









Defining genome-wide CRISPR-Cas genome-editing nuclease activity with GUIDE-seq

Nikolay L. Malinin¹, GaHyun Lee¹, Cicera R. Lazzarotto¹ , Yichao Li¹ , Zongli Zheng², Nhu T. Nguyen^{3,4}, Matthew Liebers³, Ved V. Topkar^{3,5}, A. John Iafrate³, Long P. Le³, Martin J. Aryee³ , J. Keith Joung³  and Shengdar Q. Tsai¹  

Genome-wide unbiased identification of double-stranded breaks enabled by sequencing (GUIDE-seq) is a sensitive, unbiased, genome-wide method for defining the activity of genome-editing nucleases in living cells. GUIDE-seq is based on the principle of efficient integration of an end-protected double-stranded oligodeoxynucleotide tag into sites of nuclease-induced DNA double-stranded breaks, followed by amplification of tag-containing genomic DNA molecules and high-throughput sequencing. Here we describe a detailed GUIDE-seq protocol including cell transfection, library preparation, sequencing and bioinformatic analysis. The entire protocol including cell culture can be completed in 9 d. Once tag-integrated genomic DNA is isolated, library preparation, sequencing and analysis can be performed in 3 d. The result is a genome-wide catalog of off-target sites ranked by nuclease activity as measured by GUIDE-seq read counts. GUIDE-seq is one of the most sensitive cell-based methods for defining genome-wide off-target activity and has been broadly adopted for research and therapeutic use.

Introduction

Genome-editing nucleases are a widely used technology for modifying the genome of living cells with many important applications in biology and medicine. There are four major classes of engineered nuclease: zinc finger nucleases, meganucleases, transcription activator-like effector nucleases and CRISPR–Cas nucleases¹. In particular, CRISPR–Cas nucleases have been broadly adopted, due to the ease with which they can be programmed to target new sites simply by providing a different guide RNA (gRNA)². In mammalian cells, engineered nuclease-induced DNA double-stranded breaks (DSBs) are commonly repaired by competing endogenous cellular DNA repair pathways such as error-prone nonhomologous end-joining, microhomology-mediated end-joining or relatively precise homology-directed repair. Nonhomologous end-joining is often utilized to introduce indel mutations for gene disruption, while homology-directed repair can mediate precise gene correction.

Genome-editing nucleases can, however, have unintended genome-wide ‘off-target’ activity in a wide variety of cell types and organisms^{3–5}. For basic research, one concern about off-target activity is that it may confound the interpretation of genome-editing experiments. For clinical applications, there is a risk that off-target mutations in functional genomic regions may result in cells with increased proliferative fitness or oncogenic potential⁶. Thus, sensitive, unbiased, genome-wide methods are critical for defining the global activity of genome editors.

Development of GUIDE-seq

We developed genome-wide unbiased identification of double-stranded breaks enabled by sequencing (GUIDE-seq)⁴ based on our initial discovery that blunt, phosphorylated and end-protected double-stranded oligodeoxynucleotide (dsODN) tags could be efficiently integrated into sites of nuclease-induced DSBs. We found that terminal phosphorothioate linkages were key for achieving high-frequency tag integration presumably by resisting cellular exonuclease activity⁷. 3′-end protection is

¹Department of Hematology, St. Jude Children’s Research Hospital, Memphis, TN, USA. ²Department of Biomedical Sciences, City University of Hong Kong, Hong Kong, China. ³Molecular Pathology Unit, Massachusetts General Hospital, Charlestown, MA, USA. ⁴Present address: Cutaneous Biology Research Center, Department of Dermatology, Massachusetts General Hospital, and Harvard Medical School, Boston, MA, USA. ⁵Present address: Biophysics Program, Stanford University, Stanford, CA, USA. ✉e-mail: shengdar.tsai@stjude.org

more important than 5'-end protection for achieving efficient integration⁸, and we recommend the use of the 3'-only end-protected dsODN tags in this protocol. The dsODN sequence was designed to be highly orthogonal to sequences that exist in the human genome, and we have observed that dsODN integration frequencies are strongly correlated with indel mutation frequencies⁴. To map dsODN integration sites genome-wide, we adapted anchored multiplex PCR sequencing to specifically amplify genomic DNA molecules containing the dsODN tag⁹. Finally, we found that GUIDE-seq reads have a characteristic bidirectional outward-mapping signature, originating from the site of the nuclease-induced DSB⁴. We used these characteristics to develop a streamlined easy-to-use computational pipeline for analyzing GUIDE-seq data¹⁰. The final protocol has four main parts: integrating end-protected dsODN tags into genomes of living cells, tag-specific library preparation, sequencing, and bioinformatic pipeline data processing.

Overview of the workflow

Here, we present a detailed protocol for performing GUIDE-seq (Fig. 1). The GUIDE-seq procedure is divided into five sections:

- 1 Synthesis of in vitro transcribed gRNA (Step 1);
- 2 Nucleofection and genomic DNA (gDNA) isolation (Steps 2–14);
- 3 Estimation of on-target site dsODN tag integration frequency (Steps 15–19);
- 4 GUIDE-seq library preparation (Steps 20–33);
- 5 Next-generation sequencing (NGS) and data analysis (Steps 34–54).

Using broadly adopted *Streptococcus pyogenes* CRISPR–Cas9 nuclease as an example, Cas9 and gRNA are delivered into cultured cells either as mammalian expression plasmids (Step 2A) or as a ribonucleoprotein (RNP) complex (Step 2B) along with a short dsODN tag. After delivery of CRISPR–Cas9 nuclease and dsODN tag by nucleofection, cells are typically cultured for 3 d prior to harvesting for GUIDE-seq library preparation.

Library preparation requires several straightforward steps (Fig. 1). gDNA is isolated and randomly fragmented, followed by end-repair, A-tailing and ligation of a single-tailed high-throughput sequencing adapter containing a randomized 8 bp unique molecular index (UMI). Next, two rounds of PCR are performed with primers complementary to the ligated single-tailed NGS adapter and the dsODN tag (Fig. 2).

The UMI introduced with the ligated adapter enables correction of PCR bias introduced during amplification, allowing accurate quantification of editing activity. Tag-specific amplified sequencing reads are consolidated based on their UMI and mapped to a reference genome, and on- and off-target sites are identified by characteristic bidirectional mapping read signatures.

Comparison with alternative methods

Several different techniques have been developed for defining the genome-wide activity of genome-editing nucleases^{4,8,11–17}. These methods fall into two broad categories: cell-based and biochemical methods. In general, cellular methods are the most direct while biochemical methods are the most sensitive.

GUIDE-seq is one of the most sensitive cellular genome-wide activity detection methods. Integrase defective lentivirus capture, the first method for unbiased genome-wide identification of nuclease DSBs, is based on a similar principle as GUIDE-seq but is less sensitive and has higher background due to substantial residual integrase activity of integrase defective lentivirus vectors¹³.

As GUIDE-seq is based on tagging nuclease-induced DSBs with short dsODNs, the location and frequency of tags associated with DSB repair events can be tracked even after the DNA repair event has occurred. In this way, GUIDE-seq read counts reflect the ongoing history of cellular DSB repair events. Other cellular methods such as BLESS (direct in situ breaks labeling, enrichment on streptavidin and next-generation sequencing)¹⁷, BLISS (breaks labeling in situ and sequencing)¹⁵ and DISCOVER-seq (discovery of in situ Cas off-targets and verification by sequencing)¹² detect transient DSBs as they are occurring but become insensitive to sites of off-target mutation after DSB repair is complete. These methods are complementary to GUIDE-seq and can be used to define the kinetics of off-target activity.

Complementary biochemical methods such as Digenome-seq (in vitro Cas9-digested whole-genome sequencing)¹⁴, CIRCLE-seq (circularization for in vitro reporting of cleavage effects by sequencing)¹¹ and CHANGE-seq (circularization for high-throughput analysis of nuclease genome-wide effects by sequencing)⁸ operate on the principle of detecting nuclease-induced DSBs from purified gDNA. These methods have higher sensitivity but a lower apparent validation rate due to the limits of detection of

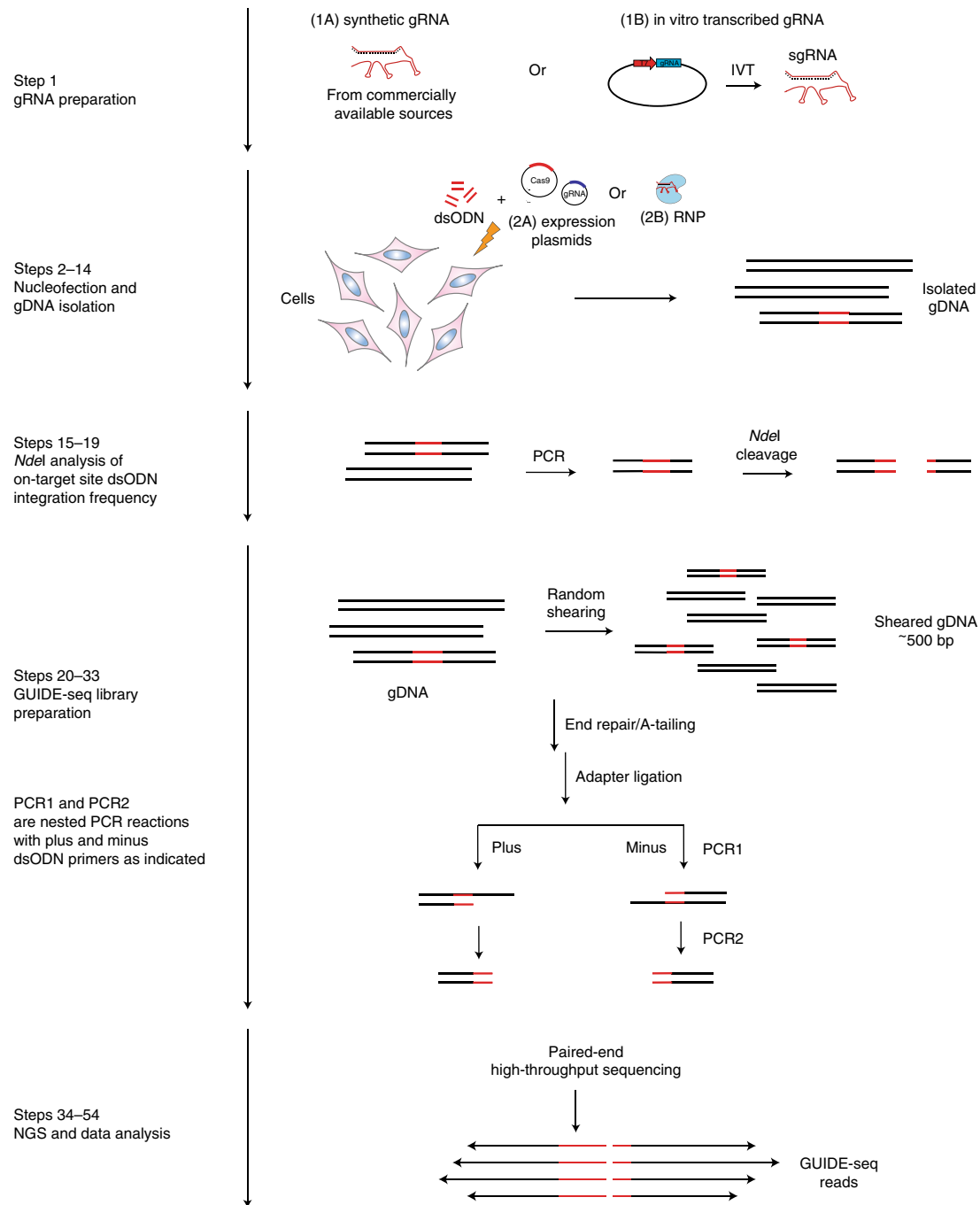


Fig. 1 | Schematic workflow of GUIDE-seq method. The GUIDE-seq procedure is divided into five sections with the protocol step numbers as shown on the left: (1) In vitro transcription (IVT) and RNP complexation; (2) nucleofection, cell culturing and gDNA isolation; (3) *NdeI* restriction and analysis of dsODN integration frequency; (4) GUIDE-seq library preparation; and (5) NGS and data analysis.

standard targeted sequencing (~0.1%). When more sensitive validation methods are used, it is clear that the complementary biochemical methods can detect some bona fide sites of off-target activity that are not picked up by GUIDE-seq. These types of methods are recommended when the highest sensitivity is required or in cell types that are sensitive to being transfected with dsODN tags.

Advantages

GUIDE-seq is a highly sensitive cell-based method that allows detection of cellular DSBs that is sequencing-efficient (requiring ~2–5 million reads per sample), unbiased and quantitative.

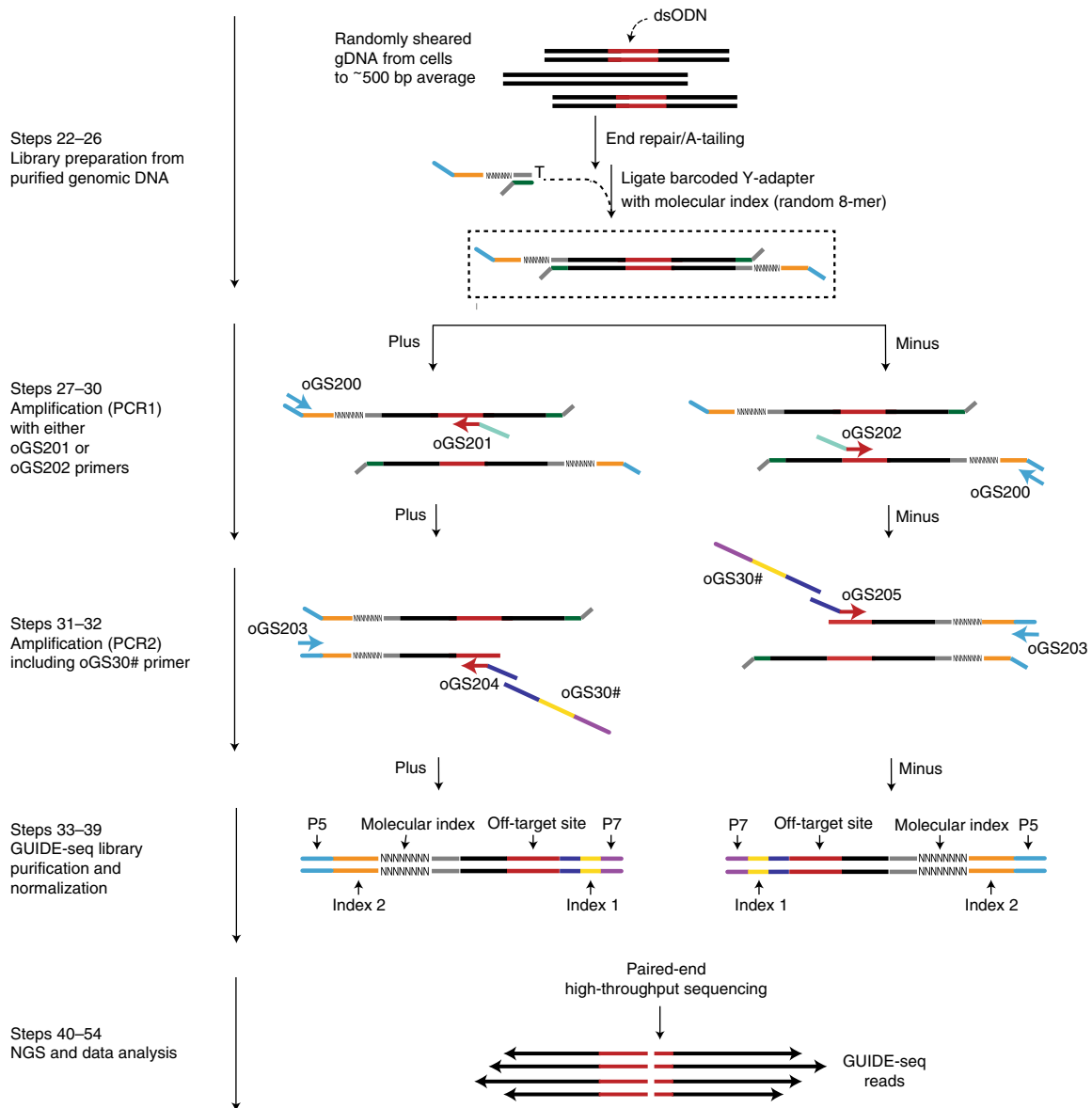


Fig. 2 | Principle of GUIDE-seq library preparation and paired-end analysis. (i) gDNA is sonicated to average size of 500 bp, followed by standard library generation steps (end-repair, A-tailing and adaptor ligation). (ii) PCR1 is carried out with a primer complementary to the ligated adaptor (oGS200) and either a sense (forward; oGS201) or an antisense (reversed; oGS202) primer complementary to dsODN sequence. dsODN sequence is red. Primers' features: i5 index/Index 2 (orange), i7 index/index 1 (yellow), Illumina adapters for sequencing P5 (blue) and P7 (purple), sequence complementary to the common adaptor oGS100 from Table 1 (dark green). The sense and antisense primers are used in separate PCR reactions (Plus and Minus). (iii) Nested PCR2 includes an additional primer containing the i7 index and complementary to the dsODN primer at its 3' end (oGS30#). oGS203 primer is complementary to the P5 adaptor but extends further on the 3' end in comparison with oGS200 used in PCR1; oGS204 and oGS205 are forward and reverse primers (used in Plus and Minus reactions, respectively) complementary to dsODN and have 3' region complementary to oGS30#. i5 index is encoded in amplified part of the adaptor upstream of UMI. Each sample has a unique combination of i5 and i7 indices, allowing for parallel sequence and analysis of numerous samples. (iv) Library normalization is required for balanced representation of samples and opposite strands for each sample (Step 39). (v) Library normalization and analysis; the reads are grouped by unique combinations of the two indices; thus, i5 and i7 indices are used to identify unique samples within a pool. The bioinformatic pipeline aligns sequences from the sample to the genome and identifies the juxtaposition of reads on opposite strands. Bidirectionally mapped loci indicate DSB locations.

GUIDE-seq can detect off-target sites with associated mutation frequencies of 0.1% and below⁴. GUIDE-seq can be easily run on benchtop sequencers such as Illumina MiSeq or MiniSeq systems. As an unbiased method, GUIDE-seq does not make a priori assumptions about sequence characteristics of potential off-target sites. All sites that fit the criteria of bidirectional mapping GUIDE-seq reads are reported, and the subset of sites with a user-specified number of mismatches to the intended target site are further labeled as high-priority off-target sites. GUIDE-seq read counts are strongly correlated

Table 1 | Sequences of dsODN-complementary oligonucleotides used in nucleofection (Step 2)

Oligo (name)	Sequence 5' to 3'	Purpose
oGS1 (dsODN FWD)	/5Phos/GTTTAATTGAGTTGTCATATGTTAATAACGGT*A*T	dsODN annealing; Reagent setup
oGS2 (dsODN REV)	/5Phos/ATACCGTTATTAACATATGACAACCTCAATTAA*A*C	dsODN annealing; Reagent setup

/5Phos/ indicates 5' phosphorylation. An asterisk indicates phosphorothioate linkage.

with indel mutation frequencies in cells edited in the absence of dsODN tag and are, therefore, a good proxy for genome-wide off-target activity⁸. Because it is a direct cell-based method, GUIDE-seq has a higher validation rate compared with *in vitro* methods such as CHANGE-seq (based on the principle of selective sequencing of nuclease-cleaved gDNA), with the trade-off that GUIDE-seq is less sensitive than CHANGE-seq, particularly for detecting low-frequency off-target activity.

Limitations

The primary limitation of GUIDE-seq is that it requires transfection with dsODN tags that may not be tolerated by all cell types. Some cell types such as human hematopoietic stem cells or induced pluripotent stem cells may have a robust DNA damage response and undergo apoptosis in response to high levels of free DNA ends. The use of 'surrogate' cell types that are more compatible for GUIDE-seq analysis may have limitations because this would not account for cell-type specific effects. If GUIDE-seq cannot be specifically optimized for a cell type of interest, it may be reasonable to try a sensitive biochemical method for defining genome-wide activity such as CHANGE-seq⁸. Finally, although possible, *in vivo* delivery of dsODNs for GUIDE-seq analysis has not yet been reported in the published literature.

Extended applications

GUIDE-seq or similar methods can also be used to characterize nucleases beyond broadly adopted Cas9, such as meganucleases¹⁸, zinc finger nucleases¹⁹, transcription activator-like effector nucleases and CRISPR-Cas nucleases that generate overhangs, such as Cas12a²⁰. Interestingly, we found that blunt end-protected dsODN tags integrate relatively efficiently into both blunt DSBs and those with 5'-overhangs. For enzymes leaving 3' overhangs such as meganucleases, others have found that using an end-protected dsODN with randomized 3' overhangs may be helpful¹⁸.

GUIDE-seq can be utilized to validate candidate sites that were detected *in silico* or by *in vitro* assays such as CHANGE-seq, Digenome-seq and DISCOVER-seq in cell types such as human primary T cells that are tolerant to transfection with GUIDE-seq dsODN tags⁸.

Variations of GUIDE-seq that utilize Tn5 tagmentation for library preparation or a longer dsODN tag have also been described recently^{16,21}. A modified version of GUIDE-seq using the Ion Torrent for sequencing instead of the Illumina platform has been used to discover off-target sites in a mouse cell line and was found to have good predictive power for unintentionally modified sites in mice embryo³. While we describe the standard GUIDE-seq approach that has been extensively validated here, we anticipate that we and others will continue to improve and adapt the method for future needs.

Experimental design

Primers and oligonucleotides

DNA oligonucleotides from Table 1 are typically synthesized at 250 nmol scale; oligonucleotides from Table 2 and Table 3 are synthesized at 100 nmol. All oligos in Tables 1, 2 and 3 are universal and can be used for different experiments with various targets. Table 1 lists two complementary oligonucleotides (oGS1 and oGS2) to be annealed to generate the standard GUIDE-seq dsODN, and Table 2 lists all nucleotides with barcodes used for ligation of Y-adapters (Step 24) prepared as described in 'Reagent setup'. Table 3 lists primers for two rounds of PCR reactions (Step 28 and 31). Where indicated, the oligos should be synthesized with degenerate nucleotides as written; these degenerate nucleotides form part of a UMI for identifying and consolidating PCR duplicates.

Table 2 | Sequences of Y-adapter nucleotides. oGS100 and oGS101-116 are Y-adapter oligonucleotides annealed as described in 'Reagent setup' and used in ligation step (Step 24)

Oligo (name)	Sequence 5' to 3'	Purpose
oGS100 (Common)	/5Phos/GATCGGAAGAGC*C*A	To generate Y adapter; Reagent setup
oGS101 (A01)	AATGATACGGCGACCACCGAGATCTACAC <u>TAGATCGC</u> NNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	Single-tailed unique barcode oligo Y adapter; Reagent setup
oGS102 (A02)	AATGATACGGCGACCACCGAGATCTACAC <u>CTCTCTAT</u> NNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	Single-tailed unique barcode oligo Y adapter; Reagent setup
oGS103 (A03)	AATGATACGGCGACCACCGAGATCTACAC <u>TATCCTCT</u> NNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	Single-tailed unique barcode oligo Y adapter; Reagent setup
oGS104 (A04)	AATGATACGGCGACCACCGAGATCTACAC <u>AGAGTAGA</u> NNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	Single-tailed unique barcode oligo Y adapter; Reagent setup
oGS105 (A05)	AATGATACGGCGACCACCGAGATCTACAC <u>GTAAGGAG</u> NNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	Single-tailed unique barcode oligo Y adapter; Reagent setup
oGS106 (A06)	AATGATACGGCGACCACCGAGATCTACAC <u>ACTGCATA</u> NNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	Single-tailed unique barcode oligo Y adapter; Reagent setup
oGS107 (A07)	AATGATACGGCGACCACCGAGATCTACAC <u>AAGGAGTA</u> NNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	Single-tailed unique barcode oligo Y adapter; Reagent setup
oGS108 (A08)	AATGATACGGCGACCACCGAGATCTACAC <u>CTAAGCCT</u> NNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	Single-tailed unique barcode oligo Y adapter; Reagent setup
oGS109 (A09)	AATGATACGGCGACCACCGAGATCTACAC <u>GACATTGT</u> NNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	Single-tailed unique barcode oligo Y adapter; Reagent setup
oGS110 (A10)	AATGATACGGCGACCACCGAGATCTACAC <u>ACTGATGG</u> NNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	Single-tailed unique barcode oligo Y adapter; Reagent setup
oGS111 (A11)	AATGATACGGCGACCACCGAGATCTACAC <u>GTACCTAG</u> NNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	Single-tailed unique barcode oligo Y adapter; Reagent setup
oGS112 (A12)	AATGATACGGCGACCACCGAGATCTACAC <u>CAGAGCTA</u> NNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	Single-tailed unique barcode oligo Y adapter
oGS113 (A13)	AATGATACGGCGACCACCGAGATCTACAC <u>CATAGTGA</u> NNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	Single-tailed unique barcode oligo Y adapter; Reagent setup
oGS114 (A14)	AATGATACGGCGACCACCGAGATCTACAC <u>TACCTAGT</u> NNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	Single-tailed unique barcode oligo Y adapter; Reagent setup
oGS115 (A15)	AATGATACGGCGACCACCGAGATCTACAC <u>CGCGATAT</u> NNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	Single-tailed unique barcode oligo Y adapter; Reagent setup
oGS116 (A16)	AATGATACGGCGACCACCGAGATCTACAC <u>TGGATTGT</u> NNWNNWNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T	Single-tailed unique barcode oligo Y adapter; Reagent setup

Adapters contain 8-mer UMIs NNWNNWNN (where N is any nucleotide; W is either A or T) and i5 index (highlighted in bold and underline). An asterisk indicates phosphorothioate linkage.

Optimization of dsODN incorporation

Here we describe GUIDE-seq protocols for analysis of genome-wide genome-editing activity in U2OS and human primary T cells. To successfully optimize GUIDE-seq for a new cell type, the critical step is to optimize the integration frequencies of dsODN tag when delivered alongside Cas9 and gRNA. This optimization is performed by a simple titration of the amount of transfected dsODN to simultaneously maximize integration frequencies while preserving cell viability and recovery. We found 100 pmol of dsODN per 20 μ l nucleofection to be a good starting point for optimization. On average, dsODN integrations should represent one-third or more of all nuclease-induced mutations. dsODN incorporation rarely exceeds 50% as it competes with indel mutations generated by nonhomologous end-joining or microhomology-mediated end-joining repair of DSBs. Empirically, we defined the minimum integration sufficient for performing GUIDE-seq as 5% (ref. ⁴). For example, if we observe 30% indel mutations at a particular target, we would anticipate a GUIDE-seq dsODN integration rate of 10%, which would exceed our absolute integration rate guideline of >5% for successful GUIDE-seq experiments. Failed GUIDE-seq experiments are most often associated with low tag integration frequencies. We provide an *NdeI* restriction enzyme analysis protocol to evaluate dsODN integration rate. Any suitable cell viability assay can be used to monitor dsODN toxicity, and routine Trypan Blue staining can provide sufficiently accurate estimates.

Table 3 | Primer sequences for two sequential PCR reactions and custom primers for Illumina NGS sequencing

Oligo (name)	Sequence 5' to 3'	Purpose
oGS200 (P5_PCR1)	AATGATACGGCGACCACCGAGATCTA	PCR1 common primer Step 28
oGS201 (P_PCR1)	GGATCTCGACGCTCTCCCTATACCGTTATTAACATATGACA	Plus PCR1 Step 28
oGS202 (M_PCR1)	GGATCTCGACGCTCTCCCTGTTTAATTGAGTTGCATATGTTAATAAC	Minus PCR1 Step 28
oGS203 (P5_PCR2)	AATGATACGGCGACCACCGAGATCTACAC	PCR2 common primer Step 31
oGS204 (P_PCR2)	CCTCTCTATGGGCAGTCGGTGATACATATGACAACCTCAATTAAC	Plus PCR2 Step 31
oGS205 (M_PCR2)	CCTCTCTATGGGCAGTCGGTGATTTGAGTTGCATATGTTAATAACGGTA	Minus PCR2 Step 31
oGS301 (P701_PCR2)	CAAGCAGAAGACGGCATAACGAGAT TCGCCTTA GTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	i7 index PCR2 Step 31
oGS302 (P702_PCR2)	CAAGCAGAAGACGGCATAACGAGAT CTAGTACG GTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	i7 index PCR2 Step 31
oGS303 (P703_PCR2)	CAAGCAGAAGACGGCATAACGAGAT TTCTGCCT GTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	i7 index PCR2 Step 31
oGS304 (P704_PCR2)	CAAGCAGAAGACGGCATAACGAGAT GCTCAGGA GTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	i7 index PCR2 Step 31
oGS305 (P705_PCR2)	CAAGCAGAAGACGGCATAACGAGAT AGGAGTCC GTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	i7 index PCR2 Step 31
oGS306 (P706_PCR2)	CAAGCAGAAGACGGCATAACGAGAT CATGCCTA GTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	i7 index PCR2 Step 31
oGS307 (P707_PCR2)	CAAGCAGAAGACGGCATAACGAGAT GTAGAGAG GTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	i7 index PCR2 Step 31
oGS308 (P708_PCR2)	CAAGCAGAAGACGGCATAACGAGAT CCTCTCTG GTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	i7 index PCR2 Step 31
oGS309 (Index1_SEQ)	ATCACCGACTGCCCATAGAGAGGACTCCAGTCAC	Index 1 sequencing primer Step 47
oGS310 (Read2_SEQ)	GTGACTGGAGTCCTCTCTATGGGCAGTCGGTGA	Read 2 sequencing primer Step 47

oGS200 and oGS203 are common primers for both direction-specific reactions in PCR1 and PCR2, respectively (Step 28 and Step 31). oGS201, oGS202 and oGS204, oGS205 are direction-specific primers for PCR1 and PCR2, respectively. PCR2 contains additional primers: oGS301–oGS308. These primers contain i7 Illumina index highlighted in bold. One primer with a unique i7 index is added to each sample in PCR2 as they anneal to the region common for both oGS204 and oGS205. oGS309 and oGS310 are sequencing primers used in Step 47.

Designing PCR primers for *NdeI* restriction analysis

For quick validation and optimization of dsODN tag integration, the GUIDE-seq dsODN tag contains an *NdeI* site, allowing for tag detection by *NdeI* digestion of PCR amplicons. It is important to design PCR primers for amplifying genomic sequence surrounding the gRNA target by applying the following guidelines:

(1) the amplicon should be 300–500 bp long without containing additional *NdeI* sites other than that in the dsODN sequence, and (2) the Cas9 cut site should be located asymmetrically so that the two cleaved products can be clearly resolved by electrophoresis on the TapeStation (Step 18). Indel rate and dsODN integration can also be evaluated by targeted amplicon sequencing analysis. Primers for targeted amplicon sequencing can be designed by following Illumina guidelines. An alternative for estimating the dsODN incorporation rate for Cas9 may be the ICE tool for homology-directed repair efficiency (<https://ice.synthego.com>), although we have not tested this ourselves.

Delivery of editing components and dsODN

We have been most successful in delivering Cas9 and gRNA as plasmids or RNP alongside dsODN using electroporation by Lonza Nucleofection. Others have also reported successful dsODN delivery into HEK293 cells using Mirus TransIT-X2 transfection reagent²². Many cationic lipids are not capable of delivering small DNA molecules to cells, so they will need to be tested on a cell-type specific basis. Therefore, we generally recommend use of electroporation-based methods to deliver editing components and dsODN as a starting point for optimizing GUIDE-seq in new cell types.

Transfection of ~600,000–900,000 cells will usually ensure that sufficient gDNA can be isolated to run the GUIDE-seq protocol. This can be achieved with one larger transfection or by pooling multiple smaller transfections together.

Required controls

To exclude random DSBs not associated with Cas9 activity, control gDNA from cells transfected with dsODN only is required. Ideally, GUIDE-seq requires a dsODN-only transfected control for each cell

type and independent set of samples. This control is required to utilize the standard GUIDE-seq analytical pipeline.

Assignment of sample barcodes

Each GUIDE-seq sample should be assigned a unique combination of i5 and i7 indexing barcodes each added during adapter ligation (Step 24) and PCR2 (Step 31), respectively. In addition to the Illumina barcodes, there are built-in UMIs in the Y-adaptor oligonucleotide, downstream of the i5 indexes. The 8-mer UMI has sequence NNWNNWNN (where N is any nucleotide and W is either A or T, as shown in Table 2). This allows for calculation of relative frequencies of the detected editing events in the bioinformatic pipeline (Step 53).

Library quantification and sequencing

Completed GUIDE-seq libraries should be quantified by qPCR, prior to pooling and sequencing. We empirically established that six serial 1:10 dilutions cover the range of concentrations seen across different samples for subsequent pooling (Step 39).

Sequencing depth

GUIDE-seq runs should be sequenced to a depth of at least 2–5 million reads per sample, achieving on-target read counts above 500 to maximize sensitivity.

Overage

For GUIDE-seq transfections and for all enzymatic reactions, include 10–20% overage to account for pipetting losses.

Materials

Biological materials

- U2OS cell line RRID [CVCL_0042](https://web.expasy.org/cellosaurus/CVCL_0042) (https://web.expasy.org/cellosaurus/CVCL_0042) (ATCC cat. no. HTB-96) **! CAUTION** Cell lines used in your research should be regularly checked to ensure they are authentic and are not infected with mycoplasma.
- Human primary CD4⁺/CD8⁺ T cells (purified from peripheral white blood cells (Stem Express, cat. no. LE010F, or similar)
- Human serum (Fisher Scientific, cat. no. MT35060CI)

Reagents

gRNA preparation

- RNase ZAP solution (Fisher Scientific, cat. no. AM9780)
- Plasmid pCRL1 for T7 in vitro transcription of gRNA (Addgene, cat. no. 153997)
- *Hind*III-HF restriction enzyme (New England Biolabs, cat. no. R3104S)
- HiScribe T7 RNA Synthesis Kit (New England Biolabs, cat. no. E2050S)
- MEGAclear Transcription Clean-Up Kit (Fisher Scientific, cat. no. AM1908)
- RNA High Sensitivity Screen Tape (Agilent, cat. no. 5067-5579)
- RNA High Sensitivity Sample Buffer (Agilent, cat. no. 5067-5580)

Mammalian cell culture and nucleofection

- Cas9 expression construct pSQT817 (Addgene, cat. no. 53373)
- gRNA expression construct, backbone BPK1520 (Addgene, cat. no. 65777)
- SE nucleofection kit (Lonza, cat. no. V4SC-1096)
- Advanced Dulbecco's Modified Eagle Medium (DMEM) (Fisher Scientific, cat. no. 12491023)
- PBS (Fisher Scientific, cat. no. 10010049)
- Trypan Blue (Fisher Scientific, cat. no. 15250061)
- FBS (Fisher Scientific, cat. no. 10-082-147)
- Trypsin-EDTA 0.05% wt/vol (Fisher Scientific, cat. no. 25300120)
- Penicillin–streptomycin (Fisher Scientific, cat. no. 15-070-063)
- GlutaMax 100× (Fisher Scientific, cat. no. 35-050-061)
- X-vivo 15 culture medium (Lonza, BEBP02-054Q)
- P3 nucleofection kit (Lonza, V4SP-3096) **▲ CRITICAL** Protocol for T-cell transfection has been optimized with this specific reagent.

- IL-15 (Miltenyi Biotec, cat. no. 170-076-114)
- IL-7 (Miltenyi Biotec, cat. no. 170-076-111)
- SpCas9 NLS recombinant protein (New England Biolabs, cat. no. M0646)

gDNA isolation

- DNAdvance purification kit (Beckman Coulter, cat. no. A48705), including Lysis LBH, Proteinase K, Pre-Bind BBA, Bind BBE and Elution EBA
- DTT solution 1 M (Sigma, cat. no. 43816)
- Qubit dsDNA BR Kit (Fisher Scientific, cat. no. Q32850) **▲ CRITICAL** This assay can distinguish dsDNA versus ssDNA and is essential for accurate dsDNA quantification.

DNA shearing

- 1× Tris-EDTA (TE) buffer (Fisher Scientific, cat. no. AM9848)

GUIDE-seq library preparation and NGS

- Sodium chloride 5 M (VWR, cat. no. 97062-858)
- Tris-HCl 1 M pH 8.0 (Fisher Scientific, cat. no. AM9855G)
- EDTA 0.5 M pH 8.0 (Fisher Scientific, cat. no. PAV4233)
- UltraPure DNase/RNase-free distilled water (Fisher Scientific, cat. no. 10977015)
- Ethanol (Sigma, cat. no. E7023) **! CAUTION** Flammable; keep away from open flame.
- *NdeI* restriction enzyme (New England Biolabs, cat. no. R0111S)
- AMPure XP PCR purification kit (Beckman Coulter, cat. no. A63881)
- End-repair mix (Fisher Scientific, cat. no. 50305934)
- T4 DNA ligase (Fisher Scientific, cat. no. 50305904)
- 25 mM dNTP solution mix (Fisher Scientific, cat. no. 50305950)
- Taq polymerase (non-hot start) (Fisher Scientific, cat. no. 10342-053) **▲ CRITICAL** Non-hot start Taq is critical for the procedure as it does not require heat activation that will inactivate the other enzymes in the reaction (Steps 22–23).
- 10× PCR Rxn Buffer (-MgCl₂) (Fisher Scientific), supplied with Taq polymerase
- Platinum Taq polymerase (Fisher Scientific, cat. no. 10966-026), including 10× PCR Rxn Buffer (-MgCl₂) and 50 mM MgCl₂
- Tetramethylammonium chloride (TMAC) 5 M (Sigma, cat. no. T3411-500ML) **! CAUTION** Toxic if swallowed or in contact with skin.
- DNA High Sensitivity D1000 Screen Tape (Agilent, cat. no. 5067-5582)
- DNA High Sensitivity D1000 Ladder (Agilent, cat. no. 5067-5586)
- DNA High Sensitivity D1000 Sample Buffer (Agilent, cat. no. 5067-56020)
- KAPA Library Quantification Kit (Illumina) Universal qPCR Mix (Kapa, cat. no. KK4824)
- Phusion Hot Start Flex 2× Mix (New England Biolabs, cat. no. M0536L)
- MiSeq reagent kit v3 600 cycles (Illumina, cat. no. MS102-3003)
- Phix control V3 kit (Illumina, cat. no. FC-110-3001)
- Tween-20 (Sigma, cat. no. P7949)
- Custom oligonucleotides (Integrated DNA Technologies)

Equipment

- 24-well cell culture plate, flat bottom with lid, tissue culture treated, non-pyrogenic, polystyrene (Corning Life Sciences, cat. no. 3524)
- 24-well non-tissue culture treated plate, flat bottom with low evaporation lid (Corning Life Sciences, cat. no. 351147)
- 96-well non-tissue culture treated plate, flat bottom with low evaporation lid (Corning Life Sciences, cat. no. 351172)
- 96-well clear V-bottom 2 ml polypropylene deep well plate (Corning Life Sciences, cat. no. 3960)
- Disposable hemocytometer (Fisher Scientific, cat. no. 22600100)
- Steriflip sterile disposable vacuum filter units (Fisher Scientific, cat. no. SCGP00525)
- T cell TransAct beads (Miltenyi Biotec, cat. no. 130-019-011)
- Qubit assay tubes (Fisher Scientific, cat. no. Q32856)
- microTUBE AFA pre-slit snap-cap 6 × 16 mm (Covaris, cat. no. 520045)

- Adhesive PCR plate seals (Fisher Scientific, cat. no. AB0558)
- Optical tube 8× strip (Agilent, cat. no. 401428)
- Optical cap 8× strip (Agilent, cat. no. 401425)
- Hard-shell 96-well PCR plates, thin wall (Bio-Rad, cat. no. HSP9601)
- Microseal adhesive qPCR film (Bio-Rad, cat. no. MSB1001)
- Universal magnetic plate (Alpaqua, cat. no. A00400)
- Qubit Fluorometer 4.0 (Fisher Scientific, cat. no. q33226)
- Acoustic focused ultrasonicator (Covaris, cat. no. E220)
- Nucleofector 4-D (Lonza, cat. no. AAF-1002X)
- Veriti 96-well thermocycler or equivalent (Fisher Scientific, cat. no. 4375786)
- CFX96 real-time PCR thermocycler or equivalent (Bio-Rad, cat. no. 1855195)
- Incubating microplate shaker (Fisher Scientific, cat. no. 02217759)
- Eppendorf 5424 micro-centrifuge or equivalent (Eppendorf, cat. no. 022620401)
- Eppendorf 5920 centrifuge or equivalent (Eppendorf, cat. no. 2231010155)
- 4200 TapeStation Instrument and accessories (Agilent, cat. no. G2991AA)
- Microtiter shaker IKA MS3 (Fisher Scientific, cat. no. 1476321)
- MiSeq (Illumina, cat. no. SY-410-1003)
- Standard inverted microscope
- Tissue culture biosafety level 2 cabinet
- Cell culture CO₂ incubator

Software

- Python v.2.7 or 3: <https://www.python.org/downloads/>
- Burrows–Wheeler Aligner (BWA): <https://github.com/lh3/bwa>
- Bedtools: <http://bedtools.readthedocs.io/en/latest/content/installation.html>
- GUIDE-seq: <https://github.com/tsailabSJ/guideseq>

Reagent setup

Ethanol, 70%

Prepare fresh 70% (vol/vol) ethanol by diluting 200-proof ethanol with nuclease-free water for purifications. **!CAUTION** Ethanol is highly flammable.

Bind BBE

Reagent is from DNAdvance purification kit. Equilibrate at room temperature (RT, 24 °C), and vortex vigorously immediately before use. When not in use, store at 4 °C. Note that expiration date is printed on the label.

Culture medium for U2OS cells

Supplement Advanced DMEM with FBS to 10% (vol/vol) final concentration, antibiotics (penicillin–streptomycin) and GlutaMAX, each from 100× stocks to 1× final dilutions (i.e., 5 ml per 500 ml of DMEM). Store at 4 °C for up to 3 months.

Culture medium for T cells

Supplement X-vivo 15 medium with HSA to 5% (vol/vol) (or 20% (vol/vol) where indicated), filter sterilize and add IL-15 with IL-7 to the final concentration of 10 ng/ml each. Prepare fresh every time.

10× STE solution

500 mM NaCl, 100 mM Tris-HCl (pH 8.0) and 10 mM EDTA. To prepare 10 ml of 10× STE stock solution, combine 1 ml of 5 M NaCl, 1 ml of 1 M Tris HCl (pH 8.0) and 0.2 ml of 0.5 M EDTA, and bring the total volume to 10 ml with molecular biology grade water. Store at RT for 1 year or longer.

GUIDE-seq oligonucleotides

Resuspend the lyophilized GUIDE-seq oligos *oGS1* and *oGS2* (Table 1) in TE buffer at a concentration of 250 μM each. Mix 40 μl of each oligo, 10 μl water and 10 μl of 10× STE buffer, and anneal with the following thermocycler program:

Step	Temperature	Time
Denaturation	95 °C	5 min
70 cycles	−0.1 °C /cycle	30 s
End	4 °C	Hold

Annealed oligonucleotides can be stored at −20 °C for up to 3 months in aliquots of 20 µl.

Library Y-adapter

Anneal the MiSeq common adapter *oGS100* with one of the *oGS101* to *oGS116* barcoded oligos to generate Y-adapters required for each sample (Table 2). Resuspend lyophilized oligos in TE buffer to a concentration of 100 µM. Combine 10 µl of each oligo, and bring the total volume to 100 µl with final buffer composition 1× STE. Anneal using thermocycler program as above for annealing GUIDE-seq oligos. Annealed oligonucleotides can be aliquoted into 10 µl aliquots and stored at −20 °C for up to 3 months.

KAPA qPCR kit

Prepare for use by adding the primers supplied in the kit to 2× master mix according to manufacturer’s instructions. The reagent can be stored at −20 °C for several months.

Procedure

Preparation of gRNA for RNP complexation ● Timing 1 week

1 Obtain gRNA for use in Step 2B either as synthetic chemically modified gRNA²³ (option A) or by in vitro transcription reaction with T7 driven template such as targets cloned into pCRL1 or equivalent (option B).

(A) Synthetic gRNA

(i) Synthetic gRNA is commercially available from several sources (e.g., IDT, Synthego or Trilink).

(B) In vitro transcribed gRNA

(i) As template for gRNA synthesis, clone gRNA target sequence into pCRL1 plasmid as previously described²³ and verify by sequencing.

▲ CRITICAL STEP Use RNase-free grade water and reagents. Clean pipettors with RNase ZAP spray.

(ii) To linearize the template plasmid using a unique enzyme immediately downstream of gRNA scaffold sequence, set up a reaction to digest up to 5 µg of plasmid using the *HindIII*-HF restriction enzyme, as follows:

Component	Volume (µl)	Final concentration
CutSmart 10× buffer	5	1×
gRNA plasmid (3–5 µg)	Variable	Up to 0.1 µg/µl
<i>HindIII</i> -HF (20 U/µl)	4	1.6 U/µl
Nuclease-free H ₂ O	to 50	
Total	50	

(iii) Incubate at 37 °C for 3 h, and heat inactivate at 80 °C for 20 min.

(iv) Purify the DNA with AMPure XP beads following the manufacturer’s protocol. Briefly, add 90 µl bead suspension (1.8× volume of sample), and mix well by pipetting up and down ten times. Incubate at RT for 5 min, and separate beads on magnetic plate for 5 min. Aspirate supernatant without disturbing the pellet, and discard. Add 200 µl of 70% ethanol, incubate for 30 s and then aspirate the solution and discard. Repeat 70% ethanol wash, and air dry the pellet for 3 min. Remove the plate from the magnetic rack, and resuspend the beads in 15 µl TE buffer by pipetting up and down ten times. Incubate at RT for 2 min. Move the plate back to the magnetic rack, allow beads to settle for 1 min and carefully transfer supernatant to a new plate.

▲ CRITICAL STEP Maintain an RNase-free environment; use RNase-free water and plastic consumables. Pipette accurately to avoid cross-contamination from bead purification. Do not over-dry beads to prevent loss of yield and samples cross-contamination.

- (v) Set up the *in vitro* transcription reaction with the HiScribe T7 RNA Synthesis Kit as follows:

Component	Volume (μl)	Final concentration
<i>Hind</i> III digested linearized plasmid DNA (1 μg)	Variable	0.05 μg/μl
NTP Buffer Mix	10	10 mM each NTP
T7 RNA polymerase	2	
Nuclease-free H ₂ O	to 20	
Total	20	

- (vi) Incubate the reaction at 37 °C for 16 h.
- (vii) Add 2 μl of DNase I (2 U/μl).
- (viii) Incubate at 37 °C for an additional 15 min.
- (ix) Purify the *in vitro* synthesized gRNA with MEGAclear purification kit according to the manufacturer's instructions.
- (x) Elute with preheated (95 °C) 50 μl elution solution, and collect gRNA by centrifuging for 1 min at 15,000g, RT. To increase the yield, repeat elution with the flow through from first elution.
- (xi) Dilute an aliquot of *in vitro* transcribed gRNA sample 10× with nuclease-free water.
- ▲ CRITICAL STEP** Dilution is required for reproducible measurements in the next step.
- (xii) Quantify the RNA using a NanoDrop spectrophotometer. Choose RNA mode to use correct extinction coefficient. The RNA sample should have 260/280 ratio of ~2. Undiluted sample should be at least 2 μg/μl to satisfy the limits for the samples volume used in Step 2B(ix).
- (xiii) Check gRNA integrity on Agilent TapeStation using RNA High Sensitivity Screen Tape and Buffer according to the manufacturer's instructions. There should be only one band close to 100 bp size; an abundance of smaller products indicates RNA degradation.
- PAUSE POINT** Make 5 μl aliquots, and freeze at –80 °C. gRNA aliquots can be stored at –80 °C for several months.

Culture, nucleofect, and harvest cells ● Timing 7 d

- 2 Here we provide two alternative protocols as examples of the approaches and cells that may be used: nucleofection of U2OS cells with SpCas9 and gRNA expression plasmids (option A) and nucleofection of primary CD4⁺/CD8⁺ T cells with reconstituted Cas9:gRNA RNP (option B), as well as a guideline for a optimizing the procedure with a new cell line (option C).

(A) U2OS cells

▲ CRITICAL Culture U2OS cells under standard conditions (37 °C, 5% vol/vol CO₂) in complete Advanced DMEM (see 'Reagent setup').

- (i) Plate cells at 40% confluency in 100 mm dishes 1 d before nucleofection.

▲ CRITICAL STEP Trypsinizing and replating cells 24 h before nucleofection is critical; it should not be omitted as doing otherwise will result in suboptimal growth and DNA recovery. Culturing cells for a longer time before nucleofection may result in lower cell viability and editing efficiency after transfection.

- (ii) Prior to transfection, pre-equilibrate a 24-well recovery plate(s) with 500 μl of complete DMEM medium in each well by incubating in 37 °C, 5% vol/vol CO₂ incubator.
- (iii) To collect cells for nucleofection, wash cells with PBS, add up to 2 ml per 100 mm dish of 0.05% wt/vol trypsin to fully cover the cell's monolayer, and incubate for 5 min in 37 °C, 5% vol/vol CO₂ incubator.
- (iv) Resuspend cells by adding the same volume of medium used for culturing to the plate. Collect the resuspended cells into a new tube, and count with hemocytometer using Trypan Blue. Cell viability should be >90%.
- (v) Collect cells needed for nucleofections (2 × 10⁵ cells per nucleofection) by centrifugation for 1 min at 1,200g, RT, and discard the supernatant.

- (vi) For each nucleofection, resuspend 2×10^5 cells in 20 μl of SE buffer with 100 pmol of dsODN (i.e., 1 μl of annealed dsODN, see 'Reagent setup'). We recommend making a single pool of cells resuspended in SE buffer with dsODN for all nucleofections. For instance, for ten samples, resuspend 2×10^6 cells in 200 μl of SE buffer and add 10 μl of annealed dsODN.
- (vii) In a 96-well plate, combine 20 μl of cell suspension with dsODN from Step 2A(vi), 500 ng of Cas9 expression plasmid SQT817 and 250 ng gRNA expression plasmid. Control samples should include cell suspension with dsODN only.
 - ▲ CRITICAL STEP** Volume of DNA added should be 2 μl or less for optimal nucleofection efficiency. Control samples should be included in each experiment to monitor random dsODN incorporation as well as cytotoxicity of nucleofection procedure.
- (viii) From the plate, mix and transfer 20 μl of the mixture to the Lonza 4-D cuvette.
- (ix) Nucleofect the samples using DN-100 pulse code on a Lonza 4-D nucleofector following the manufacturer's instructions.
- (x) Incubate at RT for 10 min.
 - ▲ CRITICAL STEP** Do not omit RT incubation as transfection efficiency can drop substantially.
- (xi) Add 80 μl medium from the pre-incubated 24-well plate (prepared in Step 2A(ii)) to the nucleofection cuvette, and gently resuspend cells two to three times by pipetting.
- (xii) Transfer the contents of the nucleofection cuvette back to the plate from Step 2A(xi), and return the plate to the CO₂ incubator for 72 h before harvesting.
- (xiii) After 72 h, harvest the cells by centrifugation for 1 min at 1,200g, RT, discard the medium and store the pellet at $-20\text{ }^\circ\text{C}$ until gDNA isolation.

(B) CD4⁺/CD8⁺ T cells

▲ CRITICAL Thaw, then activate and culture CD4⁺/CD8⁺ T cells from frozen stock before nucleofection by following the steps below. We plate cells in non-TC-treated 24-well plates.

- (i) Thaw cells and pipette into 50 ml tube. Slowly add 10 ml of X-Vivo 15 medium prewarmed to 37 $^\circ\text{C}$ without supplements. The number of cells needed for this step should be determined by calculating the following:
 - Total number of cells = (number of GUIDE-seq samples) \times (3 nucleofections per GUIDE-seq sample) \times (3×10^5 cells per nucleofection).
- (ii) Centrifuge for 10 min at 300g at RT, discard the supernatant, and resuspend cells at 1×10^6 cells/ml in T-cell medium prepared as described in 'Reagent setup'.
- (iii) Activate T cells with TransAct polymeric nanomatrix by adding the TransAct reagent directly to the cell culture at the specific dilution ratio specified by the manufacturer (1:100), and culture in a non-TC-treated 24-well plate for 72 h at 37 $^\circ\text{C}$, 5% vol/vol CO₂.
- (iv) Prior to nucleofection, pre-equilibrate a non-TC 96-well recovery plate(s) with 250 μl per well of filter-sterilized complete X-vivo 15 medium with 20% vol/vol HSA and 10 ng/ml of IL-7 and IL-15 in 37 $^\circ\text{C}$, 5% vol/vol CO₂ incubator.
- (v) For each GUIDE-seq sample, set up three nucleofections for pooling to ensure that sufficient tag-integrated cells are available for the analysis. For each nucleofection, use 75 pmol of the Cas9 protein and threefold molar excess of gRNA (from Step 1) in a final total volume of 5 μl . Prepare RNP by combining the Cas9 protein and gRNA accounting for the required molar amount of each nucleofection, pooling and overage. Incubate at RT for 20 min.
- (vi) Pool all the activated T cells from Step 2B(iii) after 72 h, and count the cells with hemocytometer. Cell viability needs to be >90%, cell count $1\text{--}2 \times 10^6$ cells/ml.
- (vii) Determine the total number of cells needed for the nucleofections as follows: number of nucleofections $\times 3 \times 10^5$. Collect cells by centrifugation 300g for 5 min at RT. For each nucleofection, resuspend 3×10^5 cells (based on the cell count from Step B(vi)) in 20 μl of P3 buffer with 100 pmol of dsODN (i.e., 1 μl of annealed dsODN, see 'Reagent setup').
 - ▲ CRITICAL STEP** We recommend making a single pool of cells resuspended in P3 buffer with dsODN for all nucleofections. For example, for ten nucleofections, resuspend 3×10^6 cells in 200 μl of P3 buffer and add 10 μl of annealed dsODN.
- (viii) In a non-TC 96-well plate, for each nucleofection combine 20 μl of the cell suspension and dsODN from Step 2B(vii) with 5 μl of the RNP complex from Step 2B(v). Control samples should include cell suspension with dsODN only.
- (ix) From the plate in the previous step, for each nucleofection mix by pipetting and transfer 20 μl of the mixture to one well of a 16-well Lonza 4-D strip. Nucleofect the

samples using EO-115 pulse code on a Lonza 4-D nucleofector following the manufacturer's instructions.

- (x) Incubate at RT for 7 min.
- (xi) Add 100 μ l of medium from the plate prepared in Step 2B(iv) to each nucleofected well.
- (xii) Transfer 120 μ l the contents of the nucleofection cuvette back to the plate from Step 2B(xi), and culture in 37 °C, 5% (vol/vol) CO₂ incubator for 72 h before harvesting.
- (xiii) After 72 h, harvest the cells by first pooling three nucleofections for each sample. Then centrifuge the pooled cells for 10 min at 300g, RT, discard the medium and store the pellet at -20 °C until gDNA isolation (Steps 3–14).

(C) Optimize procedure for new cell type

- (i) Screen for an optimal nucleofection program for a new cell type by repeating the Procedure above (Step 2A(i–xiii)) with an expression plasmid of a fluorescent protein and varying the nucleofection programs in Step 2A(ix). Lonza nucleofection kits are supplied with control EGFP plasmid ready for nucleofection. Identify the program with both good cell viability and the highest mean fluorescent intensity by flow cytometry.
- (ii) Further optimize the efficiency of editing in a new cell type using a well-characterized target by repeating Step 2A(i–xiii), varying the number of cells per nucleofection and the amount of RNP. Targets used in the original GUIDE-seq report can be used as controls⁴.
- (iii) Using the optimal editing conditions determined in the previous step, titrate the amount of GUIDE-seq dsODN used per nucleofection. For small-scale nucleofections in 20 μ l, 5–200 pmol is a good starting range.
- (iv) Analyze the rate of tag integrations compared with overall indel mutation frequencies (Steps 15–19), and choose a GUIDE-seq dsODN amount that results in integration rates >5% with integrations optimally representing one-third of overall indel mutations observed.

▲ CRITICAL STEP For highest nucleofection efficiency, the volume of added RNP should be <5 μ l.

gDNA isolation ● Timing 3 h

- 3 Three days after nucleofection, isolate gDNA with DNAdvance kit according to manufacturer's instructions:

Prepare lysis mixture for required number of wells as follows:

Component	Volume (μ l)
Lysis LBH	188
1 M DTT	5
Proteinase K	7
Total	200

- 4 Aspirate medium from the plate, and directly add 200 μ l of lysis mixture to wells of the 24-well plate.
- 5 Seal the plate with PCR film, and incubate at 55 °C for 2–16 h, with gentle 250 rpm shaking.
- 6 Transfer all recovered volume of the lysate to Costar 96 deep-well block.
- 7 Add 100 μ l Pre-Bind PBBA, and pipette ten times to mix.
- 8 Add 170 μ l Bind BBE (see 'Reagent setup'), pipette 15 times to mix and incubate at RT for 5 min.
- 9 Incubate on universal magnetic plate for 5 min at RT to separate the magnetic beads. Aspirate the supernatant without disturbing the beads.
- 10 Remove the 96-well block from the magnet. Add 340 μ l of freshly made 70% vol/vol ethanol to each well. Resuspend beads by pipetting 20 times or until fully suspended.
- 11 Return block to the magnetic plate for 1 min or until beads settle, and discard the supernatant.
- 12 Repeat the ethanol wash (Steps 10–11) two more times for a total of three ethanol washes.
- 13 Elute by resuspending pellet in 200 μ l Elution EBA and pipette ten times, then separate beads on magnetic plate for 5 min. Collect 190 μ l of the supernatant.

? TROUBLESHOOTING

- 14 Measure DNA concentration with Qubit Broad Range kit on Fluorometer 3.0 according to manufacturer's instructions. An acceptable concentration to be used is 10–50 ng/ μ l.

▲ **CRITICAL STEP** Use a fluorescent quantification method like Qubit to avoid overestimating double-stranded DNA concentration.

■ **PAUSE POINT** At this point, samples can be stored at $-20\text{ }^{\circ}\text{C}$ for several weeks.

? **TROUBLESHOOTING**

Estimating dsODN integration frequency ● Timing 4 h

- 15 Assemble PCR reactions for each experimental sample (gDNA from nuclease and dsODN-treated cells from Step 13) and corresponding control sample (gDNA from cells transfected only with dsODN) as follows:

Please refer to the experimental design section of the introduction for general considerations for primer design:

Component	Volume (μl)	Final concentration
2 \times Phusion Mix	25	1 \times
Forward primer (10 μM)	2.5	0.5 μM
Reverse primer (10 μM)	2.5	0.5 μM
gDNA (100 ng) from Step 13	Variable	2 ng/ μl
Nuclease-free water	to 50	
Total	50	

- 16 Run standard thermocycler program (make adjustments as needed for primer-specific conditions):

Step	Temperature	Time
Denaturation	98 $^{\circ}\text{C}$	5 min
35 cycles	98 $^{\circ}\text{C}$	15 s
	68 $^{\circ}\text{C}$	20 s
	72 $^{\circ}\text{C}$	30 s
	72 $^{\circ}\text{C}$	5 min
Extension	72 $^{\circ}\text{C}$	5 min
End	4 $^{\circ}\text{C}$	Hold

- 17 Purify PCR product using 1.8 \times volume (90 μl) of AMPure XP beads according to the manufacturer’s instructions as described in Step 1B(iv). Elute in 30 μl of TE buffer. Measure DNA concentration with nanodrop.

- 18 Analyze tag integration frequency by performing either an *NdeI* assay (option A) or by targeted sequencing (option B).

(A) ***NdeI* assay**

- (i) Digest 200 ng of PCR product in a volume of 20 μl with 10 units *NdeI* enzyme at 37 $^{\circ}\text{C}$ for 1 h as follows.

Component	Volume (μl)	Final concentration
10 \times CutSmart buffer	2.0	1 \times
<i>NdeI</i> 20 U/ μl	1.0	1 U/ μl
PCR product (200 ng) from Step 17	Variable	10 ng/ μl
Nuclease-free water	to 20	
Total	20	

- (ii) Purify *NdeI*-treated DNA using 1.8 \times volume (36 μl) of AMPure XP beads according to manufacturer’s instructions as described in Step 1B(iv). Elute in 20 μl of TE buffer.

- (iii) Dilute each sample with TE buffer to a final volume of 5 μl . The samples need to be diluted to fall within the concentration of the Agilent High Sensitivity D1000 ScreenTape. Mix 2 μl of the High Sensitivity D1000 Ladder or diluted sample with 2 μl of the sample buffer, and analyze on Agilent TapeStation using D1000 High Sensitivity Screen Tape. Determine fraction of DNA cleaved in each sample, using software provided with the instrument. An example is shown in Fig. 3.

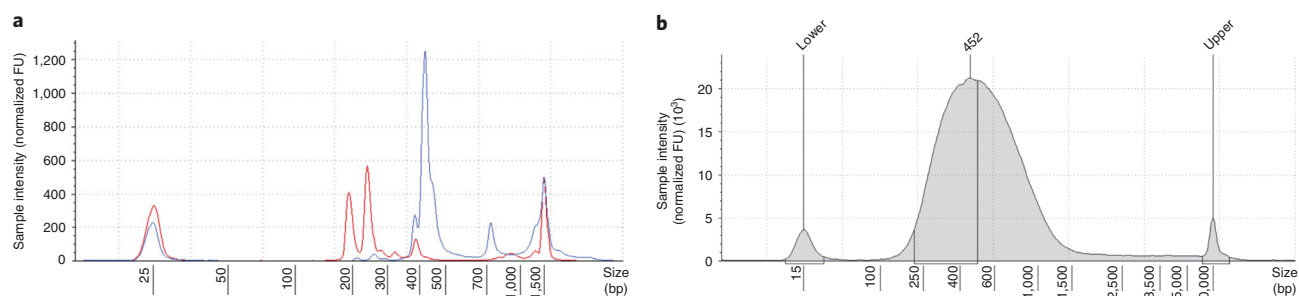


Fig. 3 | Quality control steps in the GUIDE-seq protocol. a, *NdeI* analysis of dsODN incorporation on Agilent TapeStation. Profile of uncut PCR amplicon (blue) is superimposed with profile of the same PCR product digested with *NdeI* (red). Two flanking peaks mark resolution thresholds: low (25 bp) and high (1,500 bp). Positions of the molecular size marker are indicated at the bottom. The uncut (blue peak size 400) converted to two smaller peaks around 200 bp marker (red). The targeted PCR is encompassing *B2M* site 6 as described by Lazzarotto et al.⁸ **b**, Agilent TapeStation analysis of DNA shearing distribution that shows a peak with median size centered around ~450 bp. Upper and lower refers to the upper and lower range of the ladder.

(B) Targeted sequencing analysis

- (i) Design primers for targeted sequencing following Illumina guidelines.
 - (ii) Proceed with dual-barcoded two-step PCR Amplicon sequencing following Illumina protocols for multiplex amplicon sequencing.
 - (iii) Analyze sequence data using publicly available tools such as CRISPResso2 (ref. ²⁴): <https://github.com/pinellolab/CRISPResso2>.
- 19 Check that the percentage of tag integration is >5% and represents >30% of detectable mutagenesis frequencies.

▲ **CRITICAL STEP** GUIDE-seq assay sensitivity is dependent on efficient tag integration.

? **TROUBLESHOOTING**

GUIDE-seq library preparation ● Timing 8 h

- 20 Load 400 ng of quantified gDNA from Step 14 into Covaris DNAmicroTUBE tube supplementing with TE buffer to total volume of 130 μ l, and shear gDNA to average size of 500 bp using the following Covaris E220 settings: peak incident power 105 W, duty factor 5%, cycles per burst 200, treatment time 80 s. Manufacturer's recommended settings for shearing should be followed if using a different model instrument. An example of sheared DNA is shown in Fig. 3.
- ▲ **CRITICAL STEP** Shear in 130 μ l volume to prevent formation of air bubbles that may interfere with the process.
- ▲ **CRITICAL STEP** Note the Covaris E220 shearing instrument can only handle 96 samples at a time (in contrast to other library preparation steps where samples can be processed in parallel with multichannel pipettors or automated liquid handling); thus, time should be allotted on a per-sample basis for this step.
- 21 Purify sheared DNA using 1 \times volume (130 μ l) of AMPure XP beads according to manufacturer's instructions as described in Step 1B(iv). Elute in 15 μ l of TE buffer.
- 22 Assemble single-tube end-repair/A-tailing reaction as follows:

Component	Volume (μ l)
dNTP mix (5 mM each)	1
10 \times ligation buffer	2.5
End-repair mix	2
PCR Rxn Buffer (-MgCl ₂) (10 \times)	2
Taq polymerase (non-hot start)	0.5
Nuclease-free water	0.5
DNA sample from Step 21	14
Total	22.5

23 Perform thermocycler program for end-repair/A-tailing as follows:

Step	Temperature	Time
1	12 °C	15 min
2	37 °C	15 min
3	72 °C	15 min

24 Add 1 µl of annealed Y-adapter (oGS100 with one of oGS101-116 (providing sample-specific barcodes from 01 to 16); see Table 2 and ‘Reagent setup’), and 2 µl of T4 DNA ligase to the reaction mix from the previous step as follows:

Component	Volume (µl)
Reaction from Step 22	22.5
Y-adapter	1
T4 ligase	2
Total volume	25.5

▲ CRITICAL STEP Each test or control sample should have a unique i5/i7 sample index combination. The i5 sample index is added in this step by selecting one of annealed adapters (oGS101 through 116) that contains the index sequence highlighted in bold in Table 2. The i7 sample index is added in Step 31.

25 Perform ligation in thermocycler as follows:

Step	Temperature	Time
1	16 °C	30 min
2	22 °C	30 min
End	4 °C	Hold

26 Purify ligated product using 0.9× volumes (23 µl) of AMPure XP beads according to the manufacturer’s instructions as described in Step 1B(iv). Elute in 22 µl of TE buffer.

27 Split adapter ligated product for two PCR1 reactions, 10 µl each; label one as Plus, the other Minus.

▲ CRITICAL STEP From this point forward, each sample is processed with parallel Plus (P) and Minus (M) reactions. The products from Plus PCR1 are used for Plus PCR2, and Minus PCR1 for Minus PCR2. It is critical to keep track of the P and M reactions throughout two subsequent PCR reactions and to keep them separate.

28 Assemble Plus and Minus PCR1 reactions for each sample, as follows. The only difference is the sequence of tag-specific primers (shown in bold):

Plus PCR1		Minus PCR1	
Component	Volume (µl)	Component	Volume (µl)
Nuclease-free water	11.9	Nuclease-free water	11.9
Platinum Taq PCR Rxn Buffer (-MgCl ₂) (10×)	3	Platinum Taq PCR Rxn Buffer (-MgCl ₂) (10×)	3
dNTP mix (10 mM)	0.6	dNTP mix (10 mM)	0.6
MgCl ₂ (50 mM)	1.2	MgCl ₂ (50 mM)	1.2
oGS201 primer (10 µM)	1	oGS202 primer (10 µM)	1
TMAC (500 mM)	1.5	TMAC (500 mM)	1.5
oGS200 primer (10 µM)	0.5	oGS200 primer (10 µM)	0.5
Platinum Taq polymerase	0.3	Platinum Taq polymerase	0.3
Template DNA (from Step 27)	10	Template DNA (from Step 27)	10
Total	30	Total	30

- 29 Run the PCR1 thermocycler program as follows (70 °C Δ –1.0 °C/cycle), for 15 cycles; decrease the temperature by 1 °C per cycle, starting with 70 °C:

Step	Temperature	Time
Denaturation	95 °C	5 min
15 cycles:	95 °C	30 s
	70 °C, Δ –1.0 °C/cycle	2 min
	72 °C	30 s
10 cycles:	95 °C	30 s
	55 °C	1 min
	72 °C	30 s
Extension	72 °C	5 min
End	4 °C	Hold

- 30 Keep Plus and Minus PCR1 reactions separate. Purify DNA from each reaction using 1.2 \times volumes (36 μ l) of AMPure XP beads according to manufacturer's instructions as described in Step 1B(iv). Elute in 15 μ l TE buffer, and keep the beads in the sample.
- ▲ CRITICAL STEP** For higher yield and library complexity, keep the magnetic beads in the sample for the next step.
- 31 Assemble PCR2 reactions using PCR1 products as template. The Plus and Minus reactions remain separated. The same oGS30# primer (where # is 1 through 8; see Table 3) is added to Plus and Minus reactions. oGS203 primer is universal for both; the critical difference is the tag-specific primers, i.e., oGS204 for Plus and oGS205 for Minus reactions (highlighted in bold):

Plus PCR2		Minus PCR2	
Component	Volume (μ l)	Component	Volume (μ l)
Nuclease-free water	5.4	Nuclease-free water	5.4
PCR Rxn Buffer (-MgCl ₂) (10 \times)	3	PCR Rxn Buffer (-MgCl ₂) (10 \times)	3
dNTP mix (10 mM)	0.6	dNTP mix (10 mM)	0.6
MgCl ₂ (50 mM)	1.2	MgCl ₂ (50 mM)	1.2
oGS204 (10 μM)	1	oGS205 (10 μM)	1
TMAC (500 mM)	1.5	TMAC (500 mM)	1.5
oGS203 (10 μ M)	0.5	oGS203 (10 μ M)	0.5
oGS30# (10 μ M)	1.5	oGS30# (10 μ M)	1.5
Platinum Taq polymerase	0.3	Platinum Taq polymerase	0.3
Template DNA from Step 30	15	Template DNA from Step 30	15
Total	30	Total	30

- ▲ CRITICAL STEP** Each sample should have a unique combination of i5/i7 indexes, where both P and M reactions of each sample have identical i5/i7 combination. The i5 index, encoded in the Y-adaptor, was added earlier at Step 24. i7 index is added in this step, when the same i7 primer (one of oGS301-308) is added to both P and M reactions. Thus, similar to PCR1, the only difference between the P and M reactions for each sample is presence of either P_PCR2 (oGS204) or M_PCR2 (oGS205) primers:
- 32 Run the PCR program:

Step	Temperature	Time
Denaturation	95 °C	5 min
15 cycles:	95 °C	30 s
	70 °C, Δ –1.0 °C/cycle	2 min
	72 °C	30 s

Table continued

(continued)

Step	Temperature	Time
10 cycles:	95 °C	30 s
	55 °C	1 min
	72 °C	30 s
Extension	72 °C	5 min
End	4 °C	Hold

- 33 Purify product of the PCR2 reaction using 0.7× volumes (21 µl) of AMPure XP beads according to the manufacturer’s instructions as described in Step 1B(iv). Elute in 30 µl of TE buffer.
▲ CRITICAL STEP Keep all PCR products separate; do not pool together until after quantification.
■ PAUSE POINT Samples can be stored at –20 °C at this point.

Library quantification and high-throughput sequencing ● Timing 24 h

- 34 Prepare tenfold serial dilutions of the libraries from 10⁻¹ to 10⁻⁶ with TE buffer. Typically, the 10⁻⁵ dilution range is optimal for quantification by qPCR. It is convenient to prepare 50 µl dilutions to leave sufficient material for sample pooling and sequencing. The samples, no template control, and standards are commonly used in triplicate for each qPCR run. The standards are provided with the qPCR kit.
■ PAUSE POINT Samples can be stored at –20 °C at this point up to 1 year.
- 35 Assemble SYBR qPCR reactions for the diluted Plus and Minus reaction of each sample from Step 34. Normally, 10⁻⁴ and 10⁻⁵ dilutions are appropriate for qPCR analysis. Use the standards provided with the kit, and have ‘no template control’ with water as well. Run all reactions in triplicate. Set up reactions in the hard-shell 96-well plate with reconstituted 2× KAPA master mix (see ‘Reagent setup’) as follows:

Component	Volume (µl)
Water	4
Diluted template from Step 34 /standard/water	4
KAPA qPCR MM with primer (2×)	12
Total	20

- 36 In the plate layout, mark standards and indicate corresponding concentrations (expressed as number of molecules per microliter).
- 37 Run qPCR program as follows:

Step	Temperature	Time
Initial denaturation	95 °C	5 min
	95 °C	30 s
35 cycles	60 °C	45 s
	Plate read	
Melt curve	65–95 °C Δ0.5 °C	5 s

- 38 qPCR data will be expressed both as a table with concentration as defined in Step 36 as well as a standard curve plot with marked individual samples (y axis is Cq, x axis is logarithmic scale of standards’ quantity). When standards are expressed as number of molecules per reaction, the unknown samples concentration will be automatically calculated in the same units. On the plot, choose a sample dilution that falls well within the middle of standards’ range. Use the table below for Kapa standards to generate data in molecules per microliter:

Standard	Concentration (pM)	dsDNA molecules/µL
1	20	12,000,000
2	2	1,200,000

Table continued

(continued)		
Standard	Concentration (pM)	dsDNA molecules/ μ L
3	0.2	120,000
4	0.02	12,000
5	0.002	1,200
6	0.0002	120

39 Based on calculated concentrations of the undiluted Plus and Minus reactions, choose an appropriate dilution for pooling. For each sample, pool Plus and Minus reactions from Step 34 to achieve an equal number of molecules of the Plus and Minus reactions to a total of 4.5×10^9 molecules in the volume of 5 μ L, resulting in a final library concentration of 1.5 nM, required for MiSeq sequencing. Adjustment for average library fragment size is not typically required.

? TROUBLESHOOTING

- 40 One hour before sequencing, bring the reagent cartridge to RT in a water bath to thaw all the buffers and mix well.
- 41 Prepare 4 nM PhiX control using 10 mM Tris HCl 0.1% Tween-20 (vol/vol) as diluent: 2 μ L of 10 nM PhiX stock plus 3 μ L of 10mM Tris HCl 0.1% Tween.
- 42 Denature 5 μ L of DNA library from Step 39 by adding 5 μ L of 0.2 N NaOH (prepared fresh), vortex, collect by brief centrifugation, and incubate for 5 min at RT. In a separate tube, denature 5 μ L of 4 nM PhiX prepared in Step 41.
- 43 Add 990 μ L of HT1 buffer from the kit to each of the two tubes containing denatured library and PhiX control. The concentration of diluted library is now 7.5 pM, and diluted PhiX control is now 20 pM.
- 44 Further dilute PhiX with HT1 buffer to 12.5 pM by adding 225 μ L HT1 buffer to 375 μ L of 20 pM PhiX.
- 45 Add 50 μ L of the diluted PhiX control (12.5 pM PhiX) to 1 ml of diluted library (7.5 pM) from Step 43.
- 46 In the sequencing cartridge, use separate fresh P1000 pipette tips to pierce the foil sealing wells 13, 14 and 17.
- 47 Add 3 μ L of 100 μ M Index1 (oGS309 sequencing primer from Table 3) to well 13; 3 μ L of 100 μ M Read2 (oGS310 sequencing primer from Table 3) to well 14. Mix each well carefully using 1 ml serological pipette. Add 600 μ L of the library from Step 45 to well 17.
- ▲ CRITICAL STEP** Double-check the custom primers added; if omitted, the sequencing will still be performed but without the correct reads.
- 48 For NGS analysis, GUIDE-seq specific options need to be defined in the Sample Sheet: add blank row, fill in sample ID with '1', two adapters with N701 and N501. Save and open as text. Enter AAAAAAAAAA after N501 sequence and save. See example sample sheet (Supplementary Data).
- ▲ CRITICAL STEP** In the sample sheet, do not check the boxes to use custom sequencing primers. Custom primers used in this protocol are spiked into the standard wells to enable reading of the PhiX control along with library.
- 49 Load and sequence the library, using a MiSeq Reagent Kit v3 and MiSeq system according to the manufacturer's instructions or alternative Illumina sequencer of choice.

Post-run analysis ● Timing 2 h

- 50 *Install GUIDE-seq software pipeline*¹⁰. The GUIDE-seq code is compatible with both Python 2 and Python 3. It requires BWA, bedtools and several other Python packages (see <https://github.com/tsalabSJ/guideseq#dependencies> for details). Thus, the recommended way to install GUIDE-seq pipeline is via conda, as follows:
- ```
conda create -n guideseq -c conda-forge -c bioconda -c anaconda -c tsalabSJ guide_seq
source activate guideseq
guideseq.py -h
```
- 51 *Set up reference genome*. Download the appropriate reference genome in FASTA format. For example, human genome reference Hg19 can be obtained at [http://www.broadinstitute.org/ftp/pub/seq/references/Homo\\_sapiens\\_assembly19.fasta](http://www.broadinstitute.org/ftp/pub/seq/references/Homo_sapiens_assembly19.fasta). Make an index of the reference genome once it is downloaded, for example:
- ```
bwa index Homo_sapiens_assembly19.fasta
```


- 52 Generate a manifest file (.yaml) referencing the program dependencies and samples FASTQ files paths. A detailed example and description of this file is at https://github.com/tsailabSJ/guideseq#write_manifest

Briefly, create a text file with .yaml extension. In the file, follow the example in the link above, indicate the reference genome file, analysis folder and path of bwa and bedtools command. Next, indicate the path for undemultiplexed fastq files and the barcode and target sequences for each experiment.

▲ CRITICAL STEP Read 2 FASTQ files must be reverse complemented if sequenced on a NextSeq or MiniSeq before running the analysis pipeline using a tool such as seqtk (<https://github.com/lh3/seqtk>).

- 53 Run the analysis with the command line:

```
guideseq.py all -m manifest_name.yaml
```

- 54 To check results, follow the description on GitHub: https://github.com/tsailabSJ/guideseq#pipeline_output?

? TROUBLESHOOTING

Troubleshooting

Troubleshooting advice can be found in Table 4.

Table 4 | Troubleshooting table

Step	Problem	Possible reason	Possible solution
13	Beads clumping during gDNA purification. Beads present in eluted sample	Cold reagents or excessive amount of DNA	Place beads on magnet while pipetting for next step. Use beads suspension equilibrated at 24 °C Consider using less starting material
14	Amount of DNA recovered is lower than the 400 ng of DNA required for the assay	High rate of cell death after nucleofection	U2OS cells: trypsinize and plate cells 1 d before transfection to maximize numbers of dividing cells and minimize toxicity CD4/CD8 T cells: recover with 20% HSA (vol/vol) and pool cells from two or more technical replicates if necessary
19	dsODN incorporation rate is <5%	Low efficiency of integration Low activity of Cas9	Optimize amount of dsODN used in the procedure. Titrate between 5 and 100 pmol for 200,000 cells Optimize editing activity using a control gRNA with validated high efficiency of Cas9 cutting
39	Volume is too large to pool library in 5 µl	Low yield of the library	Use undiluted library components, and reduce the volume by concentrating in speed-vac
54	High-frequency unmatched off-targets observed	Low-level cross-contamination of gRNA or gRNA-expressing plasmid	Ensure that gRNA and gRNA-expressing plasmid are purified carefully to minimize cross-contamination

Timing

- Step 1, gRNA synthesis: 1 week
- Step 2, cell culture and nucleofection: 1 week
- Steps 3–14, gDNA isolation: 4–16 h
- Steps 15–19, estimate dsODN integration frequency: 4 h
- Steps 20–33, GUIDE-seq library preparation: 1 d
- Steps 34–49, library quantification and sequencing: 2 d
- Steps 50–54, post-run analysis: 2 h

Note that physical random shearing of DNA for GUIDE-seq library preparation is performed sequentially on a Covaris E220 shearing instrument (Step 20) that can handle 96 samples at a time. Time should be allotted on a per-sample basis for this step, in contrast to other library preparation steps where samples can be processed in parallel with multichannel pipettors or automated liquid handling.

Anticipated results

Three days after nucleofection, U2OS cells should reach near confluency. Purified gDNA concentration is expected to be 6–10 µg. CD4⁺/CD8⁺ T cells tend to be more sensitive to dsODN toxicity than U2OS cells, which might lead to lower gDNA yield (see Troubleshooting (Table 4, Step 14)).

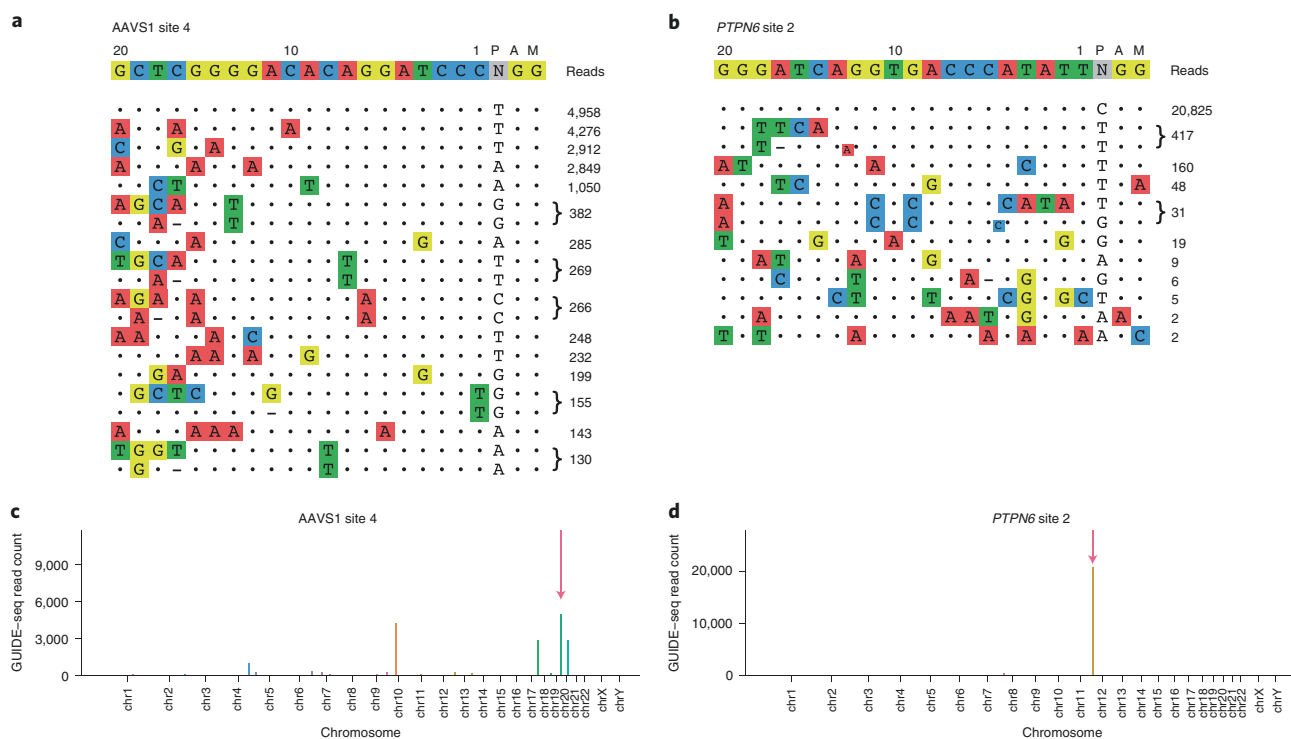


Fig. 4 | GUIDE-seq example output. **a,b**, GUIDE-seq on-target and off-target detailed visualizations are shown from previously published experiments⁸: low-specificity gRNA targeting AAVS1 site 4 (**a**); high-specificity gRNA targeting *PTPN6* site 2 (**b**). Each row represents an on-target or off-target site and GUIDE-seq read counts corresponding to that site are shown to the right. A dot represents matches with respect to the intended target sequence, and colored nucleotides indicate the sequence of mismatches. Visualizations are ranked by GUIDE-seq read counts. **c,d**, Manhattan plots for AAVS1 site 4 (**c**) and *PTPN6* (**d**) corresponding to the same GUIDE-seq data are shown; x axis represents chromosomal location, and y axis represents GUIDE-seq read counts. Red arrow shows on-target edit site.

Optimized dsODN integration frequency is expected to be >5% and at least one-third of the total indel mutagenesis rate by Cas9:gRNA as measured by *NdeI* restriction digest assay or targeted sequencing (Fig. 3a). Control DNA from unedited cells should always be analyzed and processed within each set of samples throughout the library preparation. In addition to dsODN incorporation by *NdeI* digest, another helpful quality control is estimate of DNA fragmentation size on TapeStation (Fig. 3b).

After high-throughput sequencing data are analyzed through GUIDE-seq pipeline, the results are stored in an output folder with separate subfolders for individual steps.

These two folders contain the main GUIDE-seq output.

- *identified* contains text files created with each site listed on a separate line; the list includes both on- and off-targets. It is most useful to examine the *identified* files, sorted by sequence and GUIDE-seq read count (bi.sum.mi)
- *visualization* contains final output product as an *.svg image file for each target depicting on- and off-target sites and number of reads for each. This is a vector image file with *.svg extension
- *filtered* contains text files that list DSB sites that are identified as background based on comparison with unedited control samples

The remaining folders contain intermediate reads and alignments that can be consulted for further analysis.

- *demultiplexed* contains four sequence reads (forward, reversed, index1, index2) for each sample
- *umitagged* contains sequences with UMI tags added to each read for consolidation of duplicate reads
- *consolidated* contains consolidated reads in FASTQ format
- *aligned* contains consolidated read sequences mapped to the genome

Examples of the visualization output are shown in Fig. 4 for two targets edited by Cas9 and gRNA as RNP with low (Fig. 4a) and high (Fig. 4b) specificity of editing, i.e., low and high off-target numbers and frequencies. The Manhattan plot below each visualization output represents number of reads for on and off targets as distributed across chromosomal locations. The number of

off-target sites detected may vary substantially for different gRNAs, and this can be used to pick highly specific gRNA.

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Sequencing data from experiments originally described in Lazzarotto et al.⁸ are available in NCBI Sequence Read Archive PRJNA625995.

Code availability

Open-source GUIDE-seq analysis code is available at <https://github.com/tsailabSJ/guideseq>.

References

1. Maeder, M. L. & Gersbach, C. A. Genome-editing technologies for gene and cell therapy. *Mol. Ther.* **24**, 430–446 (2016).
2. Wang, H., Russa, M. L. & Qi, L. S. CRISPR/Cas9 in genome editing and beyond. *Annu. Rev. Biochem.* **85**, 1–38 (2015).
3. Anderson, K. R. et al. CRISPR off-target analysis in genetically engineered rats and mice. *Nat. Methods* <https://doi.org/10.1038/s41592-018-0011-5> (2018).
4. Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
5. Xie, K. & Yang, Y. RNA-guided genome editing in plants using a CRISPR–Cas system. *Mol. Plant* **6**, 1975–1983 (2013).
6. Cheng, Y. & Tsai, S. Q. Illuminating the genome-wide activity of genome editors for safe and effective therapeutics. *Genome Biol.* **19**, 226 (2018).
7. Putney, S., Benkovic, S. & Schimmel, P. A DNA fragment with an alpha-phosphorothioate nucleotide at one end is asymmetrically blocked from digestion by exonuclease III and can be replicated in vivo. *Proc. Natl Acad. Sci.* **78**, 7350–7354 (1981).
8. Lazzarotto, C. R. et al. CHANGE-seq reveals genetic and epigenetic effects on CRISPR–Cas9 genome-wide activity. *Nat. Biotechnol.* **38**, 1317–1327 (2020).
9. Zheng, Z. et al. Anchored multiplex PCR for targeted next-generation sequencing. *Nat. Med.* **20**, 1479–1484 (2014).
10. Tsai, S. Q., Topkar, V. V., Joung, K. J. & Aryee, M. J. Open-source guideseq software for analysis of GUIDE-seq data. *Nat. Biotechnol.* **34**, 483–483 (2016).
11. Tsai, S. Q. et al. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat. Methods* **14**, 607–614 (2017).
12. Wienert, B. et al. Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science* **364**, 286–289 (2019).
13. Wang, X. et al. Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotechnol.* **33**, 175–178 (2015).
14. Kim, D. et al. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* **12**, 237–243 (2015).
15. Yan, W. X. et al. BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat. Commun.* **8**, 15058 (2017).
16. Schmid-Burgk, J. L. et al. Highly parallel profiling of Cas9 variant specificity. *Mol. Cell* **78**, 794–800.e8 (2020).
17. Crosetto, N. et al. Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat. Methods* **10**, 361–365 (2013).
18. Wang, L. et al. Meganuclease targeting of PCSK9 in macaque liver leads to stable reduction in serum cholesterol. *Nat. Biotechnol.* **36**, 717–725 (2018).
19. Miller, J. C. et al. Enhancing gene editing specificity by attenuating DNA cleavage kinetics. *Nat. Biotechnol.* **37**, 945–952 (2019).
20. Kleinstiver, B. et al. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat. Biotechnol.* **34**, 869.
21. Nobles, C. L. et al. iGUIDE: an improved pipeline for analyzing CRISPR cleavage specificity. *Genome Biol.* **20**, 14 (2019).
22. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* <https://doi.org/10.1126/science.aba8853> (2020).
23. Hendel, A. et al. Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat. Biotechnol.* **33**, 985–989 (2015).
24. Fiume, M. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).

Acknowledgements

This work was supported by St. Jude Children's Research Hospital and ALSAC, National Institutes of Health (NIH) Office Of The Director (OD) Somatic Cell Genome Editing (SCGE) initiative grant U01AI157189 (to S.Q.T.), St. Jude Children's Research Hospital Collaborative Research Consortium on Novel Gene Therapies for Sickle Cell Disease (SCD), the Doris Duke Charitable Foundation (2017093 and 2020154) (to S.Q.T.), an (NIH) Directors Pioneer Award (DP1 GM105378) (to J.K.J.), NIH R01 GM088040 (to J.K.J.), NIH R01 AR063070 (to J.K.J.), and the Jim and Ann Orr Massachusetts General Hospital (MGH) Research Scholar Award (to J.K.J.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Author contributions

N.L.M., C.R.L., G.L., Y.L. and S.Q.T. drafted the manuscript. S.Q.T. and J.K.J. conceived the GUIDE-seq method. S.Q.T. and Z.Z. developed the GUIDE-seq protocol in the laboratories of J.K.J. and A.I. M.L. and N.N. performed experiments during original development of GUIDE-seq under supervision of S.Q.T. and Z.Z. S.Q.T., V.T. and M.A. contributed to development of the GUIDE-seq analysis software.

Competing interests

S.Q.T. is a co-inventor on patents covering the GUIDE-seq method. S.Q.T. is a member of the scientific advisory board of Kromatid and Twelve Bio. M.J.A. and J.K.J. hold equity in SeQure Dx, Inc. J.K.J. holds equity in Chroma Medicine. J.K.J. has financial interests in Beam Therapeutics, Editas Medicine, Excelsior Genomics, Pairwise Plants, Poseida Therapeutics, Transposagen Biopharmaceuticals, and Verve Therapeutics (f/k/a Endcadia). J.K.J.'s interests were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies. J.K.J. is a co-inventor on various patents and patent applications that describe gene editing and epigenetic editing technologies, including the GUIDE-seq method.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41596-021-00626-x>.

Correspondence and requests for materials should be addressed to Shengdar Q. Tsai.

Peer review information *Nature Protocols* thanks Ciaran M. Lee and the other, anonymous reviewer(s) for their contribution to the peer review of this work.

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

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

Received: 31 October 2020; Accepted: 2 September 2021;

Published online: 12 November 2021

Related links

Key references using this protocol

Tsai, S. Q. et al. *Nat. Biotechnol.* **33**, 187 (2015): <https://doi.org/10.1038/nbt.3117>

Lazzarotto, C. R. et al. *Nat. Biotechnol.* **38**, 1317 (2020): <https://doi.org/10.1038/s41587-020-0555-7>

Kleinstiver P. B. et al. *Nat. Biotechnol.* **34** 869 (2016): <https://doi.org/10.1038/nbt.3620>

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection N/A not data were collected in the present manuscript

Data analysis GUIDE-seq analysis pipeline is freely available at GitHub, Code availability statement included in the manuscript

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequencing data from experiments originally described in Lazzarotto et al. are available in NCBI Sequence Read Archive PRJNA625995

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Not applicable The manuscript provides detailed protocol and does not deal with new data sets
Data exclusions	Not applicable The manuscript provides detailed protocol and does not deal with new data sets
Replication	Not applicable The manuscript provides detailed protocol and does not deal with new data sets
Randomization	Not applicable The manuscript provides detailed protocol and does not deal with new data sets
Blinding	Not applicable The manuscript provides detailed protocol and does not deal with new data sets

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	ATCC
Authentication	The cell line has not been authenticated after purchasing from ATCC
Mycoplasma contamination	Cell lines are routinely tested for mycoplasma contamination with MycoAlert Plus detection kit
Commonly misidentified lines (See ICLAC register)	N/A