

# Labeling Endogenous Proteins Using CRISPR-mediated Insertion of Exon (CRISPIE)

Evan A. Wilson, Tianyi Mao and Haining Zhong\*

Vollum Institute, Oregon Health & Sciences University, Portland, OR, USA

\*For correspondence: [zhong@ohsu.edu](mailto:zhong@ohsu.edu)

## Abstract

The CRISPR/Cas9 technology has transformed our ability to edit eukaryotic genomes. Despite this breakthrough, it remains challenging to precisely knock-in large DNA sequences, such as those encoding a fluorescent protein, for labeling or modifying a target protein in post-mitotic cells. Previous efforts focusing on sequence insertion to the protein coding sequence often suffer from insertions/deletions (INDELS) resulting from the efficient non-homologous end joining pathway (NHEJ). To overcome this limitation, we have developed CRISPR-mediated insertion of exon (CRISPIE). CRISPIE circumvents INDELS and other editing errors by inserting a designer exon flanked by adjacent intron sequences into an appropriate intronic location of the targeted gene. Because INDELS at the insertion junction can be spliced out, “CRISPIED” genes produce precisely edited mRNA transcripts that are virtually error-free. In part due to the elimination of INDELS, high-efficiency labeling can be achieved *in vivo*. CRISPIE is compatible with both N- and C-terminal labels, and with all common transfection methods. Importantly, CRISPIE allows for later removal of the protein modification by including exogenous single-guide RNA (sgRNA) sites in the intronic region of the donor module. This protocol provides the detailed CRISPIE methodology, using endogenous labeling of  $\beta$ -actin in human U-2 OS cells with enhanced green fluorescent protein (EGFP) as an example. When combined with the appropriate gene delivery methods, the same methodology can be applied to label post-mitotic neurons in culture and *in vivo*. This methodology can also be readily adapted for use in other gene editing contexts.

**Keywords:** CRISPR/Cas9, Genome editing, Non-homologous end joining (NHEJ), Fluorescence microscopy, Somatic cell editing, Transfection, Knock-in, Fluorescent labeling of endogenous proteins

**This protocol was validated in:** eLife (2021), DOI:10.7554/eLife.64911

## Background

Epitope tagging is widely used to label or modify a protein of interest to study its function. However, such tagging has been traditionally done via overexpression of the tagged protein from an exogenous plasmid. Protein overexpression often causes undesirable side effects, such as altered protein localization, stoichiometry, and cell physiology. To epitope-tag proteins at their endogenous levels, mouse conditional knock-in methods have been developed (e.g., Fortin *et al.*, 2014; Melander *et al.*, 2021). However, the cost and inflexibility of mouse genetics have limited their broad application. Knock-ins in other species also remain difficult.

To overcome these issues, recent efforts have employed the emerging CRISPR/Cas9 technology to introduce protein epitope tags. However, while successful labeling can be achieved in a fraction of cells, the previous methods have limitations. Although methods utilizing the homology-directed repair (HDR) mechanism hold the promise of precise DNA insertions (Nishiyama *et al.*, 2017; Roberts *et al.*, 2017), HDR is inefficient in post-mitotic cells, and is still sensitive to INDEL formation resulting from the competing non-homologous end joining (NHEJ) pathway. On the other hand, while NHEJ can achieve high repair efficiency across diverse cell types, INDEL formation at the repaired junction of double-strand breaks (DSBs) poses a significant challenge because INDELS in protein coding sequences can cause frameshift and nonsense mutations. Although this can be overcome by selecting cells with the correct editing events where growing/cloning is appropriate, its application in post-mitotic cells, such as neurons, can be problematic.

To circumvent the negative impacts of INDEL formation in exons, CRISPIE targets intronic sequences (Zhong *et al.*, 2021). The desired epitope tag is included in a designer exon module, flanked by intronic sequences at both ends of the exon. When this donor module is inserted into the intronic targeting site, the epitope tag is included in the endogenous transcript via RNA splicing. The same splicing event also eliminates INDEL events at the junction of insertion, except for the low-probability event that the INDEL is very big and it interferes with RNA splicing. Therefore, the CRISPIE process results in epitope tagging of endogenous proteins that is nearly error-free at the mRNA level, and thus at the protein level.

Compared to previous NHEJ-based epitope insertion methods, the main differences of CRISPIE are that the donor module is a functional exon, and the targeting site is within the intron. We have also empirically found that different intronic locations give highly heterogeneous labeling efficiency. While future bioinformatic approaches could better predict labeling efficiencies, it is currently necessary to screen several editing locations within a given intron. We suggest targeting at least 100 bp away from the intron-exon junction, to avoid potential interference with the splicing of endogenous exons. When optimized, we can achieve ~40% labeling efficiency *in vivo*, corresponding to ~22% per allele in a diploid genome.

Here, we detail a step-by-step protocol for achieving CRISPIE labeling of human  $\beta$ -actin (ACTB) at its N-terminus in human U-2 OS cells. Labeling of other genes can be similarly carried out using gene-specific sgRNA targeting sequences and the appropriate donors. Although not described, labeling in post-mitotic neurons follows an essentially identical procedure barring the transfection method. This protocol should be readily adapted for other proteins, cell types, and epitope tags.

## Materials and Reagents

1. 10 cm surface treated tissue culture dishes (Fisherbrand™, catalog number: FB012924)
2. 1.7 mL microcentrifuge tube (Thermo Scientific, catalog number: 21-402-903)
3. 35 mm cell culture dishes (Falcon™, catalog number: 353001)
4. Sterile polystyrene disposable serological pipets (Fisherbrand™, catalog numbers: 13-678-11 [25 mL]; 13-678-11E [10 mL]; 13-678-11D [5 mL])
5. Disposable borosilicate glass Pasteur pipets (Fisherbrand™, catalog number: 13-678-20C)
6. Human U-2 OS cells (ATCC, catalog number: HTB-96)
7. pX330-U6-Chimeric\_BB-CBh-hSpCas9 (Addgene, catalog number: 42230)
8. pKanCMV\_mRuby3 (Addgene, catalog number: 172854)
9. Lipofectamine 2000 (Invitrogen, catalog number: 11668019)

10. McCoy's 5A medium (ATCC, catalog number: 30-2007)
11. Opti-MEM (Gibco, catalog number: 31985062)
12. Trypsin/EDTA (Gibco, catalog number: 25300062)
13. Fetal bovine serum (FBS) (Hyclone, catalog number: 26140079)
14. 70% ethanol in deionized water
15. BbsI-HF (New England Biolabs, catalog number: R3539S)
16. Poly nucleotide kinase (PNK) (New England Biolabs, catalog number: M0201S)
17. Calf Intestinal Phosphatase (CIP) (New England Biolabs, catalog number: M0290S)
18. 10× Cut Smart Buffer (New England Biolabs, catalog number: B6004S)
19. T4 DNA ligase buffer (New England Biolabs, catalog number: B0202S)
20. Quick Ligation Kit (New England Biolabs, catalog number: M2200S)
21. Molecular grade water (HyClone, catalog number: SH30538.03)
22. Carbenicillin (Sigma, catalog number: C1389-5G)
23. Ampicillin (Sigma, catalog number: A9518-25G)
24. Kanamycin (Sigma, catalog number: 11815024)
25. DH5α competent cells (ThermoFisher Scientific, catalog number: 18265017)
26. Luria-Bertani (LB) agar capsules (for making plates) (Sigma, catalog number: 113002231)
27. Kanamycin LB agar plates (30 µg/mL; made in-house)
28. Ampicillin LB agar plates (50 µg/mL; made in-house)
29. Tryptone (Fisher Scientific, catalog number: BP9726-500)
30. Sodium chloride (Thermo Scientific, catalog number: J2161836)
31. Yeast extract (Gibco, catalog number: 211929)
32. Terrific Broth (TB) powder (Sigma, catalog number: 1016290500)
33. Glycerol (Sigma, catalog number: G5516-500ML)
34. Glass beads (Fisher Scientific, catalog number: 11-312C)
35. Ice bucket
36. QiaQuick Purification Kit (Qiagen, catalog number: 28706)
37. Qiagen Miniprep Kit (Qiagen, catalog number: 27104)
38. SOC Medium (see Recipes)
39. TB Medium (see Recipes)
40. McCoy's 5A Medium supplemented with FBS (see Recipes)

## Equipment

1. Water baths set to 37°C and 42°C, respectively
2. Tissue culture incubator
3. Bacterial incubator
4. Ice machine
5. Vortex (Fisher)
6. Pipetaid (Drummond)
7. Biological safety cabinet
8. Vacuum source
9. Vacuum manifold (Vac-Man™ Vacuum Manifold, Promega, catalog number: A7231)
10. Thermocycler
11. Tabletop microcentrifuge
12. Fluorescence microscope

## Software

1. SnapGene (SnapGene software; <https://snapgene.com>)
2. CRISPick (Broad Institute, <https://portals.broadinstitute.org/gppx/crispick/public>)

## Procedure

### A. Cas9 sgRNA design using CRISPick

1. The first CRISPIE plasmid needed is a Cas9 expression plasmid that also expresses the target-specific sgRNA. This allows the targeted genomic location to be specifically cut by Cas9. After designing the sgRNA *in silico*, it will be incorporated into the blank Cas9-sgRNA expression vector pX330 (Addgene #42230).
2. To select an sgRNA targeting sequence, we use Broad Institute CRISPick (<https://portals.broadinstitute.org/gppx/crispick/public>; note that other sgRNA design websites/software may also work).
3. On the starting page, select the desired reference genome, CRISPR mechanism, and origin species of the Cas9 enzyme (Figure 1).

*Note: In our protocol, we use the 'Human GRCh38' reference genome, CRISPRko, and SpyoCas9.*

Reference Genome	Mechanism	Enzyme
<input checked="" type="radio"/> Human GRCh38 (NCBI RefSeq v.109.20210514)	<input checked="" type="radio"/> CRISPRko	<input checked="" type="radio"/> SpyoCas9 (NGG)
<input type="radio"/> Human GRCh38 (Ensembl v.104)	<input type="radio"/> CRISPRa	<input type="radio"/> SaurCas9 (NNGRR)
<input type="radio"/> Human GRCh37 (NCBI RefSeq v.105.20201022)	<input type="radio"/> CRISPRi	<input type="radio"/> AsCas12a (TTTV)
<input type="radio"/> Mouse GRCm38 (NCBI RefSeq v.108)		<input type="radio"/> enAsCas12a
<input type="radio"/> Mouse GRCm38 (Ensembl v.102)		
<input type="radio"/> Rat Rnor_6.0 (NCBI RefSeq v.106)		

**Figure 1. Specifying CRISPick parameters used in CRISPIE.**

4. Click 'bulk mode' and paste the target sequence into the 'Target(s)' box, or reference the target using the accepted target formats listed on the website.
  - a. Bulk mode allows one to paste in large/multiple target sequences (Figure 2).
  - b. We recommend creating SnapGene files containing the target sequence with annotated sgRNA binding sites. This is especially useful when designing multiple sgRNAs to screen, and when analyzing the final insertion sequences.
  - c. Include an additional ~200 bp of sequence to both the 5' and 3' end of the target sequence, to ensure all sgRNA options are considered.
  - d. A target sequence needs to be located within the intron of a gene. The inherent advantage of CRISPIE is that INDELS at the insertion junction within the intronic region are spliced out, which is not possible when the INDELS are located within an exon. When targeting introns of a gene, it is important to choose an intron that won't disrupt the structure of a protein after a donor module is introduced. N- and C-terminal labels are incorporated into the first or last intron of a gene, respectively. If a large stretch of coding sequence is present in the first exon of a gene, CRISPIE would have difficulty tagging its N terminus. Targeting within the gene or at the C terminus can be used as alternatives. In addition, in order to label a gene of interest at the very C-terminus, the coding sequence of the last

exon of a gene needs to be included in the donor module. These are the constraints that need to be considered using CRISPIE. However, the sequence that needs to be included for CRISPIE is typically much shorter than those for HDR-based insertions.

### Target(s)

Quick lookup  Bulk  Upload file

```
gtaagcccggccagccgaccggggcaggcggctcacggcccggcgcaggcggcggcccctcgcccgtgcagag
ccgccgtctggccgcagcggggggcgcatggggggggaaccggaccgccgtggggggcgccgggagaagccctggg
cctccggagatgggggacaccccacgccagttcggaggcgcaggccgcctcgggaggcgcgctccgggggtccgc
tctcggggcggggcaaccggcgggtcttgtctgagccggctcttgccaatggggatcgagggtggcgcgcg
agccccccagggcccgggtggggctggggcggcattgcgctgcgctggtccttggcgctaactgctgcgct
gggaattggcgctaattgcgctgcgctgggactcaaggcgtaactgcgctgcttctggggcccgggggtccgcg
gcctgggtggggcgaaggcgggctcggccggaagggtggggctgcccggcgtcccggcgcttgcgcacttctg
cccagccgctggcccgggggtggccgctgctgcccgcgcccggggcgtgttgaaccggcggaggc
ggggctggcgcccgggtgggagggggtggggcctggcttctgcccgcgcccggggacgctccgaccagtgttgc
ctttatggtaataacgcgcccggccttcttctgcccgaatctggcgcgcccggcgcccctggcgccctaagg
actcggcgcccgggaagtggccaggcggggcgacctggctcacagcgcggcctattctcgag
```

Figure 2. Pasting target sequence (FASTA format) into the bulk mode input box.

5. Select the desired ‘CRISPick Quota’.

*Note: If one wishes to manually select the desired sgRNAs from the output, skip this step. Otherwise, this tool uses an algorithm to recommend the top sgRNAs based on the standard criteria, such as off target effects, cut positions, etc.*

6. Check the box ‘report unpicked sequences?’ (Figure 3).

### CRISPick Quota ⓘ

This tool will recommend the top N candidates according to:

- Raw ranking
- Cut position
- Mutual spacing

Report unpicked sequences?

Figure 3. Specifying the number of potential sgRNA sites by setting the CRISPick Quota.

Check “Report unpicked sequences” in order to have a file of all potential sgRNAs and their characteristics (as determined by the CRISPick algorithm).

7. After the output is displayed, download the desired files:
  - a. ‘Picking Results’ displays the best sgRNAs according to the ‘CRISPick Quota’ algorithm.
  - b. ‘Picking Summary’ provides a comprehensive list of possible sgRNAs in a table format with relevant sgRNA information.
  - c. ‘Input Target Sequence’ contains the input file you used, for your records.
8. Note that if the selected sequence will be incorporated into the donor plasmid (for excising the module so it can be inserted into the targeted genomic region), it may be desirable to include the ‘context sequence’ (also listed in the Picking Results). The context sequence is the set of sequences surrounding the sgRNA-targeted sequence that may affect Cas9 binding and the endonuclease activity at the target sequence (Figure 4).



**Figure 4. An example sgRNA target sequence including its context sequence.**

We have empirically found that the specific context sequences shown in the figure give high Azimuth scores based on CRISPick.

- Using a DNA design software (we use SnapGene), create a file with the desired gene of interest to be edited using CRISPIE. Annotate the target sites in the target gene.

### B. *In silico* construction of Cas9-containing plasmid

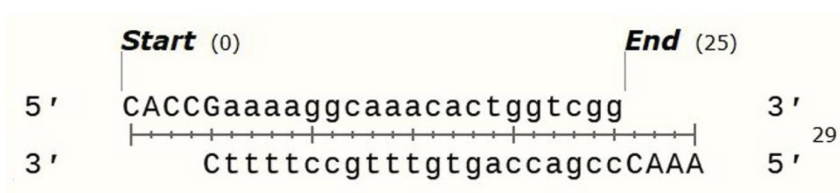
- Select the sgRNA sequence chosen from CRISPick (or other sgRNA design software).
- To incorporate the sgRNA into the pX330 plasmid via the built-in BbsI site, add the 4 bp sequence: CACC to the 5' end of the Forward oligo and AAAC to the 5' end of the Reverse oligo. This is necessary to create compatible overhangs for directional cloning of the oligos into the BbsI-digested pX330 vector.
- Additionally, if the sgRNA does not start with a G, add a G at the 5' end of the sgRNA, which is required for transcription initiation by the U6 promoter. An example is provided below for clarity (Figure 5B).

**A.**



**B.**

ACTB\_F: CACCGaaaaggcaaacactggtcgg  
 ACTB\_R: AAACccgaccagtgtttgcctttC



**Figure 5. *In silico* design of sgRNAs.**

(A) Schematic of the human  $\beta$ -actin gene with exons numbered 1–6. The red bar indicates the approximate targeting site of the CRISPIE sgRNA. The coding sequence of *ACTB* is highlighted in orange. (B) shows both the forward and reverse primer sequences for cloning the sgRNA plasmid targeting intron 1 of human *ACTB*. The sgRNA sequences (black, lower-case letters), adjacent sequences for ligating into the target vector (blue capitalized letters), and the transcription initiating G (magenta) are differentially highlighted. The lower panel shows the annealed primers, which will be ligated into the pX330 vector.

### C. Cloning the sgRNA sequence into pX330

1. Digestion, dephosphorylation, and purification of the vector
  - a. Digest pX330 (Addgene #42230) in a 25  $\mu$ L reaction at 37°C for 3 h or overnight:
 

pX330	~3 $\mu$ g
10 $\times$ NEB Cut Smart buffer	2.5 $\mu$ L
BbsI-HF	1 $\mu$ L (20 units)
Molecular-grade water	Bring to 25 $\mu$ L
  - b. After digestion, add 0.6  $\mu$ L of calf intestinal phosphatase (CIP) and incubate at 37°C for 30 min.
  - c. Purify the digested plasmid. We typically use the Qiaquick purification kit.

*Note: We recommend the HF version of BbsI. The non-HF version may lose potency over time if stored at -20°C.*

2. Phosphorylation and annealing of sgRNA oligos
  - a. Set up the following reaction in a PCR tube, so the polynucleotide kinase (PNK) reaction and annealing can both be performed in one step in a thermocycler:
 

Forward sgRNA Oligo (100 $\mu$ M stock)	1 $\mu$ L
Reverse sgRNA Oligo (100 $\mu$ M stock)	1 $\mu$ L
10 $\times$ T4 DNA Ligase Buffer	2 $\mu$ L
Molecular-grade water	15.5 $\mu$ L
PNK	0.6 $\mu$ L
Total	20 $\mu$ L
  - b. Carry out the following reaction protocol:
 

Cycle	Notes
37°C for 30 min	PNK reaction
95°C for 5.5 min	Inactivate PNK and denature oligos
-1°C every 30 s for 70 cycles	Anneal oligos
Hold at 4°C	
  - c. Dilute 2  $\mu$ L of the annealing reaction in 198  $\mu$ L of molecular-grade water for use in the ligation.
3. Ligation of annealed oligos into the digested pX330 vector
 

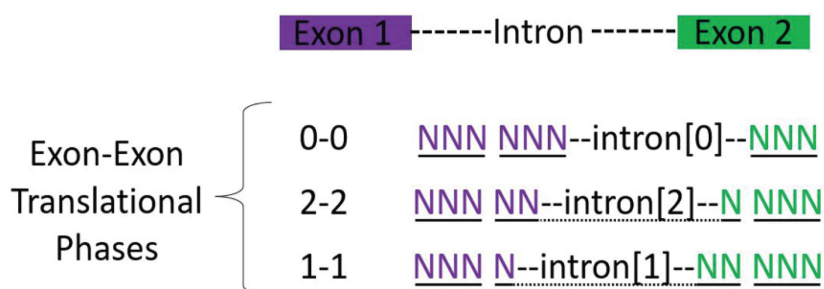
Using the NEB Quick Ligation kit, create the following reactions and incubate at room temperature for 25 min:

	Vector + Ligase Control	Insert + Vector
NEB 2 $\times$ Quick Ligase Buffer	5.5 $\mu$ L	5.5 $\mu$ L
Digested, dephosphorylated pX330	1.2 $\mu$ L	1.2 $\mu$ L
1:100 diluted annealing reaction	–	3.8 $\mu$ L
Molecular-grade water	3.8 $\mu$ L	–
NEB Quick Ligase	0.5 $\mu$ L	0.5 $\mu$ L

4. Transformation and amplification of DH5 $\alpha$  cells with ligated plasmid using standard protocols.
  - a. We use home-made DH5 $\alpha$  competent cells. With all tubes pre-chilled on ice, add 1–2  $\mu$ L of ligation reaction (above) to 30  $\mu$ L of DH5 $\alpha$  cells. Leave on ice 20–30 min. Incubate tubes in a water bath at 42°C for exactly 1 min, then immediately place on ice for 5 min. Add 300  $\mu$ L of SOC medium and incubate the tubes at 37°C shaking at 250 rpm for 1 h. Plate ~200  $\mu$ L of the transformation reaction on LB agar plates (50  $\mu$ g/mL Carbenicillin) and spread with glass beads. Incubate overnight at 37°C.
  - b. Select colonies to be amplified using the desired miniprep technique (e.g., Qiagen Miniprep Kit).
  - c. Sequence the miniprep plasmid to check for the proper insertion of the sgRNA. Our lab uses Azenta Life Sciences (previously Genewiz) sanger sequencing services.

#### D. Selecting CRISPIE Donor Plasmid and Transfection Marker

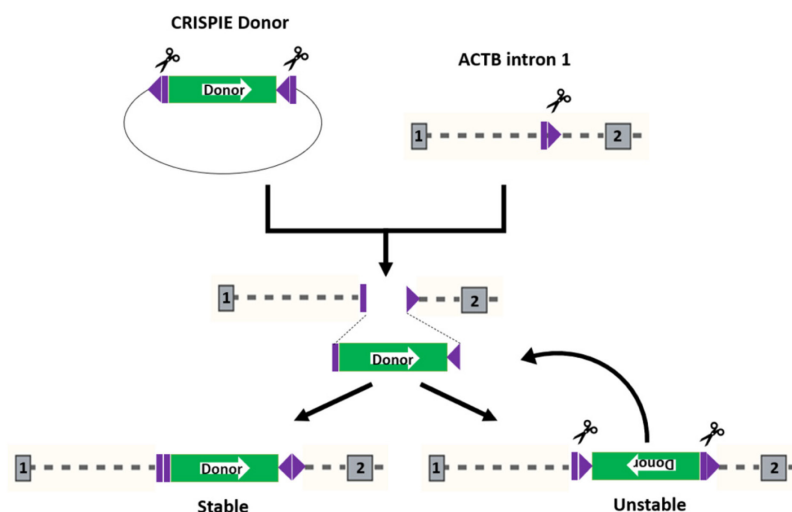
1. The choice of which CRISPIE donor to select from the Addgene repository depends on the translational reading frame of the exons flanking the targeted intron. RNA splicing can occur both between and within individual codons of a gene. Exons have splice sites that can be located before the first, second, or third base of a codon (notated as 0, 1, and 2, respectively; Figure 6).  
It is important to select a donor module that matches the reading frame of the inserted intron. Depending on the target site, one would need either a 0-0, 1-1, or 2-2 translational phase donor exon. In the case of *ACTB*, the start codon resides on exon 2 in the first (0) reading frame. Therefore, the 0-0 donor module is the desired plasmid for intron 1 of *ACTB*.
  - a. Addgene #172848 (Donor B7) contains a mEGFP with 0-0 translational phase
  - b. Addgene #172849 (Donor B8) contains a mEGFP with 1-1 translational phase
  - c. Addgene #172850 (Donor B9) contains a mEGFP with 2-2 translational phase
2. The donor plasmids listed above are self-excising plasmids: they express a designed sgRNA whose target site flanks the donor module within the same plasmid. However, we have empirically found that a donor module flanked by sgRNA sites identical to the genomic editing site, but positioned in an inverse orientation, gives a ~50–100% higher labeling efficiency compared to the generic donors. Presumably, this is in part because the target-specific sequence, when positioned in an inverse direction, can facilitate the insertion of donor in the forward direction, thus increasing the rate of successful editing (Figure 7).
  - a. Addgene #172842 is a mEGFP donor specific for the first exon of *ACTB* for the sgRNA shown in Figure 2.
  - b. The original donor plasmids were generated by gene synthesis (using Azenta Life Sciences, formerly Genewiz).
  - c. Other target-specific donor modules can be generated by replacing the current sgRNA targeting site of a donor plasmid with the appropriate target specific sgRNA sequence, including PAM and context sequences, using the available restriction enzyme sites. In this way, researchers can target intronic locations of other genes. Donors with different tags, such as different fluorescent proteins (FPs), can be generated by replacing the FP from existing donors.
  - d. The donor sequence include the tag-encoding exon flanked by important adjacent sequences needed for splicing and donor excision from the plasmid (Figure 8).
3. We also use an mRuby expression vector (Addgene #172854) as the transfection marker. Other transfection markers should be characterized and optimized before incorporation into the CRISPIE procedure.



**Figure 6. Reading frame of exons.**

The three possible splice junctions in terms of translational phases between two exons are shown. Each triplet codon is underlined, with the dotted underline representing a codon spanning across an intron.





**Figure 7. Opposite orientations of the sgRNA target site at the genomic locus and the donor facilitate the donor insertion in the forward orientation.**

After Cas9 creates a double-stranded break in genomic locus, the donor module can be inserted in either the forward or reverse orientation. Since the sgRNA sites flanking the donor module are in the opposite orientation relative to the sgRNA site at the genomic locus, forward insertion of the donor module destroys both sgRNA sites flanking the donor, resulting in a stable insertion. Inserting the donor in the reverse orientation creates functional sgRNA sites that can be subjected to further Cas9 cleavage, thus creating an unstable insertion (unless INDELS have occurred at the insertion junctions). The sgRNA binding sites are shown in purple, with the Cas9 cut site represented as the thin white line between the purple rectangle and triangle. The white arrow within the donor module represents the orientation of the donor module.



**Figure 8. The 5' to 3' sequence of a representative CRISPIE donor module.**

Sequence of Donor B1 (Addgene#172842; phase 0-0) is shown. Capitalized case letters: exonic sequence; lower case letters: intronic sequence; yellow: intronic sequence from the intron 17 of mouse CaMKIIα

(*Camk2a*); underlined: exonic sequence adjacent to *Camk2a* intron 17; purple: sgRNA targeting site; green: EGFP. Note, only the exonic sequence will be integrated into the final mRNA.

## E. U-2 OS Lipofectamine 2000 Transfection

1. Culture U-2 OS cells in McCoy's 5A medium (10% FBS). Set the incubator to 37°C and 5% CO<sub>2</sub>. The cells should be passaged every 2-3 days, when they are at ~90% confluency.
2. Twenty-four hours prior to transfection, seed  $\sim 1.25 \times 10^5$  cells in a 35 mm culture dish containing 2 mL of McCoy's 5A medium (10% FBS) and let attach overnight. Swirl cells clockwise five times, anti-clockwise five times, up and down five times, and left and right five times. This ensures an even coating of cells on the bottom of the dish. We do not use antibiotics, although we have not tested whether antibiotics may affect cell growth or labeling efficiency.
3. Before beginning transfection, ensure that the cells are 30–50% confluent. Growth outside this range could reduce transfection efficiency.
4. For each transfection reaction, combine 5  $\mu$ L of Lipofectamine 2000 with 145  $\mu$ L of Opti-MEM medium. Mix thoroughly by vortexing. Incubate the mixture at room temperature for 5 min.

*Note: Create a master mix when performing multiple transfections.*

5. According to the manufacture's protocol, the incubation time for this lipofectamine 2000/Opti-MEM master mix can be up to 20 min.
6. For each transfection reaction, combine 0.2  $\mu$ g of the transfection marker (mRuby3; acquired from Addgene), 1.0  $\mu$ g of the sgRNA/Cas9 plasmid (created in Procedure A–C above), and 1.0  $\mu$ g of the donor module-containing plasmid (acquired from Addgene) in 150  $\mu$ L of Opti-MEM into a separate tube. Negative controls can be made by excluding either the sgRNA/Cas9 plasmid or the donor module-containing plasmid from the reaction.

*Note: It is important to only use trace amounts (0.2  $\mu$ g) of the transfection marker to prevent masking of the fluorescence from the GFP-labeled cells by the transfection marker. Most red fluorescent proteins, including mRuby3, have bleed-through in the green channels, in part due to the bifurcating maturation of the fluorophore. At the same time, fluorescent signals of endogenously labeled proteins are often dim, and therefore can be easily obscured by the bleed-throughs from highly expressed red fluorescent proteins.*

7. After the 5 min lipofectamine/Opti-MEM incubation is complete, combine this mix with the DNA/Opti-MEM mixture, then thoroughly pipette up and down and vortex to mix. Incubate at room temperature for 20 min. Thorough mixing is important for optimal transfection efficiency.
8. Allocate each transfection mixture (300  $\mu$ L) to a 35 mm dish containing U-2 OS cells that were passaged 24 h prior.
9. Five and a half hours post-transfection, carefully aspirate off the transfection medium and add 2 mL of warmed McCoy's 5A medium (10% FBS) to each dish. Incubate at 37°C and 5.0% CO<sub>2</sub> for 30 min.

*Note: Pay attention to suction and add medium from the side of the dish to minimize perturbation of the cells at the bottom of the dish.*

10. Suction off media, add 2 mL of pre-warmed McCoy's 5A medium (10% FBS) to the cells, and place them in the incubator set at 37°C and 5.0% CO<sub>2</sub>.

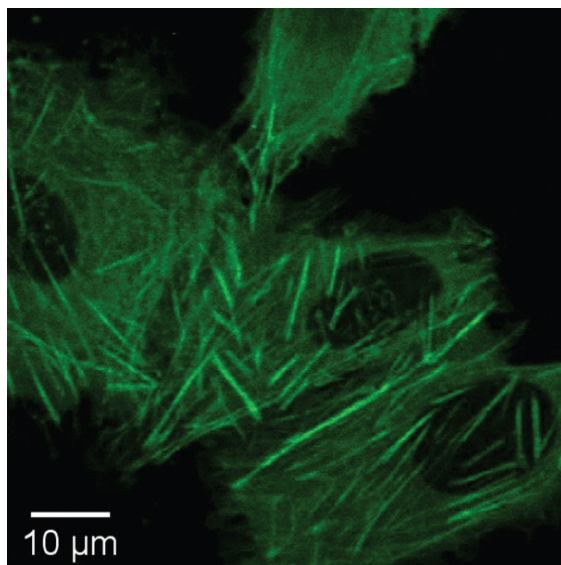
*Note: These washing steps are essential because prolonged exposure to lipofectamine 2000 impairs U-2 OS cell health.*

11. Protein expression can be assessed at 72 h post-transfection.

## F. U-2 OS cell counting used to determine transfection efficiency

1. Using a brightfield view at 100× magnification (a 10× air objective [0.3 NA] and 10× eyepieces), pick a region of cells that have regular, evenly spread growth.
2. Once the field of view (FOV) is found, zero the X and Y directions and switch to 400× magnification (40× water immersion objective [0.8 NA] and 10× eyepieces).
3. Switch to the red FP (RFP) filter and count any cells expressing mRuby. While still in the same FOV, switch to the GFP filter. Count the cells expressing GFP.

*Note: We use an Olympus BW51XI upright microscope. However, most standard fluorescence microscopes with GFP and RFP filtersets (e.g., Chroma ET49002 and ET49005) should work. It is important to only count cells that have GFP labeling of the desired protein. In the case of actin, filaments of GFP-actin should be clearly visible (see Figure 9); avoid counting cells that have non-specific GFP fluorescence (e.g., fluorescence permeating the whole cell), which occasionally happens, possibly due to finite specificity of the Cas9 editing.*



**Figure 9. Example of successfully labeled U-2 OS cell.**

Note the GFP-labeled actin filaments and the absence of GFP in the nuclei. Only count cells that have GFP fluorescence in accordance with the characteristic morphological distribution of the labeled protein.

4. Repeat this counting method for 12 FOV, using a pre-determined, raster pattern that is consistent across all samples. To reduce variations, we omit any FOV with <10 mRuby positive cells, and replace it with additional, pre-determined FOVs.
5. These cells can be imaged live or fixed, using available epi, confocal, or two-photon fluorescence microscopes.

6. Labeling efficiency is calculated by the number of successfully labeled cells (in the example, green) divided by the total number of transfected cells (red) in an FOV.

If determination of the labeling efficiency of each gene locus is needed, it can be derived from the labeling efficiency and the copy of target sites per cell using the equation below, assuming that the editing of each locus is an independent event.

$$\text{Cell labeling efficiency} = 1 - (1 - \text{allele labeling efficiency})^{\text{copy number}}$$

7. Assessment of the accuracy of CRISPIE editing in the genome can be achieved through multiple avenues. We typically use next generation sequencing (NGS). The genomic DNA and mRNA of GFP-positive cells are enriched by Fluorescence-Activated Cell Sorting (FACS) and subjected to RT-PCR and PCR, respectively, using the appropriate primers to detect the insertion junctions. Where needed, nested PCR can be used to increase specificity. The PCR amplicons are then sent for NGS (we uses Azenta Life Sciences, formerly known as Genewiz). As shown in the original publication (Zhong *et al.*, 2021), the typical result is that INDELS are abundantly present in genomic DNA sequences, but are largely absent in mRNA sequences.
  - a. The primer pairs should straddle a span of ~2–300 bp at both the 5' and 3' insertion junctions of the donor module. This ensures that the PCR will detect relatively large insertions and deletions.
  - b. Our lab uses custom MATLAB codes to align sequences and identify editing outcomes. However, this can be done by using available software packages, or by collaborating with NGS experts.

## Recipes

### 1. SOC Medium (adapted from Cold Spring Harbor, 2018)

- a. Combine the following in 450 mL of deionized water with a stir bar (recommended):
  - 10 g Tryptone
  - 2.5 g Yeast Extract
  - 0.25 g NaCl
- b. Once the components are thoroughly mixed, add 5 mL of 250 mM KCl.
- c. Adjust the pH of the solution to 7 with 10 N NaOH (roughly 100  $\mu$ L).
- d. After ensuring the total volume is 500 mL, autoclave for 30 min at 20 psi on a liquids cycle.
- e. Add 2.5 mL of filter-sterilized 2 M MgCl<sub>2</sub> solution.
- f. Add 10 mL of filter-sterilized 1 M glucose solution.

### 2. TB Medium

Following manufacturer instructions on the bottle:

- a. Combine 47.6 g TB pellets in 1 L of water.
- b. Add 4 mL of glycerol.
- c. Once dissolved, autoclave at 20 psi for 30 min.

### 3. McCoy's 5A Medium supplemented with FBS (10% v/v)

Add 55 mL of FBS to 500 mL of McCoy's 5A medium.

*Note: Although 10% of 500 mL is 50 mL, 55 mL is added to account for the extra volume of FBS in the final volume of the medium. This is important, as altered serum concentration can have significant effects on U-2 OS cell growth.*

## Acknowledgments

None of this work would have been possible without the work by the authors of the original paper (Zhong *et al.*, 2021). This work is supported by a NIH/BRAIN Initiative grant (RF1MH120119) to H.Z. and T.M., and an NIH/NINDS R01 (R01NS081071) to T.M.

## Competing interests

The authors declare no competing interests.

## References

- Cold Spring Harbor. (2018). [SOC Medium](#). Cold Spring Harbor Laboratory Press.
- Fortin, D. A., Tillo, S. E., Yang, G., Rah, J. C., Melander, J. B., Bai, S., Soler-Cedeño, O., Qin, M., Zemelman, B. V., Guo, C., *et al.* (2014). [Live imaging of endogenous PSD-95 using ENABLED: a conditional strategy to fluorescently label endogenous proteins](#). *J Neurosci* 34(50):16698-712.
- Melander, J. B., Nayebi, A., Jongbloets, B. C., Fortin, D. A., Qin, M., Ganguli, S., Mao, T. and Zhong, H. (2021). [Distinct \*in vivo\* Dynamics of Excitatory Synapses Onto Cortical Pyramidal Neurons and Inhibitory Interneurons](#). *SSRN Electron J*. doi:10.2139/ssrn.3837634.
- Nishiyama, J., Mikuni, T. and Yasuda, R. (2017). [Virus-Mediated Genome Editing via Homology-Directed Repair in Mitotic and Postmitotic Cells in Mammalian Brain](#). *Neuron* 96(4): 755-768 e755.
- Roberts, B., Haupt, A., Tucker, A., Grancharova, T., Arakaki, J., Fuqua, M. A., Nelson, A., Hookway, C., Ludmann, S. A., Mueller, I. A., *et al.* (2017). [Systematic gene tagging using CRISPR/Cas9 in human stem cells to illuminate cell organization](#). *Mol Biol Cell* 28(21): 2854-2874.
- Zhong, H., Ceballos, C. C., Massengill, C. I., Muniak, M. A., Ma, L., Qin, M., Petrie, S. K. and Mao, T. (2021). [High-fidelity, efficient, and reversible labeling of endogenous proteins using CRISPR-based designer exon insertion](#). *Elife* 10: e64911.