Integrated computational guide design, execution, and analysis of arrayed and pooled CRISPR genome editing experiments

Matthew C. Canver[1,*], Maximilian Haeussler[2,*], Daniel E. Bauer[1], Stuart H. Orkin[1,3], Neville E. Sanjana[4], Ophir Shalem[5], Guo-Cheng Yuan[6], Feng Zhang[7,8], Jean-Paul Concordet[9], Luca Pinello[10,#]

[1]Division of Hematology/Oncology, Boston Children's Hospital, Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA

[2]Santa Cruz Genomics Institute, MS CBSE, University of California, Santa Cruz, California, USA

[3]Howard Hughes Medical Institute, Boston, Massachusetts, USA

[4]New York Genome Center and Department of Biology, New York University, New York City, New York, USA

[5]Raymond G. Perelman Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Department of Genetics, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[6]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

[7]The Broad Institute, Cambridge, Massachusetts, USA.

[8]McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

[9]INSERM U1154, CNRS UMR 7196, Muséum National d'Histoire Naturelle, Paris, France.

[10]Department of Molecular Pathology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA

*Co-first authors
[#]Correspondence to Luca Pinello, lpinello@mgh.harvard.edu

Keywords: CRISPR genome editing; Cas9/Cpf1; CRISPR pooled screening; Saturating mutagenesis; sgRNA design; Lentivirus; CRISPOR; CRISPResso; Docker; Locus-specific deep sequencing; Off-target effect; On-target effect; Amplicon sequencing; Targeted sequencing; Sequence analysis; Computational tools; Software pipeline

ABSTRACT

CRISPR genome editing experiments offer enormous potential for evaluation of genomic loci using arrayed or pooled lentiviral libraries of single guide RNAs (sgRNAs). Numerous computational tools are available to help design sgRNAs with optimal on-target efficiency and minimal off-target effects. In addition, a few computational tools have been developed to analyze data resulting from genome editing experiments. However, these tools are typically developed in isolation and oftentimes are not readily translatable into laboratory-based experiments. Here we present a protocol that describes in detail both the computational and the benchtop implementation of an arrayed and/or pooled CRISPR genome editing experiment. This protocol provides instructions for sgRNA design with CRISPOR, experimental implementation, and analysis of the resulting high-throughput sequencing data with CRISPResso.

INTRODUCTION

The clustered regularly interspaced short palindromic repeats (CRISPR) nuclease system is a facile and robust genome editing system[1,2]. The CRISPR nuclease system was identified as the driver of prokaryotic adaptive immunity to allow for resistance to bacteriophages[3]. This system has been subsequently repurposed for eukaryotic genome editing by heterologous expression of the CRISPR components in eukaryotic cells. Site-specific cleavage by CRISPR nucleases such as Cas9 or Cpf1 requires an RNA molecule to guide nucleases to specific genomic loci to initiate double strand breaks (DSBs)[1,2,4]. Site-specific cleavage requires Watson-Crick base pairing of the RNA molecule to a corresponding genomic sequence upstream of a protospacer adjacent motif (PAM)[1,2]. The required RNA molecule for genome editing experiments consists of a synthetic fusion of the prokaryotic tracrRNA and crRNA to create a single chimeric guide RNA (sgRNA).

CRISPR mutagenesis relies on engagement of endogenous DNA repair pathways after nuclease-mediated DSB induction has occurred. The principal repair pathways include non-homologous end joining (NHEJ) and homology-directed repair (HDR). NHEJ repair is an error-prone pathway, which results in a heterogeneous spectrum of insertions/deletions (indels) primarily in the range of 1-10 bp[1,2,5–7]. HDR relies on the co-delivery of an extrachromosomal template to be used as a template for DNA repair as opposed to a sister chromatid. This allows for the insertion of customized sequence into the genome[1,2].

Applications of the method

A variety of computational tools have been developed for the design and analysis of CRISPR-based experiments. However, these tools are typically developed in isolation without features for facile integration with one another and/or without sufficient consideration to facilitate implementation in a laboratory setting. Here, we offer a protocol to integrate robust, publicly available tools for the design, execution, and analysis of CRISPR genome editing experiments. Specifically, we have adapted CRISPOR[8] and CRISPResso[9] to be integrated with one another as well as streamlined for experimental implementation (Fig. 1).

CRISPR mutagenesis allows for the study of both coding and non-coding regions of the genome. This involves usage of single or multiple sgRNAs via pooled screening for individual or multi-loci study. Pooled screening allows for targeting of a handful of genes up to genome-scale gene targeting[10–12]. It also allows for saturating mutagenesis (tiling of sgRNA) experiments to identify functional sequence within non-coding regions[13–15]. This protocol can be used to design and execute arrayed or pooled genome editing experiments. Furthermore, it can also be used for the design, implementation, and analysis of pooled screens for gene-targeting, saturating mutagenesis of non-coding elements, or any other targeting strategy[13,14,16].

Comparison with other methods

Numerous computational tools are freely available to aide sgRNA design for a wide spectrum of PAM sequences as well as on-target efficiency and off-target cleavage predictions: Broad GPP Portal[17], Cas-database[18], Cas-OFFinder[19], CasOT[20], CCTop[21], COSMID[22], CHOPCHOP[23,24], CRISPRdirect[25], CRISPR-DO[26], CRISPR-ERA[27], CRISPR-P[28], CROP-IT[29], DNA Striker[14], E-CRISP[30], flyCRISPR[31], GuideScan[32], GT-scan[33], MIT CRISPR design tool[7], WU-CRISPR[34], CRISPRseek[35], sgRNAcas9[36], CRISPR multiTargeter[37], as well as others offered by companies such as Deskgen[38] and Benchling[38]. CRISPOR (http://crispor.org) is a computational tool that predicts off-target cleavage sites and offers a variety of on-target efficiency scoring systems to assist sgRNA selection for more than 120 genomes using different CRISPR nucleases (Table 1)[8]. CRISPOR offers several unique advantages for designing sgRNA for genome editing experiments. First, CRISPOR integrates multiple published on target sgRNA efficiency scores including from Fusi et al[39], Chari et al[40], Xu et al[41], Doench et al (Doench 2014 and 2016)[17,42], Wang et al[43], Moreno-Mateos et al[44], Housden et al[45], Prox. GC[46], -GG[47], and Out-of-Frame[48]. It also offers previously published off-target prediction[7]. Second, CRISPOR has been optimized to facilitate experimental implementation by providing automated primer design for both on-target and off-target deep sequencing analysis. The primers and output files are further designed to be compatible for subsequent analysis by CRISPResso after the experiments have been completed. Taken together, CRISPOR provides an sgRNA design methodology to facilitate experimental execution and downstream analysis.

CRISPResso is a computational pipeline that enables accurate quantification and visualization of CRISPR genome editing outcomes, as well as comprehensive evaluation of effects on coding sequences, non-coding elements, and off-target sites from individual loci, pooled, and whole genome deep-

sequencing data[9]. The CRISPResso suite involves multiple tools for analysis, including the CRISPResso webtool and command line version of CRISPResso. There are also multiple command line tools: CRISPRessoPooled, CRISPRessoWGS, CRISPRessoCount, CRISPRessoCompare, and CRISPRessoPooledWGSCompare. The applications and features of these tools are summarized in Table 2. Alternative computational tools exist to evaluate genome editing outcomes from deep sequencing data[49–51]; however, these tools offer limited analysis functionality for pooled sequencing or WGS data as compared to the CRISPResso suite. Furthermore, CRISPResso analysis offers many unique features such as splice-site analysis or frameshift analysis to quantitative the proportion of engendered mutations that result in a frameshift when targeting coding sequence.

CRISPR genome editing reagents have taken many forms, including DNA, RNA, protein, or various combinations of each[52,53]. Furthermore, delivery of these reagents has also been attempted using a variety of methods, including electroporation, lipid-based transfection, and viral-mediated delivery[52,53]. Pooled screening relies on the ability to deliver individual reagents to individual cells in batches. Electroporation and lipid-based transfection methods offer limited ability to control the number of reagents (i.e., sgRNA) delivered per cell; however, lentiviral transduction at low transduction rates (~30-50%) results in single lentiviral integrants per cell in the majority of cases[13]. Furthermore, lentivirus offers stable integration of the CRISPR reagents into each cell's genome. These features of lentivirus allow for pooled CRISPR experiments. Therefore, this protocol uses lentivirus for both arrayed and pooled CRISPR experiments.

Limitations of on-target and off-target prediction

As described above, there are many tools available for both on-target sgRNA efficiency and off-target cleavage prediction. Progress has been made towards enhancing the predictive value of these scores; however, while these predictions are useful to focus sgRNA selection for experimental design, experimental validation provides the definitive analysis of on-target and off-target mutagenesis. Continued investigation is necessary to more completely understand the rules governing sgRNA efficiency and off-target mutagenesis, such as more fully exploring the role of epigenetics in sgRNA targeting for example.

Target identification and nuclease choice

CRISPR genome editing experiments require appropriate target identification to fit experimental objectives, which can include gene- or non-coding targeting (e.g., enhancers, CTCF or other transcription factor binding sites). Each CRISPR nuclease offers a unique PAM sequence, which have varying frequencies within the genome[14]. Depending on the location of targeting such as exons, introns, promoters, DNase hypersensitive sites, enhancers, or repressed regions, PAM frequency can change[14]. The optimal nuclease can be chosen based on density of available PAMs such as for saturating mutagenesis or proximity of PAMs to a particular genomic position such a transcription factor binding motif. Therefore, choosing the optimal nuclease depends on the region(s) of interest to be analyzed. Finally, high fidelity nucleases can be used to minimize the probability of off-target mutagenesis[54,55].

sgRNA and PCR primer design for arrayed and pooled screen experiments using CRISPOR

Arrayed experiments involve the use of an sgRNA for editing of a single target. CRISPOR offers a variety of on- and off-target prediction scores that can aide in the optimal sgRNA selection (Box 1)[8]. Analysis of on-target sgRNA efficiency can be predicted based on available scores and/or investigated experimentally by analysis of editing frequency. In addition to the numerous sgRNA efficiency prediction scores, CRISPOR offers automated primer design to facilitate polymerase chain reaction (PCR) amplification of regions for deep sequencing analysis to quantitate editing frequency. This involves PCR amplification of sequences flanking the DSB site for a given sgRNA. Similarly, CRISPOR offers computational prediction of off-target sites as well as PCR primers for deep sequencing analysis of mutagenesis at these predicted sites.

CRISPOR can also be used for the design of gene-targeted pooled screens by inputting exonic regions for sgRNA design as well as saturating mutagenesis screens. Saturating mutagenesis involves utilizing all PAM-restricted sgRNA within a given region(s) in a pooled screening format to identify functional sequences[13–15]. Saturating mutagenesis can be used to analyze coding and non-coding elements in the genome or a combination of the two. Screen resolution is a function of PAM frequency and can be enhanced by PAM choice and/or combination of nucleases with unique PAM sequences[14]. CRISPOR can be utilized to design saturating mutagenesis libraries by simply selecting all sgRNA within

the inputted region(s). It is particularly important to consider off-target prediction scores offered by CRISPOR for saturating mutagenesis screens as repetitive sequences can confound screen results[14]. sgRNAs with high probability of off-target mutagenesis can be excluded at the library design stage or can be appropriately handled at the analysis stage.

Analysis of deep sequencing from arrayed or pooled sgRNA experiments using CRISPResso

CRISPResso offers the ability to quantitate and visualize genome editing outcomes at individual loci[9]. CRISPResso provides a variety of features to offer users the opportunity to optimize analysis of sequencing data (Box 2). This analysis can be performed using the CRISPResso webtool or command line version (Table 2). CRISPResso analysis of an individual locus requires PCR amplification of the sequences flanking the genomic position of the DSB for a given sgRNA. The resulting deep sequencing FASTQ file can be analyzed by CRISPResso to quantitate the indel spectrum (Fig. 2). This analysis can be performed when targeting coding (Fig. 2a-h) or non-coding sequences (Fig. 2i,j). When targeting exonic sequence, CRISPResso can determine the frequency of in-frame and out-of-frame (frameshift) mutations produced. In the absence of selective pressure, it is expected that two-thirds of mutations introduced should cause frameshift and the remaining one-third should retain the reading if the mutation profile is random. Therefore, this analysis can be used to indicate deviation from this pattern to identify essential/toxic genes, which are expected to result in a higher frequency of in-frame mutations[14] (Fig. 2a-h). Two separate CRISPResso analyses with the same amplicon can be analyzed using the CRISPRessoCompare tool (Table 2). CRISPRessoCompare is useful in situations such as a comparing “treated” and “untreated” groups as well as to compare different experimental conditions. It can also be used to compare indel distributions created by two different sgRNA within the same region/amplicon[13] (Fig. 3b).

CRISPResso offers two methodologies to analyze deep sequencing data from CRISPR pooled screens. The first is to enumerate sgRNAs present in samples via CRISPRessoCount for the purposes of calculating enrichment and/or dropout of guides based on differing experimental conditions[13,14] (Fig. 3a). The second is an extension of CRISPResso analysis from a single amplicon to multiple amplicons using the CRISPRessoPooled tool (Table 2). This is useful for individual locus experiments that require multiple amplicons for analysis of the full region as well as for pooled screen experiments involving multiple loci/amplicons. The CRISPResso suite can also be used to analyze whole genome sequencing (WGS) data through CRISPRessoWGS. This requires pre-aligned WGS data in BAM format, which can be created using publicly available aligners (e.g., Bowtie2[56] or Burrows-Wheeler Aligner[57,58]). Similar to CRISPRessoCompare, two analyses using either CRISPRessoPooled or CRISPRessoWGS can be directly compared using CRISPRessoPooledWGSCompare (Table 2). This can be particularly useful when using multiple amplicon sequencing data or WGS data to evaluate off-target cleavages by two different sgRNA to identify sgRNAs with lower off-target activity.

MATERIALS

REAGENTS

Cloning/transformation

E. cloni 10G ELITE electrocompentent cells with recovery medium (Lucigen, cat. no. 60052)
Endura electrocompetent cells with recovery medium (Lucigen, cat. no. 60242)
Gibson Assembly Master Mix (New England Biolabs, cat. no. E2611S)
Gene Pulser/MicroPulser Electroporation Cuvettes, 0.1 cm gap (Biorad, cat. no. 1652089)
Corning Untreated 245mm Square BioAssay Dishes (Fisher Scientific, cat. no. 431111)
Fast Digest Esp3I (ThermoFisher, cat. no. FD0454)
NEB Stable Competent E. coli (High Efficiency) (New England Biolabs, cat. no. C3040H)
Pyrex solid glass beads (Fisher Scientific, cat. no. 11-312-10B)
LB broth base (ThermoFisher, cat. no. 12780052)
BD bacto dehydrated agar (Fisher Scientific, cat. no. DF0140-01-0)
Corning Falcon Bacteriological Petri Dishes with Lid (Fisher Scientific, cat. no. 08-757-100D)
TSAP thermosensitive alkaline phosphatase (Promega, cat. no. M9910)
Nuclease-Free Water (Fisher Scientific, AM9937)
Falcon Cell Scraper with 40cm Handle and 3.0cm Blade (Corning, cat. no. 353087)
T4 polynucleotide kinase (New England Biolabs, cat. no. M0201S)
Ampicillin sodium salt (Sigma, cat. no. A9518)
S.O.C. Medium (ThermoFisher, cat. no. 15544034)
Quick Ligation Kit (New England Biolabs, cat. no. M2200S)
Individual oligonucleotides (e.g., Integrated DNA Technologies, Bio-Rad)
Oligonucleotide pool synthesis (e.g., CustomArray Inc., Twist Bioscience)
QIAquick PCR purification kit (Qiagen, cat. no. 28104)

Gel electrophoresis

SYBR Safe DNA Gel Stain (ThermoFisher, cat. no. S33102)
VWR Life Science AMRESCO Agarose I (VWR, cat. no. 97062-250)
1 Kb Plus DNA Ladder (ThermoFisher, cat. no. 10787018)
50x TAE buffer (Boston Bioproducts, cat. no. BM-250)

PCR

Phusion Hot Start Flex DNA Polymerase (New England Biolabs, cat. no. M0535S)
Q5 High-Fidelity DNA Polymerase (New England Biolabs, cat. no. M0491S)
Herculase II Fusion DNA Polymerase (Agilent Genomics, cat. no. 600675)
Dimethyl sulfoxide (Sigma, cat. no. D8418)
QIAquick Gel Extraction Kit (Qiagen, cat. no. 28704)
MinElute PCR Purification Kit (Qiagen, cat. no. 28004)
Qubit dsDNA HS Assay Kit (ThermoFisher, cat. no. Q32854)
DNeasy Blood & Tissue Kit (Qiagen, cat. no. 69504)

Plasmids/plasmid preparation

lentiGuide-Puro (Addgene plasmid ID: 52963)[16]
lentiCas9-Blast (Addgene plasmid ID: 52962)[16]
lenti-Cas9-VQR-Blast (Addgene plasmid ID: 87155)[14]
Qiagen Plasmid Maxi Kit (Qiagen, cat. no. 12163)
AccuPrep Plasmid Mini Extraction Kit (Bioneer, cat. no. K-3030)

Lentivirus Production

pCMV-VSV-G (Addgene plasmid ID: 8454)
psPAX2 (Addgene plasmid ID: 12260)
Polyethylenimine, Branched (Sigma, cat no. 408727)
Sucrose (Sigma, cat no. S0389)
Steriflip-HV, 0.45 µm, PVDF, radio-sterilized (Millipore, cat. no SE1M003M00)
Stericup-GP, 0.22 µm, polyethersulfone, 500 mL, radio-sterilized (Millipore, cat no. SCGPU05RE)

Phosphate buffered saline 1x, w/o calcium, magnesium (Lonza, cat. no 17-516F)
DMEM (Life Technologies, cat no. 11995-073)
Ultracentrifuge Tube, Thinwall, Polypropylene, 38.5 mL, 25 x 89 mm (Beckman Coulter, cat no. 326823)
Falcon 50mL Conical Centrifuge Tubes (Fisher Scientific, cat no. 14-432-22)
Corning TC-Treated Culture Dishes, 15cm, round (Fisher, cat no. 08-772-24)
Penicillin-streptomycin (Life Technologies, cat no. 15140122)

REAGENT SETUP
Preparing sgRNA cloning, sequencing, and primer oligos: Dissolve oligonucleotides in nuclease-free water at a concentration of 100 µM. Store at −20 °C for up to two years.
TAE electrophoresis solution: Dilute 50x TAE buffer stock with dH_2O for a 1x working solution. Store at room temperature (25 °C) for up to 6 months.
Ampicillin solution: Dissolve ampicillin in a one-to-one mix of pure ethanol and dH_2O to a final concentration of 100 mg/mL. Store at -20 ^{o}C.
HEK293 cell culture medium: DMEM with glucose/sodium pyruvate supplemented with 10% fetal bovine serum (FBS) and 1% penicillin–streptomycin. Sterilize with 0.22 µm filter store it at 4 °C for up to 4 weeks.

CAUTION: The HEK293 cells used should be regularly checked to ensure that they are not infected by mycoplasma.

Polyethylenimine solution: Dissolve in nuclease free water at 10 µg/µL with pH 7.4.
20% sucrose solution: Dissolve sucrose in PBS to create 20% solution by mass. Sterilize with 0.22 µm filter store it at 4 °C for up to 1 year.

EQUIPMENT
Sequence analysis software (e.g. Thermo Scientific Vector NTI, Lasergene DNAstar, CLC Main Workbench, ApE, SerialCloner, or MacVector)
Electroporation system (e.g., Bio-Rad's Gene Pulser MXcel, Gene Pulser Xcell, and MicroPulser Electroporators)
MiSeq or HiSeq sequencing system (Illumina)
Qubit Fluorometer (Qubit)
UV-Vis spectrophotometer (NanoDrop)
Gel visualization system (Alpha Innotech)
Thermocycler (Bio-Rad)
UV-light transilluminator and UV-filter face-mask
CAUTION: Wear gloves, lab coat and face shield to avoid harm to eyes and skin caused by UV light.
VWR Microcentrifuge Tubes, Polypropylene (VWR, cat. no. 87003-294)
8-Strip PCR Tubes, 0.2mL, (Fisher Scientific, cat. no. 14-222-250)
Sterile Culture Tube with Attached Dual Position Cap (VWR, cat. no. 89497-812)
Ultracentrifuge (Beckman Coulter)
SW32Ti rotor with compatible swinging buckets (Beckman Coulter, cat no. 369694)

COMPUTING EQUIPMENT
Computer with at least 8 GB of RAM and 1TB of disk space (command line version only). Windows, OSX, or Linux platforms supported.

PROTOCOL

sgRNA design using CRISPOR | Timing 1-4 h

1) The CRISPOR webtool allows for the design sgRNAs from a single genomic locus (option A). sgRNA can be selected manually from the CRISPOR output using on- and off-target predictions (Box 1). If multiple genomic loci are required for sgRNA design or command line is preferred, the command line version of CRISPOR can be used (option B).

(A) sgRNA design using the CRISPOR webtool | Timing 1-4 h

i) Use the CRISPOR webtool: http://crispor.org/

ii) Input target DNA sequence. The webtool requires sequence input (<1 kilobase) or genomic coordinates. The input sequence must be a genome sequence, not a cDNA which can include sequence that is not in the genome due to splicing. You can obtain the sequence using a website like the UCSC Genome Browser (https://genome.ucsc.edu, click "View - DNA") or Ensembl (http://www.ensembl.org, click "Export data – Text"). From the UCSC Genome Browser, the current sequence in view can also be sent directly to CRISPOR via the menu entry "View – In external tools"

? TROUBLESHOOTING

iii) Select the relevant assembly (e.g. hg19, mm9).

CRITICAL STEP: It is important to pick the appropriate assembly to be consistent with the genomic coordinates of the DNA sequences provided in Step 1A(ii) and it also essential for accurate prediction of off-target sites.

iv) Select a PAM sequence for the relevant nuclease (e.g. NGG for *S. pyogenes* Cas9).

v) Run CRISPOR analysis to identify optimal sgRNA based on double strand break position as well as on- and off-target prediction (Box 1). Select CRISPOR on-target PCR primers for sequencing and subsequent indel analysis by CRISPResso.

CRITICAL STEP: CRISPOR allows for PCR amplicon lengths of 50-600 bp. Amplicon length should be selected based on deep sequencing read length. 100-200 bp amplicons are reasonable for paired-end deep sequencing with 75 or 150 bp reads. Because sequencing quality is lower towards the end of the reads, there must be some overlap of read pairs to allow reliable merging. For example, for a 200 bp amplicon it is suggested to use 125-150bp reads in order to have enough overlap. The melting temperature (Tm) is calculated for each primer pair.

vi) If you are conducting a saturation mutagenesis screen: the list of the oligonucleotide pool sequences for Gibson assembly to be ordered, sequencing primers for validation, sequencing amplicons for CRISPResso, and the full list of guide sequences can be downloaded by following the link "Saturation Mutagenesis" from the CRISPOR output page. Four files are provided:

- REGION_1_satMutOligos.tsv
  This file contains the sequences to order from a custom oligonucleotide pool supplier. Columns are:
  - guideId: the identifier of the guide sequence in the input sequence. It consists of the position of the PAM and the strand, e.g. "4rev"
  - targetSeq: the guide sequence including the PAM
  - mitSpecScore: the MIT Guide Specificity score (0-100, higher = more specific)
  - off-target count: number of predicted off-targets (by default at 4 mismatches)
  - targetGenomeGeneLocus: gene symbol and sequence location name, e.g. "exon:PITX2"
  - Doench'16EffScore: The Doench 2016 guide efficiency Score

  - Moreno-MateosEffScore: The Moreno-Mateos 2015 (crisprScan) guide efficiency score
  - OligoNucleotideAdapterHandle+PrimerFw: The forward oligonucleotide to order from the supplier
  - OligoNucleotideAdapterHandle+PrimerRev: The reverse oligonucleotide to order from the supplier

- REGION_1_ontargetPrimers.tsv
  This file contains two primers for each guide that can be ordered from an oligonucleotide supplier. The primers can be used to amplify one DNA fragment around each guide.
  - guideId: guide identifier, see above
  - forwPrimer: forward primer sequence
  - forwPrimerTm: forward primer Tm
  - revPrimer: reverse primer sequence
  - revPrimerTm: reverse primer Tm
  - ampliconSequence: the genomic sequence between forward and reverse primer

- REGION_1_ontargetAmplicons.tsv
  After sequencing using the ontargetPrimers file, this file can be used as input for CRISPRessoPooled to determine the cleavage frequency of each guide.
  - guideId: guide identifier, see above
  - ampliconSequence: the genomic sequence between forward and reverse primers, see above
  - guideSequence: the guide sequence located within the amplicon

- REGION_1_targetSeqs.tsv
  This file contains simply a list of all guide sequences, one per line. It can be used to quantify the relative abundance of these guides in a sample of sequenced cells and is one of the input files for CRISPRessoCount.

CRITICAL STEP: When designing a pooled screen, it is important to consider the inclusion of positive and negative controls (e.g., non-targeting sgRNA) in the library design. It is important to note that non-targeting sgRNAs are often genome- and PAM-specific. For example, a non-targeting sgRNA was designed for human hg19 reference genome using the NGG PAM.

If it is necessary to test a subset of guides, it is possible to filter those files using the command line procedure illustrated in step 1B xiv).

(B) sgRNA design using command line CRISPOR | Timing 1-4 h

i) Installation with Docker: Docker is a virtualization technology that allows packaging software with all dependencies into files called containers and to be executed on either Windows, Linux or OSX. This allows creating and distributing “frozen” version of the software that will always run regardless of the updates or changes of libraries or the required dependencies on the host machine. In this case, you don’t need to install any dependencies but just Docker. Download and install Docker from here: https://docs.docker.com/engine/installation/. Be sure to share some folders of your disk to be used to for the input data to analyze and for the output obtained through Docker. By default, any subfolder within your home directory is automatically shared. If a custom setup is necessary, please refer to: https://docs.docker.com/engine/tutorials/dockervolumes/

i) Type the command:

```
docker pull lucapinello/crispor_crispresso_nat_prot
```

ii) Verify that the container was downloaded successfully by running the command:

```
docker run lucapinello/crispor_crispresso_nat_prot crispor.py
```

iii) Create a folder to store the genomes to use and download a pre-indexed genome for CRISPOR. Obtain the CRISPOR assembly identifier of the genome (e.g. hg19 or mm9). If you are unsure, the full list of assemblies is available at http://crispor.tefor.net/genomes/genomeInfo.all.tab.

iv) Execute the following commands. Here we are assuming that the user will store the pre-indexed genomes in the folder: `/home/user/crispor_genomes`

```
mkdir -p /home/user/crispor_genomes

docker run -v /home/user/crispor_genomes:/crisporWebsite/genomes
lucapinello/crispor_crispresso_nat_prot downloadGenome hg19
/crisporWebsite/genomes
```

v) Prepare genomic input sequences in FASTA format (see a description of the format here https://www.ncbi.nlm.nih.gov/blast/fasta.shtml).

CRITICAL STEP: It is important to pick the appropriate assembly to be consistent with the genomic coordinates of the DNA sequences provided in Step 1B(iii) and it also essential for accurate prediction of off-target sites.

Generate FASTA file with exonic sequences (Optional) | Timing 15 min

vi) If you have a list of genes and you want to target their exons, use an internet browser and go to the UCSC Genome Browser page http://genome.ucsc.edu/cgi-bin/hgTables

vii) Set the following options (adjust entries as appropriate for each experiment):
- clade: Mammal,
- genome: Human, assembly: hg19,
- group: Genes and Gene Predictions,
- track: Gencode V19 (or preferred annotation),
- table: Basic,
- output format:sequence,
- output file: selected_genes_exons.fasta
- file type returned: "plain text"

viii) Press the button "paste list" and enter the list of gene identifiers.

ix) Press the button “get output”. A new page will open

x) Select the option “genomic” and press submit. A new page will open

xi) Select “CDS: Exons and “One FASTA record per region (exon, intron, etc.)”

xii) Press the button: "get sequence" and save the file as selected_genes_exons.fasta

Run CRISPOR analysis on single- or multi-FASTA input file | Timing 15 min

xiii) Run CRISPOR over your single- or multi-fasta input file. Let’s assume this file is stored in /home/user/crispor_data/crispor_input.fa and contains a sequence with id >REGION_1.

```
docker run \
-v /home/user/crispor_data/:/DATA \
-v /home/user/crispor_genomes:/crisporWebsite/genomes \
-w /DATA \
lucapinello/crispor_crispresso_nat_prot \
crispor.py hg19 crispor_input.fa crispor_output.tsv --
satMutDir=./
```

The sgRNA sequences will be written to the file crispor_output.tsv and 4 files for each sequence ID in the FASTA file will be created like in the web version:

- REGION_1_satMutOligos.tsv
- REGION_1_ontargetPrimers.tsv
- REGION_1_ontargetAmplicons.tsv
- REGION_1_targetSeqs.tsv

Select the primers to use for the experiment and create the files required for future CRISPResso analysis | Timing 5 min

xiv) To test only a subset of sgRNAs, using the file REGION_1_satMutOligos.tsv created in the previous step (or with the web version), and containing all the relevant information for the design, select the lines corresponding to sgRNA with desired scores and save them in a new file called REGION_1_satMutOligos_filtered.tsv (this operation can be performed for example using excel or the awk utility). Then run the following commands:

```
docker run -v $PWD/crispor_data/:/DATA -w /DATA
lucapinello/crispor_crispresso_nat_prot bash -c "join -1 1 -2 1
REGION_1_satMutOligos_filtered.tsv REGION_1_ontargetAmplicons.tsv -o 2.1,2.2,2.3 >
CRISPRessoPooled_amplicons.tsv"

docker run -v $PWD/crispor_data/:/DATA -w /DATA
lucapinello/crispor_crispresso_nat_prot bash -c "join -1 1 -2 1
REGION_1_satMutOligos_filtered.tsv REGION_1_ontargetAmplicons.tsv -o 2.3 | sed 1d
> CRISPRessoCounts_sgRNA.tsv"

docker run -v $PWD/crispor_data/:/DATA -w /DATA
lucapinello/crispor_crispresso_nat_prot bash -c "join -1 1 -2 1
REGION_1_satMutOligos_filtered.tsv REGION_1_ontargetPrimers.tsv -o
2.1,2.2,2.3,2.4,2.5,2.6,2.7 > REGION_1_ontargetPrimers_filtered.tsv"
```

This will create filtered version of the files to use to design amplicons and to perform CRISPResso analysis. If the input file to CRISPOR was a multi-fasta, repeat this step if necessary for the other regions.

Synthesis and cloning of individual sgRNA into a lentiviral vector | Timing 3 d to obtain sequenced-confirmed, cloned sgRNA lentiviral plasmid

2) When ordering individual oligonucleotides for sgRNA cloning, add “CACC” onto the sense strand of the sgRNA (5’ CACCNNNNNNNNNNNNNNNNNNN 3’) and “AAAC” onto the antisense strand of the sgRNA (5’ AAACNNNNNNNNNNNNNNNNNNN 3’). The addition of a “G” at the beginning of the sgRNA sequence can increase transcription from the U6 promoter (5’ CACCGNNNNNNNNNNNNNNNNNNN 3’ and 5’ AAACNNNNNNNNNNNNNNNNNNNC 3’); this is not required if the 5’ end of the sgRNA sequence is already a “G”. Individually resuspend lyophilized primer oligonucleotides at 100 μM. Using CRISPOR, the sequences for sgRNA cloning into lentiGuide-Puro can be downloaded from the “PCR primers” page available from the sgRNA results table. Store oligonucleotides at -20 ^{o}C.

PAUSE POINT: Resuspended oligonucleotides can be stored at -20 ^{o}C for years.

3) Set up phosphorylation reaction:

Components	Amount	Final Concentration
Oligonucleotide 1 (at 100 μM)	1 μL	-

Oligonucleotide 2 (at 100 µM)	1 µL	
10x ligation buffer	1 µL	1x
T4 polynucleotide kinase		10 units
Nuclease-free water	to 10 µL	

4) Perform phosphorylation and annealing of the oligonucleotides in a thermocycler:

Step	Temperature	Time	Comment
1	37 °C	30 min	
2	95 °C	5 min	
3	95 °C to 25 °C	30 s	Ramp down 5 °C/minute

5) Digest sgRNA expressing plasmid (lentiGuide-Puro) with Esp3I (an isoschozomer of BsmBI) restriction enzyme for 15 minutes at 37 °C.

Components	Amount	Final Concentration
lentiGuide-Puro plasmid	1 µg	-
10x FastDigest buffer	2 µL	1x
DTT (20 mM)	1 µL	1 mM
Esp3I restriction enzyme	1 µL	
Nuclease-free water	To 20 µL	

6) Add 1 µL of TSAP to dephosphorylate reaction for 15 minute-1 hour at 37 °C following by heat inactivation for 15 minutes at 74 °C.

7) Perform gel purification of the Esp3I-digested/dephosphorylated lentiGuide-Puro plasmid.

8) Perform ligation reaction for 5-15 minutes at room temperature (25 °C) using 30-70 ng of Esp3I-digested/dephosphorylated lentiGuide-Puro plasmid with the phosphorylated-annealed oligonucleotides produced in Steps 2-4 diluted 1:500 with nuclease-free water.

Components	Amount	Final Concentration
2x quick ligase reaction buffer	5 µL	1x
Diluted phosphorylated/annealed oligonucleotides	1 µL	
Esp3I-digested/dephosphorylated lentiGuide-Puro plasmid	30-70 ng	1x
Quick ligase	1 µL	10 units
Nuclease-free water	To 10 µL	

9) Transform ligation reaction in stable competent *E. coli*. Using recombination-deficient *E. coli* is important to minimize recombination events of repetitive elements in lentiviral plasmid.

10) Plate transformation on a 10-cm ampicillin-resistant (0.1 mg/mL ampicillin) agar plate. Incubate at 37 °C for 14-16 hours.

CRITICAL STEP: Growth at 32°C reduces recombination between the lentiviral long-terminal repeats, but is not essential.

11) Screen colonies for correct sgRNA by mini-scale plasmid preparation followed by Sanger sequencing with the U6 sequencing primer[52]: CGTAACTTGAAAGTATTTCGATTTCTTGGC

? TROUBLESHOOTING

PAUSE POINT: Successfully cloned sgRNA plasmids can be stored long term (years) at -20 °C prior to creating lentivirus (Steps 42-63).

Synthesis and cloning of pooled sgRNA libraries into a lentiviral vector | Timing 2 d to obtain cloned sgRNA lentiviral plasmid library

12) Use sgRNA's identified/chosen in Step 1 to design full-length oligonucleotides (96-99 bp) flanked by barcodes and homologous sequence to the lentiGuide-Puro plasmid (Table 3). Barcodes allow for multiple unique libraries to be individually amplified from the same pool of oligonucleotides. Batch sgRNA library cloning is performed using Gibson assembly, which relies on homologous flanking sequence for successful cloning.

13) Order synthesized DNA oligonucleotide pool synthesis on a programmable microarray.

14) Obtain resuspended oligonucleotides for use as the template for the PCR reaction in Step 15. Consider freezing subset of oligonucleotides to save as a backup.

    PAUSE POINT: Oligonucleotide pools can be stored at -20 ^{o}C for short term storage (months to years) and -80^{o}C for long term storage (years).

15) Set up barcode-specific PCR to amplify using barcode-specific primers (Table 4). This PCR will herein be referred to as "Library Synthesis PCR1" (lsPCR1).

    CRITICAL STEP: It is important to use a proofreading polymerase to minimize introduction of PCR errors.

Components	Amount	Final Concentration
Synthesized oligonucleotide template	1 µL	-
5x Phusion HF buffer or 5x Q5 reaction buffer	10 µL	1X
Deoxynucleoside triphosphates (dNTPs; at 10 mM)	1 µL	200 µM
Barcode-specific forward primer (at 10 µM)	2.5 µL	0.5 µM
Barcode-specific reverse primer (at 10 µM)	2.5 µL	0.5 µM
Phusion hot start flex DNA polymerase or Q5 high-fidelity DNA polymerase	0.5 µL	1.0 units/50 µl PCR
Nuclease-free water	to 50 µL	

    ? TROUBLESHOOTING

16) Perform multiple cycles (e.g., 10, 15, 20 cycles) of lsPCR1 in a thermocycler as follows (Fig. 4a):

Step	Temperature	Time	Comment
1	98 ^{o}C	30 s	
2	98 ^{o}C	10 s	
3	63 ^{o}C	30 s	
4	72 ^{o}C	30 s	Go to 2, X times (e.g., 10, 15, 20 cycles)
5	72 ^{o}C	5 min	

    PAUSE POINT: the PCR product can be stored at 4 ^{o}C in the thermocycler at the end of the PCR program in the short term (days). Longer term storage should be at -20 ^{o}C.

17) Run a fraction (e.g., 1-5 µL) of the lsPCR1 product on a 2% agarose gel to determine if the products occur at the expected size of ~100 bp (Fig. 4a). Choose the sample with the minimal number of PCR cycles that produces a visible band in order to minimize PCR bias to produce to the next step (Fig. 4a).

18) Dilute lsPCR1 reaction chosen from Step 17 1:10 using nuclease-free water.

19) Set up "Library Synthesis PCR2" (herein referred to as lsPCR2) with the following universal lsPCR2 primers (Table 5). lsPCR2 amplification removes the library-specific barcodes and adds sequence homologous to the lentiGuide-Puro plasmid for Gibson assembly-based cloning of the library.

CRITICAL STEP: It is important to use a proofreading polymerase to minimize introduction of PCR errors.

Components	Amount	Final Concentration
Diluted lsPCR1	1 µL	-
5x Phusion HF buffer or 5x Q5 reaction buffer	10 µL	1X
dNTPs (at 10 mM)	1 µL	200 µM
lsPCR2_forward primer (at 10 µM)	2.5 µL	0.5 µM
lsPCR2_reverse primer (at 10 µM)	2.5 µL	0.5 µM
Phusion hot start flex DNA polymerase or Q5 high-fidelity DNA polymerase	0.5 µL	1.0 units/50 µl PCR
Nuclease free water	To 50 µL	

20) Perform multiple cycles (e.g., 10, 15, 20 cycles) of lsPCR2 in a thermocycler as follows (Fig. 4b):

Step	Temperature	Time	Comment
1	98 ^{o}C	30 s	
2	98 ^{o}C	10 s	
3	63 ^{o}C	30 s	
4	72 ^{o}C	30 s	Go to 2, X times (e.g., 10, 15, 20 cycles)
5	72 ^{o}C	5 min	

PAUSE POINT: PCR products can be stored at 4 ^{o}C in the thermocycler at the end of the PCR program in the short term (days). Longer term storage should be at -20 ^{o}C.

21) Run lsPCR2 products on a 2% agarose gel to determine if the products occur at the expected size of ~140 bp (Fig. 4b). Choose the sample with the minimal number of PCR cycles that produces a visible band in order to minimize PCR bias (Fig. 4b). Gel purify the ~140 bp band.

CRITICAL STEP: If synthesizing multiple libraries simultaneously, leave empty lanes between different samples during gel electrophoresis to minimize risk of sample contamination during gel purification.

22) Set up Gibson assembly reaction using gel purified Esp3I-digested/dephosphorylated lentiGuide-Puro plasmid from Steps 5-7:

Components	Amount	Final Concentration
Gel-purified lsPCR2	10 ng	
Esp3I-digested/dephosphorylated lentiGuide-Puro plasmid	25 ng	
2X Gibson assembly master mix	10 µL	1x
Nuclease-free water	To 20 µL	

23) Incubate Gibson assembly reaction mix from Step 22 in a thermocycler at 50°C for 60 minutes

24) Thaw electrocompetent bacterial cells on ice until they thaw completely (~10-20 minutes).

25) Mix thawed electrocompetent bacterial cells by gently tapping on the side of the tube.

26) Aliquot 25 µL of cells to a pre-chilled microcentrifuge tube on ice.

27) Add 1 µL of the Gibson assembly reaction from Step 23 to the 25 µL of thawed electrocompetent bacterial cells from Step 26.

? TROUBLESHOOTING

28) Carefully pipet 25 µL of the electrocompetent bacterial cell/Gibson assembly reaction mixture from Step 27 into a chilled 0.1 cm gap electroporation cuvette without introducing bubbles. Quickly flick the cuvette downward to deposit the cells across the bottom of the cuvette's well.

CRITICAL STEP: Minimize introduction of air bubbles as they can interfere with plasmid electroporation efficiency.

29) Electroporate sample at 25 µF, 200 Ohms, 1500 volts using an electroporator. Expected time constant during electroporation is ~4.7 (typical range of 4.2-4.8)

? TROUBLESHOOTING

30) Add 975 µL of recovery or SOC medium to the cuvette as soon as possible after electroporation (now final volume of 1 mL in cuvette). Pipet up and down enough times to resuspend the cells within the cuvette (likely a single pipetting up and down should be sufficient). Recovery medium is preferred; however, SOC medium can be used in its place.

CRITICAL STEP: Electroporation is toxic to bacterial cells. Addition of recovery and SOC medium as quickly as possible enhances bacterial cell survival post-electroporation.

31) Transfer the cell mixture (~1 mL) to a culture tube already containing 1 mL of SOC medium.

32) Shake each culture tube (containing a total of 2 mL of recovery/SOC medium) at 250 rpm for 1 hour at 37^{o}C.

CRITICAL STEP: Growth at 32^{o}C reduces recombination between the lentiviral long-terminal repeats; however, this is not essential and 37 ^{o}C can be utilized.

33) Remove 1 uL from the 2 mL mixture (from Step 32) and add it to 500 µL of SOC medium. Mix well and plate 125 µL of the mixture (8,000x dilution) onto a pre-warmed 10-cm ampicillin-resistant (0.1 mg/mL ampicillin) agar plate and then plate the remaining 375 µL (2,667x dilution) onto a separate 10-cm ampicillin-resistant (0.1 mg/mL ampicillin) agar plate. These dilution plates will be used to estimate transformation efficiency of the sgRNA plasmid library to ensure that full library representation is preserved.

34) Plate the 2 mL transformation mixture (from Step 37) onto a pre-warmed 24.5 cm^2 ampicillin-resistant (0.1 mg/mL ampicillin) agar plate using Pyrex beads to ensure even spreading. Spread the liquid culture until it is largely absorbed into the agar and won't drip when inverted for incubation in Step 35.

35) Grow all three plates (2,667x dilution plate, 8,000x dilution plate, and 24cm^2 non-dilution plate) from Steps 33-34 inverted for 14-16 hours at 37^{o}C.

CRITICAL STEP: Growth at 32^{o}C reduces recombination between the lentiviral long-terminal repeats; however, this is not essential and 37 ^{o}C can be utilized.

36) Count the number of colonies on the two dilution plates from Step 33. Multiple this number of colonies by the dilution factor (2,667x or 8,000x) and by the increased area of the 24.5 cm2 plate (~7.6-fold increase in area) for estimation of the total number of colonies on the 24.5 cm^2 plate. It is reasonable to aim for >50x representation of the sgRNA library.

? TROUBLESHOOTING

37) Select ~10-20 colonies from the dilution plates (from Step 36) for screening by mini-scale plasmid preparation and subsequent Sanger sequencing (as in Step 11) for preliminary evaluation of library representation. The expectation is to identify unique sgRNA sequences among all sequenced mini-scale plasmid preparation given that probability of obtaining the same sgRNA from a full library of sgRNA should be low.

? TROUBLESHOOTING

38) Pipette 10 mL of LB broth onto each 24.5 cm^2 plate from Step 34 and scrape the colonies off with a cell scraper. The LB broth aides in removal of the colonies from the agar.

39) Pipette the LB broth/scraped bacterial colony mixture into a pre-weighed 50 mL tube and repeat the procedure a second time on the same plate with additional 5-10 mL of LB to maximize removal of bacterial colonies.

40) Centrifuge LB broth/scraped bacterial colony mixture at 2000 rpm for 5 minutes to pellet the bacteria and then discard the supernatant. Weigh the bacterial pellet in the tube (and subtract the pre-weighed tube to determine weight of the bacterial pellet) to determine the proper number of columns for maxi-scale plasmid preparation of the library. Each column can support ~0.45 g of bacterial pellet.

? TROUBLESHOOTING

41) Perform a sufficient number of maxi preps and combine eluted library plasmid DNA.

Lentivirus production from individual sgRNA plasmid or pooled sgRNA plasmid library | Timing 4 d

42) Passage and maintain HEK293 cells with 16 mL of HEK293 medium as previously described[52].

43) Perform transfection when HEK293 cells reach ~80% confluency in 15 cm round plate using polyethylenimine (PEI) as a transfection reagent. The plasmid:PEI ratio should be 1 µg of total transfected plasmid to 3 µg of PEI. Total plasmid consists of the sum of VSV-G µg + psPAX2 + library plasmid µg.

44) Mix PEI, VSV-G, psPAX2, and library plasmid in 1 mL of filtered DMEM without supplements in a sterile microcentrifuge tube:

? TROUBLESHOOTING

Reagent	Amount
VSV-G plasmid	8.75 µg
psPAX2 plasmid	16.25 µg
Individual sgRNA plasmid or sgRNA library plasmid	25 µg
Branched PEI (at 10 µg/µL)	150 µg

45) Invert microcentrifuge tube several times to mix. Allow tube to incubate at room temperature (25 ^oC) for 20-30 minutes.

CRITICAL STEP: the DMEM/plasmid DNA/PEI mixture should change from translucent to opaque during this incubation period.

46) Add the full volume (~1 mL) of DMEM/plasmid DNA/PEI mixture dropwise to the HEK293 cells and incubate at 37 ^oC for 16-24 hours.

47) Replace the media with 16 mL of fresh HEK293 medium 16-24 hours after transfection in Step 46.

48) Lentiviral supernatant harvest #1: 24 hours after replacing the medium in Step 46, collect the 16 mL media that now contains lentiviral particles (herein referred to as "viral supernatant") in a 50 mL falcon tube. Replace with 16 mL of fresh HEK293 medium. Store the viral supernatant at 4^{o}C.

    CAUTION: For this step and all steps remaining in the "Lentivirus production from individual sgRNA plasmid or pooled sgRNA plasmid library" section of this protocol, take all necessary precautions for the handling lentivirus and disposal of lentiviral waste. Lentivirus is capable of integrating into the genome of human cells.

49) Lentiviral supernatant harvest #2: 24 hours after lentiviral supernatant harvest #2, collect the viral supernatant and put it into the same 50 mL falcon tube from Step 48 (should now contain ~32 mL of viral supernatant).

50) Centrifuge collected viral supernatant from Step 49 at 2500-3000 rpm for 5 minutes to pellet HEK293 cells and other debris.

51) Filter the viral supernatant through a 0.45 µm 50 mL filter.

    ? TROUBLESHOOTING

    PAUSE POINT: The filtered viral supernatant can be stored at 4 ^{o}C for up to 7 days; however, storage may result in loss of viral titer. Ultracentrifugation as soon as possible is recommended for ultimate storage at -80 ^{o}C.

52) Transfer filtered viral supernatant to ultracentrifugation tubes.

53) Add 4-6 mL of 20% sucrose solution to the bottom of the ultracentrifugation tube to create a sucrose layer (also known as a "sucrose cushion") below the viral supernatant. This sucrose layer helps to remove any remaining debris from the viral supernatant as only the high-density viral particles can pass through the sucrose layer to pellet while the low-density debris remain in the supernatant.

    ? TROUBLESHOOTING

54) Add sterile PBS to the top of the ultracentrifuge tube until the final liquid volume within the ultracentrifuge tube is ~2-3 mm from the top. The volume must be within 2-3 mm of the top of the tube to prevent the tube from collapsing due to the force of ultracentrifugation.

55) Place ultracentrifuge tubes into ultracentrifuge buckets.

56) Weigh all ultracentrifuge buckets containing ultracentrifuge tubes prior to ultracentrifugation to ensure appropriate weight-based balancing. Use sterile PBS to correct weight discrepancies.

57) Centrifuge at 24,000 rpm for 2 hours in an ultracentrifuge

58) Remove ultracentrifuge tubes from ultracentrifuge buckets. The sucrose gradient should still be intact.

    ? TROUBLESHOOTING

59) Discard supernatant by inverting the ultracentrifugation tube. Allow inverted ultracentrifuge tube to stand on sterile paper towels to dry for 1-3 minutes.

    ? TROUBLESHOOTING

60) Revert ultracentrifugation tube to upright orientation and add 100 µL of sterile DMEM supplemented with 1% FBS.

61) Incubate for 3-4 hours (or overnight) shaking at 4°C to resuspend the lentiviral pellet.

62) Store concentrated lentivirus in a microcentrifuge tube at -80°C. Due to loss of lentiviral titer with repeated freeze/thaw cycles, consider aliquoting at this point. Aliquots frozen at the same time can be assumed to have the same titer, which can be advantageous for multiple experiments using the same lentivirus.

63) Experiments/screens using lentivirus should be performed as appropriate for the experimental design/objectives and the cell being utilized for study. For pooled screens, it is recommended to achieve a transduction rate between 30-50% to ensure single integrants into cells based on a Poisson distribution. After experiments/screens have been completed, cell pellets can be stored at -20 or -80 °C for weeks to months before processing with deep sequencing analysis of the samples (Steps 64-72).

Deep sequencing genome editing experiments using individual sgRNA or from pooled screens | Timing 1 d

64) Extract genomic DNA (gDNA) from fresh cell pellets after completion of genome editing screens/experiments or from frozen pellets from previous experiments.

65) Determine the concentration of each DNA sample by NanoDrop or other equivalent method.

66) Set up “Library Analysis PCR1” reaction (herein referred to as laPCR1). Use lentiGuide-Puro-specific primers (Table 6) to enumerate sgRNAs present for enrichment/dropout analysis or locus-specific primers to quantitate indels (Table 7).

CRITICAL STEP: It is important to use a proofreading polymerase to minimize introduction of PCR errors.

CRITICAL STEP: Optimization of DMSO concentration may be required (typically 1-10%). 8% is used in the reaction below.

CRITICAL STEP: Amount of gDNA for laPCR1 can vary based on experimental needs. On average, a genome from a single cell is approximately 6 picograms. Use adequate gDNA to represent desired number of cells.

? TROUBLESHOOTING

Components	Amount	Final Concentration
gDNA	X µL	-
Herculase II reaction buffer	10 µL	1X
dNTPs (at 100 mM)	1 µL	200 µM
lentiGuide_forward or Locus_forward primer (at 10 µM)	2.5 µL	0.5 µM
lentiGuide_forward or Locus_reverse primer (at 10 µM)	2.5 µL	0.5 µM
DMSO	4 µL	8%
Herculase II fusion DNA polymerase	0.5 µL	0.5 reactions
Nuclease-free water	to 50 µL	

67) Perform multiple cycles (e.g., 10, 15, 20 cycles) of laPCR1 in a thermocycler as follows. Gradient PCR may be required to determine the optimal annealing temperature (Fig. 4c).

PAUSE POINT: the PCR product can be stored at 4 °C in the thermocycler at the end of the PCR program in the short terms (days). Longer term storage should be at -20 °C.

? TROUBLESHOOTING

Step	Temperature	Time	Comment
1	95 °C	30 s	
2	95 °C	10 s	
3	60 °C	30 s	
4	72 °C	30 s	Go to 2, X times (e.g., 10, 15, 20 cycles)
5	72 °C	5 min	

68) Set up "Library Analysis PCR2" (herein referred to as laPCR2) using barcode-specific primers (Tables 8-9). Each sample will have a unique Illumina Nextera index to allow demultiplexing of multiple samples sequenced together. Set up two PCRs for each sample.

CRITICAL STEP: Two separate 10 µL reactions are performed as opposed to a single 20 µL reaction to minimize PCR bias.

Components	Amount	Final Concentration
laPCR1 (1:5 diluted)	1 µL	-
Herculase II reaction buffer	2 µL	1X
dNTPs (at 100 mM)	0.1 µL	100 µM
Forward primer (at 2 µM)	1 µL	0.2 µM
Reverse primer (at 2 µM)	1 µL	0.2 µM
Herculase II fusion DNA polymerase	0.1 µL	0.5 reactions
Nuclease free water	to 10 µL	

69) Perform multiple cycles (e.g., 10, 15, 20 cycles) of laPCR2 in a thermocycler as follows:

Step	Temperature	Time	Comment
1	95 °C	30 s	
2	95 °C	10 s	
3	60 °C	30 s	
4	72 °C	30 s	Go to 2, X times (e.g., 10, 15, 20 cycles)
5	72 °C	5 min	

70) Combine both 10 µL reaction for a total volume of 20 µL. Run laPCR2 products on a 2% agarose gel to determine if the products occur at the expected size (expected size varies based on primers used in laPCR1). Choose the sample with the minimal number of PCR cycles that produces a visible band in order to minimize PCR bias (Fig. 4d). Gel purify the relevant band of the expected size.

CRITICAL STEP: Leave empty lanes in between different samples to minimize risk of sample contamination during gel purification.

71) Quantitate DNA concentration by Qubit or other equivalent method.

72) Submit barcoded samples for deep sequencing (e.g., MiSeq or HiSeq) depending on the desired number of reads required to adequately represent the library or locus. The desired read length should be chosen by on the length of the amplicon generated by the primers used in Step 66.

CRISPResso installation and analysis of deep sequencing data | Timing 10-180 min

73) The CRISPResso utility can be used for analysis of deep sequencing data of a single locus/amplicon on a local machine or proprietary server, using a command line version or a webtool freely available at: http://crispresso.rocks (option A). The following steps will present detailed instructions for both workflows using the webtool (option A), command line (option B), and Docker (option C). Installation of CRISPResso

(A) Webtool installation | Timing 0 min

No installation is required, skip to next step.

(B) Installation of command line version of CRISPResso | Timing 15 min

To install CRISPResso on a local machine, it is necessary first to install three dependencies before running the setup script:

i. Download and install Python 2.7. Anaconda following the instruction at this link: http://continuum.io/downloads.

ii. Download and install a recent version of Java from this link: http://java.com/download.

iii. Check if a recent C compiler and make utility are available. For Linux systems those commands are usually available. For Mac with OSX 10.7 or greater, open the terminal app and execute the command:

```
make
```

This will trigger the installation of OSX developer tools. Windows is not officially supported, but we provide an alternative with the Docker installation (see point B).

iv. Open a terminal and type:

```
pip install CRISPResso --no-use-wheel –verbose
```

v. Close the terminal and open a new one, this will set the PATH variable. Now you are ready to use the command line version of CRISPResso.

Installation of CRISPRessoPooled, and CRISPRessoWGS (Optional) | Timing 15 min

vi. To use CRISPRessoPooled or CRISPRessoWGS it is necessary to install also bowtie2[56] and SAMtools[59], if they are not installed. In a terminal type the following commands:

```
conda config --add channels conda-forge
conda config --add channels defaults
conda config --add channels r
conda config --add channels bioconda
conda install samtools bowtie2
```

? TROUBLESHOOTING

(C) Installation of CRISPResso with Docker | Timing 15 min

i. Check to have installed Docker (see step 1B (i)) and then type the command:

```
docker pull lucapinello/crispor_crispresso_nat_prot
```

ii. Check if the container was download successfully running the command:

```
docker run lucapinello/crispor_crispresso_nat_prot CRISPResso --help
```

If CRISPResso was properly installed, you will see the help of the command line version of CRISPResso.

74) Analysis of deep sequencing data with CRISPResso:

(A) Analysis of deep-sequencing data using the CRISPResso webtool | Timing 15 min

i) Open a web browser and point to the page: http://crispresso.rocks.
ii) Select option for single-end or paired end reads based on the type of deep sequencing performed. Upload relevant FASTQ file(s) for analysis.

? TROUBLESHOOTING

iii) Input amplicon sequence (required), sgRNA sequence (optional), and coding sequence(s) (optional input for when targeting exonic sequence). The exonic sequence must be a subsequence of the amplicon sequence and not the sequence of the entire exon.

? TROUBLESHOOTING

iv) Adjust parameters for: window size (bp around each side of cleavage site) to quantify NHEJ edits (if sgRNA sequence provided), Minimum average read quality (phred33 scale), Minimum single bp quality (phred33 scale), Exclude bp from the left side of the amplicon sequence for the quantification of the mutations, Exclude bp from the right side of the amplicon sequence for the quantification of the mutations, and Trimming Adapter (Box 2).
v) Submit samples for analysis and download analysis reports. An illustrative video of the entire process is provided here: http://www.youtube.com/embed/dXblIIiAe00?autoplay=1

(B) Analysis of deep-sequencing data using command line CRISPResso | Timing 15 min

i) Gather the required information and files (see Step 1A(v) or 1B(xiv), the CRISPOR utility will output all the required information to run CRISPResso):

- FASTQ files to analyze, for example:
  - reads1.fastq.gz
  - reads2.fastq.gz
- Reference amplicon sequence, for example:

  AATGTCCCCCAATGGGAAGTTCATCTGGCACTGCCCACAGGTGAGGAGGTC
  ATGATCCCCTTCTGGAGCTCCCAACGGGCCGTGGTCTGGTTCATCATCTGT
  AAGAATGGCTTCAAGAGGCTCGGCTGTGGTT

- sgRNA used, for example:

  TGAACCAGACCACGGCCCGT

- Check if the files were trimmed for adapters or not.
- Expected repaired amplicon sequence (only if it is necessary to quantify HDR efficiency).
- Desired single bp or average read quality in PHRED33 score (suggested 20 for single bp, 30 for the read).
- Exonic sequence to use for the frameshift analysis (only if the reference amplicon sequence contains an exon or part of it).

- Experiment name, for example: BCL11A_exon2

ii) Open a terminal and run the CRISPResso utility. Here we are assuming that the FASTQ files, reads1.fastq.gz and reads2.fastq.gz are stored in the folder: /home/user/amplicons_data/

```
CRISPResso -r1 /home/luca/amplicons_data/reads1.fastq.gz -r2
/home/luca/amplicons_data/reads2.fastq.gz -a
AATGTCCCCCAATGGGAAGTTCATCTGGCACTGCCCACAGGTGAGGAGGTCATGATCCCCTTCTGGAG
CTCCCAACGGGCCGTGGTCTGGTTCATCATCTGTAAGAATGGCTTCAAGAGGCTCGGCTGTGGTT -g
TGAACCAGACCACGGCCCGT -s 20 -q 30 -n BCL11A_exon2
```

iii) After the execution of the command a new folder with all the results will be created, in this example the folder will be in: /home/user/amplicons_data/CRISPResso_on_BCL11A_exon2. The summary of the different events discovered in the sequencing data is presented in the file Alleles_frequency_table.txt, and in several illustrative plots (.pdf files). Refer to the online documentation for the full description of the output: https://github.com/lucapinello/CRISPResso.

? TROUBLESHOOTING

(C) Analysis of deep-sequencing data using CRISPResso with Docker | Timing 15 min

i) Gather the required information and files as in option (B) (Step 74B(i)).

ii) Assuming that the FASTQ files, reads1.fastq.gz and reads2.fastq.gz are stored in the folder: /home/user/amplicons_data/ run the command:

```
docker run -v /home/user/amplicons_data/:/DATA -w /DATA
lucapinello/crispor_crispresso_nat_prot CRISPResso -r1
reads1.fastq.gz -r2 reads2.fastq.gz -a
AATGTCCCCCAATGGGAAGTTCATCTGGCACTGCCCACAGGTGAGGAGGTCATGATCCCCTT
CTGGAGCTCCCAACGGGCCGTGGTCTGGTTCATCATCTGTAAGAATGGCTTCAAGAGGCTCG
GCTGTGGTT --g TGAACCAGACCACGGCCCGT -s 20 -q 30 -n BCL11A_exon2
```

ii) After the execution of the command a new folder with all the results will be created, in this example the folder will be in: /home/user/amplicons_data/CRISPResso_on_BCL11A_exon2. The summary of the different events discovered in the sequencing data is presented in the file Alleles_frequency_table.txt, and in several illustrative plots (.pdf files). Refer to the online documentation for the full description of the output: https://github.com/lucapinello/CRISPResso.

? TROUBLESHOOTING

CRISPResso analysis of lentiGuide-specific deep sequencing data for sgRNA enumeration by CRISPRessoCount | Timing 15 min

75) CRISPRessoCount is a utility for the enumeration of sgRNA. To use CRISPRessoCount is necessary to use the command line version of CRISPResso (see Step 73B). CRISPRessoCount can enumerate sgRNAs from a user-generated list or can empirically identify all sgRNA present in the FASTQ file.

76) Gather the required data and information (adjust entries as appropriate for experiment):

- FASTQ file containing the sequence of different sgRNAs to enumerate, for example: L3_NGG_HM_Plasmid.fastq.gz
- Optionally the file containing only the sequence of the sgRNAs to enumerate, one per line for example: library_NGG.txt (or the file created by CRISPOR, see Step 1A (vi) or 1B (xiv))
- The sgRNA scaffold sequence immediately following the spacer, for example: GTTTTAGAGCTAGAAATAGC

- The sgRNA length, for example 20.
- Desired single bp or average read quality in PHRED33 score (suggested 20 for single bp, 30 for the read).
- Experiment name, for example: NGG_HM_PLASMID

77) Run the CRISPRessoCount analysis. Here we are assuming that the FASTQ file and the optional sgRNA file are stored in the folder: /home/user/sgRNA _data/

(A) Using the command line version (see Step 73B for installation):

```
CRISPRessoCount -r /home/user/sgRNA_data/L3_NGG_HM_Plasmid.fastq.gz
-s 20 -q 30 -f library_NGG.txt -t GTTTTAGAGCTAGAAATAGC -l 20 --name
NGG_HM_PLASMID
```

(B) Using the Docker container (see Step 73C for installation):

```
docker run \
-v /home/user/sgRNA_data/:/DATA \
-w /DATA \
lucapinello/crispor_crispresso_nat_prot \
CRISPRessoCount -r L3_NGG_HM_Plasmid.fastq.gz \
-s 20 -q 30 -f library_NGG.txt \
-t GTTTTAGAGCTAGAAATAGC -l 20 --name NGG_HM_PLASMID
```

78) After the execution of the command a new folder with all the results will be created, in this example the folder will be in: /home/user/amplicons_data/CRISPRessoCount_on_NGG_HM_PLASMID. This folder contains two files, the execution log (CRISPRessoCount_RUNNING_LOG.txt) and a file (CRISPRessoCount_only_ref_guides_on_L3_NGG_HM_Plasmid.fastq.gz.txt) containing a table with the raw and normalized count for each sgRNA in the input FASTQ file:

Guide_Sequence	Read_Counts	Read_%	Reads_Per_Millions_(RPM)
CTCTGCCCTTCTGACATTGT	2592	0.312315496488	3123.15496488
ATGTGAGCATATGTATTCAT	2553	0.30761630499	3076.1630499
CCTGCTATGTGTTCCTGTTT	2455	0.2958080802	2958.080802
TTCTCGTGCCTCAGCCTCCT	2430	0.292795777957	2927.95777957
ACCCTGTGTATTTCACACAT	2349	0.283035918692	2830.35918692
TTCCATTTAATACACAATGT	2235	0.269299820467	2692.99820467

79) To calculate enrichment and/or dropout ('depletion') between two conditions, calculate the log2 ratio using the normalized sgRNA counts (that takes in account the total number of reads within each sequenced sample) using the column *Reads_Per_Millions_(RPM)* for the two conditions.

? TROUBLESHOOTING

CRISPResso analysis of using CRISPRessoPooled | Timing 20 min - 1 day

80) CRISPRessoPooled is a utility to analyze and quantify targeted sequencing CRISPR/Cas9 experiments involving sequencing libraries with pooled amplicons. To use CRISPRessoPooled is necessary to use the command line version of CRISPResso (see Step 73B). Although CRISPRessoPooled can be run in different modes (amplicon only, genome only and mixed), in this protocol we use the most reliable mode for quantification called *mixed mode*. In this mode, the amplicon reads are aligned to both reference amplicons and genome to discard ambiguous or spurious reads. For more details regarding the different running modes, consult the online help: https://github.com/lucapinello/CRISPResso.

81) Gather the required data and information. FASTQ files to analyze containing reads for a pooled experiment with multiple sgRNA, for example:

    - Reads_pooled_1.fastq.gz
    - Reads_pooled_2.fastq.gz

  - Reference amplicons sequences used in the pooled experiment
  - sgRNA sequences used, one for each amplicon
  - Check if the files were trimmed for adapters or not
  - Expected repaired amplicons sequences (only if it is necessary to quantify HDR efficiency).
  - Desired single bp or average read quality in PHRED33 score (suggested 20 for single bp, 30 for the read).
  - Exonic sequences to use for the frameshift analysis (only if the reference amplicon sequence contains an exon or part of it).
  - Amplicon names, all the names should be different
  - A global name for the pooled analysis, for example: Pooled_amplicons

82) Create a folder for the reference genome and genomic annotations (this step is necessary only one time)

```
mkdir -p /home/user/crispresso_genomes /hg19
```

83) *Download the reference genome*: Download the desired genome sequence data and precomputed index from https://support.illumina.com/sequencing/sequencing_software/igenome.html. In this example, we download the hg19 assembly of the human genome.

84) Move to the folder from Step 83 that will be used to store the genome:

```
cd /home/user/crispresso_genomes/hg19

wget ftp://igenome:G3nom3s4u@ussd-
ftp.illumina.com/Homo_sapiens/UCSC/hg1938/Homo_sapiens_UCSC_hg19.
tar.gz
```

85) Extract the data with the command:

```
tar -xvzf Homo_sapiens_UCSC_hg19.tar.gz \
Homo_sapiens/UCSC/hg19/Sequence/Bowtie2Index \
--strip-components 5
```

86) Create an amplicon description file or use the amplicon file created by CRISPOR at Step 1A(vi) or 1B (xiv). This file, is a tab delimited text file with up to 5 columns (first 2 columns required):

AMPLICON_NAME: an identifier for the amplicon (must be unique).
AMPLICON_SEQUENCE: amplicon sequence used in the design of the experiment.
sgRNA_SEQUENCE (OPTIONAL): sgRNA sequence used for this amplicon without the PAM sequence. If not available, enter NA.
EXPECTED_AMPLICON_AFTER_HDR (OPTIONAL): expected amplicon sequence in case of HDR. If more than one, separate by commas and not spaces. If not available, enter NA.
CODING_SEQUENCE (OPTIONAL): Subsequence(s) of the amplicon corresponding to coding sequences. If more than one, separate by commas and not spaces. If not available, enter NA.

A properly formatted file should look like this:

```
Site1 CACACTGTGGCCCCTGTGCCCAGCCCTGGGCTCTCTGTACATGAAGCAAC CCCTGTGCCCAGCCC NA NA
Site2 GTCCTGGTTTTTGGTTTGGGAAATATAGTCATC NA GTCCTGGTTTTTGGTTTAAAAAAATATAGTCATC NA
Site3 TTTCTGGTTTTTGGTTTGGGAAATATAGTCATC NA NA GGAAATATA
```

Here we are assuming that the file is saved in the folder /home/user/pooled_amplicons_data/ and is called AMPLICONS_FILE.txt

87) Open a terminal and run the CRISPRessoPooled utility. Here we are assuming that the FASTQ files, Reads_pooled_1.fastq.gz and Reads_pooled_1.fastq.gz are stored in the folder: /home/user/pooled_amplicons_data/

(A) Using the command line version (see Step 73B for installation):

```
CRISPRessoPooled \
-r1 /home/user/pooled_amplicons_data/Reads_pooled_1.fastq.gz \
-r2 /home/user/pooled_amplicons_data/Reads_pooled_2.fastq.gz \
-f /home/user/amplicons_data/AMPLICONS_FILE.txt \
-x /home/user/crispresso_genomes/hg19/genome \
--gene_annotations /home/user/crispresso_genomes/hg19/gencode_v19.gz \
-s 20 -q 30 \
--name Pooled_amplicons
```

(B) Using the Docker container (see Step 73C for installation):

```
docker run \
-v /home/user/amplicons_data/:/DATA -w /DATA \
-v /home/user/crispresso_genomes/:/GENOMES \
lucapinello/crispor_crispresso_nat_prot \
CRISPRessoPooled \
-r1 Reads_pooled_1.fastq.gz \
-r2 Reads_pooled_2.fastq.gz \
-f AMPLICONS_FILE.txt \
-x /GENOMES/genome \
--gene_annotations /GENOMES/hg19/gencode_v19.gz
-s 20 -q 30 \
--name Pooled_amplicons
```

88) After the execution of the command in the previous step, a new folder with all the results will be created. In this example, the folder will be in: /home/user/amplicons_data/CRISPRessoPooled_on_Pooled_amplicons. In this folder, the user can find those files:

- REPORT_READS_ALIGNED_TO_GENOME_AND_AMPLICONS.txt: this file contains the same information provided in the input description file, plus some additional columns:
  - Amplicon_Specific_fastq.gz_filename: name of the file containing the raw reads recovered for the amplicon.
  - n_reads: number of reads recovered for the amplicon.
  - Gene_overlapping: gene/s overlapping the amplicon region.
  - chr_id: chromosome of the amplicon in the reference genome.
  - bpstart: start coordinate of the amplicon in the reference genome.
  - bpend: end coordinate of the amplicon in the reference genome.
  - Reference_Sequence: sequence in the reference genome for the region mapped for the amplicon.
- MAPPED_REGIONS (folder): this folder contains all the fastq.gz files for the discovered regions.
- A set of folders with the CRISPResso report on the amplicons with enough reads.
- SAMPLES_QUANTIFICATION_SUMMARY.txt: this file contains a summary of the quantification and the alignment statistics for each region analyzed (read counts and percentages for the various classes: Unmodified, NHEJ, point mutations, and HDR).

- CRISPRessoPooled_RUNNING_LOG.txt: execution log and messages for the external utilities called.

? TROUBLESHOOTING

CRISPResso analysis using CRISPRessoWGS | Timing 1 h

89) CRISPRessoWGS is a utility for the analysis of genome editing experiment from whole genome sequencing (WGS) data. To use CRISPRessoWGS is necessary to use the command line version of CRISPResso (see Step 73B).

90) Gather the required data and information:
- Genome aligned WGS data in BAM format. For example: SAMPLE_WGS.bam
- The sequence of the reference genome used for the alignment, in FASTA format (see Step 83).
- Optionally a gene annotation file (see Step 1B(vi-xii))
- Coordinates of the regions to analyze
- Expected repaired sequences for each region analyzed (only if it is necessary to quantify HDR efficiency).
- Desired single bp or average read quality in PHRED33 score (suggested 20 for single bp, 30 for the read).
- Exonic sequences to use for the frameshift analysis (only if the region contains an exon or part of it).
- Region names, all the names should be different
- A global name for the WGS analysis, for example: WGS_regions

91) Create a regions description file. This file is a tab delimited text file with up to 7 columns (4 required) and contains the coordinates of the regions to analyze and some additional information:
- chr_id: chromosome of the region in the reference genome.
- bpstart: start coordinate of the region in the reference genome.
- bpend: end coordinate of the region in the reference genome.
- REGION_NAME: an identifier for the region (must be unique).
- sgRNA_SEQUENCE (OPTIONAL): sgRNA sequence used for this genomic segment without the PAM sequence. If not available, enter NA.
- EXPECTED_SEGMENT_AFTER_HDR (OPTIONAL): expected genomic segment sequence in case of HDR. If more than one, separate by commas and not spaces. If not available, enter NA.
- CODING_SEQUENCE (OPTIONAL): Subsequence(s) of the genomic segment corresponding to coding sequences. If more than one, separate by commas and not spaces. If not available, enter NA.

A properly formatted file should look like this:

```
chr1 65118211 65118261 R1 CTACAGAGCCCCAGTCCTGG NA NA
chr6 51002798 51002820 R2 NA NA NA
```

Here we are assuming that the file is saved in the folder /home/user/wgs_data/ and it is called REGIONS_FILE.txt

92) Open a terminal and run the CRISPRessoWGS utility. Here we are assuming that the BAM is stored in the folder: /home/user/wgs_data/ and was aligned using the human hg19 reference genome.

(A) Using the command line version of CRISPRessoWGS:

```
CRISPRessoWGS\
-b /home/user/wgs_data/SAMPLE_WGS.bam\
-f /home/user/wgs_data/REGIONS_FILE.txt \
-r /home/user/crispresso_genomes/hg19/hg19.fa \
```

```
--gene_annotations /home/user/crispresso_genomes/hg19/gencode_v19.gz
\
-s 20 -q 30 \
--name WGS_regions
```

(B) Using the Docker container version of CRISPRessoWGS:

```
docker run \
-v /home/user/wgs_data/:/DATA -w /DATA \
-v /home/user/crispresso_genomes/:/GENOMES \
lucapinello/crispor_crispresso_nat_prot \
CRISPRessoWGS \
-b SAMPLE_WGS.bam\
-f REGIONS_FILE.txt \
-r /GENOMES/hg19.fa \
--gene_annotations /GENOMES/gencode_v19.gz \
-s 20 -q 30 \
--name WGS_regions
```

After the execution of the command a new folder with all the results will be created, in this example the folder will be in: /home/user/wgs_data/CRISPRessoWGS_on_WGS_regions. In this folder, the user can find these files:

- REPORT_READS_ALIGNED_TO_SELECTED_REGIONS_WGS.txt: this file contains the same information provided in the input description file, plus some additional columns:
  - sequence: sequence in the reference genome for the region specified.
  - gene_overlapping*:* gene/s overlapping the region specified.
  - n_reads: number of reads recovered for the region.
  - bam_file_with_reads_in_region: file containing only the subset of the reads that overlap, also partially, with the region. This file is indexed and can be easily loaded for example on IGV for visualization of single reads or for the comparison of two conditions.
  - fastq.gz_file_trimmed_reads_in_region: file containing only the subset of reads fully covering the specified regions, and trimmed to match the sequence in that region. These reads are used for the subsequent analysis with CRISPResso.
- ANALYZED_REGIONS (folder): this folder contains all the BAM and FASTQ files, one for each region analyzed.
- A set of folders with the CRISPResso report on the regions provided in input.
- *CRISPRessoWGS_RUNNING_LOG.txt*: execution log and messages for the external utilities called.

? TROUBLESHOOTING

CRISPResso analysis of using CRISPRessoCompare and CRISPRessoPooledWGSCompare | Timing 30 min

93) CRISPRessoCompare and CRISPRessoPooledWGSCompare are two utilities allowing the user to compare two or several regions in two different conditions, as in the case of the CRISPResso, CRISPRessoPooled or CRISPRessoWGS. To use those utilities, it is necessary to use the command line version of CRISPResso (see Step 73B).

94) Gather the required data and information: Two CRISPResso, CRISPRessoPooled or CRISPRessoWGS completed analysis. For example:
  - /home/user/results/CRISPRessoPooled_on_sample1
  - /home/user/results/CRISPRessoPoooled_on_sample2

A name to use for the report, for example: comparison_sample1_sample2

95) Open a terminal and run the CRISPRessoCompare or CRISPRessoPooledCompare utility:

(A) Using the command line version (see Step 73B for installation):

```
CRISPRessoPooledWGSCompare \
/home/user/results/CRISPRessoPooled_on_sample1 \
/home/user/results/CRISPRessoPooled_on_sample2 \
-n comparison_sample1_sample2
```

(B) Using the Docker container (see Step 73C for installation):

```
docker run \
-v /home/user/results/:/DATA -w /DATA \
lucapinello/crispor_crispresso_nat_prot \
CRISPRessoPooledWGSCompare \
CRISPRessoPooled_on_sample1 \
CRISPRessoPooled_on_sample2 \
-n comparison_sample1_sample2
```

The syntax is exactly the same for results obtained with CRISPRessoWGS. For results obtained with CRISPResso, is necessary to use instead the utility CRISPRessoCompare.

After the execution of the command `CRISPRessoPooledWGSCompare` a new folder with all the results will be created, in this example the folder will be in: `/home/user/results/CRISPRessoPooledWGSCompare_on_comparison_sample1_sample2`

In this folder, the user will see those files:

- COMPARISON_SAMPLES_QUANTIFICATION_SUMMARIES.txt: this file contains a summary of the quantification for each of the two conditions for each region and their difference (read counts and percentages for the various classes: Unmodified, NHEJ, MIXED NHEJ-HDR and HDR).
- A set of folders with CRISPRessoCompare reports on the common regions with enough reads in both conditions.
- CRISPRessoPooledWGSCompare_RUNNING_LOG.txt: detailed execution log.

Inside each CRISPRessoCompare output folder the user will find:

- Comparison_Efficiency.pdf: a figure containing a comparison of the edit frequencies for each category (NHEJ, MIXED NHEJ-HDR and HDR) and as well the net effect subtracting the second sample (second folder in the command line) provided in the analysis from the first sample (first folder in the command line).
- Comparison_Combined_Insertion_Deletion_Substitution_Locations.pdf: a figure showing the average profile for the mutations for the two samples in the same scale and their difference with the same convention used in the previous figure (first sample – second sample).
- CRISPRessoCompare_RUNNING_LOG.txt: detailed execution log.

TROUBLESHOOTING

Table 10 | Troubleshooting

Step	Problem	Possible Reason	Solution
1A(ii)	Regions >1 kb or multiple regions	Large region or multi-loci design	For batch mode or genomic regions >1 kilobase, the command line version may be utilized (Step 1B)
1A(ii)	Genome of interest not available	-	If the genome of interest is not available in crispor.org, contact crispor@tefor.net and send a description of where to download the genome and optionally a pointer to the gene transcript models
11	Low cloning efficiency	Inadequate lentiGuide-Puro plasmid dephosphorylation and/or Esp3I restriction enzyme digestion	Repeat lentiGuide-Puro plasmid dephosphorylation and/or Esp3I restriction enzyme digestion. May increase duration of dephosphorylation and/or restriction enzyme digestion.
15	PCR failure	Too much oligonucleotide template	Dilute subset of oligonucleotide pool from Step 13 based on manufacturer's recommendations or try different dilutes (eg, 1:10 dilution with nuclease-free water)
15	PCR bias	Suboptimal DNA polymerase	Q5 high-fidelity DNA polymerase has been shown to result in less PCR bias as compared to Phusion hot start flex DNA polymerase
27, 29, 36	Low transformation efficiency/Abnormal electroporation time constant	Abnormal reaction chemistry from too much Gibson assembly reaction	Minimize Gibson assembly reaction volume added (<2 µL, but 0.5-1 µL is recommended)
36	Low cloning/transformation efficiency	Recombination of lentiviral long-terminal repeats	Incubation at 32 ^{o}C instead of 37 ^{o}C can minimize recombination events and result in increased transformation efficiency
36	Low cloning/transformation efficiency	Low efficiency of Gibson assembly reaction or no identifiable cause	Perform multiple Gibson assembly reactions (>1). Combine Gibson assembly reactions and concentrate using minimum elution kit to elute in 10 µL of nuclease-free water
36	Low cloning/transformation efficiency	Suboptimal electrocompetent cells	Endura electrocompetent cells have demonstrated a higher efficiency; however, consider performing direct head-to-head comparison experiment with E. Cloni electrocompetent cells
37	Overrepresentation of single or few sgRNA sequences	This likely results from PCR biases	Repeat library synthesis beginning with lsPCR1 (Step 15) and ensure usage of the correct primers for lsPCR1 and lsPCR2 reactions. It is possible that the full-length oligonucleotides have been incorrectly synthesized; however, this possibility is less likely.
37	sgRNA from different libraries identified	This likely results from using the incorrect barcode specific lsPCR1 reaction	Repeat lsPCR1 reaction with correct barcode-specific primers. Separate lsPCR2 samples with multiple empty

		(Step 21) or contamination from other libraries during gel purification (Step 27)	lanes and use separate gel purification supplies for each sample.
37	Identified sgRNA sequence not present library or any other barcoded libraries	This likely represents error in the oligonucleotide synthesis pipeline	Proceed with protocol as long as majority (50-75%) of sgRNAs identified match sequences within library
40	Low yield from maxi-scale plasmid preparation	This likely results from overloading the maxi-scale plasmid preparation column	Ensure that <0.45 g of bacterial pellet are loaded into each column
44	Low lentiviral titer	Mixing of plasmid DNA and PEI in DMEM with supplements	Charged proteins (e.g., from FBS) can interfere with the charge-based complexation of PEI with plasmid DNA. Repeat with filtered DMEM without supplements.
51	Viral supernatant stuck in the filter	This is due to a large amount of HEK293 cells and other debris in the viral supernatant clogging the filters	After Step 50, transfer viral supernatant to fresh 50 mL tube and repeat centrifugation from Step 50 to further remove HEK293 cells and debris.
53	Disrupted or mixed sucrose/viral supernatant layers in ultracentrifugation tube	This is likely due to releasing air bubbles while creating the layers. Alternatively, the layers can be disrupted if the pipette tip is not at the bottom of the tube while dispensing the sucrose	Minimize air bubbles while layering sucrose. Ensure pipette tube is at the bottom of the ultracentrifugation tube when layering sucrose.
58	Debris at bottom of virus	Absent, disrupted, or inappropriate sucrose gradient	See Troubleshooting for Step 53. Handle lentivirus-containing ultracentrifuge tubes and buckets carefully to avoid disrupting the layers. Ensure the sucrose solution is appropriately 20% by mass.
59	Low lentiviral titer	Loss of lentiviral pellet	Invert tube as quickly as possible in a single movement. Once tube is inverted, do not revert to the upright orientation to avoid resuspension of viral pellet and subsequent loss of resuspended viral particles.
66-67	PCR failure	Suboptimal DMSO concentration and/or annealing temperature	Separately perform DMSO concentration gradient and annealing temperature gradient for each primer set used
73-74	Multiple sequencing files	Examination of multiple unique amplicons	Use command line CRISPResso as opposed to the webtool for batch mode to expedite analysis of all files
73B(vi)	Pooled experiments; WGS data	-	Use CRISPRessoPooled or CRISPRessoWGS for analysis
74	Web version fails to complete the analysis complaining about the size of the file	The combined maximum size for the FASTQ file(s) must be less than 100MB.	To analyze bigger file(s), it is necessary to use the command line version of CRISPResso (skip to Step 73B).

	uploaded.		
74	Low number of reads analyzed by CRISPResso despite high number of reads in the input file	If using paired end reads, check that the reads overlap and that fully cover the reference amplicon sequence provided.	See considerations on amplicon design presented in Step 1A(vi)).
74	Low number of reads analyzed by CRISPResso despite high number of reads in the input file	This likely results from low sequence quality reads (low Phred33 scores) being removed from the analysis	Phred33-based read filtering is too stringent based on quality of sequencing data. Adjust the following two parameters: “Minimum average read quality (phred33 scale)” and “Minimum single bp quality (phred33 scale)”
74	Many indels created far from the predicted DSB position	This likely results from low sequencing quality whereby sequencing errors are interpreted as indels	Utilize the “Window size (bp around each side of cleavage site) to quantify NHEJ edits (if sgRNA sequence provided)” feature to restrict analysis of indels to a set interval centered around the prediction DSB position (Box 2)
73B, 74	The command line version of CRISPResso is not running correctly	Be sure to have install Anaconda Python 2.7 and not Python 3.x, since CRISPResso is not compatible with Python 3.	Install Anaconda Python 2.7, then run again the setup

73B, 74	The setup of the command line version of CRISPResso cannot compile the required dependencies.	A C compiler is not installed on your Linux system	Install the required packages with the following command:
			
			RHEL/CentOS:
			`sudo yum install gcc gcc-c++ autoconf automake`
			
			Ubuntu/Debian:
			`sudo apt-get install build-essential`
			
			Then run again the setup.

73B, 74	The setup of the command line version of CRISPResso cannot compile the dependencies.	The developer tools are not installed on your OSX system and typing *make* in a terminal window doesn’t trigger the installation.	If, you can install them manually creating a free account at this link: https://developer.apple.com/downloads/ and then search for "Command Line Tools" to find and download the version compatible with you OSX version. Double click on the downloaded .dmg file to mount the disk and complete the installation. Then run again the setup.
73B, 74	CRISPResso cannot find the required dependencies	The PATH variable for the installed dependencies cannot be set in the same terminal used for the installation.	If CRISPResso complains about *needle* or *flash* not present in your path, please be sure to use a new terminal after the installation as described in Step 73B
74	CRISPResso cannot align any sequence to the reference amplicon	The amplicon sequence provided may be not correct.	Inspect the first few lines of your FASTQ file - the start of the amplicon sequence should match the start of your sequences (or their reverse

			complement).
74	CRISPResso reports spurious indels localized at the end(s) of the amplicon.	The reads may be not trimmed for adapters/barcodes.	If reads are not already trimmed, select the adapters used for trimming under the 'Trimming Adapter' heading under the 'Optional Parameters'. Failure to trim adaptors may result in false positives. For the command line version, it is necessary to add the option `--trim sequences` (see Box 2).
74	CRISPResso cannot align any sequence to the reference amplicon for paired end reads if provided as a single FASTQ file	The paired end reads are provided in a single file instead of two required (a format usually referred as interleaved).	In this case, it is necessary to use the command line version of CRISPResso and add the option `--split_paired_end`.
74	CRISPResso analysis is taking too much time.	The high number or reads analyzed are slowing down the computation.	Consider to speed-up the computation adding the option `-p <number_or_processes>`, to enable the parallel option feature. For example, in a machine with 4 cores, you will add `-p 4`
74	CRISPResso is not performing the frameshift analysis	The subsequence of the amplicon corresponding to the exon(s) is not provided.	To enable the frameshift analysis, it is necessary to add the option -e to the command line with the (sub)sequence of the amplicon corresponding to the exon (see Box 2)
79	The enrichment score is not defined, contains NAN or INF	If one of the two condition has a value of 0, the ratio may be not defined,	Add a pseudo-count to both the conditions before taking the ratio.
88	Some regions are not analyzed and reported using CRISPSRessoPooled	By default, a region is analyzed and reported if it has at least 1000 reads.	If some amplicon is not reported in the analysis, check if the sequence used in the amplicon description file is correct. Check also the parameter `--min_reads_to_use_region`, this parameter allows to control which amplicons have sufficient reads to be analyzed (default 1000).
92	Some regions are not analyzed and reported using CRISPRessoWGS	By default, a region is analyzed and reported if it has at least 10 reads.	Adjust the parameter `--min_reads_to_use_region`, this parameter allows to control which regions have sufficient reads to be analyzed (default 10).

Timing

Step 1, sgRNA design using CRISPOR: 1-4 h

Step 1A, sgRNA design using the CRISPOR webtool: 1-4 h

Step 1B, sgRNA design using command line CRISPOR: 1-4 h

Steps 1B(vi-xii), Generate FASTA file with exonic sequences: 15 min

Step 1B(xiii), Run CRISPOR analysis on single- or multi-FASTA input file: 15 min

Step 1C(xiv), Create primers and files for future CRISPResso analysis: 15 min

Steps 2-11, Synthesis and cloning of individual sgRNA into a lentiviral vector: 3 d

Steps 12-41, Synthesis and cloning of pooled sgRNA libraries into a lentiviral vector: 2 d

Steps 42-63, Lentivirus production from individual sgRNA plasmid or pooled sgRNA plasmid library: 4 d

Steps 64-72, Deep sequencing genome editing experiments using individual sgRNA or from pooled screens: 1 d

Steps 73-79, CRISPResso installation and analysis of deep sequencing data: 10-180 min

Step 73A, Webtool installation: 0 min

Step 73B, Installation of command line version of CRISPResso: 15 min

Step 73B(vi), Installation of CRISPRessoPooled, and CRISPRessoWGS: 15 min

Step 73C, Installation of CRISPResso with Docker: 15 min

Step 74A, Analysis of deep-sequencing data using the CRISPResso webtool: 15 min

Step 74B, Analysis of deep-sequencing data using command line CRISPResso: 15 min

Step 74C, Analysis of deep-sequencing data using CRISPResso with Docker: 15 min

Steps 75-79, CRISPResso analysis of lentiGuide-specific deep sequencing data for sgRNA enumeration by CRISPRessoCount: 15 min

Steps 80-88, CRISPResso analysis using CRISPRessoPooled: 20 min - 1 day

Steps 89-92, CRISPResso analysis using CRISPRessoWGS: 1 h

Steps 93-95, CRISPResso analysis of using CRISPRessoCompare and CRISPRessoPooledWGSCompare: 30 min

ANTICIPATED RESULTS

Locus-specific deep sequencing analysis using CRISPResso can quantify indel frequency of coding targets (Fig. 2a-h) and non-coding targets (Fig. 2i,j). Quantification of editing (Fig. 2a) and distribution of combined insertions/deletions/substitutions (Fig. 2b) demonstrates successful editing of *BCL11A* exon 2 with indels flanking the DSB site. Frameshift analysis reveals that BCL11A loss is well-tolerated given frameshift and in-frame mutations consistent with the expected proportions (Fig. 2c,d). Quantification of editing (Fig. 2e) and distribution of combined insertions/deletions/substitutions (Fig. 2f) demonstrates low editing frequency of *MYB* exon 5 with indels flanking the DSB site. This result is consistent with toxicity associated with MYB loss or poor sgRNA activity. Frameshift analysis provides further evidence of toxicity resulting from MYB loss given deviation from expected proportion of frameshift and in-frame mutations (Fig. 2g,h). Taken together, the low editing rates and increased rates of in-frame mutations are consistent with toxicity due to MYB loss.

Saturating mutagenesis of *BCL11A* exon 2 and the functional core of the *BCL11A* enhancer[13,14] demonstrates enrichment of sgRNA based on fetal hemoglobin (HbF) levels as compared to non-targeting controls (Fig. 3a). Enrichment was determined by enumerating sgRNA read counts within deep sequenced samples (Fig. 3a). Enrichment of indels based on HbF levels as shown by a heatmap was determined by indel quantification using CRISPResso followed by CRISPRessoCompare analysis (Fig. 3b).

Box 1 | Selection of sgRNA based on on-target and off-target prediction scores.

CRISPOR provides two separate types of predictions: on-target and off-target scores. Off-target scores try to estimate the strength of off-target cleavage which depends on the sequence similarity between the on-target sequence and the off-target sequence. Predictions are first calculated on the level of the individual putative off-target site using the cutting frequency determination (CFD) off-target score. An off-target with a high off-target score is more likely to be cleaved than one with a low score. The scores of all off-target sites of a guide are then summarized into the "guide specificity score". It ranges from 0 to 100; a guide with a high specificity score is expected to have fewer off-targets and/or off-targets with a lower cleavage frequencies than one with a lower score. It was previously shown that guides with specificity scores >50 never lead to more than around a dozen off-targets, and none had a total off-target cleavage >10% in whole-genome assays[8]. These guides are shown in green by CRISPOR.

On the level of the individual site, off-targets with a CFD score < 0.02 are very unlikely to be cleaved, only a few cases out of ~200 known off-targets had such a score. For off-target sites that are cleaved, the CFD score is correlated well with cleavage strength but predicted off-target sites are known to be primarily false positives. As such, for most guides, more than 95% of predicted off-targets will not show any cleavage detectable with the sequencing depths in use today. In addition, as summarized in Haeussler et al.[8], a few guides did not lead to more than 3-5 detected off-targets, in spite of very low specificity scores in the range 10-30.

The other type of prediction is on-target scoring, which tries to estimate on-target cleavage and depends only on the target sequence. CRISPOR offers eight algorithms in total, but only two were most predictive when compared on independent datasets[8]. The algorithm by Doench et al[17] (herein referred to as the "Doench 2016 score") is optimal for cell-culture based assays where the guide is expressed from a U6 promoter, like in this protocol. The score by Moreno-Mateos et al[44] (CRISPRscan) is best for guides expressed in-vitro with a T7 promoter (e.g. for mouse/rat/zebrafish oocyte injections, not discussed further here). On-target scoring is not very reliable, the correlation between cleavage and the Doench 2016 score is only around 0.4. However, there certainly is an enrichment, e.g. among guides with the top 20% of Doench 2016 scores, 40% are in the top efficiency quartile, whereas with a random choice one would expect 25%.

By default, guides in CRISPOR are sorted by guide specificity score and off-targets are sorted by CFD score. The relative importance of off-target specificity versus on-target efficiency depends on the experiment. In some cases, e.g. when targeting a short region, one has little choice and a guide with low specificity (i.e., many/high-probability off-targets) is the only one available, in which case more time and efforts may be needed to screen for off-targets later. To give some idea of the number of off-targets to test, we found that most off-targets of guides with a specificity of > 50 are detected by screening the 96 predicted off-targets with the highest CFD scores[8].

Both off- and on-target scores have been pre-calculated for most common model organisms and are shown when hovering with the mouse over a guide in the UCSC Genome Browser track "Genes" - "CRISPR" at http://genome.ucsc.edu. The sequence currently shown in the UCSC Genome Browser can be sent directly to CRISPOR by selecting "View – In external tools" in the Genome Browser menu. In the UCSC track, guides are colored by predicted on-target efficiency whereas they are colored by off-target specificity on Crispor.org.

It can be particularly important to use off-target scoring prediction to aide in saturating mutagenesis experiments for sgRNA depletion or dropout (Fig. 1b, 2)[14]. sgRNA with highly probable off-target sites may deplete/dropout from screens due to cellular toxicity resulting from multiple cleavages as opposed to a biological effect from mutagenesis of the on-target site.

Box 2 | Optimizing parameters for CRISPResso analysis.

CRISPResso allows to set many parameters depending on the analysis to perform. Details on all the parameters are available in the online help section. Here we discuss the key parameters that can significantly affect the quantification.

Amplicon sequence (-a): This is the sequence expected to be observed without any edits. The sequence must be reported without adapters and barcodes.

sgRNA sequence *(-g):* This is the sequence of the sgRNA used and it should be a subsequence of the amplicon sequence (or its reverse complement). Although this parameter is optional, it is fundamental to specify it to enable the window mode (see the parameter--window_size, in order to reduce false positive in the quantification). It is important to remember that the sgRNA needs to be input as the guide RNA sequence (usually 20 nt) immediately 5' of the PAM sequence (usually NGG for SpCas9). For other nucleases, such as the Cpf1 system, enter the sequence (usually 20 nt) immediately 3' of the PAM sequence and explicitly set the cleavage offset (see the parameter: --cleavage_offset).

Coding sequence *(-c):* the subsequence/s of the amplicon sequence covering one or more coding sequences. Without this sequence, the frameshift analysis cannot be performed.

Window size (-w): This parameter allows for the specification of a window(s) in bp around each sgRNA to quantify the indels. This can help limit sequencing or amplification errors or non-editing polymorphisms from being inappropriately quantified in CRISPResso analysis. The window is centered on the predicted cleavage site specified by each sgRNA. Any indels not overlapping or substitutions not adjacent to the window are excluded. A value of 0 will disable this filter (default: 1).

Cleavage offset (--cleavage_offset): This parameter allows for the specification of the cleavage offset to use with respect to the provided sgRNA sequence. The default is -3 and is suitable for the SpCas9 system. For alternate nucleases, other cleavage offsets may be appropriate, for example, if using Cpf1 set this parameter to 1.

Average read and single bp quality (--min_average_read_quality and --min_single_bp_quality): These parameters allows for the specification of the minimum average quality score or the minimum single bp score to include a read for the analysis. The scale used is PHRED33 (default: 0, minimum: 0, maximum: 40). The PHRED score represent the confidence in assign a particular nucleotide in a read. The maximum score of 40 corresponds to an error rate of 0.01%. This average quality of a read is helpful to filter out low quality reads and a reasonable value for this parameter is 30. Use also the single bp quality for a more aggressive filtering; any read with a single bp below the threshold will be discarded. A reasonable value for this parameter is greater than 20.

Identity score: (--min_identity_score): This parameter allows for the specification of the min identity score for the alignment (default: 60.0). In order for a read to be considered properly aligned, it should pass this threshold. We suggest to lower this threshold only if really large insertions or deletions are expected in the experiment (>40% of the amplicon length).

Exclude ends of the read: (--exclude_bp_from_left and --exclude_bp_from_right): Sometime artifacts are present at the ends of the reads due to imperfect trimming or drop in quality scores. To exclude those regions those parameters, allow to exclude few bp from the left and/or right of the amplicon sequence for the quantification of the indels (default: 15).

Trimming of adapters (--trim_sequences) This parameter enables the trimming of Illumina adapters with Trimmomatic (default: False). For custom adapters the user can customize the Trimmomatic execution using the parameter --trimmomatic_options_string.

ACKNOWLEDGMENTS

M.C.C. is supported by a National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Award (F30DK103359-01A1). M.H. is funded by grants NIH/NHGRI 5U41HG002371-15 and NIH/NCI 5U54HG007990-02 and by a grant from the California Institute of Regenerative Medicine, CIRM GC1R-06673C.). D.E.B. is supported by NIDDK (K08DK093705, R03DK109232), NHLBI (DP2OD022716), Burroughs Wellcome Fund, Doris Duke Charitable Foundation Innovations in Clinical Research Award, ASH Scholar Award, Charles H. Hood Foundation Child Health Research Award, and Cooley's Anemia Foundation fellowship. S.H.O. is supported by an award from the NHLBI award (P01HL032262) and an award from the NIDDK (P30DK049216, Center of Excellence in Molecular Hematology). N.E.S. is supported by the NIH through NHGRI (R00-HG008171). L.P. is supported by a National Human Genome Research Institute (NHGRI) Career Development Award (R00HG008399).

AUTHOR CONTRIBUTIONS

M.C.C., M.H., and L.P. conceived this project. M.H. created CRISPOR. L.P. created CRISPResso. M.C.C., M.H., and L.P. wrote the manuscript with input from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests.

FIGURE LEGENDS

Fig. 1: Design, experimental preparation, and data analysis pipelines for arrayed and pooled sgRNA experiments.

Fig. 2: Individual locus-specific deep sequencing analysis of coding and non-coding targeting by CRISPResso. a, Frequency distribution of alleles with indels (shown in blue) and without indels (shown in red) for an sgRNA *BCL11A* exon 2. **b**, All reads with sequence modifications (insertions, deletions, and substitutions) with modification mapped to position within the *BCL11A* exon 2 reference amplicon. The vertical dashed line indicates the predicted Cas9 cleavage. **c**, Frameshift analysis of *BCL11A* exon 2 coding sequence targeted reads. Frameshift mutations are shown in red and in-frame mutations are shown in tan. **d**, Frameshift and in-frame mutagenesis profiles indicating position affected by modification for *BCL11A* exon 2 targeting. **e**, Frequency distribution of alleles with indels (shown in blue) and without indels (shown in red) for an sgRNA *MYB* exon 5. **f**, All reads with sequence modifications (insertions, deletions, and substitutions) with modification mapped to position within the *MYB* exon 5 reference amplicon. The vertical dashed line indicates the predicted Cas9 cleavage. **g**, Frameshift analysis of *MYB* exon 5 coding sequence targeted reads. Frameshift mutations are shown in red and in-frame mutations are shown in tan. **h**, Frameshift and in-frame mutagenesis profiles indicating position affected by modification for *MYB* exon 5 targeting. **i**, Frequency distribution of alleles with indels (shown in blue) and without indels (shown in red) for the top-scoring sgRNA targeting the *BCL11A* enhancer. **j**, All reads with sequence modifications (insertions, deletions, and substitutions) with modification mapped to position within the *BCL11A* enhancer reference amplicon. The vertical dashed line indicates the predicted Cas9 cleavage.

Fig. 3: Analysis of saturating mutagenesis screen of the functional core of the *BCL11A* enhancer (chr2:60,722,330-60,722,452 (hg19)). a, sgRNA enrichment based on analysis of fetal hemoglobin (HbF) levels when performing saturating mutagenesis of *BCL11A* exon 2 and the functional core of the *BCL11A* enhancer using NGG- and NGA-restricted sgRNA using data from two previously published studies[13,14]. Non-targeting sgRNA are pseudo-mapped with 5 bp spacing. **b**, Heatmap of sgRNA enrichment based on analysis of HbF levels when performing saturating mutagenesis within the *BCL11A* enhancer. HbF indel enrichment on per nucleotide based on CRISPResso analysis of amplicon genomic sequencing of sorted cells exposed to two different sgRNAs.

Fig. 4: Pooled sgRNA library preparation and analysis. a, Representative results for lsPCR1. **b**, Representative results for lsPCR2. **c**, Gradient PCR for locus-specific primers for laPCR1. **d**, Representative results for laPCR2.

TABLES

Table 1 | List of available PAM sequences. R = A or G; N = A, C, G or T.

Nuclease	Species	PAM	PAM position	References
Cas9	*Streptococcus pyogenes*	NGG	3'	[1,2]
Cas9	*Streptococcus pyogenes* HF1/ eSpCas9	NGG	3'	[54,55]
Cas9	*Streptococcus pyogenes* variant	NGA	3'	[60]
Cas9	*Streptococcus pyogenes* variant	NGCG	3'	[60]
Cas9	*Staphylococcus aureus*	NNGRRT	3'	[61]
Cas9	*Staphylococcus aureus KKH* variant	NNNRRT	3'	[62]
Cas9	*Streptococcus thermophilus ST1*	NNAGAA	3'	[63]
Cas9	*Streptococcus thermophilus A*	NGGNG	3'	[64]
Cas9	*Neisseria meningitides*	NNNNGATT	3'	[63]
Cas9	*Campylobacter jejuni*	NNNNACA	3'	[64,65]
Cpf1	*Acidaminococcus sp. BV3L6; Lachnospiraceae bacterium ND2006*	TTTN	5'	[4]

Table 2 | CRISPResso analysis suite

Name	Format	Purpose	Input File Formats	Comments
CRISPResso	Webtool	Analysis of single amplicon/locus deep sequencing	.fastq or .fastq.gz	crispresso.rocks
CRISPResso	Command line	Analysis of single amplicon/locus deep sequencing	.fastq or .fastq.gz	Large file support; batch mode capability
CRISPRessoPooled	Command line	Analysis of pooled amplicon experiments	fastq or .fastq.gz	-
CRISPRessoWGS	Command line	Analysis of WGS data or prealigned reads	.bam	Useful to interrogate any region of the genome for off-targets effects
CRISPRessoCompare	Command line	Comparison of two CRISPResso analyses	Output for CRISPResso analysis on two different samples	Useful for example to compare treated and untreated samples or to compare different experimental conditions
CRISPRessoPooledWGSCompare	Command line	Compare experiments involving several regions analyzed by either CRISPRessoPooled or CRISPRessoWGS	Output from CRISPRessoPooled or CRISPResso WGS analysis on two different samples	-
CRISPRessoCount	Command line	Enumerate sgRNA's present within a given sample	.fastq or .fastq.gz	Useful to perform enrichment or dropout analysis

Table 3 | Barcoding strategy for oligonucleotide pool synthesis. The sgRNA sequence is denoted with N's

Barcode Number	Forward barcode	Reverse barcode	Homologous sequence upstream of sgRNA	Homologous sequence downstream of sgRNA	Full oligonucleotide sequence
1	CGGGTTCCGT	GCTTAGAATAGAA	GGAAAGGACGAAACACCG	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	CGGGTTCCGTGGAAAGGACGAAACACCGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCGCTTAGAATAGAA
2	GTTTATCGGGC	ACTTACTGTACC	GGAAAGGACGAAACACCG	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	GTTTATCGGGCGGAAAGGACGAAACACCGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCACTTACTGTACC
3	ACCGATGTTGAC	CTCGTAATAGC	GGAAAGGACGAAACACCG	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	ACCGATGTTGACGGAAAGGACGAAACACCGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCCTCGTAATAGC
4	GAGGTCTTTCATGC	CACAACATA	GGAAAGGACGAAACACCG	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	GAGGTCTTTCATGCGGAAAGGACGAAACACCGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCCACAACATA
5	TATCCCGTGAAGCT	TTCGGTTAA	GGAAAGGACGAAACACCG	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	TATCCCGTGAAGCTGGAAAGGACGAAACACCGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTTCGGTTAA
6	TAGTAGTTCAGACGC	ATGTACCC	GGAAAGGACGAAACACCG	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	TAGTAGTTCAGACGCGGAAAGGACGAAACACCGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCATGTACCC
7	GGATGCATGATCTAG	CATCAAGC	GGAAAGGACGAAACACCG	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	GGATGCATGATCTAGGGAAAGGACGAAACACCGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCCATCAAGC
8	ATGAGGACGAATCT	CACCTAAAG	GGAAAGGACGAAACACCG	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	ATGAGGACGAATCTGGAAAGGACGAAACACCGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCCACCTAAAG
9	GGTAGGCACG	TAAACTTAGAACC	GGAAAGGACGAAACACCG	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	GGTAGGCACGGGAAAGGACGAAACACCGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAAACTTAGAACC
10	AGTCATGATTCAG	GTTGCAAGTCTAG	GGAAAGGACGAAACACCG	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	AGTCATGATTCAGGGAAAGGACGAAACACCGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCGTTGCAAGTCTAG

Table 4 | Primers for lsPCR1 library preparation

Primer	Sequence (5' to 3')
Barcode1F	CGGGTTCCGTGGAAAGG
Barcode1R	TTCTATTCTAAGCGCCTTATTTTAACTTGC
Barcode2F	GTTTATCGGGCGGAAAGG
Barcode2R	GGTACAGTAAGTGCCTTATTTTAACTTGC
Barcode3F	ACCGATGTTGACGGAAAGG
Barcode3R	GCTATTACGAGGCCTTATTTTAACTTGC
Barcode4F	GAGGTCTTTCATGCGGAAAGG
Barcode4R	TATGTTGTGGCCTTATTTTAACTTGC
Barcode5F	TATCCCGTGAAGCTGGAAAGG
Barcode5R	TTAACCGAAGCCTTATTTTAACTTGC
Barcode6F	TAGTAGTTCAGACGCGGAAAGG
Barcode6R	GGGTACATGCCTTATTTTAACTTGC
Barcode7F	GGATGCATGATCTAGGGAAAGG
Barcode7R	GCTTGATGGCCTTATTTTAACTTGC
Barcode8F	ATGAGGACGAATCTGGAAAGG
Barcode8R	CTTTAGGTGGCCTTATTTTAACTTGC
Barcode9F	GGTAGGCACGGGAAAGG
Barcode9R	GGTTCTAAGTTTAGCCTTATTTTAACTTGC
Barcode10F	AGTCATGATTCAGGGAAAGG
Barcode10R	CTAGACTTGCAACGCCTTATTTTAACTTGC

Table 5 | Primers for lsPCR2 library preparation

Primer	Sequence (5' to 3')
lsPCR2_forward	TAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCG
lsPCR2_reverse	ACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC

Table 6 | Primers for laPCR1 for pLentiGuide-specific deep sequencing for sgRNA enumeration. Blue sequence is Illumina Nextera handle sequence. Red sequence is specific to the lentiGuide-Puro plasmid.

Primer	Sequence (5' to 3')
lentiGuide_forward	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCG
lentiGuide_reverse	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTTTAGTTTGTATGTCTGTTGCTATTATGTCTACTATTCTTTCCC

Table 7 | Primers for laPCR1 for locus-specific deep sequencing. Blue sequence is Illumina Nextera handle sequence. Recommend 20 bp of locus-specific sequence.

Primer	Sequence (5' to 3')
Locus_forward	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-Locus-Specific-Sequence
Locus_reverse	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-Locus-Specific-Sequence

Table 8 | Illumina forward sequencing primers for PCR2 (i5-**Index**-Handle)

Primer	Sequence (5' to 3')
F501	AATGATACGGCGACCACCGAGATCTACAC**TAGATCGC**TCGTCGGCAGCGTC
F502	AATGATACGGCGACCACCGAGATCTACAC**CTCTCTAT**TCGTCGGCAGCGTC
F503	AATGATACGGCGACCACCGAGATCTACAC**TATCCTCT**TCGTCGGCAGCGTC
F504	AATGATACGGCGACCACCGAGATCTACAC**AGAGTAGA**TCGTCGGCAGCGTC
F505	AATGATACGGCGACCACCGAGATCTACAC**GTAAGGAG**TCGTCGGCAGCGTC
F506	AATGATACGGCGACCACCGAGATCTACAC**ACTGCATA**TCGTCGGCAGCGTC
F507	AATGATACGGCGACCACCGAGATCTACAC**AAGGAGTA**TCGTCGGCAGCGTC
F508	AATGATACGGCGACCACCGAGATCTACAC**CTAAGCCT**TCGTCGGCAGCGTC
F517	AATGATACGGCGACCACCGAGATCTACAC**GCGTAAGA**TCGTCGGCAGCGTC

Table 9 | Illumina reverse primers for deep sequencing (Reverse Primers (i7-**Index**-Handle)

Primer	Sequence (5' to 3')
R701	CAAGCAGAAGACGGCATACGAGAT**TCGCCTTA**GTCTCGTGGGCTCGG
R702	CAAGCAGAAGACGGCATACGAGAT**CTAGTACG**GTCTCGTGGGCTCGG
R703	CAAGCAGAAGACGGCATACGAGAT**TTCTGCCT**GTCTCGTGGGCTCGG
R704	CAAGCAGAAGACGGCATACGAGAT**GCTCAGGA**GTCTCGTGGGCTCGG
R705	CAAGCAGAAGACGGCATACGAGAT**AGGAGTCC**GTCTCGTGGGCTCGG
R706	CAAGCAGAAGACGGCATACGAGAT**CATGCCTA**GTCTCGTGGGCTCGG
R707	CAAGCAGAAGACGGCATACGAGAT**GTAGAGAG**GTCTCGTGGGCTCGG
R708	CAAGCAGAAGACGGCATACGAGAT**CCTCTCTG**GTCTCGTGGGCTCGG
R709	CAAGCAGAAGACGGCATACGAGAT**AGCGTAGC**GTCTCGTGGGCTCGG
R710	CAAGCAGAAGACGGCATACGAGAT**CAGCCTCG**GTCTCGTGGGCTCGG
R711	CAAGCAGAAGACGGCATACGAGAT**TGCCTCTT**GTCTCGTGGGCTCGG
R712	CAAGCAGAAGACGGCATACGAGAT**TCCTCTAC**GTCTCGTGGGCTCGG

REFERENCES

1. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339,** 823–6 (2013).
2. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–23 (2013).
3. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–1712 (2007).
4. Zetsche, B. *et al.* Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163,** 759–71 (2015).
5. Canver, M. C. *et al.* Characterization of genomic deletion efficiency mediated by clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *J. Biol. Chem.* **289,** 21312–21324 (2014).
6. Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154,** 1380–9 (2013).
7. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31,** 827–32 (2013).
8. Haeussler, M. *et al.* Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* **17,** 148 (2016).
9. Pinello, L. *et al.* Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat. Biotechnol.* **34,** 695–697 (2016).
10. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343,** 84–7 (2014).
11. Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32,** 267–73 (2014).
12. Zhou, Y. *et al.* High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509,** 487–91 (2014).
13. Canver, M. C. *et al.* BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527,** 192–7 (2015).
14. Canver, M. C. *et al.* Variant-aware saturating mutagenesis using multiple Cas9 nucleases identifies regulatory elements at trait-associated loci. *Nat. Genet.* (2017).
15. Sanjana, N. *et al.* High-resolution interrogation of functional elements in the noncoding genome. *Science* **353,** 1545–1549 (2016).
16. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11,** 783–784 (2014).
17. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34,** 184–91 (2016).
18. Park, J., Kim, J. & Bae, S. Cas-Database: Web-based genome-wide guide RNA library design for gene knockout screens using CRISPR-Cas9. *Bioinformatics* **32,** 2017–23 (2016).
19. Bae, S., Park, J. & Kim, J. S. Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30,** 1473–1475 (2014).
20. Xiao, A. *et al.* CasOT□: a genome-wide Cas9 / gRNA off-target searching tool. **30,** 3–5 (2014).
21. Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J. & Mateo, J. L. CCTop: An intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLoS One* **10,** 1–11 (2015).
22. Cradick, T. J., Qiu, P., Lee, C. M., Fine, E. J. & Bao, G. COSMID: A Web-based Tool for Identifying and Validating CRISPR/Cas Off-target Sites. *Mol. Ther. Acids* **3,** e214 (2014).
23. Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M. & Valen, E. CHOPCHOP: A CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* **42,** 401–407 (2014).
24. Labun, K., Montague, T. G., Gagnon, J. A., Thyme, S. B. & Valen, E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res.* **44,** W272–W276 (2016).
25. Naito, Y., Hino, K., Bono, H. & Ui-Tei, K. CRISPRdirect: Software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics* **31,** 1120–1123 (2015).
26. Ma, J. *et al.* CRISPR-DO for genome-wide CRISPR design and optimization. *Bioinformatics* **32,** 3336–3338 (2016).
27. Liu, H. *et al.* CRISPR-ERA: A comprehensive design tool for CRISPR-mediated gene editing,

repression and activation. *Bioinformatics* **31,** 3676–3678 (2015).
28. Lei, Y. *et al.* CRISPR-P: A web tool for synthetic single-guide RNA design of CRISPR-system in plants. *Mol. Plant* **7,** 1494–1496 (2014).
29. Singh, R., Kuscu, C., Quinlan, A., Qi, Y. & Adli, M. Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. *Nucleic Acids Res.* **43,** 1–8 (2015).
30. Heigwer, F., Kerr, G. & Boutros, M. E-CRISP: fast CRISPR target site identification. *Nat. Methods* **11,** 122–123 (2014).
31. Gratz, S. J. *et al.* Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. *Genetics* **196,** 961–971 (2014).
32. Perez, A. R. *et al.* GuideScan software for improved single and paired CRISPR guide RNA design. *Nat. Biotechnol.* (2017). doi:10.1038/nbt.3804
33. O'Brien, A. & Bailey, T. L. GT-Scan: Identifying unique genomic targets. *Bioinformatics* **30,** 2673–2675 (2014).
34. Wong, N., Liu, W. & Wang, X. WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biol.* **16,** 218 (2015).
35. Zhu, L. J., Holmes, B. R., Aronin, N. & Brodsky, M. H. CRISPRseek: A Bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. *PLoS One* **9,** (2014).
36. Xie, S., Shen, B., Zhang, C., Huang, X. & Zhang, Y. SgRNAcas9: A software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites. *PLoS One* **9,** 1–9 (2014).
37. Prykhozhij, S. V., Rajan, V., Gaston, D. & Berman, J. N. CRISPR multitargeter: A web tool to find common and unique CRISPR single guide RNA targets in a set of similar sequences. *PLoS One* **10,** 1–18 (2015).
38. Tycko, J., Myer, V. E. & Hsu, P. D. Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity. *Mol. Cell* **63,** 355–370 (2016).
39. Fusi, N., Smith, I., Doench, J. & Listgarten, J. In Silico Predictive Modeling of CRISPR / Cas9 guide efficiency. *bioRxiv* (2015).
40. Chari, R., Mali, P., Moosburner, M. & Church, G. M. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat. Methods* **12,** 823–826 (2015).
41. Xu, H. *et al.* Sequence determinants of improved CRISPR sgRNA design. *Genome Res.* **25,** 1147–57 (2015).
42. Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* **32,** 1262–7 (2014).
43. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343,** 80–4 (2014).
44. Moreno-Mateos, M. A. *et al.* CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat. Methods* **12,** 982–8 (2015).
45. Housden, B. E. *et al.* Identification of potential drug targets for tuberous sclerosis complex by synthetic screens combining CRISPR-based knockouts with RNAi. *Sci. Signal.* **8,** rs9 (2015).
46. Ren, X. *et al.* Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in Drosophila. *Cell Rep.* **9,** 1151–1162 (2014).
47. Farboud, B. & Meyer, B. J. Dramatic enhancement of genome editing by CRISPR/cas9 through improved guide RNA design. *Genetics* **199,** 959–971 (2015).
48. Bae, S., Kweon, J., Kim, H. S. & Kim, J.-S. Microhomology-based choice of Cas9 nuclease target sites. *Nat Methods* **11,** 705–706 (2014).
49. Güell, M., Yang, L. & Church, G. M. Genome editing assessment using CRISPR Genome Analyzer (CRISPR-GA). *Bioinformatics* **30,** 1–3 (2014).
50. Park, J., Lim, K., Kim, J.-S. & Bae, S. Cas-Analyzer: an online tool for assessing genome editing results using NGS data. *Bioinformatics* btw561 (2016). doi:10.1093/bioinformatics/btw561
51. Xue, L. J. & Tsai, C. J. AGEseq: Analysis of genome editing by sequencing. *Mol. Plant* **8,** 1428–1430 (2015).
52. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8,** 2281–308 (2013).
53. Nelson, C. E. & Gersbach, C. A. Engineering Delivery Vehicles for Genome Editing. *Annu. Rev. Chem. Biomol. Eng.* **7,** 637–62 (2016).

54. Kleinstiver, B. P. *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529,** 490–495 (2016).
55. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351,** 84–88 (2016).
56. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
57. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
58. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26,** 589–595 (2010).
59. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
60. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523,** 481–485 (2015).
61. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520,** 186–91 (2015).
62. Kleinstiver, B. P. *et al.* Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat. Biotechnol.* **33,** 1293–1298 (2015).
63. Esvelt, K. M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* **10,** 1116–21 (2013).
64. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513,** 569–73 (2014).
65. Kim, E. *et al.* In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. *Nat. Commun.* **8,** 1–12 (2017).

Figure 1

a

Arrayed sgRNA experiment (One sgRNA in one experiment)

sgRNA design by CRISPOR

Design sgRNA using CRISPOR webtool or command line version. sgRNA chosen based on on- and off-target prediction or as appropriate for the experimental design (Step 1; Box 1)

On-target analysis

Off-target analysis

Experimental preparation

1. Clone individual sgRNA (Steps 2-11)
2. Produce lentivirus (Steps 42-63)
3. Locus-specific deep sequencing using on-target, locus-specific primers (Steps 64-72)

1. Clone individual sgRNA (Steps 2-11)
2. Produce lentivirus (Steps 42-63)

OPTION A
3a. Locus-specific deep sequencing using locus-specific primers for each off-target site (Steps 64-72)

OPTION B
3b. Whole genome sequencing

Data analysis by CRISPResso

1. Analysis of FASTQ files using CRISPResso webtool or command line version (Steps 73-74) (Fig. 2a,b,e,f,i,j)
2. Consider frameshift or splice site analysis (if applicable; Fig. 2c,d,g,h)

OPTION A
1. Analysis of FASTQ files using CRISPResso webtool or command line version (Steps 73-74). If analyzing multiple off-target loci, use CRISPRessoPooled (Steps 80-88).
2. Consider frameshift or splice site analysis (if applicable)

OPTION B
1. Align whole genome sequencing to relevant genome
2. CRISPRessoWGS analysis using aligned reads (BAM format) (Steps 89-92)

b

Pooled sgRNA experiment (Multiple sgRNA in one experiment)

Saturating mutagenesis library: Design all sgRNA in region(s) for saturating mutagenesis of coding and/or non-coding elements (Step 1; Box 1)

Gene library: design multiple sgRNA for each gene of interest (Step 1; Box 1)

Any other customized library of sgRNA for a pooled experiment (Step 1; Box 1)

1. Batch clone pool of sgRNA oligonucleotides (Steps 12-41)
2. Produce lentivirus (Steps 42-63)

OPTION A
3a. Deep sequencing using lentiGuide plasmid--specific primers (Steps 64-72)

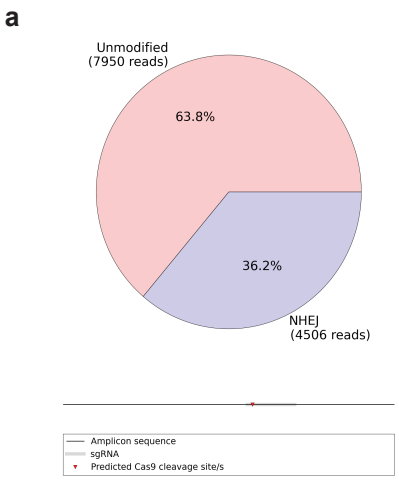
OPTION B
3b. Locus-specific deep sequencing using on-target, locus-specific primers tiling all region(s) of interest (Steps 64-72)

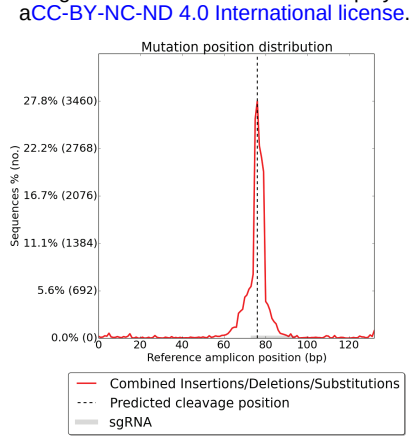
OPTION A
1. Analysis of FASTQ files using CRISPRessoCounts (Steps 75-79)
2. Normalize reads across all experimental samples
3. Calculate enrichment or depletion based on experimental design (Fig. 2)

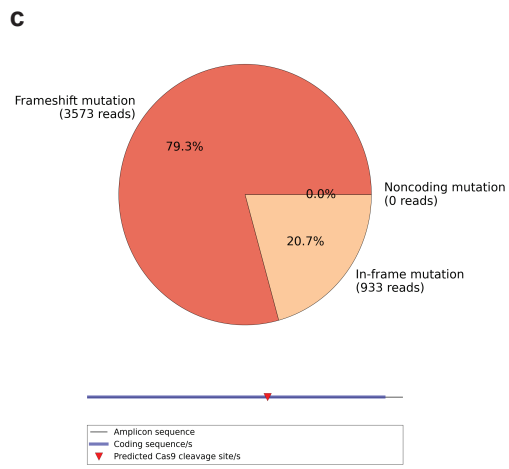
OPTION B
1. Analysis of FASTQ files using CRISPRessoPooled (Steps 80-88)

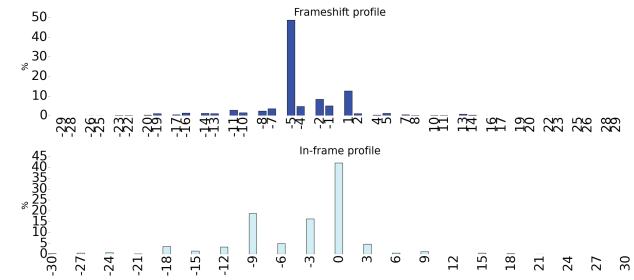
Figure 2

***BCL11A* exon 2**

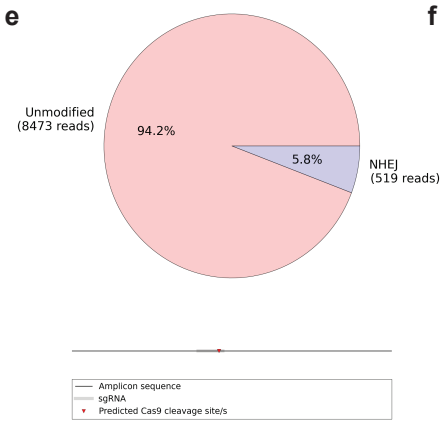


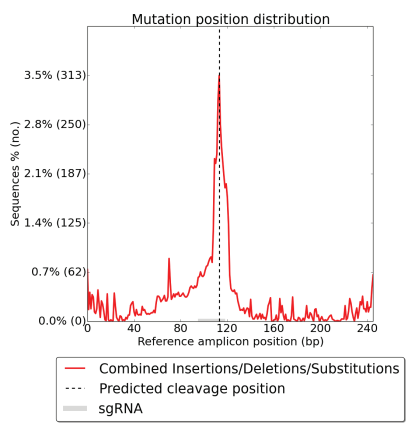


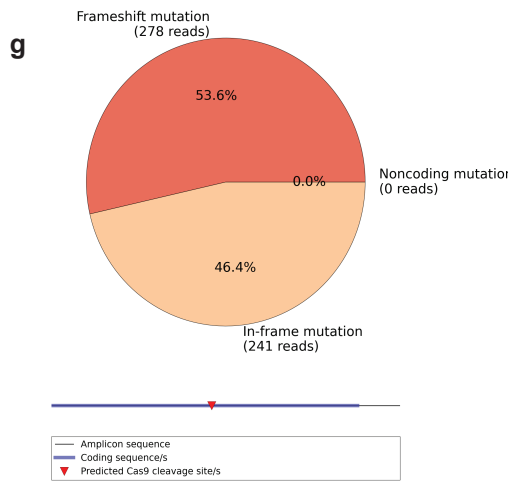


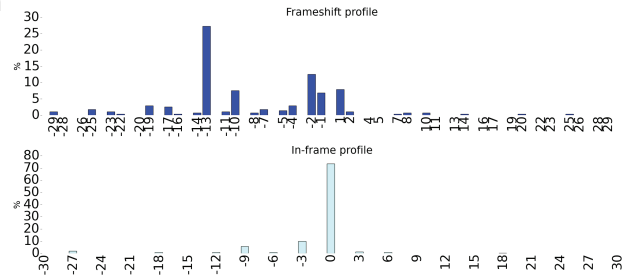


***MYB* exon 5**

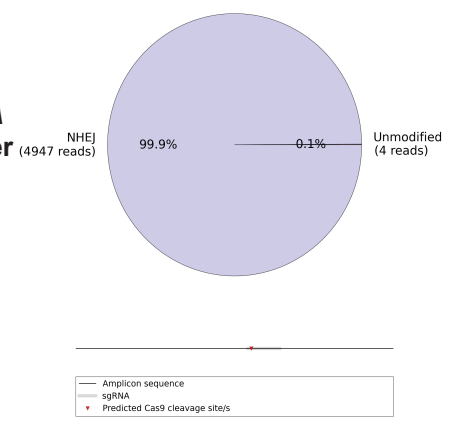








***BCL11A* Enhancer**



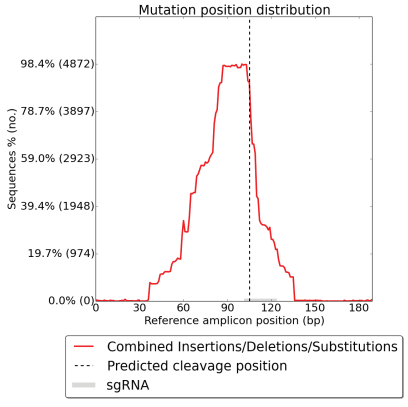


Figure 3

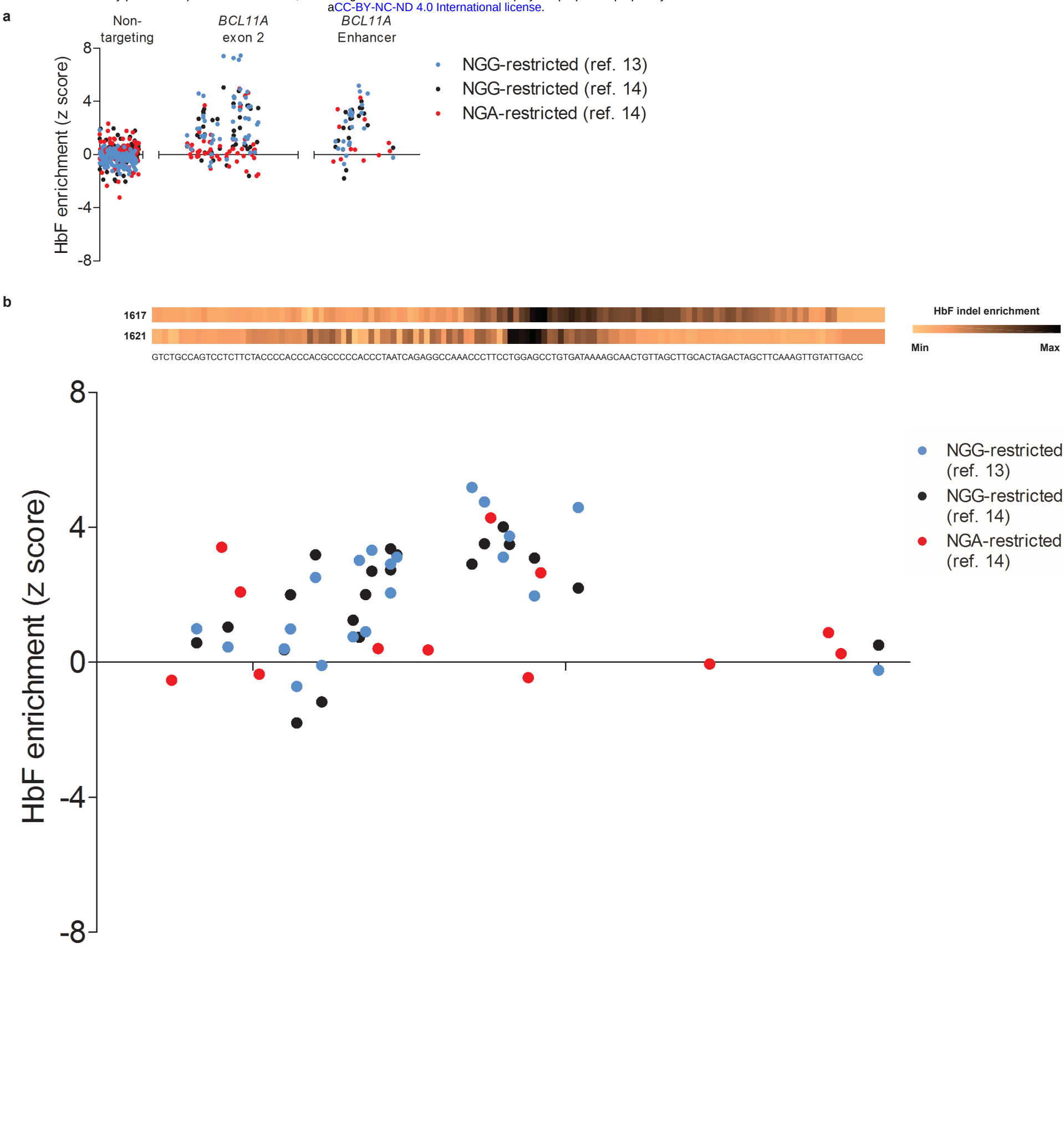


Figure 4