during isogenic cell line generation (Yu et al., 2020). It is, therefore, of the utmost importance to include proper controls to confirm that observed differences in phenotypes or protein activity are due to the mutation of interest. When generating isogenic lines, transfections should contain a negative control sample that lacks a gRNA or has a non-targeting gRNA sequence that is absent from the human genome (Table 1, #6). For each set of transfections to generate isogenic cell lines (Basic Protocol 4), single cell clones should be isolated from this negative control sample alongside the on-target gRNA samples. Using these clones as the 'wild-type' cell lines for phenotyping and SNV characterization can control for the process of clonal

expansion and potential effects due to DNA damage introduction by the BEs. Additionally, for each genotype of interest (wild-type, heterozygous, and homozygous for each SNV of interest), it is crucial to generate at least three lines derived from different clones. We also recommend generating an additional “control” line for each genotype that is a mixture of each of the individual clones with that specific genotype. Whole-genome sequencing or whole-exome sequencing of the resultant cell lines is highly encouraged to determine whether other genomic modifications might have occurred during the process of clonal expansion.

Quantification of base editing efficiency in bulk cells

After base editing has had 3-5 days to occur, select the appropriate endpoint analysis for your experiment. For bulk Sanger sequencing, EditR can be used to reduce background signal and generate editing percentages based on a p-value with a detection limit for base editing of ~7% (Kluesner et al., 2018). However, the reliability of this method is highly dependent on the quality of the Sanger sequencing read. Low quality Sanger sequencing reads may mislead the researcher into incorrectly concluding that their base editing efficiency is high enough to proceed to cell line generation (see Figure 6C and 6E for an example). Genomic DNA samples can also be prepared for Illumina NGS to robustly quantitate editing efficiencies. Targeted amplicon sequencing is the most common method with which to do this, and has been previously described (Gaudelli et al., 2017; Veeranagouda et al., 2018; Yang et al., 2014). We recommend the use of CRISPResso2 as an open-access software to quantify base editing efficiencies from fastq files (Clement et al., 2019). However, it is important to note that over-amplification during either round of PCR can create PCR bias, which will result in inaccurate quantification; care should be taken to use the fewest possible PCR cycles during gDNA amplification and barcoding.

Cell line considerations

HEK293T cells are easily transfected and robust enough to tolerate clonal expansion. This provides researchers with a relatively 'well-behaved' cell line to optimize experimental conditions and assess the feasibility of generating a cell line with a given SNV before moving forward with other mammalian cell lines. The ultimate cell line to use will be entirely dependent on downstream experimental goals and should be thoroughly investigated before embarking on isogenic cell line generation. Furthermore, immortalized cell lines, such as HEK293T cells, often harbor gene duplications, chromosomal rearrangements, and mutations that allow them to effectively propagate in tissue culture. The exact genomic modifications may even vary from laboratory to laboratory for a given cell line. Additionally, these modifications can differ greatly from the reference genome. As such, it is prudent to first sequence the target locus of interest for any mutations or variation before designing gRNA sequences. Additionally, if the data is available for that specific cell line, check the ploidy at each locus to determine the copy number of the gene of interest (Lin et al., 2014).

TROUBLESHOOTING:

Low or no cell viability following transfection.

A common problem is low cell viability following transfection. This may be solved by using less cationic lipid reagent, but usually at the expense of a lower transfection efficiency. However, the most common cause of low cell viability is transfecting under-confluent cells (see Figure 5B). Determining the confluency of cells is highly subjective; we have included pictures of cells with the optimal density for transfection, in Figure 5A.

Low cell viability following transfection, with all surviving cells concentrated in one area

Another common problem after transfection is low cell viability, with “islands” of cells or cells only on the periphery of the wells. Additionally, cells in the middle of the islands are not transfected. This is usually due to an uneven distribution of cells when plating. This can be solved by using a cell strainer during plating and/or gently tapping the plate repeatedly against the hood surface after plating to halt circular movement of media (eddies) inside the well, which can push the cells to the edges of the wells.

Low editing efficiency after transfection.

Sometimes observing an editing efficiency of <10% is due to low transfection efficiency. This can be due to transfecting over-confluent cells (see Figure 5C), which results in high cell viability but low transfection and editing efficiencies. If over-confluent cells are transfected and bulk sequencing shows some editing efficiency, using FACS to enrich for transfected cells prior to re-assessing editing efficiency in bulk is advised. However, it is generally recommended for the researcher to re-transfect the cells when the confluency is ideal, as in Figure 5A.

If the transfection efficiency was >80%, and low or no base editing efficiency was observed, there could be an issue with the gRNA, or the target sequence may be incompatible with the base editor that was used. To assess if there is an issue with the gRNA, we suggest repeating the same experiment, but using wild-type Cas9 instead of the base editor, and then monitor the target site for indel formation. This will reveal if the Cas9:gRNA complex can access the target site. If low or no indels are formed, we recommend trying a different Cas enzyme such as the SaCas9-KKH, or SpRY Cas9 and a different spacer sequence. If the gRNA results in high levels of indel introduction, the

problem may be with the sequence context of the target, in which case we suggest using a different deaminase (see Table 2) and quantifying editing efficiency in bulk again. If even after optimization, editing efficiency is less than 10%, generating SNV-containing isogenic cell lines may still be feasible, but will require selection strategies, and dilution plating should not be used. See the Troubleshooting section on *“Certain genotypes are not observed after clonal expansion.”*

Accidentally disrupted or washed away a vast majority of the cell monolayer.

There are two instances in the protocol where indelicate techniques can result in loss of cells and be detrimental to the experiment if actions are not taken to correct for them. During harvesting of genomic DNA for genotyping (Basic Protocol 2), one might lose cells while washing. In this case, even a small number of cells remaining can be sufficient for PCR amplification of the targeted site, albeit a higher volume of cell lysate will have to be used (please refer to the Troubleshooting genomic PCR section). In this case, however, zero base editing efficiency may be due to only the presence of untransfected cells remaining. During sub-culturing of single cell-derived clones (Basic Protocol 5) this can happen as well. If some cells still remain adhered, add the appropriate amount of prewarmed culture media and place back into the incubator.

Multiple PCR bands observed during gDNA amplification.

If your primers are specific (i.e. designed with no matches to other targets) and still result in off-target amplification, increasing the annealing temperature or decreasing primer concentration can improve specificity. Furthermore, a high primer concentration can result in primer dimers and no band of the desired size. We have observed that using a lower concentration of primer than the manufacturer’s recommendation (up to 10-fold less) usually generates cleaner on-target product bands. See Figure 4B for examples.

No PCR bands observed during gDNA amplification.

Increasing DMSO concentration or trying other additives may help with the genomic DNA PCR. Additionally, the amount of genomic DNA added to the PCR mixture can have drastic effects on PCR success. The recommended volume suggested in Basic Protocol 2 is based on 90-100% confluency pre-harvesting, but sometimes that may be too much (in certain cases, this results in smears on the gel). We recommend first trying 10-fold less genomic DNA template in the PCR reaction, by diluting the genomic DNA mixture in water. If fewer cells were harvested, increasing the amount of cell lysate up to 10-fold may be required. Decreasing primer concentration may improve the specificity and strength of amplification, especially with NGS. Additionally, gDNA can also be harvested using a kit, which can help with difficult gDNA PCRs.

No colony expansion is observed after sorting for single cells.

Sorting via FACS can be stressful for the cells, but there are some modifications that can help improve the success rate of clonal expansion. Reducing the time between harvesting and sorting can improve the outcome of colony formation after sorting. Using a higher concentration of FBS in the collection media can also promote single cell survival and proliferation. Another suggestion is adding cell-conditioned media as a supplement to the collection media to facilitate cell growth. This can be done by collecting the supernatant from cells that have been passaged 24 hours prior, spinning it down, and adding it to the collection media through a 0.45 μm filter. Dilution plating is less harsh on the cells and can have higher success rates of clonal expansion, but as there is no selection for transfected cells, the percentage of cell lines with the SNV of interest will be lower. One option is to use FACS to enrich for transfected cells, allow them to recover, and then use dilution plating to isolate and clonally expand single cells.

Certain genotypes are not observed after clonal expansion

If no homozygous clones are obtained, it is possible that the homozygous edit with the target SNV is lethal, especially if the gene of interest is essential. We recommend using tools such as ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>) to assess the pathogenicity of the target SNV in humans (if the SNV has only been observed in the heterozygous form in humans, the homozygous genotype may be lethal), and DepMap (<https://depmap.org/portal/>) to determine gene essentiality in the specific cell type. If the culprit is that initial base editing efficiencies are low, then additional selection strategies (beyond just selecting for cells with high base editor expression levels) can be employed. These selection strategies make use of reporter systems that involve the modulation of the presence or color of a fluorescent protein by base editing activity concurrently with editing of a genomic locus (e.g. ACE reporter, BE-FLARE, BIG-TREE, GO reporter) (Brookhouser et al., 2020; Coelho et al., 2018; Katti et al., 2020; St. Martin et al., 2018). In these cases, the reporter plasmid and a gRNA targeting the fluorescent protein are co-transfected with the BE and target gRNA. Individual cells with active base editor can be isolated by FACS, to enrich for cells with active base editor.

Conversely, if only homozygous clones are obtained, less active base editor variants can be used to reduce activity (see Table 2). Or, different populations of cells can be isolated with FACS, such as cells with intermediate base editor expression or activity levels. Additionally, recent work has demonstrated that gRNA mismatches can titrate the activity of other genome editing agents. While this strategy has not been explicitly tested with base editors, it may act to reduce base editing activity as well (Jost et al., 2020).

The single cell-derived clones have a different morphology than the examples provided, or ones shown in literature.

Although generally consistent, the morphology of single cell-derived clones can vary from clone to clone and from cell line to cell line. For this reason, we recommend aiming to generate at least three clones for each SNV in its respective cell line to confirm the reproducibility of your results. The resulting cell lines can be subjected to whole-genome or whole-exome genotyping to confirm that only the point mutations of interest have been introduced.

UNDERSTANDING RESULTS:

Having the capability to generate both homozygous and heterozygous clones harboring SNVs highlights the importance of using base editing technologies; generating matched wild-type, heterozygous, and homozygous knock-in clones with traditional genome editing methods is typically quite inefficient or impossible without the use of “blocking mutations” (Paquet et al., 2016) or multiple clonal expansion steps (Kim et al., 2018). Using this protocol, we have found that base editors typically introduce their respective SNVs with >10-fold higher efficiency and >100-fold higher precision than traditional genome editing methods without the use of potentially undesired “blocking mutations”, as quantified by NGS, and without having to physically manipulate cells into phase-enriched populations (physical fractionization) or having to add exogenous chemicals to block cells into specific phases of the cell cycle (chemical blockade). This is in direct contrast to traditional genome editing methods, where typically, accurate SNV introduction is less than 1% (Coggins et al., 2017; Ousterout et al., 2015; Paquet et al., 2016; Ran et al., 2013; Riesenberber et al., 2019; Yeh et al., 2019).

When using FACS to isolate and clonally expand single cells, we have obtained homozygous edits with a success rate of approximately 25%, and heterozygous edits with a success rate of approximately 22%. This is measured from an average 8 clones that we typically obtain per 96-well plate. Additionally, in these cases, initial base editing efficiencies (when measured in bulk, as in Basic

Protocol 3) were estimated at around 23%. When using the dilution method, we have obtained homozygous edits with a success rate of approximately 5%, and heterozygous edits with a success rate of approximately 14%. This is measured from an average of 11 clones that we typically obtain per 96-well plate. Additionally, in these cases, initial base editing efficiencies (when measured in bulk, as in Basic Protocol 3) were estimated at around 38%.

TIME CONSIDERATIONS:

Generating the gRNA plasmids will take 1-2 days, followed by another 1-2 days to sequence the plasmid and verify accurate spacer replacement. Preparing high quality, endotoxin-free plasmids for transfections will take 1 day. Time considerations for the initial assessment of base editing efficiencies (feasibility of generating an isogenic cell line) are as follows: 3-5 days to transfect the BE:gRNA combination(s) and allow for editing, 1 day to harvest the gDNA and PCR-amplify the locus of interest, followed by 1-2 days to determine the percentage desired SNVs introduction by Sanger sequencing. Once > 10% editing is confirmed in bulk, transfections of the same BE:gRNA combination(s) and preparation for single cell isolation requires 3 days. Regardless of which isolation method the researcher uses (FACS or dilution plating), expansion of single cell clones can span from 1-4 weeks depending on the growth rate of the cell line and the phenotypic effects of the introduced SNV. For HEK293T cells, single cell-derived colonies are generally seen 4-7 days after single cell isolation and are ready to be sub-cultured and/or harvested 1 to 2 weeks thereafter. Subsequent sequencing of final clones will take 1-2 days; clonal expansion of the desired cell lines for cryo-storage will take 1-3 weeks depending on the clonal cell line's proliferation rate. Downstream experimental work (researcher-specific assays) can be simultaneously performed during clonal expansion for storage.

ACKNOWLEDGEMENTS:

This article is protected by copyright. All rights reserved.

C.A.V. was supported by an institutional Ruth L. Kirschstein National Research Service Award, NIH Grant T32 CA009523. Q.T.C. was supported by the Molecular Biophysics Training Grant, NIH Grant T32 GM008326. A.C.K. is partially funded by NIH grant R21 GM135736. We are grateful to Jesus Olvera, Cody Fine, and Vu Nguyen of the UCSD Human Embryonic Stem Cell Core Facility for technical assistance of flow cytometry experiments. This work was made possible in part by the CIRM Major Facilities grant (FA1-00607) to the Sanford Consortium for Regenerative Medicine.

LITERATURE CITED:

- 1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., & Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. <https://doi.org/10.1038/nature15393>
- Ahmed, Z., Zeeshan, S., Mendhe, D., & Dong, X. (2020). Human gene and disease associations for clinical-genomics and precision medicine research. *Clinical and Translational Medicine*, *10*(1), 297–318. <https://doi.org/10.1002/ctm2.28>
- Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, *576*(7785), 149–157. <https://doi.org/10.1038/s41586-019-1711-4>
- Arbab, M., Shen, M. W., Mok, B., Wilson, C., Matuszek, Ž., Cassa, C. A., & Liu, D. R. (2020). Determinants of Base Editing Outcomes from Target Library Analysis and Machine Learning. *Cell*, *182*(2), 463-480.e30. <https://doi.org/10.1016/j.cell.2020.05.037>

- Brookhouser, N., Tekel, S. J., Standage-Beier, K., Nguyen, T., Schwarz, G., Wang, X., & Brafman, D. A. (2020). BIG-TREE: Base-Edited Isogenic hPSC Line Generation Using a Transient Reporter for Editing Enrichment. *Stem Cell Reports*, 14(2), 184–191. <https://doi.org/10.1016/j.stemcr.2019.12.013>
- Clement, K., Rees, H., Canver, M. C., Gehrke, J. M., Farouni, R., Hsu, J. Y., Cole, M. A., Liu, D. R., Joung, J. K., Bauer, D. E., & Pinello, L. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nature Biotechnology*, 37(3), 224–226. <https://doi.org/10.1038/s41587-019-0032-3>
- Coelho, M. A., Li, S., Pane, L. S., Firth, M., Ciotta, G., Wrigley, J. D., Cuomo, M. E., Maresca, M., & Taylor, B. J. M. (2018). BE-FLARE: A fluorescent reporter of base editing activity reveals editing characteristics of APOBEC3A and APOBEC3B. *BMC Biology*, 16. <https://doi.org/10.1186/s12915-018-0617-1>
- Coggins, N. B., Stultz, J., O’Geen, H., Carvajal-Carmona, L. G., & Segal, D. J. (2017). Methods for Scarless, Selection-Free Generation of Human Cells and Allele-Specific Functional Analysis of Disease-Associated SNPs and Variants of Uncertain Significance. *Scientific Reports*, 7(1), 15044. <https://doi.org/10.1038/s41598-017-15407-4>
- Dang, Y., Jia, G., Choi, J., Ma, H., Anaya, E., Ye, C., Shankar, P., & Wu, H. (2015). Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biology*, 16(1), 280. <https://doi.org/10.1186/s13059-015-0846-3>
- Di Nicolantonio, F., Arena, S., Gallicchio, M., Zecchin, D., Martini, M., Flonta, S. E., Stella, G. M., Lamba, S., Cancelliere, C., Russo, M., Geuna, M., Appendino, G., Fantozzi, R., Medico, E., & Bardelli, A. (2008). Replacement of normal with mutant alleles in the genome of normal

human cells unveils mutation-specific drug responses. *Proceedings of the National Academy of Sciences of the United States of America*, 105(52), 20864–20869.
<https://doi.org/10.1073/pnas.0808757105>

Doman, J. L., Raguram, A., Newby, G. A., & Liu, D. R. (2020). Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nature Biotechnology*, 38(5), 620–628. <https://doi.org/10.1038/s41587-020-0414-6>

Elbing, K. L., & Brent, R. (2019). Recipes and Tools for Culture of *Escherichia coli*. *Current Protocols in Molecular Biology*, 125(1), e83. <https://doi.org/10.1002/cpmb.83>

Evanoff, M., & Komor, A. C. (2019). Base editors: Modular tools for the introduction of point mutations in living cells. *Emerging Topics in Life Sciences*, 3(5), 483–491.
<https://doi.org/10.1042/ETLS20190088>

Fox, K., Rallapalli, K. L., & Komor, A. C. (2020). Rewriting Human History and Empowering Indigenous Communities with Genome Editing Tools. *Genes*, 11(1), 88.
<https://doi.org/10.3390/genes11010088>

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

Gaudelli, N. M., Lam, D. K., Rees, H. A., Solá-Esteves, N. M., Barrera, L. A., Born, D. A., Edwards, A., Gehrke, J. M., Lee, S.-J., Liquori, A. J., Murray, R., Packer, M. S., Rinaldi, C., Slaymaker, I. M., Yen, J., Young, L. E., & Ciaramella, G. (2020). Directed evolution of adenine base editors with increased activity and therapeutic application. *Nature Biotechnology*.
<https://doi.org/10.1038/s41587-020-0491-6>

- Gehrke, J. M., Cervantes, O., Clement, M. K., Wu, Y., Zeng, J., Bauer, D. E., Pinello, L., & Joung, J. K. (2018). An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nature Biotechnology*, *36*(10), 977–982. <https://doi.org/10.1038/nbt.4199>
- Giuliano, C. J., Lin, A., Girish, V., & Sheltzer, J. M. (2019). Generating Single Cell–Derived Knockout Clones in Mammalian Cells with CRISPR/Cas9. *Current Protocols in Molecular Biology*, *128*(1), e100. <https://doi.org/10.1002/cpmb.100>
- Grav, L. M., Sergeeva, D., Lee, J. S., Marin de Mas, I., Lewis, N. E., Andersen, M. R., Nielsen, L. K., Lee, G. M., & Kildegaard, H. F. (2018). Minimizing Clonal Variation during Mammalian Cell Line Engineering for Improved Systems Biology Data Generation. *ACS Synthetic Biology*, *7*(9), 2148–2159. <https://doi.org/10.1021/acssynbio.8b00140>
- Gray, S. J., Blake, B. L., Criswell, H. E., Nicolson, S. C., Samulski, R. J., & McCown, T. J. (2010). Directed Evolution of a Novel Adeno-associated Virus (AAV) Vector That Crosses the Seizure-compromised Blood–Brain Barrier (BBB). *Molecular Therapy*, *18*(3), 570–578. <https://doi.org/10.1038/mt.2009.292>
- Grünewald, J., Zhou, R., Garcia, S. P., Iyer, S., Lareau, C. A., Aryee, M. J., & Joung, J. K. (2019). Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature*, *569*(7756), 433–437. <https://doi.org/10.1038/s41586-019-1161-z>
- Grünewald, J., Zhou, R., Iyer, S., Lareau, C. A., Garcia, S. P., Aryee, M. J., & Joung, J. K. (2019). CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nature Biotechnology*, *37*(9), 1041–1048. <https://doi.org/10.1038/s41587-019-0236-6>
- Guidelines for base editing in mammalian cells · Benchling.* (n.d.). Retrieved June 19, 2020, from <https://benchling.com/pub/liu-base-editor>

- H, G., Yr, Z., & K, R. (1993, June 18). *Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting*. *Cell*; *Cell*.
[https://doi.org/10.1016/0092-8674\(93\)90644-6](https://doi.org/10.1016/0092-8674(93)90644-6)
- Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., & Taipale, J. (2018). CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nature Medicine*, *24*(7), 927–930.
<https://doi.org/10.1038/s41591-018-0049-z>
- Hess, G. T., Tycko, J., Yao, D., & Bassik, M. C. (2017). Methods and applications of CRISPR-mediated base editing in eukaryotic genomes. *Molecular Cell*, *68*(1), 26–43.
<https://doi.org/10.1016/j.molcel.2017.09.029>
- Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, *157*(6), 1262–1278. <https://doi.org/10.1016/j.cell.2014.05.010>
- Hu, Z., Wang, S., Zhang, C., Gao, N., Li, M., Wang, D., Wang, D., Liu, D., Liu, H., Ong, S.-G., Wang, H., & Wang, Y. (2020). A compact Cas9 ortholog from *Staphylococcus Auricularis* (SauriCas9) expands the DNA targeting scope. *PLoS Biology*, *18*(3), e3000686.
<https://doi.org/10.1371/journal.pbio.3000686>
- Hwang, G.-H., Park, J., Lim, K., Kim, S., Yu, J., Yu, E., Kim, S.-T., Eils, R., Kim, J.-S., & Bae, S. (2018). Web-based design and analysis tools for CRISPR base editing. *BMC Bioinformatics*, *19*(1), 542. <https://doi.org/10.1186/s12859-018-2585-4>
- Jiang, F., Taylor, D. W., Chen, J. S., Kornfeld, J. E., Zhou, K., Thompson, A. J., Nogales, E., & Doudna, J. A. (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science (New York, N.Y.)*, *351*(6275), 867–871. <https://doi.org/10.1126/science.aad8282>

- Jin, S., Zong, Y., Gao, Q., Zhu, Z., Wang, Y., Qin, P., Liang, C., Wang, D., Qiu, J.-L., Zhang, F., & Gao, C. (2019). Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science (New York, N.Y.)*, *364*(6437), 292–295. <https://doi.org/10.1126/science.aaw7166>
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, *337*(6096), 816–821. <https://doi.org/10.1126/science.1225829>
- Jost, M., Santos, D. A., Saunders, R. A., Horlbeck, M. A., Hawkins, J. S., Scaria, S. M., Norman, T. M., Hussmann, J. A., Liem, C. R., Gross, C. A., & Weissman, J. S. (2020). Titrating gene expression using libraries of systematically attenuated CRISPR guide RNAs. *Nature Biotechnology*, *38*(3), 355–364. <https://doi.org/10.1038/s41587-019-0387-5>
- Katti, A., Foronda, M., Zimmerman, J., Diaz, B., Zafra, M. P., Goswami, S., & Dow, L. E. (2020). GO: A functional reporter system to identify and enrich base editing activity. *Nucleic Acids Research*, *48*(6), 2841–2852. <https://doi.org/10.1093/nar/gkaa124>
- Kim, S.-I., Matsumoto, T., Kagawa, H., Nakamura, M., Hirohata, R., Ueno, A., Ohishi, M., Sakuma, T., Soga, T., Yamamoto, T., & Woltjen, K. (2018). Microhomology-assisted scarless genome editing in human iPSCs. *Nature Communications*, *9*. <https://doi.org/10.1038/s41467-018-03044-y>
- Kim, Y. B., Komor, A. C., Levy, J. M., Packer, M. S., Zhao, K. T., & Liu, D. R. (2017). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nature Biotechnology*, *35*(4), 371–376. <https://doi.org/10.1038/nbt.3803>

- Kluesner, M. G., Nedveck, D. A., Lahr, W. S., Garbe, J. R., Abrahante, J. E., Webber, B. R., & Moriarity, B. S. (2018). EditR: A Method to Quantify Base Editing from Sanger Sequencing. *The CRISPR Journal*, 1(3), 239–250. <https://doi.org/10.1089/crispr.2018.0014>
- Koblan, L. W., Doman, J. L., Wilson, C., Levy, J. M., Tay, T., Newby, G. A., Maianti, J. P., Raguram, A., & Liu, D. R. (2018). Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nature Biotechnology*, 36(9), 843–846. <https://doi.org/10.1038/nbt.4172>
- Kohlbrener, E., Aslanidi, G., Nash, K., Shklyae, S., Campbell-Thompson, M., Byrne, B. J., Snyder, R. O., Muzyczka, N., Warrington, K. H., & Zolotukhin, S. (2005). Successful Production of Pseudotyped rAAV Vectors Using a Modified Baculovirus Expression System. *Molecular Therapy*, 12(6), 1217–1225. <https://doi.org/10.1016/j.ymthe.2005.08.018>
- Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, 533(7603), 420–424. <https://doi.org/10.1038/nature17946>
- Kosicki, M., Tomberg, K., & Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nature Biotechnology*, 36(8), 765–771. <https://doi.org/10.1038/nbt.4192>
- Kwart, D., Paquet, D., Teo, S., & Tessier-Lavigne, M. (2017). Precise and efficient scarless genome editing in stem cells using CORRECT. *Nature Protocols*, 12(2), 329–354. <https://doi.org/10.1038/nprot.2016.171>
- Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., Karapetyan, K., Katz, K., Liu, C., Maddipatla, Z., Malheiro, A.,

- McDaniel, K., Ovetsky, M., Riley, G., Zhou, G., ... Maglott, D. R. (2018). ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, *46*(D1), D1062–D1067. <https://doi.org/10.1093/nar/gkx1153>
- Lappalainen, T., Scott, A. J., Brandt, M., & Hall, I. M. (2019). Genomic Analysis in the Age of Human Genome Sequencing. *Cell*, *177*(1), 70–84. <https://doi.org/10.1016/j.cell.2019.02.032>
- Lee, H. K., Smith, H. E., Liu, C., Willi, M., & Hennighausen, L. (2020). Cytosine base editor 4 but not adenine base editor generates off-target mutations in mouse embryos. *Communications Biology*, *3*. <https://doi.org/10.1038/s42003-019-0745-3>
- Li, X., Wang, Y., Liu, Y., Yang, B., Wang, X., Wei, J., Lu, Z., Zhang, Y., Wu, J., Huang, X., Yang, L., & Chen, J. (2018). Base editing with a Cpf1-cytidine deaminase fusion. *Nature Biotechnology*, *36*(4), 324–327. <https://doi.org/10.1038/nbt.4102>
- Lin, Y.-C., Boone, M., Meuris, L., Lemmens, I., Van Roy, N., Soete, A., Reumers, J., Moisse, M., Plaisance, S., Drmanac, R., Chen, J., Speleman, F., Lambrechts, D., Van de Peer, Y., Tavernier, J., & Callewaert, N. (2014). Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nature Communications*, *5*. <https://doi.org/10.1038/ncomms5767>
- Liu, M., Rehman, S., Tang, X., Gu, K., Fan, Q., Chen, D., & Ma, W. (2019). Methodologies for Improving HDR Efficiency. *Frontiers in Genetics*, *9*. <https://doi.org/10.3389/fgene.2018.00691>
- Liu, Z., Chen, S., Shan, H., Jia, Y., Chen, M., Song, Y., Lai, L., & Li, Z. (2020). Efficient base editing with high precision in rabbits using YFE-BE4max. *Cell Death & Disease*, *11*(1), 1–11. <https://doi.org/10.1038/s41419-020-2244-3>

- Liu, Z., Shan, H., Chen, S., Chen, M., Song, Y., Lai, L., & Li, Z. (2019). Efficient base editing with expanded targeting scope using an engineered Spy-mac Cas9 variant. *Cell Discovery*, 5, 58. <https://doi.org/10.1038/s41421-019-0128-4>
- Malone, E. R., Oliva, M., Sabatini, P. J. B., Stockley, T. L., & Siu, L. L. (2020). Molecular profiling for precision cancer therapies. *Genome Medicine*, 12. <https://doi.org/10.1186/s13073-019-0703-1>
- Martin, R. M., Ikeda, K., Cromer, M. K., Uchida, N., Nishimura, T., Romano, R., Tong, A. J., Lemgart, V. T., Camarena, J., Pavel-Dinu, M., Sindhu, C., Wiebking, V., Vaidyanathan, S., Dever, D. P., Bak, R. O., Laustsen, A., Lesch, B. J., Jakobsen, M. R., Sebastiano, V., ... Porteus, M. H. (2019). Highly Efficient and Marker-free Genome Editing of Human Pluripotent Stem Cells by CRISPR-Cas9 RNP and AAV6 Donor-Mediated Homologous Recombination. *Cell Stem Cell*, 24(5), 821-828.e5. <https://doi.org/10.1016/j.stem.2019.04.001>
- Molla, K. A., & Yang, Y. (2019). CRISPR/Cas-Mediated Base Editing: Technical Considerations and Practical Applications. *Trends in Biotechnology*, 37(10), 1121-1142. <https://doi.org/10.1016/j.tibtech.2019.03.008>
- Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., Noda, T., Abudayyeh, O. O., Gootenberg, J. S., Mori, H., Oura, S., Holmes, B., Tanaka, M., Seki, M., Hirano, H., Aburatani, H., Ishitani, R., Ikawa, M., Yachie, N., ... Nureki, O. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science (New York, N.Y.)*, 361(6408), 1259-1262. <https://doi.org/10.1126/science.aas9129>

- Norris, A. D., Kim, H.-M., Colaiácovo, M. P., & Calarco, J. A. (2015). Efficient Genome Editing in *Caenorhabditis elegans* with a Toolkit of Dual-Marker Selection Cassettes. *Genetics*, *201*(2), 449–458. <https://doi.org/10.1534/genetics.115.180679>
- Ochiai, H., Miyamoto, T., Kanai, A., Hosoba, K., Sakuma, T., Kudo, Y., Asami, K., Ogawa, A., Watanabe, A., Kajii, T., Yamamoto, T., & Matsuura, S. (2014). TALEN-mediated single-base-pair editing identification of an intergenic mutation upstream of BUB1B as causative of PCS (MVA) syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(4), 1461–1466. <https://doi.org/10.1073/pnas.1317008111>
- Ousterout, D. G., Kabadi, A. M., Thakore, P. I., Majoros, W. H., Reddy, T. E., & Gersbach, C. A. (2015). Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. *Nature Communications*, *6*, 6244. <https://doi.org/10.1038/ncomms7244>
- Paquet, D., Kwart, D., Chen, A., Sproul, A., Jacob, S., Teo, S., Olsen, K. M., Gregg, A., Noggle, S., & Tessier-Lavigne, M. (2016). Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature*, *533*(7601), 125–129. <https://doi.org/10.1038/nature17664>
- Phelan, M. C. (2006). Techniques for mammalian cell tissue culture. *Current Protocols in Protein Science*, Appendix 3, Appendix 3C. <https://doi.org/10.1002/0471140864.psa03cs46>
- Popejoy, A. B., & Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature News*, *538*(7624), 161. <https://doi.org/10.1038/538161a>

- Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, 8(11), 2281–2308. <https://doi.org/10.1038/nprot.2013.143>
- Rees, H. A., & Liu, D. R. (2018). Base editing: Precision chemistry on the genome and transcriptome of living cells. *Nature Reviews Genetics*, 19(12), 770–788. <https://doi.org/10.1038/s41576-018-0059-1>
- Richter, M. F., Zhao, K. T., Eton, E., Lapinaite, A., Newby, G. A., Thuronyi, B. W., Wilson, C., Koblan, L. W., Zeng, J., Bauer, D. E., Doudna, J. A., & Liu, D. R. (2020). Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nature Biotechnology*. <https://doi.org/10.1038/s41587-020-0453-z>
- Riesenberg, S., Chintalapati, M., Macak, D., Kanis, P., Maricic, T., & Pääbo, S. (2019). Simultaneous precise editing of multiple genes in human cells. *Nucleic Acids Research*, 47(19), e116. <https://doi.org/10.1093/nar/gkz669>
- S, S., R, P., F, B., C, R., La, D., Kw, K., B, V., & N, P. (2009, March 10). *A panel of isogenic human cancer cells suggests a therapeutic approach for cancers with inactivated p53*. Proceedings of the National Academy of Sciences of the United States of America; Proc Natl Acad Sci U S A. <https://doi.org/10.1073/pnas.0813333106>
- Seidman, C. E., Struhl, K., Sheen, J., & Jessen, T. (2001). Introduction of Plasmid DNA into Cells. *Current Protocols in Molecular Biology*, 37(1), mb0108s37. <https://doi.org/10.1002/0471142727.mb0108s37>
- Sherry, S. T., Ward, M.-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: The NCBI database of genetic variation. *Nucleic Acids Research*, 29(1), 308–311.

- Shin, H. Y., Wang, C., Lee, H. K., Yoo, K. H., Zeng, X., Kuhns, T., Yang, C. M., Mohr, T., Liu, C., & Hennighausen, L. (2017). CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. *Nature Communications*, *8*, 15464. <https://doi.org/10.1038/ncomms15464>
- Smits, A. H., Ziebell, F., Joberty, G., Zinn, N., Mueller, W. F., Clauder-Münster, S., Eberhard, D., Fälth Savitski, M., Grandi, P., Jakob, P., Michon, A.-M., Sun, H., Tessmer, K., Bürckstümmer, T., Bantscheff, M., Steinmetz, L. M., Drewes, G., & Huber, W. (2019). Biological plasticity rescues target activity in CRISPR knock outs. *Nature Methods*, *16*(11), 1087–1093. <https://doi.org/10.1038/s41592-019-0614-5>
- Stacey, G. N., & Masters, J. R. (2008). Cryopreservation and banking of mammalian cell lines. *Nature Protocols*, *3*(12), 1981–1989. <https://doi.org/10.1038/nprot.2008.190>
- St. Martin, A., Salamango, D., Serebrenik, A., Shaban, N., Brown, W. L., Donati, F., Munagala, U., Conticello, S. G., & Harris, R. S. (2018). A fluorescent reporter for quantification and enrichment of DNA editing by APOBEC–Cas9 or cleavage by Cas9 in living cells. *Nucleic Acids Research*, *46*(14), e84. <https://doi.org/10.1093/nar/gky332>
- Sur, I. K., Hallikas, O., Vähärautio, A., Yan, J., Turunen, M., Enge, M., Taipale, M., Karhu, A., Aaltonen, L. A., & Taipale, J. (2012). Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. *Science (New York, N.Y.)*, *338*(6112), 1360–1363. <https://doi.org/10.1126/science.1228606>
- Tan, J., Zhang, F., Karcher, D., & Bock, R. (2019). Engineering of high-precision base editors for site-specific single nucleotide replacement. *Nature Communications*, *10*(1), 439. <https://doi.org/10.1038/s41467-018-08034-8>

- Tycko, J., Myer, V. E., & Hsu, P. D. (2016). Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity. *Molecular Cell*, *63*(3), 355–370. <https://doi.org/10.1016/j.molcel.2016.07.004>
- Veeranagouda, Y., Debono-Lagneaux, D., Fournet, H., Thill, G., & Didier, M. (2018). CRISPR-Cas9-Edited Site Sequencing (CRES-Seq): An Efficient and High-Throughput Method for the Selection of CRISPR-Cas9-Edited Clones. *Current Protocols in Molecular Biology*, *121*(1), 31.14.1-31.14.11. <https://doi.org/10.1002/cpmb.53>
- Walton, R. T., Christie, K. A., Whittaker, M. N., & Kleinstiver, B. P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science (New York, N.Y.)*, *368*(6488), 290–296. <https://doi.org/10.1126/science.aba8853>
- Wang, D., Zhang, C., Wang, B., Li, B., Wang, Q., Liu, D., Wang, H., Zhou, Y., Shi, L., Lan, F., & Wang, Y. (2019). Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. *Nature Communications*, *10*(1), 4284. <https://doi.org/10.1038/s41467-019-12281-8>
- Watkins, J. A., Irshad, S., Grigoriadis, A., & Tutt, A. N. (2014). Genomic scars as biomarkers of homologous recombination deficiency and drug response in breast and ovarian cancers. *Breast Cancer Research : BCR*, *16*(3), 211. <https://doi.org/10.1186/bcr3670>
- Webber, B. R., Lonetree, C.-L., Kluesner, M. G., Johnson, M. J., Pomeroy, E. J., Diers, M. D., Lahr, W. S., Draper, G. M., Slipek, N. J., Smeester, B. A., Lovendahl, K. N., McElroy, A. N., Gordon, W. R., Osborn, M. J., & Moriarity, B. S. (2019). Highly efficient multiplex human T cell engineering without double-strand breaks using Cas9 base editors. *Nature Communications*, *10*(1), 5222. <https://doi.org/10.1038/s41467-019-13007-6>

- Weisheit, I., Kroeger, J. A., Malik, R., Klimmt, J., Crusius, D., Dannert, A., Dichgans, M., & Paquet, D. (2020). Detection of Deleterious On-Target Effects after HDR-Mediated CRISPR Editing. *Cell Reports*, 31(8), 107689. <https://doi.org/10.1016/j.celrep.2020.107689>
- Yang, L., Yang, J. L., Byrne, S., Pan, J., & Church, G. M. (2014). CRISPR/Cas9-Directed Genome Editing of Cultured Cells. *Current Protocols in Molecular Biology*, 107(1), 31.1.1-31.1.17. <https://doi.org/10.1002/0471142727.mb3101s107>
- Yeh, C. D., Richardson, C. D., & Corn, J. E. (2019). Advances in genome editing through control of DNA repair pathways. *Nature Cell Biology*, 21(12), 1468–1478. <https://doi.org/10.1038/s41556-019-0425-z>
- Yokoyama, W. M., Thompson, M. L., & Ehrhardt, R. O. (2012). Cryopreservation and Thawing of Cells. *Current Protocols in Immunology*, 99(1), A.3G.1-A.3G.5. <https://doi.org/10.1002/0471142735.ima03gs99>
- Yu, Y., Leete, T. C., Born, D. A., Young, L., Barrera, L. A., Lee, S.-J., Rees, H. A., Ciaramella, G., & Gaudelli, N. M. (2020). Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity. *Nature Communications*, 11(1), 2052. <https://doi.org/10.1038/s41467-020-15887-5>
- Zafra, M. P., Schatoff, E. M., Katti, A., Foronda, M., Breinig, M., Schweitzer, A. Y., Simon, A., Han, T., Goswami, S., Montgomery, E., Thibado, J., Kasthuber, E. R., Sánchez-Rivera, F. J., Shi, J., Vakoc, C. R., Lowe, S. W., Tschaharganeh, D. F., & Dow, L. E. (2018). Optimized base editors enable efficient editing in cells, organoids and mice. *Nature Biotechnology*, 36(9), 888–893. <https://doi.org/10.1038/nbt.4194>

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

Zuo, E., Sun, Y., Wei, W., Yuan, T., Ying, W., Sun, H., Yuan, L., Steinmetz, L. M., Li, Y., & Yang, H. (2019). Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science (New York, N.Y.)*, *364*(6437), 289–292. <https://doi.org/10.1126/science.aav9973>

Zuo, E., Sun, Y., Yuan, T., He, B., Zhou, C., Ying, W., Liu, J., Wei, W., Zeng, R., Li, Y., & Yang, H. (2020). A rationally engineered cytosine base editor retains high on-target activity while reducing both DNA and RNA off-target effects. *Nature Methods*, *17*(6), 600–604. <https://doi.org/10.1038/s41592-020-0832-x>

FIGURE LEGENDS:

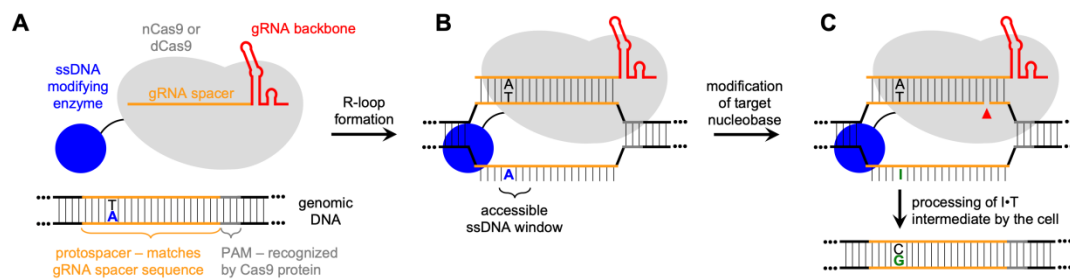


Figure 1. Adenine base editor (ABE) schematic. **(A)** nCas9 or dCas9 binds to a target genomic locus through canonical Watson-Crick base pairing between the spacer sequence of the gRNA and the protospacer of the DNA. The presence of a protospacer adjacent motif (PAM) is also required for Cas9 binding. The gRNA is comprised of a user-defined ~20 nucleotide (nt) spacer that designates the genomic target to be modified (orange) and a ~80 nt scaffold sequence necessary for Cas-binding (red). **(B)** Upon R-loop formation, a small window of about 5 nucleotides on the bottom strand become accessible to a Cas9-tethered ssDNA modifying enzyme. **(C)** In ABE-based editing, this enzyme deaminates adenine nucleobases within this window into inosines, while the top strand is cleaved if nCas9 is used. The I•T intermediate is permanently converted into a G•C base pair following DNA replication or repair. The cytosine base editor (CBE) works analogously, by converting C•G base pairs into T•A, through a U•G intermediate.

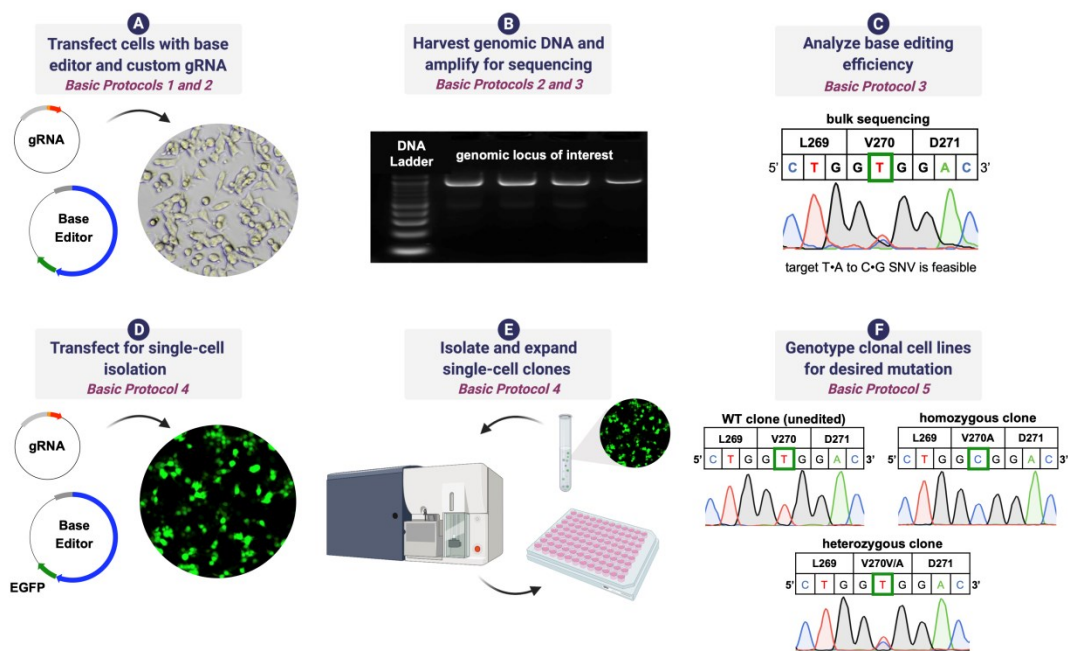


Figure 2. Generating isogenic cell lines harboring a target SNV, using the MUTYH V270A variant as an example. **(A)** Transfection-quality plasmids encoding a custom-designed gRNA and base editor (BE) are transfected into mammalian cells. **(B)** 3 days post-transfection, genomic DNA is harvested, and the genomic locus of interest is amplified via PCR and then prepared for sequencing. **(C)** The resulting PCR products are sequenced to assess the feasibility of generating an isogenic cell harboring the target SNV. Shown is bulk sequencing results of the total cell population transfected with a BE/gRNA combination to introduce the V270A variant in MUTYH. The sequencing chromatogram shows base editing activity on the target base. **(D)** Following confirmation of >10% editing efficiency, step A is repeated. The BE plasmid also encodes for an EGFP fluorescent marker, allowing for the use of FACS to isolate individual cells expressing a base editor. **(E)** FACS is utilized to sort single, EGFP-positive cells into individual wells of a 96-well plate containing culture media. **(F)** Single cell-derived colonies clonally expand for 1-2 weeks. Sequencing the resulting clonal cell lines confirms the generation of an isogenic cell lines harboring the target SNV. Data from three individual isogenic cell lines of the V270A MUTYH variant are shown, containing wild-type, homozygous, and heterozygous genotypes. The target nucleotide is highlighted in green.

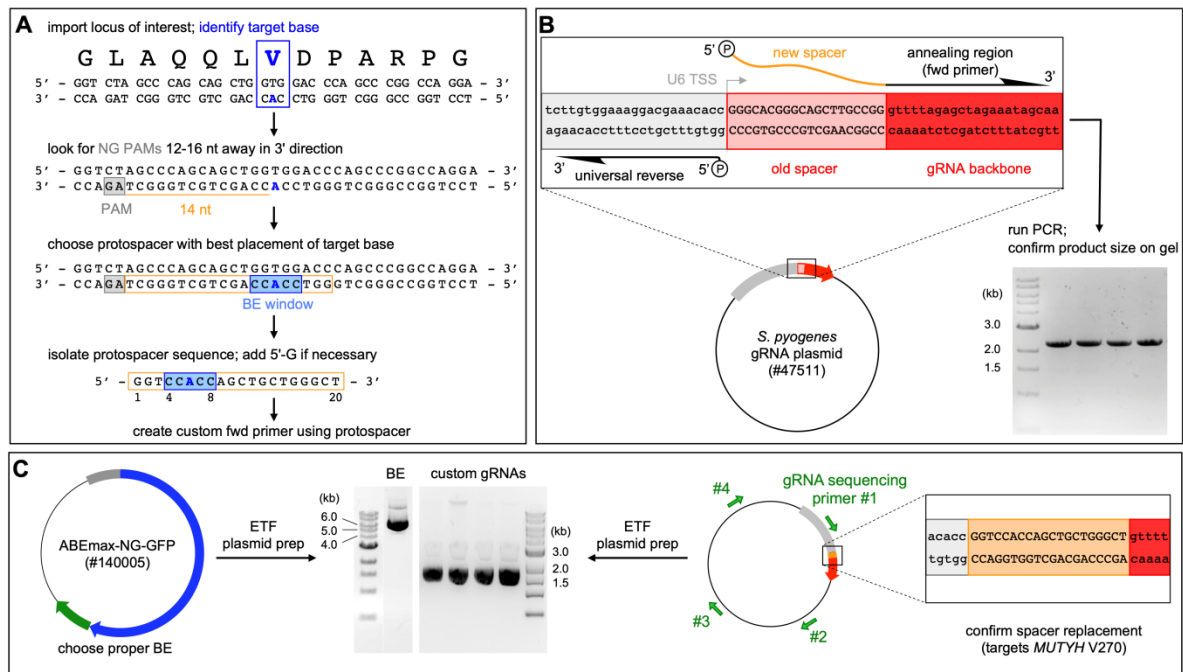


Figure 3. Overview of Basic Protocol 1 **(A)** To design a custom gRNA (steps 1-5), first import the target locus into a viewing software of choice, being mindful of the isoform and reference genome used to annotate the SNV of interest (shown is the *MUTYH* gene sequence surrounding residue 270, using the hg38 reference genome). Identify the target amino acid and nucleotide (blue, V270 codon, target adenine is on the template strand), and identify potential PAMs that are positioned 12-16 nt downstream, in the 3' direction, on the same DNA strand as the target nucleotide. Select the protospacer that positions the target nucleotide closest to the center of the editing window (position 6), or that positions potential bystander edits outside of the editing window. The gray "AG" PAM optimally positions the target base, while simultaneously pushing a potential bystander base to position 9, outside the canonical editing window. Order universal reverse and custom forward primers to replace the spacer sequence on a compatible gRNA expression plasmid (*S. pyogenes*, such as Addgene #47511). **(B)** Schematic showing the site-directed mutagenesis or around-the-horn cloning strategy for gRNA spacer replacement, and agarose gel showing confirmation of a PCR product band with the correct size (steps 8-10). **(C)** Preparation of endotoxin-free (ETF) midi- or max-preps of gRNA and BE plasmids (steps 19-24) with a representative agarose gel showing proper size and quality of supercoiled ETF plasmids. Additionally, in step 18, use Sanger sequencing to verify the sequence of the full gRNA plasmid.

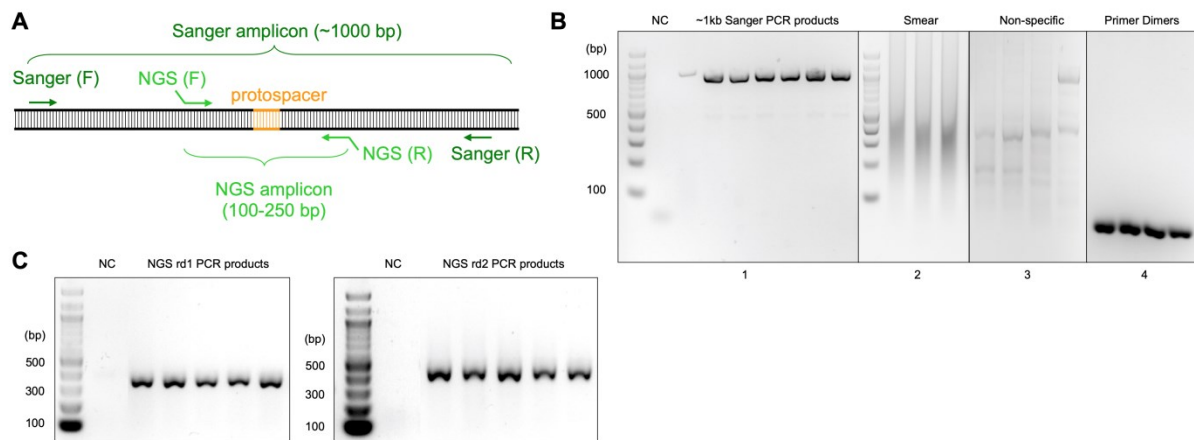


Figure 4. Overview of gDNA amplification (Basic Protocol 2, steps 10-14) **(A)** When designing primers for Sanger sequencing, ensure there is at least 50 bp between the protospacer and the beginning of the sequencing read. Positioning the protospacer in the center of the read will give the best chance for high quality data within the protospacer region. NGS amplicons should amplify a gDNA region based on the number of cycles in the NGS run (i.e. for a 300-cycle run, the amplicon should be ~250 bp, and for a 150-cycle run, the amplicon should be ~100 bp). **(B)** For *MUTYH* V270 Sanger sequencing, primers #11 and 12 (Table 1) were used for amplification of a ~1 kb amplicon. Panel 1 displays a 2% agarose gel showing a clean negative control (NC, lane 1) and ideal gDNA amplification products (lanes 3-8). Additional 2% gels show common issues we have experienced. Specifically, panel 2 shows a PCR product following the use of too much gDNA template (as evidence by smears), panel 3 shows a PCR product with off-target bands, and panel 4 shows a PCR outcome that predominantly consists of primer dimer bands. **(C)** Representative 2% agarose gels of round 1 and 2 PCR products for NGS using a ~280 bp amplicon as an example. The negative control samples contain water (round 1) or the previous negative control PCR product (round 2) as a template instead of gDNA. Round 1 primer adapter sequences will add 66 bp to the length of the amplicon, creating a ~350 bp product in this example. Round 2 primer sequences add an additional 74 bp to the length of the amplicon, resulting in a ~420 bp product.

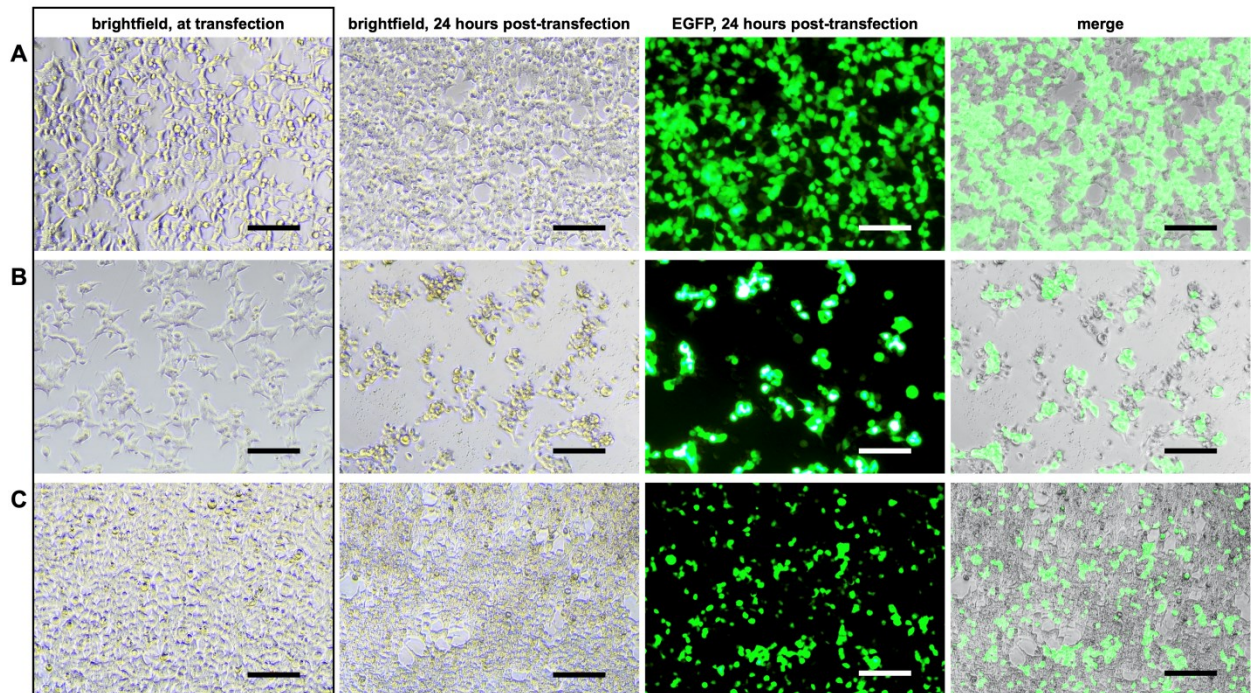


Figure 5. Representative images of successful and unsuccessful transfection experiments (Basic Protocol 2, steps 1-9). Shown are brightfield images of HEK293T cell confluency at the time of transfection (left most column, in box), as well as brightfield (middle left column), GFP fluorescence (middle right column), and merged (right most column) images of the same cells 24 hours post-transfection with gRNA and BE plasmids to generate the V270A MUTYH variant. Shown are cells that were transfected at **row (A)** – ideal confluency (70-85%), resulting in high transfection efficiency and high cell viability across the surface of the plate, **row (B)** – under confluency (<70%), resulting in high transfection efficiency but low cell viability, and **row (C)** – over confluency (>85%), resulting in low transfection efficiency but high cell viability. The latter two conditions are illustrative of poor base editing experiments. Scale bars = 166 μm .

Author

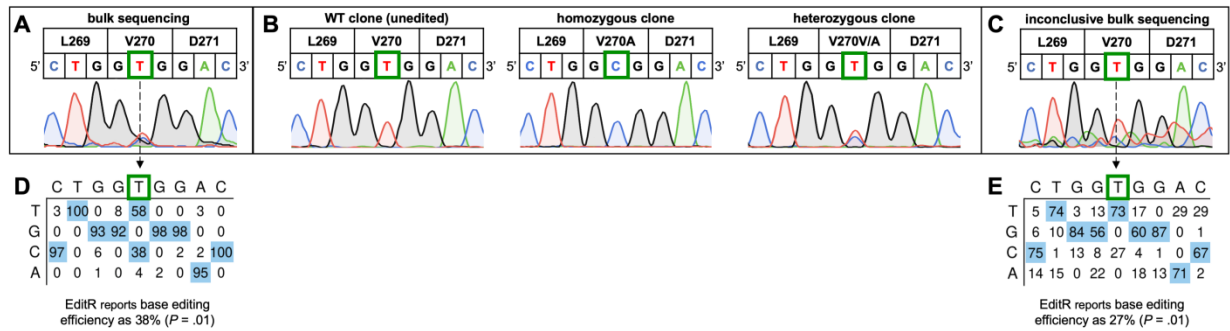


Figure 6. Representative Sanger sequencing traces from genotyping of cells in bulk and after clonal expansion (Basic Protocols 3 and 5). **(A)** Sanger sequencing trace of the bulk population of total cells after transfecting with a BE:gRNA combination to introduce the MUTYH V270A mutation. **(B)** Following re-transfection and clonal expansion, single cell-derived colonies potentially harboring the MUTYH V270A variant were then sequenced. Sanger sequencing chromatograms show the generation of three individual isogenic cell lines, in this instance, with genotypes of wild-type (left), homozygous (middle), and heterozygous (right). **(C)** A low-quality chromatogram with a high degree of background noise obtained after the bulk population of total cells were sequenced via Sanger sequencing. It is recommended to re-sequence rather than make a decision about isogenic cell line generation from such a trace. **(D)** Analysis of the .ab1 file of the bulk population of total cells, as described in panel **A**. EditR quantifies a base editing efficiency of 38% ($P = .01$) The observed base editing efficiency at the target nucleotide indicates the feasibility of generating the targeted V270A variants, since the editing efficiency is $>10\%$. **(E)** Analysis of the .ab1 file of the bulk population of total cells with a degree of background noise, as described in panel **C**. EditR quantifies a base editing efficiency of 27% ($P = .01$), which may mislead the researcher into incorrectly concluding a successful initial base editing experiment. Please refer to Basic Protocol 3 for information on EditR.

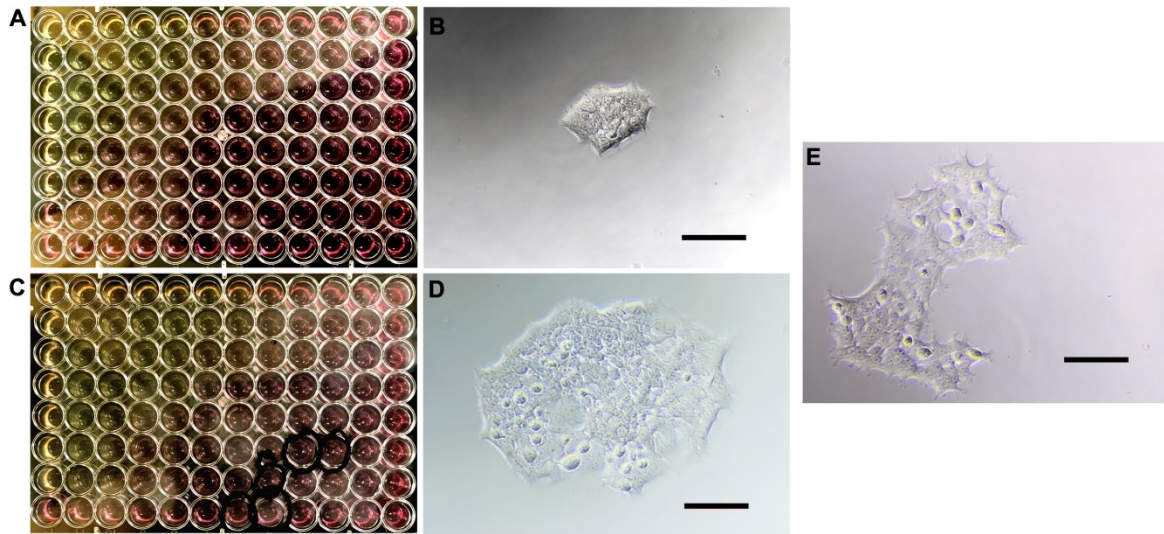


Figure 7. Representative images from dilution plating method (Alternate Protocol 2). **(A)** Image of a 96-well plate taken four days post-dilution plating. The top left wells have a higher concentration of cells as indicated by the low pH level (yellow-colored culture medium). **(B)** 20X magnification image of a well from **(A)** with a single cell-derived colony, taken four days post-dilution plating. **(C)** Image of the 96-well plate from **(A)**, taken seven days post-dilution plating. Wells which are harboring single cell-derived clones are marked with black circles. **(D)** 20X magnification image of the same colony from **(B)**, taken seven days post-dilution plating. **(E)** 20X magnification image of a doublet cell-derived colony, taken seven days post-dilution plating. Scale bars = 166 μm .

Author

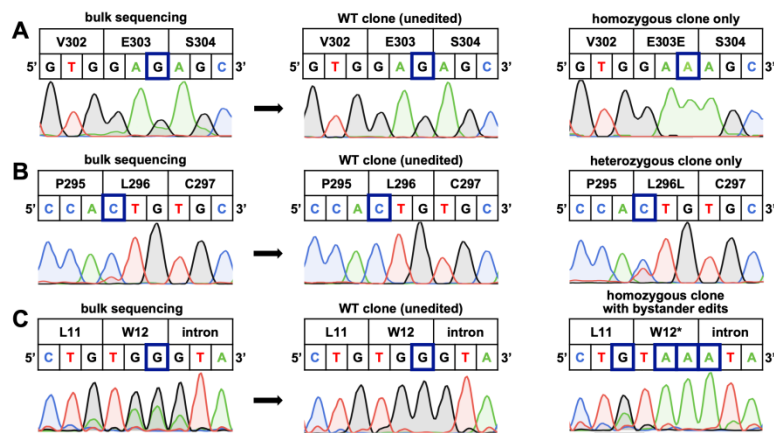


Figure 8. Potential pitfalls in using Base Editing to generate isogenic cell lines. **(A)** Attempted introduction of the MUTYH E303E mutation resulted in only WT and homozygous cell lines, with no heterozygous clones observed. Experiment can be repeated with a less active BE variant, or cells with lower GFP signal isolated. **(B)** Attempted introduction of the MUTYH L296L mutation resulted in only a WT and heterozygous cell lines, with no homozygous clones observed. Check gene and mutation for lethality, and repeat experiment using base editor activity selection scheme. **(C)** Attempted introduction of the MUTYH W12* mutation resulted in bystander edits of non-target bases in the protospacer region. Because a premature stop codon is being introduced, bystander edits are acceptable.

TABLES:

Table 1. Primers and additional sequences used to generate custom gRNAs or amplify the target locus. Underlined sequences anneal to the gRNA plasmid backbone; Bold sequences are protospacers (replace this with your custom protospacer); Italics represent 8 nt NGS barcode sequences (order multiple rd2 fwd and rev primers with difference barcodes), a unique fwd and rev barcode combination should be used for each sample.

Usage	#	Primer	Sequence	Notes
gRNA cloning	1	universal reverse	<u>GGTGTTCGTCCTTCCACAAG</u>	Site-directed mutagenesis (of spacer region)
	2	V270 gRNA forward	GGTCCACCAGCTGCTGGGCT <u>GTTTTAGAGCTAGAAATAGCA</u>	
	3	E303 gRNA forward (adds 5'G)	GACAGGCTCTCCACAGGGCAC <u>GTTTTAGAGCTAGAAATAGCA</u>	
	4	L296 gRNA forward (adds 5'G)	GCCACTGTGCAGCCAGTGCCC <u>GTTTTAGAGCTAGAAATAGCA</u>	
	5	W12 (exon 1 pmSTOP) gRNA reverse	GTACCCACAGACGACTCAGG <u>GTTTTAGAGCTAGAAATAGCA</u>	
	6	non-targetin	GGTATTACTGATATTGGTGGG	Spacer sequence

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/cpmb.129](https://doi.org/10.1002/cpmb.129).

		g gRNA spacer		only
Sanger sequencing	7	gRNA sequencing 1 (U6)	<u>TACGTGACGTAGAAAGTAAT</u>	To verify gRNA clones
	8	gRNA sequencing 2 (ColE1)	<u>TTTCTCCCTTCGGGAAGCGT</u>	
	9	gRNA sequencing 3 (backbone)	<u>GAGATTATCAAAAAGGATCT</u>	
	10	gRNA sequencing 4 (AmpR)	<u>TTGTGCAAAAAGCGGTTAG</u>	
	11	Sanger forward * 1 <i>MUTYH</i> codons 193-332	AGGAGGTGAATCAACTCTGGGC	*Used to sequence V270, L296, E303
	12	Sanger reverse 2 <i>MUTYH</i> codons 193-332	CCGAACCCTACTCAAGCCAAGA	
NGS	13	NGS rd1 adapter forward	ACACTCTTCCCTACACGACGCTCTTCCGATCTNNNN	Add sequence to 5' end of target-specific primers
	14	NGS rd1 adapter	TGGAGTTCAGACGTGTGCTCTTCCGATCT	

	reverse		
1	NGS rd1 <i>MUTYH</i> codons 264-291 forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGCAGCTC TGGTAGGATGTTGG	Rd1 primers for V270 example
5	NGS rd1 <i>MUTYH</i> codons 264-291 reverse	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCAGTAGGCTTACTCT CTGGC	

Table 2. Suggested base editors designed to maximize on-target editing or minimize off-target (OT) editing.

	High-efficiency	Addgene #	Reduced off-target editing	Notes	Addgene #
CB E	BE4max-NG-P2A-EGFP	125616, 140001	BE4max-NG (YE1)	Less processive, narrowed window	138159
	RrAPOBEC3F	138340	RrAPOBEC3F (F130L)	Retains high on-target activity	138341
	PpAPOBEC1	138349	PpAPOBEC1 (H122A)	Minimal OTs, slightly reduced on-target activity	138345, 138338
	SsAPOBEC3B	138343	SsAPOBEC3B (R54Q)	Minimal OTs, "BC" preference	138344
	AncBE4max-P2A-EGFP	112100	eA3A-BE3 (N57G)	"(A)UC" preference	131315
A BE	ABEmax-NG-P2A-EGFP	140005	evo-TadA (V106W)	Inactivated or deleted wt-TadA	125647, 138495
	NG-ABE8e	138491	evo-TadA (F148A)		<i>Zhou et al. 2019</i>

ABE8.20-m	136300	evo-TadA (V82G)	131313
-----------	--------	-----------------	--------

Table 3. Next-generation sequencing barcodes or round 2 PCR primers.

#	Name	Primer sequence	Barcode
1	Fwd-A	AATGATACGGCGACCACCGAGATCTACACATTACTCGACTCTTTCCCTACACG AC	ATTACTC G
2	Fwd-B	AATGATACGGCGACCACCGAGATCTACACTCCGGAGAACAACACTCTTTCCCTACAC GAC	TCCGGAG A
3	Fwd-C	AATGATACGGCGACCACCGAGATCTACACCGCTCATTACACTCTTTCCCTACACG AC	CGCTCAT T
4	Fwd-D	AATGATACGGCGACCACCGAGATCTACACGAGATTCCACACTCTTTCCCTACACG AC	GAGATTC C
5	Fwd-E	AATGATACGGCGACCACCGAGATCTACACATTAGAAACAACACTCTTTCCCTACACG AC	ATTGAGA A
6	Fwd-F	AATGATACGGCGACCACCGAGATCTACACGAATTCGTACACTCTTTCCCTACACG AC	GAATTGC T
7	Fwd-G	AATGATACGGCGACCACCGAGATCTACACCTGAAGCTACACTCTTTCCCTACACG AC	CTGAAGC T
8	Fwd-H	AATGATACGGCGACCACCGAGATCTACACTAATGCGCACACTCTTTCCCTACACG AC	TAATGCG C
9	Fwd-I	AATGATACGGCGACCACCGAGATCTACACCGGCTATGACACTCTTTCCCTACACG AC	CGGCTAT G
10	Fwd-J	AATGATACGGCGACCACCGAGATCTACACTCCGCGAAACAACACTCTTTCCCTACACG AC	TCCGCGA A
11	Fwd-K	AATGATACGGCGACCACCGAGATCTACACTCTCGCGCACACTCTTTCCCTACACG AC	TCTCGCG C
12	Fwd-L	AATGATACGGCGACCACCGAGATCTACACAGCGATAGACACTCTTTCCCTACAC GAC	AGCGATA G

1 3	Rev-1	CAAGCAGAAGACGGCATAACGAGATATCGTGATGTGACTGGAGTTCAGACGTGT GCTCTTC	ATCACGA T
1 4	Rev-2	CAAGCAGAAGACGGCATAACGAGATATACATCGGTGACTGGAGTTCAGACGTGT GCTCTTC	CGATGTA T
1 5	Rev-3	CAAGCAGAAGACGGCATAACGAGATATGCCTAAGTGACTGGAGTTCAGACGTGT GCTCTTC	TTAGGCA T
1 6	Rev-4	CAAGCAGAAGACGGCATAACGAGATATTGGTCAGTGACTGGAGTTCAGACGTGT GCTCTTC	TGACCAA T
1 7	Rev-5	CAAGCAGAAGACGGCATAACGAGATATCACTGTGTGACTGGAGTTCAGACGTGT GCTCTTC	ACAGTGA T
1 8	Rev-6	CAAGCAGAAGACGGCATAACGAGATATATTGGCGTGACTGGAGTTCAGACGTGT GCTCTTC	GCCAATA T
1 9	Rev-7	CAAGCAGAAGACGGCATAACGAGATATGATCTGGTGACTGGAGTTCAGACGTGT GCTCTTC	CAGATCA T
2 0	Rev-8	CAAGCAGAAGACGGCATAACGAGATATTCAAGTGTGACTGGAGTTCAGACGTGT GCTCTTC	ACTTGAA T
2 1	Rev-9	CAAGCAGAAGACGGCATAACGAGATATCTGATCGTGACTGGAGTTCAGACGTGT GCTCTTC	GATCAGA T
2 2	Rev-10	CAAGCAGAAGACGGCATAACGAGATATAAGCTAGTGACTGGAGTTCAGACGTGT GCTCTTC	TAGCTTA T
2 3	Rev-11	CAAGCAGAAGACGGCATAACGAGATATGTAGCCGTGACTGGAGTTCAGACGTGT GCTCTTC	GGCTACA T
2 4	Rev-12	CAAGCAGAAGACGGCATAACGAGATATTACAAGGTGACTGGAGTTCAGACGTGT GCTCTTC	CTTGTA T
2 5	Rev-13	CAAGCAGAAGACGGCATAACGAGATTGTTGACTGTGACTGGAGTTCAGACGTGT GCTCTTC	AGTCAAC A
2 6	Rev-14	CAAGCAGAAGACGGCATAACGAGATACGGAAGTGTGACTGGAGTTCAGACGTGT GCTCTTC	AGTTCCG T
2 7	Rev-15	CAAGCAGAAGACGGCATAACGAGATTCTGACATGTGACTGGAGTTCAGACGTGT GCTCTTC	ATGTCAG A
2	Rev-	CAAGCAGAAGACGGCATAACGAGATCGGGACGGGTGACTGGAGTTCAGACGTG	CCGTCCC

8	16	TGCTCTTC	G
2 9	Rev- 18	CAAGCAGAAGACGGCATAACGAGATGTGCGGACGTGACTGGAGTTCAGACGTGT GCTCTTC	GTCCGCA C
3 0	Rev- 19	CAAGCAGAAGACGGCATAACGAGATCGTTTCACGTGACTGGAGTTCAGACGTGT GCTCTTC	GTGAAAC G
3 1	Rev- 20	CAAGCAGAAGACGGCATAACGAGATAAGGCCACGTGACTGGAGTTCAGACGTGT GCTCTTC	GTGGCCT T
3 2	Rev- 21	CAAGCAGAAGACGGCATAACGAGATTCCGAAACGTGACTGGAGTTCAGACGTGT GCTCTTC	GTTTCGG A
3 3	Rev- 22	CAAGCAGAAGACGGCATAACGAGATTACGTACGGTGACTGGAGTTCAGACGTGT GCTCTTC	CGTACGT A
3 4	Rev- 23	CAAGCAGAAGACGGCATAACGAGATATCCACTCGTGACTGGAGTTCAGACGTGT GCTCTTC	GAGTGG AT
3 5	Rev- 25	CAAGCAGAAGACGGCATAACGAGATAAAGGAATGTGACTGGAGTTCAGACGTGT GCTCTTC	ATTCCTTT
3 6	Rev- 27	CAAGCAGAAGACGGCATAACGAGATATATCAGTGTGACTGGAGTTCAGACGTGT GCTCTTC	ACTGATA T

Author Manuscript