



# HHS Public Access

Author manuscript

*Nat Protoc.* Author manuscript; available in PMC 2017 October 01.

Published in final edited form as:

*Nat Protoc.* 2017 April ; 12(4): 828–863. doi:10.1038/nprot.2017.016.

## Genome-scale CRISPR-Cas9 Knockout and Transcriptional Activation Screening

**Julia Joung**<sup>1,2,3,4,\*</sup>, **Silvana Konermann**<sup>2,3,4,\*</sup>, **Jonathan S. Gootenberg**<sup>2,3,4,5</sup>, **Omar O. Abudayyeh**<sup>2,3,4,6</sup>, **Randall J. Platt**<sup>2,3,4</sup>, **Mark D. Brigham**<sup>2,3,4</sup>, **Neville E. Sanjana**<sup>2,3,4</sup>, and **Feng Zhang**<sup>1,2,3,4,†</sup>

<sup>1</sup>Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

<sup>2</sup>Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA

<sup>3</sup>McGovern Institute for Brain Research at MIT, Cambridge, Massachusetts 02139, USA

<sup>4</sup>Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>5</sup>Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA

<sup>6</sup>Department of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

### Abstract

Forward genetic screens are powerful tools for the unbiased discovery and functional characterization of specific genetic elements associated with a phenotype of interest. Recently, the RNA-guided endonuclease Cas9 from the microbial CRISPR (clustered regularly interspaced short palindromic repeats) immune system has been adapted for genome-scale screening by combining Cas9 with pooled guide RNA libraries. Here we describe a protocol for genome-scale knockout and transcriptional activation screening using the CRISPR-Cas9 system. Custom- or ready-made guide RNA libraries are constructed and packaged into lentiviral vectors for delivery into cells for screening. As each screen is unique, we provide guidelines for determining screening parameters and maintaining sufficient coverage. To validate candidate genes identified from the screen, we further describe strategies for confirming the screening phenotype as well as genetic perturbation through analysis of indel rate and transcriptional activation. Beginning with library design, a genome-scale screen can be completed in 9–15 weeks followed by 4–5 weeks of validation.

---

<sup>†</sup>Correspondence should be addressed to zhang@broadinstitute.org (F.Z.).

<sup>\*</sup>These authors contributed equally to this work.

**Author contributions** J.J., S.K., J.S.G., O.O.A., R.J.P., M.D.B., N.S., and F.Z. designed and performed the experiments. J.J., S.K., and F.Z. wrote the manuscript with help from all authors.

**Competing financial interests** The authors declare competing financial interests. F.Z. is a founder and scientific advisor for Editas Medicine and a scientific advisor for Horizon Discovery.

## INTRODUCTION

Systematic and high-throughput genetic perturbation technologies within live model organisms are necessary for fully understanding gene function and epigenetic regulation<sup>1-3</sup>. Forward genetic screens allow for a “phenotype-to-genotype” approach to mapping specific genetic perturbations to a phenotype of interest. Generally, this involves perturbing many genes at once, selecting cells or organisms for a desired phenotype, and then sequencing the perturbations to identify the genetic features involved in the phenotypic change. Early screening approaches relied on chemical DNA mutagens to induce genetic changes, but this process was inefficient and mutations were costly to identify. More recently, tools that utilize the RNA interference (RNAi) pathway, specifically short hairpin RNAs (shRNAs)<sup>4-7</sup>, to perturb transcript levels have revolutionized screening approaches<sup>8-13</sup>. ShRNAs exploit the endogenous RNAi machinery to knock down sequence-complementary mRNAs (Fig. 1). Despite the contribution of RNAi screens to many biological advances, this approach is hampered by incomplete knockdown of transcripts and high off-target activity, resulting in low signal-to-noise and limited interpretations<sup>14-16</sup>.

### Cas9 as a tool for precise genome editing

Programmable nucleases have emerged as a promising new genetic perturbation technology capable of precisely recognizing and cleaving target DNA<sup>17-19</sup>. Particularly, the RNA-guided endonuclease Cas9 from the microbial CRISPR (clustered regularly interspaced short palindromic repeat) immune system has proved powerful for precise DNA modifications<sup>20-25</sup>. Cas9 is guided to specific genomic targets by short RNAs that form Watson-Crick base pairs with the DNA. Thus Cas9 is easily retargetable.

Cas9 generates precise double-strand breaks (DSBs) at target loci that are repaired through either homology-directed repair (HDR) or, more often, non-homologous end-joining (NHEJ)<sup>26</sup>. HDR precisely repairs the DSB using a homologous DNA template, whereas NHEJ is error-prone and introduces indels. When Cas9 is targeted to a coding region, loss-of-function (LOF) mutations can occur as a result of frameshifting indels that produce a premature stop codon and subsequent nonsense-mediated decay of the transcript or generate a non-functional protein (Fig. 1)<sup>24,25</sup>. These features make Cas9 ideal for genome editing applications.

### Transcriptional activation with Cas9

In addition to generating LOF mutations, Cas9 can modulate transcription without modifying the genomic sequence through fusing catalytically inactive Cas9 (dCas9) to transcriptional activation and repression domains<sup>27-34</sup>. CRISPR activation (CRISPRa) and CRISPR inhibition (CRISPRi) can be achieved by direct fusion or recruitment of activation and repression domains, such as VP64 and KRAB, respectively<sup>29,35</sup>. CRISPRa in particular offers a significant improvement as a screening platform over other activation approaches. Previously gain-of-function (GOF) screens were primarily limited to cDNA overexpression libraries, which suffered from incomplete representation, overexpression beyond physiological levels and endogenous regulation, lack of isoform diversity, and high cost of construction. CRISPRa overcomes these limitations because it activates gene transcription at

the endogenous locus and simply requires the synthesis and cloning of RNA guides, making it much more affordable.

The first generation of CRISPRa fused dCas9 to a VP64 or p65 activation domain to produce modest transcriptional upregulation, the range of which was not suitable for genome-scale screening<sup>29,31–33,35</sup>. Second generation CRISPRa designs produced more robust upregulation by recruiting multiple activation domains to the dCas9 complex. For instance, SunTag recruits multiple VP64 activation domains via a repeating peptide array of epitopes paired with single-chain variable fragment antibodies<sup>28</sup>. Another activation method, VPR, uses a tandem fusion of three activation domains, VP64, p65, and Rta to dCas9 to enhance transcriptional activation<sup>27</sup>. We devised an alternative approach to CRISPRa that involved incorporating MS2 binding loops in the sgRNA backbone to recruit two different activation domains, p65 and HSF1, to a dCas9-VP64 fusion (Fig. 1)<sup>30</sup>. By recruiting three distinct transcriptional effectors, this synergistic activation mediator (SAM) complex could robustly and reliably drive transcriptional upregulation. A comparison of SunTag, VPR, and SAM across various cell types and species suggested that SAM induced more potent activation in some contexts, but further analysis is needed to determine which approach is most effective for GOF screening<sup>36</sup>.

### Applications of CRISPR-Cas9 screening

Together with large pooled single guide RNA (sgRNA) libraries, Cas9 can mediate high-throughput LOF and GOF dissection of many selectable phenotypes and elucidate complex biological questions. As a proof of principle to demonstrate the CRISPR-Cas9 system's utility for screening, we constructed genome-scale CRISPR-Cas9 knockout (GeCKO) and SAM libraries to identify genes that, upon knockout or activation, confer resistance to the BRAF-inhibitor vemurafenib in a melanoma cell line<sup>30,37</sup>. In addition to vemurafenib resistance, CRISPR-Cas9 screens have provided insight into the molecular basis of gene essentiality, drug and toxin resistance, the hypoxia response, and the role of flavivirus host factors in infection<sup>28,30,37–52</sup>. Although most screens have been performed in *in vitro* systems, the Cas9 system has also been applied *ex vivo* in dendritic cells to study the bacterial lipopolysaccharides response and *in vivo* to identify key factors that allow a non-metastatic lung cancer cell line to metastasize<sup>42,44</sup>. CRISPR-Cas9 screens have also been expanded to the noncoding genome through saturated mutagenesis by tiling sgRNAs across a noncoding locus to uncover functional elements in the *BCL11A* enhancer, *POU5F1* locus, and *CUL3* locus, as well as p53 and ESR1 transcription factor binding sites<sup>53–57</sup>. For reference, we have compiled a table of previously published screens (Table 1).

Here we explain in detail how to set up and perform pooled genome-scale knockout and transcriptional activation screens using Cas9. We describe protocols for designing and cloning an sgRNA library, packaging lentiviral vectors for transduction, analyzing screening results, and validating candidate genes identified from the screen (Fig. 2). Although we specifically focus on knockout and activation screening using the GeCKO and SAM systems, the protocol can be applied to other types of screens (e.g. other CRISPRa systems, Cas9 knockdown, and saturated mutagenesis).

## Comparison with other pooled screening technologies

Although both shRNA and Cas9 are easily retargetable, extensive characterization has shown that Cas9 is much more robust and specific than shRNA<sup>37,58–61</sup>. A comparison of GeCKO with shRNA screening indicated that for guides targeting the same gene, GeCKO results were more consistent and had higher validation rates<sup>37</sup>. For LOF screening to identify essential genes, Cas9 knockout screening has been shown to be more consistent and effective than shRNA screening<sup>61</sup>. Cas9 knockout screening consistently identified more lethal genes than shRNA, indicating a lower false negative rate and suggesting that identification of cellular dependencies may require complete gene inactivation<sup>62</sup>.

At the same time, however, in amplified regions of cancer cell genomes, DSBs generated by the Cas9 nuclease can produce gene-independent DNA damage phenotypes and false positive results<sup>45,62,63</sup>. Cas9 knockout screens that target the 5' exons of candidate genes may produce in-frame variants that retain functionality and obscure genetic dependencies, a limitation that may be overcome by targeting functional protein domains<sup>43</sup>. Furthermore, there is evidence that Cas9 and shRNA screens identify non-overlapping gene sets and may have different false positive or false negative results<sup>64</sup>. While Cas9 knockout screening is an enormously valuable LOF screening method, shRNA and CRISPRi screening may complement Cas9 knockout screening to further elucidate complex biological questions.

## Experimental Design

**Screening strategies**—In general, there are two formats for conducting a screen: arrayed and pooled. For arrayed screens, individual reagents are aliquoted into separate wells in multi-well plates. This format allows for a diverse range of measured phenotypes such as fluorescence, luminescence, or even direct imaging of cellular phenotypes<sup>2,65–67</sup>, but it is costly and time-consuming. An alternative format, and one that has been widely used for Cas9-based screens, is pooled screening in which pooled lentiviral libraries are transduced at a low multiplicity of infection (MOI) to ensure that most cells receive only one stably-integrated RNA guide. After the screen is complete, deep sequencing of the sgRNAs in the bulk genomic DNA identifies changes in the sgRNA distribution due to the applied screening selection pressure. As a result, pooled screens are less expensive and time-intensive than arrayed screens, but are generally limited to growth phenotypes or to fluorescence-activated cell sorting (FACS)-selectable phenotypes. These can be categorized as positive (e.g. resistance to a drug, toxin, or pathogen), negative (e.g. essential genes, toxicity), or marker gene selection (e.g. reporter gene expression) (Box 1).

Regardless of the type of screening selection, NGS is used to compare the number of reads for each sgRNA in the perturbed experimental condition relative to a control to identify candidate genes for validation. For positive and negative selection screens, the experimental and control conditions may be the same infection replicate treated with drug and vehicle respectively or analyzed at two different time points. For marker gene selection screens, during FACS the cells with the highest and lowest marker gene expressions are selected for the experimental and control conditions.

**Design and selection of the sgRNA library:** Although each sgRNA library is computationally designed for a specific purpose, the basic design process is consistent across libraries. First, the genomic regions of interest for targeting the sgRNA library are identified based on known sgRNA targeting rules (e.g. 5' conserved exons for gene knockout, upstream or downstream of the transcriptional start site for transcriptional activation or repression respectively). Second, all possible sgRNA targets with the Cas9 ortholog-specific protospacer adjacent motif (PAM) are identified and selected based on four criteria: (i) minimization of off-target activity, (ii) maximization of on-target activity, (iii) avoidance of homopolymer stretches (e.g. AAAA, GGGG), and (iv) GC content. Recent work has begun to elucidate the features that govern sgRNA specificity and efficiency<sup>41,47</sup>. Although specificity and efficiency will likely vary across experimental settings, false positive sgRNAs in screens can still be mitigated by including redundant sgRNAs in the library and requiring multiple distinct sgRNAs targeting the same gene to display the same phenotype when identifying screening hits. Once the targeting sgRNAs have been chosen, additional non-targeting guides that do not target the genome should be included as negative controls. Non-targeting guides are critical for evaluating the noise and success of a screen. At the end of the screen, top hit-targeting guides in the experimental condition should be significantly enriched or depleted compared to those in the control condition, while the non-targeting guides should remain relatively unchanged between experimental and control conditions.

We provide several genome-scale libraries for knockout and activation screening through Addgene (see REAGENTS). For knockout screening, the GeCKO v2 libraries target the 5' conserved coding exons of 19,050 human or 20,611 mouse coding genes with 6 sgRNAs per gene (Fig. 3a)<sup>68</sup>. In addition to targeting coding genes, the GeCKO v2 libraries also target 1,864 human miRNAs or 1,175 mouse miRNAs with 4 sgRNAs per miRNA. Each species-specific library contains 1,000 non-targeting control sgRNAs. The GeCKO library is available in a 1 vector (lentiCRISPR v2) or 2 vector (lentiCas9-Blast and lentiGuide-Puro) format (Fig. 3b). For activation screening, the SAM libraries target the 200bp region upstream of the transcriptional start site of 23,430 human or 23,439 mouse RefSeq coding isoforms with 3 sgRNAs per isoform (Fig. 3c)<sup>30</sup>. The library has to be combined with additional SAM effectors in a 2 vector (lentiSAM v2 and lenti MS2-P65-HSF1\_Hygro) or 3 vector (lenti dCAS-VP64\_Blast, lenti MS2-P65-HSF1\_Hygro, and lenti sgRNA(MS2)\_Puro or lenti sgRNA(MS2)\_Zeo) format (Fig. 3d). Both GeCKO v2 and SAM libraries prioritize sgRNAs with minimal off-target activity.

For designing custom libraries, we have provided a python script (Supplementary Data 1) that generates sgRNAs targeting a set of genes at specified genomic regions. The script prioritizes sgRNAs with fewer potential off-target sites by accounting for position and distribution of mismatches between the candidate spacer sequence and similar sites in the genome<sup>59</sup>. One can readily adapt this python script to design libraries for different genomes, nucleases, or regions of interest, such as noncoding regions for saturated mutagenesis screening or protein functional domains. In cases when a subset of genes is known to be involved in the screening phenotype and/or when the cell number is limited, one can consider performing a targeted screen that captures a subset of the genes in the genome-scale

screens provided. We have included a python script (Supplementary Data 2) for isolating the sgRNA target sequences corresponding to the genes in the targeted screen and adding flanking sequences for cloning. Additionally, one can consider adapting the sgRNA library plasmid backbone to the needs of the screen. For instance, when screening *in vivo* in complex tissues, one can use a cell-type specific promoter to ensure that only the cell type of interest is perturbed. To select for successful transduction by FACS, one can replace the antibiotic selection marker with a fluorescent marker. For these situations, we provide below a protocol for cloning a custom sgRNA library.

**Approaches for sgRNA library construction and delivery:** Throughout the sgRNA library cloning and amplification process, it is important to minimize any potential bias that may affect screening results. For example, the number of PCR cycles in the initial amplification of the pooled oligo library synthesis should be limited to prevent introducing bias during amplification. Scale each step of the cloning procedure provided according to the size of the library to reduce loss of sgRNA representation. After sgRNA library transformation, limit the growth time to avoid intercolony competition which can result in plasmid amplification bias. Here we provide a protocol and accompanying python script (Supplementary Data 3) for assessing sgRNA library distribution by next-generation sequencing (NGS) prior to screening.

Depending on the desired application, the sgRNA library can be delivered with lentivirus, retrovirus, or adeno-associated virus (AAV). Lentivirus and retrovirus integrate into the genome, whereas AAV does not integrate and thus for screening, AAV delivery is limited to non-dividing cells. In contrast, retrovirus only transduces dividing cells. In addition, AAV has a smaller insert size capacity compared to lentivirus and retrovirus. As a result, to date most of the screens have relied on lentiviral delivery and we have provided two methods for lentivirus production and transduction.

**Selection:** Since the parameters for each screen differ according to the screening phenotype, in lieu of providing a protocol for screening selection we have outlined general considerations for setting the relevant screening parameters as well as technical advice for carrying out a screening selection in Box 2. Additional *in vivo* screening considerations are described in Box 3. We also provide guidelines for saturated mutagenesis screening design and analysis in Box 4.

**Analysis of screening results:** For examples of anticipated results, we provide data from genome-scale knockout and transcriptional activation screening for genes that confer BRAF inhibitor vemurafenib (PLX) resistance in a BRAF<sup>V600E</sup> (A375) cell line<sup>30,37</sup>. As a result of the screening selection pressure, at the end of a successful screen the sgRNA library distribution in the experimental condition should be significantly skewed compared to the baseline and control conditions, with some sgRNAs enriched and others depleted (as assayed by NGS) (Fig. 4a,b). The targeting sgRNA representation should be significantly skewed compared to the non-targeting sgRNA representation, which should be relatively unchanged. In addition, the relative enrichment or depletion of sgRNAs between the experimental and control conditions should correlate between different infection replicates. Depending on the type of screen (positive, negative, or marker gene selection), the



enrichment or depletion of sgRNAs will be used to identify candidate genes that confer the screening phenotype.

Screening analysis methods such as RNAi gene enrichment ranking (RIGER), redundant siRNA activity (RSA), Model-based Analysis of Genome-wide CRISPR/Cas9 Knockout (MAGeCK), and STARS typically select candidate genes with multiple enriched or depleted sgRNAs to reduce the possibility that the observed change in sgRNA distribution was due to off-target activity of a single sgRNA<sup>47,69–71</sup>. RIGER ranks sgRNAs according to their enrichment or depletion and for each gene, examines the positions of the sgRNAs targeting that gene in the ranked sgRNA list<sup>69</sup>. The algorithm then assesses whether the set of positions is biased towards the top of the list using a Kolmogorov-Smirnov statistic and calculates an enrichment score and gene ranking based on a permutation test. RSA is similar to RIGER, except that it assigns statistical significance based on an iterative hypergeometric distribution formula<sup>70</sup>. Another screening analysis method, MAGeCK, assesses the statistical significance of sgRNA rankings using the negative binomial model before identifying positively and negatively selected genes and pathways using the robust ranking aggregation algorithm<sup>71</sup>. STARS scores genes using the probability mass function of a binomial distribution and generates false-discovery rates<sup>47</sup>.

These screening analysis methods can be adapted for noncoding screens by partitioning the noncoding region into smaller sections and assigning sgRNAs to each section. Since indels can be of varying lengths, sections with consistent enrichment or depletion of multiple sgRNAs indicate the presence of a potentially functional regulatory element. In this protocol we describe in detail how to identify candidate genes using RIGER. Each candidate gene identified from the screening analysis should have multiple significantly enriched or depleted sgRNAs in the experimental condition relative to the control (Fig. 4c,d). The RIGER *P* values of the candidate genes should also be significantly lower than the rest of the genes (Fig. 4e,f).

**Validation of candidate genes:** Given that the screening process can be noisy and the analysis produces a ranked list of candidate genes, it is necessary to verify that perturbation of the identified candidate genes confers the phenotype of interest. For validation, each of the sgRNAs that target the candidate gene can be individually cloned into the plasmid backbone of the sgRNA library and validated for the screening phenotype. In addition, the perturbation induced by each sgRNA, indel rate and transcriptional activation for knockout and activation screening, respectively, will be quantified to establish a phenotype-to-genotype relationship.

Indel rates can be detected either by the SURVEYOR nuclease assay or by NGS. Compared to SURVEYOR, which we have described previously<sup>72</sup>, NGS is more suitable for sampling a large number of sgRNA target sites and therefore described here. For measuring indel rates, it is important to design primers situated at least 50 bp from the target cleavage site to allow for the detection of longer indels. Our protocol for targeted NGS outlines a two-step PCR in which the first step uses custom primers to amplify the genomic region of interest and the second step uses universal, barcoded primers for multiplex deep sequencing on the Illumina platform. Relative to the one-step PCR method recommended for preparing sgRNA

libraries for NGS, the two-step PCR method is more versatile and less costly for assessing many different target sites because custom primers for each target site can be readily combined with different universal, barcoded primers.

After NGS, indel rates can be calculated by running the provided python script (Supplementary Data 4) that implements two different algorithms. The first aligns reads using the Ratcliff-Obershelp algorithm and then finds regions of insertion or deletion from this alignment<sup>73</sup>. The second method, adapted from the Geneious aligner, scans k-mers across the read and maps the alignment to detect indels<sup>74</sup>. In practice, the Ratcliff-Obershelp alignment algorithm is more accurate, while the k-mer based alignment algorithm is faster. These indel rates are then adjusted to account for background indel rates via a maximum likelihood estimation (MLE) correction<sup>59</sup>. The MLE correction models the observed indel rate as a combination of the true indel rate resulting from Cas9 cleavage and a separately measured background indel rate. The true indel rate is that which maximizes the probability of the observed read counts under the assumption that they obey a binomial distribution with the background rate.

Measurement of transcriptional activation usually entails isolation of RNA, reverse transcription of the RNA to cDNA, and quantitative PCR (qPCR). Various different methods have been described for each step of the process. In this protocol, we provide a method for reverse transcription followed by qPCR that is rapid, high-throughput, and cost-effective and thus ideal for quantifying fold upregulation for validation. Our method involves direct lysis of cells grown on a 96-well plate followed by reverse transcription and TaqMan qPCR. TaqMan-based detection is more specific and reproducible than SYBR-based detection because it relies on a fluorogenic probe specific to the target gene, whereas SYBR depends on a dsDNA-binding dye. TaqMan also allows for multiplexing with control probes that measure housekeeping gene expression as a proxy for total RNA concentration.

After validation of the screening phenotype and perturbation, we recommend verifying down- or up-regulation of protein expression for knockout or transcriptional activation screening respectively. Immunohistochemistry and Western blot are two of the most common methods for verifying protein expression. Immunohistochemistry requires fixing the validation cell lines and detecting the target protein using a specific antibody, whereas Western blot involves harvesting protein and separating by electrophoresis before staining with the specific antibody. Although immunohistochemistry provides additional information on protein localization, it often requires a more specific antibody than Western blot because proteins are not separated by size. Thus, Western blot is preferable for verifying protein expression of candidate genes.

## **MATERIALS**

### **REAGENTS**

#### **SgRNA libraries and backbones**

- lentiCRISPR v2 (Addgene, cat. no. 52961)
- lentiCas9-Blast (Addgene, cat. no. 52962)



- lentiGuide-Puro (Addgene, cat. no. 52963)
- lenti dCAS-VP64\_Blast (Addgene, cat. no. 61425)
- lenti MS2-P65-HSF1\_Hygro (Addgene, cat. no. 61426)
- lenti sgRNA(MS2)\_Zeo backbone (human; Addgene, cat. no. 61427)
- lenti sgRNA(MS2)\_Puro backbone (human; Addgene, cat. no. 73795)
- lenti sgRNA(MS2)\_Puro optimized backbone (mouse; Addgene, cat. no. 73797)
- lentiSAMv2 backbone (human; Addgene, cat. no. 75112)
- Human GeCKO v2 Library, 1 plasmid system (Addgene, cat. no. 1000000048)
- Human GeCKO v2 Library, 2 plasmid system (Addgene, cat. no. 1000000049)
- Mouse GeCKO v2 Library, 1 plasmid system (Addgene, cat. no. 1000000052)
- Mouse GeCKO v2 Library, 2 plasmid system (Addgene, cat. no. 1000000053)
- Human SAM Library, Zeo, 3 plasmid system (Addgene, cat. no. 1000000057)
- Human SAM Library, Puro, 3 plasmid system (Addgene, cat. no. 1000000074)
- Mouse SAM Library, Puro optimized, 3 plasmid system (Addgene, cat. no. 1000000075)
- Human SAM library, lentiSAMv2, 2 plasmid system (Addgene, cat. no. 1000000078)

#### Custom sgRNA library cloning

- Pooled oligo library (Twist Bioscience or CustomArray)
- PCR primers for amplifying oligo library for cloning are listed in Table 2. Primers longer than 60bp can be ordered as 4-nmol ultramers (Integrated DNA technologies)
- NEBNext High Fidelity PCR Master Mix, 2× (New England BioLabs, cat. no. M0541L) **Critical** To minimize error in amplifying oligos, it is important to use a high-fidelity polymerase. Other high-fidelity polymerases, such as PfuUltra II (Agilent) or Kapa HiFi (Kapa Biosystems), may be used as a substitute. For amplification of high-diversity libraries such as sgRNA libraries, we recommend the NEBNext High-Fidelity PCR Master Mix.
- UltraPure DNase/RNase-free distilled water (Thermo Fisher, cat. no. 10977023)
- QIAquick PCR Purification Kit (Qiagen, cat. no. 28104)
- QIAquick gel extraction kit (Qiagen, cat. no. 28704)
- UltraPure TBE buffer, 10x (Thermo Fisher, cat. no. 15581028)
- SeaKem LE agarose (Lonza, cat. no. 50004)
- SYBR Safe DNA stain, 10,000x (Thermo Fisher, cat. no. S33102)

- 1-kb Plus DNA ladder (Thermo Fisher, cat. no. 10787018)
- 50bp DNA ladder (Thermo Fisher, cat. no. 10416014)
- TrackIt Cyan/Orange Loading Buffer (Thermo Fisher, cat. no. 10482028)
- FastDigest Esp3I (BsmBI; Thermo Fisher, cat. no. FD0454)
- FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher, cat. no. EF0651)
- DTT, Molecular Grade (Promega, cat. no. P1171)
- Gibson Assembly Master Mix (New England BioLabs, cat. no. E2611L)
- GlycoBlue Coprecipitant (Thermo Fisher, cat. no. AM9515)
- 2-Propanol (Sigma-Aldrich, cat. no. I9516-25ML)
- Sodium chloride solution (Sigma-Aldrich, cat. no. 71386-1L)
- Tris-EDTA buffer solution (Sigma-Aldrich, cat. no. 93283-100ML)

### SgRNA plasmid amplification

- LB Agar, Ready-Made Powder (Affymetrix, cat. no. 75851)
- LB Broth, Ready-Made Powder (Affymetrix, cat. no. 75852)
- Ampicillin, 100 mg ml<sup>-1</sup>, sterile filtered (Sigma-Aldrich, cat. no. A5354)
- Endura ElectroCompetent Cells (Lucigen, cat. no. 60242) **Critical** High efficiency competent cells (>10<sup>10</sup> cfu μg<sup>-1</sup>) reduce the number of electroporations required for obtaining sufficient sgRNA library representation.
- NucleoBond Xtra Maxi EF (Macherey-Nagel, cat. no. 740424.10) **Critical** An endotoxin-free plasmid purification kit is important for avoiding endotoxicity in virus production and mammalian cell culture. Other endotoxin-free plasmid purification kits, such as the Qiagen Plasmid Plus Midi kit, may be used as substitutes.
- One Shot Stbl3 chemically competent *E. coli* (Thermo Fisher, cat. no. C737303)
- SOC outgrowth medium (New England BioLabs, cat. no. B9020S)
- NucleoBond Xtra Midi EF (Macherey-Nagel, cat. no. 740420.50) **Critical** An endotoxin-free plasmid purification kit is important for avoiding endotoxicity in virus production and mammalian cell culture. Other endotoxin-free plasmid purification kits, such as the Qiagen Plasmid Plus Maxi kit, may be used as substitutes.

### Next-generation sequencing

- Primers for amplifying the library for NGS are listed in Table 3. Primers for amplifying indels for NGS are listed in Table 4. Primers longer than 60bp can be ordered as 4 nmol ultramers (Integrated DNA technologies)

- KAPA HiFi HotStart ReadyMix, 2× (Kapa Biosystems, cat. no. KK2602)  
**Critical** To minimize error in amplifying oligos, it is important to use a high-fidelity polymerase. Other high-fidelity polymerases, such as PfuUltra II (Agilent) or Kapa HiFi (Kapa Biosystems), may be used as a substitute. For amplification of gDNA for indel analysis, we recommend the KAPA HiFi HotStart ReadyMix.
- Qubit dsDNA HS Assay Kit (Thermo Fisher, cat. no. Q32851)
- NextSeq 500/550 High Output Kit v2 (150 cycle; Illumina, cat. no. FC-404-2002)
- MiSeq Reagent Kit v3 (150 cycle; Illumina, cat. no. MS-102-3001)
- MiSeq Reagent Kit v2 (300 cycle; Illumina, cat. no. MS-102-2002)
- PhiX Control Kit v3 (Illumina, cat. no. FC-110-3001)
- Sodium hydroxide solution, 10 N (Sigma-Aldrich, cat. no. 72068-100ML)  
**Caution** Sodium hydroxide, 10 N is very hazardous in case of contact with skin, contact with eyes, ingestion, and inhalation. Wear protective clothing and avoid contact.
- Tris, pH 7.0 (Thermo Fisher, cat. no. AM9850G)

#### Mammalian cell culture

- HEK293FT (Thermo Fisher, cat. no. R70007) **Caution** Check cell lines regularly to ensure they are authentic and are not infected with mycoplasma.
- DMEM, high glucose, GlutaMAX Supplement, pyruvate (Thermo Fisher, cat. no. 10569010)
- Penicillin-streptomycin, 100× (Thermo Fisher, cat. no. 15140122)
- Fetal bovine serum, premium grade (VWR, cat. no. 97068-085)
- TrypLE Express, no phenol red (Thermo Fisher, cat. no. 12604021)
- HUES 66 cell line (Harvard Stem Cell Science) **Caution** Check cell lines regularly to ensure they are authentic and are not infected with mycoplasma.
- Geltrex LDEV-free reduced growth factor basement membrane matrix (Thermo Fisher, cat. no. A1413202)
- mTeSR1 medium (Stemcell Technologies, cat. no. 05850)
- Normocin (InvivoGen, cat. no. ant-nr-1)
- Rho-associated protein kinase (ROCK) inhibitor (Y-27632; Millipore, cat. no. SCM075)
- Accutase (Stemcell Technologies, cat. no. 07920)
- Dulbecco's PBS (DPBS; Thermo Fisher, cat. no. 14190250)

### Lentivirus production and titer

- Opti-MEM I reduced serum medium (Thermo Fisher, cat. no. 31985062)
- pMD2.G (Addgene, cat. no. 12259)
- psPAX2 (Addgene, cat. no. 12260)
- pcDNA3-EGFP transfection control plasmid (Addgene, cat. no. 13031)
- Lipofectamine 2000 transfection reagent (Thermo Fisher, cat. no. 11668019)
- PLUS Reagent (Thermo Fisher, cat. no. 11514015)
- Polyethylenimine HCl MAX, Linear, Mw 40,000 (PEI Max; Polysciences, cat. no. 24765-1)
- Polybrene (Hexadimethrine bromide; Sigma-Aldrich, cat. no. 107689-10G)
- Blasticidin S HCl (Thermo Fisher, cat. no. A1113903)
- Puromycin dihydrochloride (Thermo Fisher, cat. no. A1113803)
- Hygromycin B (Thermo Fisher, cat. no. 10687010)
- Zeocin (Thermo Fisher, cat. no. R25001)
- CellTiter-Glo Luminescent Cell Viability Assay (Promega, cat. no. G7571)

### Screening and Validation

- Quick-gDNA MidiPrep (Zymo Research, cat. no. D3100)
- DNA Binding Buffer (Zymo Research, cat. no. D4004-1-L)
- DNA Wash Buffer (Zymo Research, cat. no. D4003-2-24)
- DNA Elution Buffer (Zymo Research, cat. no. D3004-4-4)
- Ethyl alcohol, Pure (Sigma-Aldrich, cat. no. 459844-500ML)
- Primers for cloning the validation sgRNAs are listed in Table 5. (Integrated DNA technologies)
- T4 polynucleotide kinase (New England BioLabs, cat. no. M0201S)
- T4 DNA ligase reaction buffer, 10× (New England BioLabs, cat. no. B0202S)
- T7 DNA ligase with 2× rapid ligation buffer (Enzymatics, cat. no. L6020L)
- BSA, Molecular Biology Grade (New England BioLabs, cat. no. B9000S)
- QuickExtract DNA extraction solution (Epicentre, cat. no. QE09050)
- RNase AWAY (VWR, cat. no. 53225-514)
- Proteinase K (Sigma-Aldrich, cat. no. P2308-25MG)
- Tris, 1 M, pH 8.0 (Thermo Fisher, cat. no. AM9855G)
- Deoxyribonuclease I bovine (Sigma-Aldrich, cat. no. D2821-50KU)

- UltraPure 1 M Tris-HCl Buffer, pH 7.5 (Thermo Fisher, cat. no. 15567027)
- Calcium chloride solution (Sigma-Aldrich, cat. no. 21115-1ML)
- Glycerol (Sigma-Aldrich, cat. no. G5516-100ML)
- MgCl<sub>2</sub>, 1 M (Thermo Fisher, cat. no. AM9530G)
- Triton X-114 (Sigma-Aldrich, cat. no. X114-100ML)
- Proteinase K Inhibitor (EMD Millipore, cat. no. 539470-10MG)
- Dimethyl sulfoxide (Sigma-Aldrich, cat. no. D8418-50ML)
- Ethylene glycol-bis(2-aminoethylether)-N,N,N',N'-tetraacetic acid (Sigma-Aldrich, cat. no. E3889-10G)
- RevertAid RT Reverse Transcription kit (Thermo Fisher, cat. no. K1691)
- Oligo dT (TTTTTTTTTTTTTTTTTTTTNN; Integrated DNA Technologies)
- TaqMan target probes, FAM dye (Thermo Fisher)
- Taqman endogenous control probe, VIC dye (e.g. Human GAPD, GAPDH, Endogenous Control VIC®/MGB probe, primer limited; Thermo Fisher, cat. no. 4326317E)
- TaqMan Fast Advanced Master Mix, 2x (Thermo Fisher, cat. no. 4444557)

## EQUIPMENT

- Axygen 8-Strip PCR tubes (Fisher Scientific, cat. no. 14-222-250)
- Axygen PCR plates, 96 well (VWR, cat. no. PCR-96M2-HS-C)
- 384-well optical plate (e.g. Roche, LightCycler 480 Multiwell plate 384, cat. no. 5102430001)
- Axygen 1.5 ml Boil-Proof Microcentrifuge Tubes, (VWR, cat. no. 10011-702)
- Falcon tubes, polypropylene, 15 ml (Corning cat. no. 352097)
- Falcon tubes, polypropylene, 50 ml (Corning, cat. no. 352070)
- Filtered sterile pipette tips (e.g. Rainin)
- 100 mm × 15 mm Not TC-Treated Bacteriological Petri Dish (Corning, cat. no. 351029)
- 245mm Square BioAssay Dish without Handles, not TC-Treated Culture (Corning, cat. no. 431111)
- VWR Bacti Cell Spreaders (VWR, cat. no. 60828-688)
- AirPore Tape Sheets (Qiagen, cat. no. 19571)
- Nunc EasYFlask 25cm<sup>2</sup>, Filter Cap, 7 ml working volume (T25 flask; Thermo Scientific, cat. no. 156367)

- Nunc EasYFlask 75cm<sup>2</sup>, Filter Cap, 25 ml working volume, (T75 flask; Thermo Scientific, cat. no. 156499)
- Nunc EasYFlask 225 cm<sup>2</sup>, filter cap, 70 ml working volume (T225 flask; Thermo Scientific, cat. no. 159934)
- Corning bottle-top vacuum filter system, 0.22 μM (Sigma-Aldrich, cat. no. CLS431098)
- Stericup filter unit, 0.45 μM (Millipore, cat. no. SCHVU02RE)
- Syringe filter unit, 0.45 μM (Millipore, cat. no. SLHV013SL)
- Disposable Syringes with Luer-Lok Tip (Fisher Scientific, cat. no. 14-829-45)
- Falcon tissue culture plate, 6 wells (Corning, cat. no. 353224)
- Falcon tissue culture plate, 12 wells (Corning, cat. no. 353043)
- Falcon tissue culture dish, 100mm (Corning, cat. no. 353003)
- 96 well flat clear bottom black polystyrene TC-treated microplates (Corning, cat. no. 3904)
- BD BioCoat clear Poly-D-Lysine 96 well clear flat bottom TC-treated microplate (Corning, cat. no. 356461)
- Cellometer SD100 Counting Chambers (Nexcelom Bioscience, cat. no. CHT4-SD100-002)
- Zymo-Spin V with Reservoir (Zymo Research, cat. no. C1016-25)
- Collection Tubes, 2 ml (Zymo Research, cat. no. C1001-25)
- Amicon Ultra-15 Centrifugal Filter Unit with Ultracel-100 membrane (Millipore, cat. no. UFC910008)
- Thermocycler with programmable temperature stepping functionality, 96 well (e.g. Applied Biosystems Veriti, cat. no. 4375786)
- Real-time PCR system, 384 well (e.g. Roche Lightcycler 480 cat. no. 05015243001)
- Desktop microcentrifuges (e.g. Eppendorf, cat. nos. 5424 and 5804)
- Eppendorf ThermoStat C (Eppendorf, cat. no. 5383000019)
- Gene Pulser Xcell Microbial System (Bio-Rad, cat. no. 1652662)
- Digital gel imaging system (GelDoc EZ, Bio-Rad, cat. no. 170-8270), and blue sample tray (Bio-Rad, cat. no. 170-8273)
- Blue-light transilluminator and orange filter goggles (SafeImager 2.0; Invitrogen, cat. no. G6600)
- Gel quantification software (Bio-Rad, ImageLab or open-source ImageJ from the National Institutes of Health (NIH), USA, available at <http://rsbweb.nih.gov/ij/>)



- UV spectrophotometer (e.g. NanoDrop 2000c, Thermo Scientific)
- Plate spectrophotometer (e.g. Synergy H4 Hybrid Multi-Mode Microplate Reader, BioTek)
- Qubit Assay Tubes (Thermo Fisher, cat. no. Q32856)
- Qubit Fluorometer (Thermo Fisher, cat. no. Q33216)
- MiSeq System (Illumina, cat. no. SY-410-1003)
- NextSeq 500/550 System (Illumina, cat. nos. SY-415-1001 and SY-415-1002)
- Cell counter (e.g. Cellometer Image Cytometer, Nexcelom Bioscience)
- Sorvall Legend XTR Centrifuge (Thermo Fisher, cat. no. 75004520)
- Python 2.7 (<https://www.python.org/downloads/>)
- Twobitreader (<https://pypi.python.org/pypi/twobitreader>)
- Biopython (<http://biopython.org/DIST/docs/install/Installation.html>)
- Seqmap (<http://www-personal.umich.edu/~jianghui/seqmap/>)

## REAGENT SETUP

**TBE electrophoresis solution**—Dilute TBE buffer in distilled water to a 1× working condition, and store it at room temperature (18–22 °C) for up to 6 months.

**Ethanol, 80% (vol/vol)**—Prepare 80% (vol/vol) ethanol in UltraPure water right before use.

**D10 medium**—For culture of HEK 293FT cells, prepare D10 medium by supplementing DMEM with GlutaMAX and 10% (vol/vol) FBS. For routine cell line culture and maintenance, D10 can be further supplemented with 1× penicillin-streptomycin. Store the medium at 4 °C for up to 1 month.

**mTeSR1 medium**—For culture of human embryonic stem cells (hESCs), prepare mTeSR1 medium by supplementing it with the supplement supplied with the medium and 100 µg ml<sup>-1</sup> Normocin. Prepared medium can be stored at 4 °C for up to 2 months.

**Proteinase K, 300 U ml<sup>-1</sup>**—Resuspend 25 mg of Proteinase K in 2.5 ml of 10 mM Tris, pH 8.0 for 10 mg ml<sup>-1</sup> (300 U ml<sup>-1</sup>) of Proteinase K. Store at 4 °C for up to 1 year.

**Deoxyribonuclease I, 50 KU ml<sup>-1</sup>**—Resuspend 50 KU of Deoxyribonuclease I in a solution containing 50% (vol/vol) Glycerol, 10 mM CaCl<sub>2</sub>, and 50 mM Tris-HCl (pH 7.5) for 50 KU ml<sup>-1</sup> of Deoxyribonuclease I; as follows. Store at -20 °C for up to 2 years.

Component	Amount (µl)	Final concentration
Tris-HCl, pH 7.5, 1M	250	50 mM
CaCl <sub>2</sub> , 1M	50	10 mM

Glycerol	2500	50% (vol/vol)
UltraPure Water	2200	
Total	5000	

**RNA lysis buffer**—Prepare a solution with 9.6 mM Tris-HCl (pH 7.8), 0.5 mM MgCl<sub>2</sub>, 0.44 mM CaCl<sub>2</sub>, 10 μM DTT, 0.1% (wt/vol) Triton X-114, and 3 U ml<sup>-1</sup> Proteinase K in UltraPure water, as follows. The final pH of the solution should be approximately 7.8. Store at 4 °C for up to 1 year. **Critical** Prepare solution under RNase-free conditions.

Component	Amount (ml)	Final concentration
Tris, pH 8.0, 1M	1.2	4.8 mM
Tris, pH 7.5, 1M	1.2	4.8 mM
MgCl <sub>2</sub> , 1M	0.125	0.5 mM
CaCl <sub>2</sub> , 1M	0.110	0.44 mM
Dtt, 0.1M	0.025	10 μM
Proteinase K, 300 U ml <sup>-1</sup>	2.5	3 U ml <sup>-1</sup>
Triton X-114, 10% wt/vol	2.5	0.1% (wt/vol)
UltraPure Water	242	
Total	250	

**EGTA, 0.5 M, pH 8.3**—Resuspend EGTA in UltraPure water and adjust the pH of the solution to 8.3 with NaOH, 10 N, as follows. **Critical** EGTA is light sensitive, and can be stored at 4 °C protected from light for up to 2 years. Prepare solution under RNase-free conditions. Take aliquots to measure pH in order to keep main stock from being RNase-contaminated by the pH probe.

Component	Amount	Final Concentration
EGTA	9.5 g	0.5 M
Tris, pH 8.0, 1M	3.125 ml	0.0625 M
NaOH, 10N	6.1 ml	1.22 M
UltraPure Water	to 50 ml	
Total	50 ml	

**RNA lysis stop solution**—Resuspend 10 mg of Proteinase K Inhibitor in 150 μl of DMSO for a final concentration of 100 mM. Combine with EGTA, 0.5 M, pH 8.3 for a final solution with 1 mM Proteinase K inhibitor, 90 mM EGTA, and 113 μM DTT in UltraPure water, as follows. Aliquot into 8-strip PCR tubes to avoid freeze-thaw and facilitate sample processing with multichannel pipettes. Store at -20 °C for up to 1 year. **Critical** Prepare solution under RNase-free conditions.

Component	Amount (ml)	Final Concentration
Proteinase K inhibitor, 100 mM	0.150	1 mM

EGTA, 0.5 M, pH 8.3	2.694	90 mM
Dtt, 0.1 M	0.017	113 $\mu$ M
UltraPure Water	12.14	
Total	15	

---

**Oligo dT, 100  $\mu$ M**—Resuspend oligo dT to 100  $\mu$ M in UltraPure water. Aliquot and store at  $-20$  °C for up to 2 years.

## EQUIPMENT SETUP

**Large LB agar plates (245 mm square bioassay dish, ampicillin)**—Reconstitute the LB Broth with agar at a concentration of  $35 \text{ g L}^{-1}$  in deionized water and swirl to mix. Autoclave to sterilize. Allow the LB agar to cool to  $55$  °C before adding ampicillin to a final concentration of  $100 \mu\text{g ml}^{-1}$  and swirl to mix. On a sterile bench area, pour  $\sim 300$  ml of LB agar per 245 mm square bioassay dish. Place the lids on the plates and allow them to cool for 30–60 min until solidified. Invert the plates and let sit for several more hours or overnight. Agar plates can be stored in plastic bags or sealed with parafilm at  $4$  °C for up to 3 months.

**Standard LB agar plates (100 mm Petri dish, ampicillin)**—Preparation of standard LB agar plates is similar to large LB agar plates, except pour  $\sim 20$  ml of LB agar per 100 mm Petri dish. Store at  $4$  °C for up to 3 months.

## PROCEDURE

### Designing a custom sgRNA library o TIMING 3–5 w; 1 w hands-on

- 1 Construct a pooled sgRNA library by designing and cloning a custom sgRNA library (Steps 1–17) or amplifying a ready-made library from Addgene (Skip to Step 18). We have provided python scripts for designing a library that targets any set of genomic coordinates (option A, Supplementary Data 1) or a subset of an existing library (option B, Supplementary Data 2).
  - A. Generation of a library targeting a custom set of genomic coordinates
    - i. *Install requirements for library generation python script.* The python script `design_library.py` generates sgRNAs that target a set of specified genomic coordinates (Supplementary Data 1). Install python 2.7, twobitreader, biopython, and seqmap. For seqmap, install the version 1.0.13 source code for all platforms and compile with `g++ -O3 -m64 -o seqmap match.cpp`. Place seqmap in the same folder as the python script `design_library.py`.
    - ii. *Input target genomic coordinates for library design.* Once a set of genes and coordinates for the custom sgRNA library has been identified, prepare a target genes csv file containing the gene name, chromosome, start of the targeted region, and end of the targeted region in each column from left to right.

The target genes csv file should contain the headers name, chrom, start, and end. The python script provided will identify potential sgRNAs that target each gene within the genomic region as described in the target genes csv file. Refer to the table below for a sample input file:

name	chrom	start	end
EGFR	chr7	55086525	55086725
LPAR5	chr12	6745297	6745497
GPR35	chr2	241544625	241544825

- iii. *Design custom library.* Download the genome 2bit file that the target gene coordinates corresponds to from the UCSC Genome Browser (<http://hgdownload.cse.ucsc.edu/downloads.html>). The genome 2bit file will be used to construct a database of off-target scores based on the position and distribution of mismatches between each spacer sequence and similar sequences in the genome. For each region in the target genes csv file, the python script will identify potential sgRNAs and select a specified number of sgRNAs with fewer potential off-target sites using this database for the custom library. To design a custom library, run python `design_library.py` with the following optional parameters:

Flag	Description	Default
-o	Output csv file with names for target genes, spacer sequences, spacer orientations, chromosome locations, cleavage site locations, off-target scores, and oligo library sequences in columns from left to right	final_guides.csv
-i	Prefix of input genome 2bit file	hg19
-g	Target genes csv file with the gene name, chromosome, start of the targeted region, and end of the targeted region in columns from left to right	genes.csv
-gc	Minimum GC content required for an sgRNA spacer sequence	25
-s	Minimum spacing required between cleavage sites of sgRNAs targeting the same genomic region	20
-n	Maximum number of guides selected targeting each gene in the target genes csv file	3
-db	Use an existing off-target database constructed from a previous custom library design for a new library	False
-gecko or -sam	Specify the type of library and add the respective flanking sequences to the spacers for the oligo library synthesis	Neither

When designing sgRNAs against large genomic regions (>50kb), we recommend splitting the target genes csv file into several files containing different subsets of the target genes to parallelize the library design process and minimize run time. After running `design_library.py`, the spacers that target the specified genomic coordinates will be written to an output csv file. When designing a new custom library targeting the same genomic regions as a previous custom library, using the previously constructed off-target database can significantly reduce the script execution time. If `-gecko` or `-sam` is specified, the full oligo library sequence containing the spacers and respective flanking sequences for synthesis will be in the last column.

## B. Generation of a targeted library from an existing library

- i. *Input target genes for library design.* The python script `design_targeted_library.py` (Supplementary Data 2) extracts the sgRNA spacers from an existing library that target a specified set of genes. Install python 2.7 (<https://www.python.org/downloads/>). Once a set of genes for the targeted screen has been identified, prepare a csv file containing the names of the target genes with each line corresponding to one gene. Prepare another csv file for the annotated genome-scale library with the names of each gene in the first column and respective spacer sequences in the second column. Each line contains a different spacer sequence. The gene names in the target genes file should be in the same format as the names of the annotated library file.
- ii. *Design targeted custom library.* Isolate the subset of spacers from the genome-scale library that correspond to the target genes by running python `design_targeted_library.py` with the following optional parameters:

Flag	Description	Default
<code>-o</code>	Output csv file with names for target genes, corresponding spacer sequences, and oligo library sequences in columns from left to right	<code>oligos.csv</code>
<code>-l</code>	Annotated library csv file with names in the first column and corresponding spacer sequences in the second column	<code>annotated_library.csv</code>
<code>-g</code>	Target genes csv file with names of target genes	<code>target_genes.csv</code>
<code>-gecko</code> or <code>-sam</code>	Specify the type of library and add the respective flanking sequences to the spacers for the oligo library synthesis	Neither

After running `design_targeted_library.py`, the subset of spacers for the target genes will be written to an output csv file. If `-gecko` or `-sam` is specified, the full oligo library sequence containing the spacers and respective flanking sequences for synthesis will be in the last column.

- 2 Synthesize the oligo library as a pool on an array through a DNA synthesis platform such as Twist Bioscience or CustomArray. Synthesis typically requires 2–4 weeks depending on the size of the oligo library. Parafilm and store pooled oligos at  $-20\text{ }^{\circ}\text{C}$ .

### Cloning a custom sgRNA library o TIMING 2 d

- 3 *PCR amplification of pooled oligo library.* Throughout the sgRNA library cloning process, refer to the following table or the number of reactions recommended at each cloning step for a library size of 100,000 sgRNAs and scale the number of reactions according to the size of the custom sgRNA library:

Steps	Cloning process	Number of reactions
3–7	PCR amplification of pooled oligo library	12
8–10	Restriction digest of plasmid backbone	16
11–12	Gibson assembly	10 with sgRNA insert, 5 control
13–17	Isopropanol precipitation	10 with sgRNA insert, 5 control

Amplify the pooled oligo library from Step 2 using the Oligo-Fwd and Oligo-Rev primers (Table 2). Prepare a master mix using the reaction ratios outlined below:

Component	Amount per reaction ( $\mu\text{l}$ )	Final concentration
NEBNext High Fidelity PCR Master Mix, 2 $\times$	12.5	1 $\times$
Pooled oligo library template from Step 2	1	0.04 ng $\mu\text{l}^{-1}$
Oligo-Fwd primer (Universal)	1.25	0.5 $\mu\text{M}$
Oligo-Knockout-Rev or Oligo-Activation-Rev primer	1.25	0.5 $\mu\text{M}$
UltraPure water	9	
Total	25	

**Critical Step** To minimize error in amplifying oligos, it is important to use a high-fidelity polymerase. Other high-fidelity polymerases, such as PfuUltra II (Agilent) or Kapa HiFi (Kapa Biosystems), may be used as a substitute.

- 4 Aliquot the PCR master mix into 25  $\mu\text{l}$  reactions and perform a PCR using the following cycling conditions:

Cycle number	Denature	Anneal	Extend
1	98 $^{\circ}\text{C}$ , 30 s		



2–21	98 °C, 10 s	63 °C, 10 s	72 °C, 15 s
22			72 °C, 2 min

**Critical Step** Limit the number of PCR cycles to 20 cycles during amplification to reduce potential biases introduced during amplification.

- 5 After the reaction is complete, pool the PCR reactions and purify the PCR product using the QIAquick PCR purification kit according to the manufacturer's directions. Quantify the product by NanoDrop.
- 6 Run the PCR purified oligo library from Step 5 on a gel along with a 50bp ladder: cast a 2% (wt/vol) agarose gel in TBE buffer with SYBR Safe dye. Run half of the oligo library in the gel at 15 V cm<sup>-1</sup> for 45 min.

**Critical Step** Run on a 2% (wt/vol) agarose gel for long enough to separate the target library (140bp) from a possible primer dimer of ~120bp. Under the optimized PCR conditions suggested above the presence of primer dimers should be minimal.

- 7 Gel extract the purified PCR product using the QIAquick gel extraction kit according to the manufacturer's directions and quantify the final product by NanoDrop.
- 8 *Restriction digest of plasmid backbone.* Digest the desired library plasmid backbone with the restriction enzyme Esp3I (BsmBI) that cuts around the sgRNA target region. Refer to the master mix set up below for the reaction ratios:

Component	Amount per reaction (μl)	Final concentration
FastDigest Buffer, 10×	2	1×
Library Plasmid Backbone	1	50 ng μl <sup>-1</sup>
FastDigest Esp3I (BsmBI)	1	
FastAP Thermosensitive Alkaline Phosphatase	1	
DTT, 100 mM	0.2	1 mM
UltraPure water	14.8	
Total	20	

- 9 Aliquot 20 μl reactions from the master mix and incubate the restriction digest reaction at 37 °C for 1 h.
- 10 After the reaction has completed, pool the restriction digest reactions from Step 9 and run the entire pooled restriction digest reaction on a gel. Cast a 2% (wt/vol) agarose gel in TBE buffer with SYBR Safe dye and run the reaction in the gel at 15 V cm<sup>-1</sup> for 30 min. Gel extract the library plasmid backbone using the QIAquick gel extraction kit according to the manufacturer's protocol and quantify by NanoDrop. Note that the Gecko library backbones contain a 1880bp filler sequence which should be visible as a dropout. The SAM library

backbones do not contain a filler sequence and the expected dropout of 20bp is usually not readily visible.

- 11** *Gibson Assembly.* Set up a master mix for the Gibson reactions on ice according to the reaction ratios below. Be sure to include reactions without the sgRNA library insert as a control.

Component	Amount per reaction	Final concentration
Gibson Assembly Master Mix, 2×	10 $\mu$ l	1×
Digested library Plasmid Backbone from Step 10	330 ng	16.5 ng $\mu$ l <sup>-1</sup>
SgRNA library insert from Step 7 or UltraPure water control	50 ng	2.5 ng $\mu$ l <sup>-1</sup>
UltraPure water	To 20 $\mu$ l	
Total	20 $\mu$ l	

- 12** Aliquot 20  $\mu$ l reactions from the master mix and incubate the Gibson reaction at 50 °C for 1h.

**Pause Point** Completed Gibson reactions can be stored at –20 °C for at least 1 week.

- 13** *Isopropanol precipitation.* Pool cloning and control reactions separately. Purify and concentrate the sgRNA library by mixing the following:

Component	Amount per reaction ( $\mu$ l)	Final concentration
Gibson Assembly Reaction	20	
Isopropanol	20	
GlycoBlue Coprecipitant	0.2	0.075 $\mu$ g $\mu$ l <sup>-1</sup>
NaCl solution, 5 M	0.4	50 mM
Total	~40	

**Critical Step** In addition to concentrating the library, purification by isopropanol precipitation removes salts from the Gibson reaction that can interfere with electroporation.

- 14** Vortex and incubate at room temperature for 15 min and centrifuge at >15,000  $\times$  g for 15 min to precipitate the plasmid DNA. The precipitated plasmid DNA should appear as a small light blue pellet at the bottom of the microcentrifuge tube.
- 15** Aspirate the supernatant and gently wash the pellet twice without disturbing it using 1 ml of ice-cold (–20 °C) 80% (vol/vol) ethanol in UltraPure water.
- 16** Carefully remove any residual ethanol and air dry for 1 min.
- 17** Resuspend the plasmid DNA pellet in 5  $\mu$ l of TE per Gibson reaction by incubating at 55 °C for 10 min and quantify the custom sgRNA library by NanoDrop. Isopropanol-purified sgRNA libraries can be stored at –20 °C for several months.

**Amplification of pooled sgRNA library o TIMING 2 d**

- 18** *Pooled sgRNA library transformation.* Electroporate the library at 50–100 ng  $\mu\text{l}^{-1}$  using Endura ElectroCompetent cells according to the manufacturer's directions. If amplifying a ready-made genome-scale library from Addgene, repeat for a total of 1 electroporation per 10,000 sgRNAs in the library. If amplifying a custom sgRNA library, repeat for a total of 1 electroporation per 5,000 sgRNAs in the library and include an additional electroporation for the control Gibson reaction.
- 19** Pre-warm 1 large LB agar plate (245 mm square bioassay dish, ampicillin) per electroporation of the sgRNA library at 37 °C. Each large LB agar plate can be substituted with 10 standard LB agar plates. Pre-warm 1 standard LB agar plate (100 mm Petri dish, ampicillin) for calculating electroporation efficiency at 37 °C. For amplification of a custom sgRNA library, include an additional standard LB agar plate for the control Gibson reaction.
- 20** After the 1 h recovery period, pool electroporated cells and mix well by inverting.
- 21** Prepare a dilution for calculating transformation efficiency. To prepare the dilution mix, add 10  $\mu\text{l}$  of the pooled electroporated cells to 990  $\mu\text{l}$  of LB medium for a 100-fold dilution and mix well. Then add 100  $\mu\text{l}$  of the 100-fold dilution to 900  $\mu\text{l}$  of LB medium for a 1,000-fold dilution and mix well.
- 22** Plate 100  $\mu\text{l}$  of the 1,000-fold dilution onto a pre-warmed standard LB agar plate (100 mm Petri dish, ampicillin from Step 19). This is a 10,000-fold dilution of the full transformation that will be used to estimate the transformation efficiency.
- 23** If amplifying a custom sgRNA library, repeat Steps 21–22 for the control Gibson reaction.
- 24** To plate pooled electroporated cells, add 1 volume of LB medium to the pooled electroporated cells from Step 20, mix well, and plate on large LB agar plates (option A) or standard LB agar plates (option B).
- A.** Plating on large LB agar plates
    - i.** Plate 2 ml of electroporated cells on each of the pre-warmed large LB agar plates from Step 19 using a cell spreader. Spread the liquid culture until it is largely absorbed into the agar and does not drip when the plate is inverted. At the same time, make sure the liquid culture does not completely dry out as this will lead to poor survival.
  - B.** Plating on standard LB agar plates
    - i.** Alternatively, plate 200  $\mu\text{l}$  of electroporated cells on each of the pre-warmed standard LB agar plates from Step 19 using the same technique as described in Step 24 Ai.

**Critical Step** Plating the electroporated cells evenly is important for preventing intercolony competition that may skew the sgRNA library distribution.

- 25 Incubate all LB agar plates overnight at 37 °C for 12–14 h.

**Critical Step** Limiting the bacterial growth time to 12–14 h ensures that there is sufficient growth for sgRNA library amplification without potentially biasing the sgRNA library distribution through intercolony competition or differences in colony growth rates.

- 26 *Calculate electroporation efficiency.* Count the number of colonies on the 10,000-fold dilution plate. Multiply the number of colonies by 10,000 and the number of electroporations to obtain the total number of colonies on all plates. If amplifying a ready-made sgRNA library from Addgene, proceed only if the total number of colonies is greater than 100 colonies per sgRNA in the library. If amplifying a custom sgRNA library, proceed only if there are more than 500 colonies per sgRNA in the library.

**Critical Step** Obtaining a sufficient number of colonies per sgRNA is crucial for ensuring that the full library representation is preserved and that sgRNAs did not drop out during amplification.

#### **Troubleshooting**

- 27 In addition, for amplification of a custom sgRNA library, calculate the electroporation efficiency for the control Gibson reaction and proceed only if there are at least 20 times more colonies per electroporation in the sgRNA library condition compared to the control Gibson reaction.

#### **Troubleshooting**

- 28 *Harvest colonies from the LB agar plates.* Pipette 10 ml of LB medium onto each large LB agar plate or 1 ml of LB medium onto each standard LB agar plate. Gently scrape the colonies off with a cell spreader and transfer the liquid with scraped colonies into a 50-ml Falcon tube.
- 29 For each LB agar plate, repeat Step 28 for a total of 2 LB medium washes to capture any remaining bacteria.
- 30 Calculate the number of maxipreps needed by measuring the OD600 of the harvested bacterial suspension as follows: Number of maxipreps =  $OD600 \times (\text{total volume of suspension}) / 1200$ . Maxiprep the amplified sgRNA library by using the Macherey-Nagel NucleoBond Xtra Maxi EF kit according to the manufacturer's directions.

**Critical Step** Using an endotoxin-free plasmid purification kit is important for avoiding endotoxicity in virus production and mammalian cell culture. To ensure that the plasmid preparation is endotoxin-free, it is important to dilute the bacterial suspension to an OD600 within the linear range of the spectrophotometer, typically around 0.1–0.5, and measure the OD600 of the

dilution. Then multiply the OD600 by the dilution factor to obtain the OD600 of the bacterial suspension. Approximately 1 maxiprep is needed for 2 densely plated large LB agar plates.

- 31 Pool the resulting plasmid DNA and quantify by NanoDrop. Maxipreped sgRNA library can be aliquoted and stored at  $-20\text{ }^{\circ}\text{C}$ .

### Next-generation sequencing of the amplified sgRNA library to determine sgRNA distribution o TIMING 3–5 d

- 32 *Library PCR for NGS.* We have provided NGS primers that amplify the sgRNA target region with Illumina adapter sequences (Table 3). To prepare the sgRNA library for NGS, set up a reaction for each of the 10 NGS-Lib-Fwd primers and 1 NGS-Lib-KO-Rev or NGS-Lib-SAM-Rev barcode primer as follows:

Component	Amount per reaction ( $\mu\text{l}$ )	Final concentration
NEBNext High Fidelity PCR Master Mix, 2 $\times$	25	1 $\times$
Pooled sgRNA library template from Step 31	1	0.4 ng $\mu\text{l}^{-1}$
NGS-Lib-Fwd primer (unique)	1.25	0.25 $\mu\text{M}$
NGS-Lib-KO-Rev or NGS-Lib-SAM-Rev primer (barcode)	1.25	0.25 $\mu\text{M}$
UltraPure water	21.5	
Total	50	

**Critical step** Using a different reverse primer with a unique barcode for each library allows for pooling and sequencing of different libraries in a single NextSeq or HiSeq run. This is more efficient and cost-effective than running the same number of libraries on multiple Miseq runs.

**Critical Step** To minimize error in amplifying sgRNAs, it is important to use a high-fidelity polymerase. Other high-fidelity polymerases, such as PfuUltra II (Agilent) or Kapa HiFi (Kapa Biosystems), may be used as a substitute.

- 33 Perform a PCR using the following cycling conditions:

Cycle number	Denature	Anneal	Extend
1	95 $^{\circ}\text{C}$ , 5 min		
2–13	98 $^{\circ}\text{C}$ , 20 s	60 $^{\circ}\text{C}$ , 15 s	72 $^{\circ}\text{C}$ , 15 s
14	72 $^{\circ}\text{C}$ , 1 min		

- 34 After the reaction is complete, pool the PCR reactions and purify the PCR product by using the QIAquick PCR purification kit according to the manufacturer's directions.
- 35 Quantify the purified PCR product and run 2  $\mu\text{g}$  of the product on a 2% (wt/vol) agarose gel. Successful reactions should yield a  $\sim 260$ – $270$ bp product for the knockout library and a  $\sim 270$ – $280$ bp product for the activation library. Gel extract using the QIAquick gel extraction kit according to the manufacturer's directions.

**Pause Point** Gel-extracted samples can be stored at  $-20^{\circ}\text{C}$  for several months.

- 36** Quantify the gel-extracted samples using the Qubit dsDNA HS Assay Kit according to the manufacturer's instructions.
- 37** Sequence the samples on the Illumina MiSeq or NextSeq according to the Illumina user manual with 80 cycles of read 1 (forward) and 8 cycles of index 1. We recommend sequencing with 5% PhiX control on the MiSeq or 20% PhiX on the NextSeq to improve library diversity and aiming for a coverage of  $>100$  reads per sgRNA in the library.
- 38** *Analyze sequencing data with count\_spacers.py.* We provide a python script for analyzing the NGS results for the sgRNA representation (Supplementary Data 3). Install python 2.7 (<https://www.python.org/downloads/>) and biopython (<http://biopython.org/DIST/docs/install/Installation.html>). Prepare a csv file containing the guide spacer sequences with each line corresponding to one sequence.
- 39** To determine the spacer distribution, run python count\_spacers.py with the following optional parameters:

Flag	Description	Default
-f	Fastq file containing NGS data for analysis	NGS.fastq
-o	Output csv file with guide spacer sequences in the first column and respective read counts in the second column	library_count.csv
-i	Input csv file with guide spacer sequences	library_sequences.csv
-no-g	Indicate absence of a guanine before the guide spacer sequence	guanine is present

After running count\_spacers.py, spacer read counts will be written to an output csv file. Relevant statistics including the number of perfect guide matches, non-perfect guide matches, sequencing reads without key, the number of reads processed, percentage of perfectly matching guides, percentage of undetected guides, and skew ratio will be written to statistics.txt. An ideal sgRNA library should have more than 70% perfectly matching guides, less than 0.5% undetected guides, and a skew ratio of less than 10.

**Critical Step** The human SAM libraries do not have a guanine before the guide spacer sequence, so make sure to run the script with the parameter -no-g when analyzing those libraries.

### Troubleshooting

#### Lentivirus production and titer o TIMING 1–2 w

- 40** *Perform an antibiotic kill curve.* Prior to lentivirus production and titer, we recommend performing a kill curve for the antibiotic used to select the sgRNA library and additional necessary components on the relevant cell line for screening. To do so, seed cells at 10% confluency in media containing a range of antibiotic concentrations typically used for selection.



- 41** Refresh media with antibiotic every 3 days. After 4–7 days, choose the lowest concentration of antibiotic sufficient to kill all cells.

**Critical Step** It is important to use the lowest concentration of antibiotic to avoid excessively stringent selection that biases selection for cells transduced with multiple sgRNAs.

- 42** *HEK 293FT maintenance.* Culture cells in D10 medium at 37 °C with 5% CO<sub>2</sub> and maintain according to the manufacturer's recommendation.

- 43** To passage, aspirate the medium and rinse the cells by gently adding 5 ml of TrypLE to the side of the T225 flask, so as not to dislodge the cells. Remove the TrypLE and incubate the flask for 4–5 min at 37 °C until the cells have begun to detach. Add 10 ml of warm D10 to the flask and dissociate the cells by pipetting them up and down gently, and transfer the cells to a 50-ml Falcon tube.

**Critical Step** We typically passage cells every 1–2 d at a split ratio of 1:2 or 1:4, never allowing cells to reach more than 70% confluency. For lentivirus production, we recommend using HEK 293FT cells with a passage number less than 10.

- 44** *Preparation of cells for transfection.* Seed the well-dissociated cells into T225 flasks 20–24 h before transfection at a density of  $1.8 \times 10^7$  cells per flask in a total volume of 45 ml D10 medium. Refer to the following table for the number of T225 flasks recommended for each plasmid construct:

Plasmid construct	Number of T225 flasks
sgRNA library	4
Additional Cas9 nuclease or activator components	2 per component
GFP transfection control	1

**Critical Step** Do not plate more cells than the recommended density, as doing so may reduce transfection efficiency.

- 45** *Lentivirus plasmid transfection.* Cells in T225 flasks from Step 44 are optimal for transfection at 80–90% confluency. We outline below a transfection method using Lipofectamine 2000 and PLUS reagent. Alternatively, we describe a cost-effective method for lentivirus transfection with Polyethylenimine (PEI) in Box 5. For each lentiviral target, combine the following lentiviral target mix in a 15-ml or 50-ml Falcon tube and scale up accordingly:

Component	Amount per T225 flask
Opti-MEM	2250 $\mu$ l
pMD2.G (lentiviral helper plasmid)	15.3 $\mu$ g
psPAX (lentiviral helper plasmid)	23.4 $\mu$ g
Lentiviral target plasmid	30.6 $\mu$ g

**Critical Step** Transfecting at the recommended cell density is crucial for maximizing transfection efficiency. Lower densities can result in Lipofectamine 2000 toxicity for cells, while higher densities can reduce transfection efficiency.

- 46 Prepare the PLUS reagent mix as follows and invert to mix:

Component	Amount per T225 flask
Opti-MEM	2250 $\mu$ l
PLUS reagent	297

- 47 Add the PLUS reagent mix to the lentiviral target mix, invert, and incubate at room temperature for 5 min.

- 48 Prepare the Lipofectamine reagent mix as follows and invert to mix:

Component	Amount per T225 flask
Opti-MEM	4500 $\mu$ l
Lipofectamine 2000	270

- 49 Add the lentiviral target and PLUS reagent mix to the Lipofectamine reagent mix, invert, and incubate at room temperature for 5 min.

- 50 Pipette 9 ml of the lentiviral transfection mix from Step 49 into each T225 flask from Step 44 and shake gently to mix. Return the T225 flasks to the incubator.

- 51 After 4 h, replace the medium with 45 ml of pre-warmed D10 medium. The constitutive GFP expression plasmid transfection control indicates transfection efficiency.

### Troubleshooting

- 52 *Harvest and store lentivirus.* 2 d after the start of lentiviral transfection, pool the lentivirus supernatant from the T225 flasks transfected with the same plasmid construct and filter out cellular debris using Millipore's 0.45  $\mu$ m Stericup filter unit.

**Pause Point** The filtered lentivirus supernatant can be aliquoted and stored at  $-80^{\circ}\text{C}$ . Avoid freeze-thawing lentivirus supernatant.

- 53 *Determine the lentiviral titer through transduction.* The CRISPR-Cas9 system has been used in a number of mammalian cell lines. Conditions may vary for each cell line. Lentiviral titer should be determined using the relevant cell line for the screen. Below we detail transduction conditions and calculation of viral titer for HEK 293FT cells (option A) and hESC HUES66 cells (option B).

A. Lentiviral transduction and titer for HEK 293FT cells by spinfection

- i. *HEK 293FT maintenance and passaging.* Refer to Steps 42–43.

- ii. *Preparation of cells for spinfection.* For each lentivirus, seed 6 wells of a 12-well plate at a density of  $3 \times 10^6$  cells in 2 ml D10 medium per well with  $8 \mu\text{g ml}^{-1}$  of polybrene. In each well, add 400  $\mu\text{l}$ , 200  $\mu\text{l}$ , 100  $\mu\text{l}$ , 50  $\mu\text{l}$ , 25  $\mu\text{l}$ , or 0  $\mu\text{l}$  of lentivirus supernatant from Step 52. Mix each well thoroughly by pipetting up and down.
- iii. Spinfect the cells by centrifuging the plates at  $1000 \times g$  for 2 h at 33 °C. Return the plates to the incubator after spinfection.
- iv. *Replating spinfection for calculation of viral titer.* 24 h after the end of spinfection, remove the medium, gently wash with 400  $\mu\text{l}$  TrypLE per well, add 100  $\mu\text{l}$  of TrypLE, and incubate at 37 °C for 5 min to dissociate the cells. Add 2 ml of D10 medium per well and resuspend the cells by pipetting up and down.
- v. Determine the cell concentration using the Cellometer Image Cytometer for the 0  $\mu\text{l}$  lentivirus supernatant condition.
- vi. For each virus condition, seed 4 wells of a 96-well clear bottom black tissue culture plate at a density of  $4 \times 10^3$  cells (from Step 53Aiv) based on the cell count determined in Step 53Av in 100  $\mu\text{l}$  of D10 medium. Add an additional 100  $\mu\text{l}$  of D10 medium with the corresponding selection antibiotic for the virus at an appropriate final concentration to 2 wells and 100  $\mu\text{l}$  of regular D10 medium to the other 2 wells.
- vii. 72–96 h after replating, when the no virus conditions contain no viable cells and the no antibiotic selection conditions are at 80–90% confluency, quantify the cell viability for each condition using CellTiter Glo according to the manufacturer's protocol.

**Critical Step** We have found that Cell Titer Glo can be diluted 1:4 in PBS to reduce cost while still achieving optimal results.

- viii. For each virus condition, calculate the multiplicity of infection (MOI) as the average luminescence, or viability, of the 2 wells in the condition with antibiotic selection divided by the average luminescence of the 2 wells in the condition without antibiotic selection. A linear relationship between lentivirus supernatant volume and MOI is expected at lower volumes, with saturation achieved at higher volumes.
- B.** Lentiviral transduction and titer for hESC HUES66 cells by mixing
- i. *HUES66 maintenance.* We routinely maintain HUES66 cells (a hESC cell line) in feeder-free conditions with mTeSR1

medium on GelTrex-coated tissue culture plates. To coat a 100-mm tissue culture dish, dilute cold GelTrex 1:100 in 5 ml of cold DMEM, cover the entire surface of the culture dish, and place the dish in the incubator for at least 30 min at 37 °C. Aspirate the GelTrex mix prior to plating. During passaging and plating, mTeSR1 medium is supplemented further with 10 µM ROCK inhibitor. Refresh the mTeSR1 medium daily.

- ii. *Passaging HUES66.* Aspirate the medium and rinse the cells once by gently adding 10 ml of DPBS to the side of the 100-mm tissue culture dish, so as not to dislodge the cells. Dissociate the cells by adding 2 ml of Accutase and incubate at 37 °C for 3–5 min until the cells have detached. Add 10 ml of DMEM, resuspend the dissociated cells, and pellet the cells at  $200 \times g$  for 5 min. Remove the supernatant, resuspend the cells in mTeSR1 medium with 10 µM ROCK inhibitor and replat the cells onto GelTrex-coated plates. Replace with normal mTeSR1 medium 24 h after plating.

**Critical Step** We typically passage cells every 4–5 d at a split ratio of 1:5 or 1:10, never allowing cells to reach more than 70% confluency.

- iii. *Preparation of cells for lentiviral transduction.* For each lentivirus, plate 6 wells of a Geltrex-coated 6-well plate at a density of  $5 \times 10^5$  cells in 2 ml mTeSR1 medium per well. In each well, add 400 µl, 200 µl, 100 µl, 50 µl, 25 µl, or 0 µl of lentivirus supernatant, fill to a total volume of 3 ml with DPBS, and supplement with 10 µM ROCK inhibitor. Plate an additional no antibiotic selection control well at the same seeding density without virus. Mix each well thoroughly by pipetting up and down.
- iv. 24 h after lentiviral transduction, replace the medium with mTeSR1 containing the relevant antibiotic selection. For the no antibiotic selection control well, replace the medium with normal mTeSR1 medium. Refresh the mTeSR1 medium with and without antibiotic selection every day until the plate is ready for the next step.
- v. *Calculation of viral titer.* 72–96 h after starting the antibiotic selection, when the no virus condition contains no viable cells and the no antibiotic selection control condition is 80–90% confluent, rinse the cells with 2 ml of DPBS, add 500 µl of Accutase, and incubate at 37 °C for 3–5 min to dissociate the cells. Add 2 ml of DMEM and mix well.

- vi. Count and record the number of cells in each well using the Cellometer Image Cytometer.
- vii. For each virus condition, calculate the MOI as the number of cells in the antibiotic selection condition divided by the number of cells in the no antibiotic selection control. A linear relationship between lentivirus supernatant volume and MOI is expected at lower volumes, with saturation achieved at higher volumes.

### Troubleshooting

**Lentiviral transduction and screening o TIMING 3–6 w—CRITICAL:** Skip to Step 56 if screening using a 1-vector format.

**54** *Generation of a cell line with stably expressed Cas9 components.* Prior to transducing the sgRNA library, transduce the relevant cell line with the additional Cas9 components that are not present in the sgRNA library backbone at an MOI < 0.7. If two additional Cas9 components are required, for instance dCas9-VP64 and MS2-p65-HSF1, both components can be transduced at the same time. Scale up as necessary to generate sufficient cells for maintaining sgRNA representation after sgRNA library transduction and selection. We have found that generation of a clonal line with Cas9 or SAM components is not necessary for successful screening. The cells can therefore be transduced and selected as a bulk population at the desired scale. Below we describe lentiviral transduction and cell line generation methods for HEK 293FT cells (option A) and hESC HUES66 cells (option B).

#### A. Generation of HEK 293FT cell lines

- i. Seed cells in 12-well plates at a density of  $3 \times 10^6$  cells in 2 ml D10 medium per well with  $8 \mu\text{g ml}^{-1}$  of polybrene. Add the appropriate volume of lentivirus supernatant from Step 52 to each well, and make sure to include a no virus control. Mix each well thoroughly by pipetting up and down.
- ii. Spinfect the cells by spinning the plates at  $1000 \times g$  for 2 h at  $33^\circ\text{C}$ . Return the plates to the incubator after spinfection.
- iii. 24 h after the end of spinfection, remove the medium, gently wash with  $400 \mu\text{l}$  TrypLE per well, add  $100 \mu\text{l}$  of TrypLE, and incubate at  $37^\circ\text{C}$  for 5 min to dissociate the cells. To each well, add 2 ml of D10 medium with the appropriate selection antibiotic for the lentivirus and resuspend the cells by pipetting up and down.
- iv. Pool the resuspended cells from the wells with virus and seed the cells into T225 flasks at a density of  $9 \times 10^6$  cells per flask in 45 ml of D10 medium with selection antibiotic.

- v. Transfer the resuspended cells from the no virus control into a T75 and add 13 ml of D10 medium with selection antibiotic.
- vi. Refresh the selection antibiotic every 3 d and passage as necessary for 4–7 d and until there are no viable cells in the no virus control.

**B. Generation of HUES66 cell lines**

- i. Seed cells in Geltrex-coated 6-well plates at a density of  $5 \times 10^5$  cells in of 2 ml mTeSR1 medium per well. Add the appropriate volume of lentivirus supernatant from Step 52 to each well, and make sure to include a no virus control. Fill up the total volume to 3 ml with DPBS, and supplement with 10  $\mu$ M ROCK inhibitor. Mix each well thoroughly by pipetting up and down.
- ii. 24 h after lentiviral transduction, replace the medium with mTeSR1 containing the relevant selection antibiotic. Refresh the mTeSR1 medium with selection antibiotic every day and passage as necessary for 4–7 d and until there are no viable cells in the no virus control.

**Critical Step** The lentiviral transduction method for generating a cell line for screening should be consistent with the method for titering the virus to ensure that cells are transduced at the appropriate MOI.

- 55** After selecting for successfully transduced cells, allow the cells to recover from selection by culturing in normal medium for 2–7 d before transducing with the sgRNA library. If culturing for more than 7 d after selection or after freezing cells, re-select the Cas9 component cell line with the appropriate selection antibiotic to ensure expression of the Cas9 components and allow the cells to recover before sgRNA library transduction.

**Pause Point** Cells can be frozen according to the manufacturer's protocol.

- 56** *Transduction of cells with the sgRNA library.* Repeat Steps 54–55 for lentiviral sgRNA library transduction at the appropriate MOI and selection of transduced cell lines. To ensure that most cells receive only one genetic perturbation, transduce the sgRNA library at an MOI < 0.3. Scale up the transduction such that the sgRNA library has a coverage of >500 cells expressing each sgRNA. For example, for a library size of 100,000 unique sgRNAs, transduce  $1.67 \times 10^8$  cells at an MOI of 0.3. After the appropriate selection for 4–7 days, the cells are ready for screening. For knockout screening, we have found that maximal knockout efficiency is achieved 7 days after sgRNA transduction and therefore recommend selecting for 7 days before starting the screen selection. In contrast, maximal SAM activation is achieved as early as 4 days after sgRNA transduction. If selection is complete based on the no virus control, gain-of-function screening can be started 5 days after transduction. We generally



recommend to perform 4 independent screening bioreps (i.e. 4 separate sgRNA library infections followed by separate screening selection). Multiple screening bioreps are critical for determining screening hits with a high rate of validation. Since the parameters of each screen depends on the screening phenotype of interest, we provide guidelines and technical considerations for the screening selection process (Box 2–4).

**Critical Step** It is important to aim for a coverage of >500 cells per sgRNA to guarantee that each perturbation will be sufficiently represented in the final screening readout. Increase the coverage as necessary if the screening selection pressure is not very strong or if performing a negative selection screen. Transducing the sgRNA library at an MOI < 0.3 ensures that most cells receive at most one genetic perturbation. Transducing at higher MOI's may confound screening results.

### Harvest genomic DNA for screening analysis o TIMING 5–7 d

- 57 *Harvest genomic DNA.* At the end of the screen, harvest genomic DNA (gDNA) from a sufficient number of cells to maintain a coverage of >500. For a library size of 100,000 unique sgRNAs, harvest gDNA from at least  $5 \times 10^7$  cells for downstream sgRNA analysis using the Zymo Research Quick-gDNA MidiPrep according to the manufacturer's protocol. Elution should be performed twice with 150–200  $\mu$ l each for maximum recovery of gDNA.

**Critical Step** Make sure to tighten the connection between the reservoir and the column and centrifuge at a sufficient speed and time to remove any residual buffer. Addition of a final dry spin is recommended to remove residual wash buffer.

**Pause Point** Frozen cell pellets or isolated gDNA can be stored at  $-20^\circ\text{C}$  for several months.

- 58 *Preparation of the gDNA for NGS analysis.* Refer to Steps 32–33 for how to amplify the sgRNA for NGS. Scale up the number of reactions such that all of the gDNA harvested from the screen is amplified. Each 50  $\mu$ l reaction can hold up to 2.5  $\mu$ g of gDNA. Barcoded NGS-Lib-Rev primers enable sequencing of different screening conditions and bioreps (i.e. experimental condition biorep 1 and control condition biorep 1) in a pooled sequencing run.

#### Troubleshooting

- 59 *Purification of amplified screening NGS library.* For large-scale PCR purification, we recommend using the Zymo-Spin V with Reservoir. Add 5 volumes of DNA Binding Buffer to the PCR reaction, mix well, and transfer to Zymo-Spin V with Reservoir in a 50 ml Falcon tube. Each Zymo-Spin V column can hold up to 12 ml.

**Critical Step** Make sure to tighten the connection between the reservoir and the Zymo-Spin V column.

- 60** Centrifuge at  $500 \times g$  for 5 min at room temperature. Discard the flow-through.
- 61** Add 2 ml of DNA Wash Buffer and centrifuge at  $500 \times g$  for 5 min at room temperature. Discard the flow-through and repeat for an additional wash.
- 62** Remove the reservoir from the Zymo-Spin V column and transfer the column to a 2 ml collection tube. In a microcentrifuge, spin at the maximum speed ( $>12,000 \times g$ ) for 1 min at room temperature to remove residual wash buffer.
- 63** Transfer the Zymo-Spin V column to a 1.5 ml microcentrifuge tube. Add 150  $\mu$ l of elution buffer, wait for 1 min, and spin at the maximum speed ( $>12,000 \times g$ ) for 1.5 min at room temperature to elute the purified PCR reaction.
- 64** Pool the purified PCR reactions and quantify. Refer to Steps 37–39 for NGS analysis of the sgRNA distribution. For screening NGS analysis, we recommend aiming for a coverage of  $>500$  reads per sgRNA in the library.
- 65** *Analysis of screening results with RNAi gene enrichment ranking (RIGER).* Prior to RIGER analysis, determine the sgRNA fold change due to screening selection. For each biorep of screening experimental or control condition, add a pseudocount of 1 to the NGS read count of each sgRNA and normalize by the total number of NGS read counts for that condition. To obtain the sgRNA fold change, divide the experimental normalized sgRNA count by the control and take the base 2 logarithm.
- 66** Prepare a RIGER input csv file with the column headers WELL\_ID, GENE\_ID, and biorep 1, biorep 2, etc from left to right. WELL\_ID is a list of sgRNA identification numbers, GENE\_ID the genes that the sgRNAs target, and biorep columns the sgRNA fold change. Each row contains a different sgRNA in the library.
- 67** RIGER is launched through GENE-E from the Broad Institute (<http://www.broadinstitute.org/cancer/software/GENE-E/download.html>). Start GENE-E and import the input csv file by navigating to File > Import > Ranked Lists. Click the table cell containing the first data row and column as instructed. Launch RIGER by going to Tools > RIGER. Adjust the RIGER settings to the following recommended values:

---

Number of permutations: 1,000,000

---

Method to convert hairpins to genes: Kolmogorov-Smirnov

---

Gene rank order: Positive to negative for positive selection screens; negative to positive for negative selection screens

---

Select adjust gene scores to accommodate variation in hairpin set size

---

Select hairpins are pre-scored

---

Hairpin Id: WELL\_ID

---

Convert hairpins to: GENE\_ID

---

- 68** Once the RIGER analysis has completed, export the gene rank dataset. Determine the top candidates genes based on either the overlap or the average ranking between the screening bioreps.

#### Validation of candidate genes for screening phenotype o TIMING 4–5 w

- 69** *Cloning validation sgRNAs into the plasmid backbone of the sgRNA library.* Design top and bottom strand primers for cloning the top 3 sgRNAs for each candidate gene individually into the plasmid backbone of the sgRNA library used for screening according to Table 5 as we have previously described<sup>72</sup>. Primers for cloning 2 non-targeting sgRNAs (NT1 and NT2) for control are also provided in Table 5.
- 70** Resuspend the top and bottom strand primers to a final concentration of 100  $\mu\text{M}$ . Prepare the following mixture for phosphorylating and annealing the top and bottom strand primers for each validation sgRNA:

Component	Amount ( $\mu\text{l}$ )	Final concentration
sgRNA-top, 100 $\mu\text{M}$	1	10 $\mu\text{M}$
sgRNA-bottom, 100 $\mu\text{M}$	1	10 $\mu\text{M}$
T4 ligation buffer, 10 $\times$	1	1 $\times$
T4 PNK	0.5	
UltraPure water	6.5	
Total	10	

- 71** Phosphorylate and anneal the primers in a thermocycler using the following conditions: 37  $^{\circ}\text{C}$  for 30 min; 95  $^{\circ}\text{C}$  for 5 min; ramp down to 25  $^{\circ}\text{C}$  at 5  $^{\circ}\text{C}$   $\text{min}^{-1}$ .
- 72** After the annealing reaction is complete, dilute the phosphorylated and annealed oligos 1:10 by adding 90  $\mu\text{l}$  of UltraPure water.
- Pause Point** Annealed oligos can be stored at  $-20^{\circ}\text{C}$  for at least 1 week.
- 73** Clone the annealed sgRNA inserts into the sgRNA library backbone by setting up a Golden Gate assembly reaction for each sgRNA. We have found that when cloning many sgRNAs, Golden Gate assembly is efficient and offers a high cloning success rate. Mix the following for each sgRNA:

Component	Amount ( $\mu\text{l}$ )	Final concentration
Rapid Ligase Buffer, 2 $\times$	12.5	1 $\times$
FastDigest Esp3I (BsmBI)	1	
DTT	0.25	1 mM
BSA, 20 $\text{mg ml}^{-1}$	0.125	0.1 $\text{mg ml}^{-1}$
T7 ligase	0.125	
Diluted oligo duplex from Step 72	1	0.04 $\mu\text{M}$
sgRNA library backbone	1	1 $\text{ng } \mu\text{l}^{-1}$

UltraPure water	9
Total	25

**Critical Step** We recommend using FastDigest Esp3I (Fermentas) as we have had reports of Esp3I from other vendors not working as efficiently in the Golden Gate assembly reaction setup described. It is not necessary to perform a negative control (no insert) Golden Gate assembly reaction as it will always contain colonies and therefore is not a good indicator of cloning success.

- 74** Perform a Golden Assembly reaction using the following cycling conditions:

Cycle number	Condition
1–15	37 °C for 5 min, 20 °C for 5 min

**Pause Point** Completed Golden Gate assembly reactions can be stored at –20 °C for at least 1 week.

- 75** *Transformation and midiprep.* Transform the Golden Gate assembly reaction into a competent *E. coli* strain, according to the protocol supplied with the cells. We recommend the Stb13 strain for quick transformation. Thaw the chemically competent Stb13 cells on ice, add 2 µl of the product from Step 74 into ice-cold Stb13 cells, and incubate the mixture on ice for 5 min. Heat-shock the mixture at 42 °C for 30 s and return to ice immediately for 2 min. Add 100 µl of SOC medium and plate it onto a standard LB agar plate (100 mm Petri dish, ampicillin). Incubate it overnight at 37 °C.
- 76** The next day, inspect the plates for colony growth. Typically, there should be tens to hundreds of colonies on each plate.

#### Troubleshooting

- 77** From each plate, pick 1 or 2 colonies for midiprep to check for the correct insertion of sgRNA and for downstream lentivirus production. To prepare a starter culture for midiprep, use a sterile pipette tip to inoculate a single colony into a 3-ml culture of LB medium with 100 µg ml<sup>-1</sup> ampicillin. Incubate the starter culture and shake it at 37 °C for 4–6 h at >250 rpm.
- 78** Expand each starter culture by transferring the starter culture to 2 separate 25-ml cultures of LB medium with 100 µg ml<sup>-1</sup> ampicillin in a 50-ml Falcon tube. Remove the cap and seal the top of the tube with AirPore Tape Sheets. Incubate the culture and shake it at 37 °C overnight at >250 rpm.
- 79** 12–16 h after seeding the starter culture, isolate the plasmids using an endotoxin-free midiprep kit such as the Macherey-Nagel NucleoBond Xtra Midi EF kit according to the manufacturer's protocol.

**Critical Step** Using an endotoxin-free plasmid purification kit is important for avoiding endotoxicity in virus preparation and mammalian cell culture. In our

experience, plasmids prepared with endotoxin-free midiprep have higher purity and produce higher lentiviral titers.

**Pause Point** Midiprep validation sgRNA constructs can be stored at  $-20^{\circ}\text{C}$  for at least 1 year.

- 80** *Sequence validation of sgRNA cloning.* Verify the correct insertion of the validation sgRNAs by sequencing from the U6 promoter using the U6-fwd primer. Compare the sequencing results to the sgRNA library plasmid sequence to check that the 20-nt sgRNA target sequence is properly inserted between the U6 promoter and the remainder of the sgRNA scaffold.

### Troubleshooting

- 81** *Generation of validation cell lines.* Prepare lentivirus for validation by scaling down the lentivirus production in Steps 44–51 to T25 flasks or 2 wells of a 6-well plate. Filter the lentivirus supernatant using 5-ml syringes and 0.45  $\mu\text{m}$  syringe filters.
- 82** Titer the lentivirus according to Step 53. If preparing multiple validation sgRNA lentivirus in the same plasmid backbone at the same time, titer lentivirus from 2–3 different sgRNAs and extend the average titer to the rest of the lentivirus.
- 83** Similar to during screening, transduce either naive cells or Cas9 component-expressing cells with validation sgRNA lentivirus at an MOI  $< 0.5$  according to Steps 54–55. For knockout validation, select for 7 d to allow for sufficient time for indel saturation.
- 84** *Validation of candidate genes for screening phenotype.* Once the antibiotic selection for validation cell lines is complete, verify the screening phenotype by applying the screening selection on the validation cell lines and assessing cell proliferation, death, or fluorescence for positive, negative, and marker gene selection screens respectively. In addition, determine the indel rate for knockout screens (Steps 85–99) or fold activation for activation screens (Steps 100–110).
- 85** *Indel rate analysis for validating a knockout screen.* We describe a two-step PCR for NGS in which the first step uses custom primers to amplify the genomic region of interest and the second step uses universal, barcoded primers for multiplexed sequencing of up to 96 different samples in the same NGS run. For each validation sgRNA, design custom round 1 NGS primers (NGS-indel-R1) that amplify the 100–300bp region centered around the sgRNA cut site according to Table 4. It is important to design primers situated at least 50 bp from the target cleavage site to allow for the detection of longer indels. Aim for an annealing temperature of  $60^{\circ}\text{C}$  and check for potential off-target sites using Primer-BLAST. If necessary, include a 1–10bp staggered region to increase the diversity of the library.
- 86** *Harvest gDNA from validation cell lines.* Seed the validation cells at a density of 60% confluency with 3 bioreps in a 96-well clear bottom black tissue culture plate.

- 87** 1 d after seeding, when the cells have reached confluency, aspirate the media and add 50  $\mu$ l of QuickExtract DNA Extraction Solution. Incubate at room temperature for 2–3 min.
- 88** Scrape the cells with a pipette tip, mix thoroughly by pipetting up and down, and transfer the mixture to a 96-well PCR plate.
- 89** Extract the genomic DNA by running the following cycling conditions:

Cycle number	Condition
1	65 °C, 15 min
2	68 °C, 15 min
3	98 °C, 10 min

**Pause Point** Extracted genomic DNA can be stored at –20 °C for up to several months.

- 90** *First round PCR for indel analysis by NGS.* Amplify the respective target regions for each validation and control cell line by using custom NGS-indel-R1 primers (Table 4) in the following reaction:

Component	Amount ( $\mu$ l)	Final concentration
KAPA HiFi HotStart ReadyMix, 2 $\times$	10	1 $\times$
First round PCR from Step 91	1	
NGS-indel-R2-Fwd	1	0.5 $\mu$ M
NGS-indel-R2-Rev	1	0.5 $\mu$ M
UltraPure water	7	
Total	20	

**Critical Step** To minimize error in amplifying sgRNAs, it is important to use a high-fidelity polymerase. Other high-fidelity polymerases, such as PfuUltra II (Agilent) or NEBNext (New England BioLabs), may be used as a substitute.

- 91** Perform a PCR with the following cycling conditions:

Cycle number	Denature	Anneal	Extend
1	98 °C, 3 min		
2–23	98 °C, 10 s	63 °C, 10 s	72 °C, 25 s
14	72 °C, 2 min		

- 92** *Second round PCR for indel analysis by NGS.* Barcode the first round PCR for NGS by amplifying the product with different NGS-indel-R2 primers (Table 4) in the following reaction:

Component	Amount ( $\mu$ l)	Final concentration
KAPA HiFi HotStart ReadyMix, 2 $\times$	10	1 $\times$

QuickExtract from Step 89	1	
NGS-indel-R1-Fwd	1	0.5 μM
NGS-indel-R1-Rev	1	0.5 μM
UltraPure water	7	
Total	20	

- 93 Perform a PCR using the same cycling conditions as described in Step 91.
- 94 After the reaction is complete, run 5 μl of each amplified target on a gel to verify successful amplification of a single product at the appropriate size. Cast a 2% (wt/vol) agarose gel cast in TBE buffer with SYBR Safe dye and run the gel at 15 V cm<sup>-1</sup> for 30 min.

**Troubleshooting**

- 95 Pool the PCR products and purify the pooled product using the QIAquick PCR purification kit, according to the manufacturer’s instructions.

**Critical Step** Due to variable PCR efficiencies and product lengths, pooling without normalization may result in variation in NGS representation. When pooling without normalization, aim for 20,000–40,000 reads per sgRNA during sequencing. Alternatively, if sequencing reads are limited, consider purifying each barcoded PCR product separately, quantifying by NanoDrop, and normalizing the PCR products to the same concentration prior to pooling.

- 96 Run the pooled PCR product on a 2% (wt/vol) agarose gel as described in Step 94 and gel extract the appropriately sized bands using the QIAquick gel extraction kit according to the manufacturer’s directions.

**Pause Point** Gel-extracted product can be stored at –20 °C for several months.

- 97 Sequence gel-extracted samples on the Illumina MiSeq according to the Illumina user manual with 260 cycles of read 1, 8 cycles of index 1, and 8 cycles of index 2. We recommend aiming for >10,000 reads per sgRNA.

- 98 *Indel analysis of validation sgRNAs with calculate\_indel.py.* We provide a python script for analyzing the NGS results for indel rates (Supplementary Data 4). Install python 2.7 (<https://www.python.org/downloads/>), biopython (<http://biopython.org/DIST/docs/install/Installation.html>), and SciPy (<https://www.scipy.org/install.html>). Construct a sample sheet with each line corresponding to a separate sample. Each line should contain sample name, fasta or fastq file name, guide sequence, PCR target amplicon, and experimental or control in columns from left to right. The last column is only required when performing the maximum likelihood estimate (MLE) correction. When MLE is performed, control samples which reflect the background indel rate should be labeled “Control” and experimental samples should be labeled “Experimental”. Refer to the following table for an example of a sample sheet:

example1	example1.fastq	GCCCGATCGCTATATCCACG	TGTATATACCTCGCGCCTAACTGCCAGCTGACCACGCCGTA
example2	example2.fastq	CGAGATAAGTCAGCAGGGGC	CTCTTCTGCTCAAGCGAGTTCCAGAGGTCCTTGCCGAGG



```
example3 example3.fastq CACCCACACCAACCGCAGAA CTGGGTTTAACCGAGCTAGTCCTGAAGATCTTGAGTAAC
```

- 99** If processing all files with a single command, run python calculate\_indel.py with the following optional parameters:

Flag	Description	Default
-f	Indicates input file is fasta format	Fastq file format
-a	Uses alternative hashing algorithm for calculation <sup>74</sup>	Ratcliff-Obershelp based algorithm <sup>73</sup>
-o	Output file with calculated indel rates and statistics	calc_indel_out.csv
-i	Input file with sample name, fasta or fastq file name, guide sequence, PCR target amplicon, and experimental or control in columns from left to right	sample_sheet.csv
-v or -q	Increase or decrease reporting as script runs	Standard reporting
-no-m	Does not perform MLE correction	MLE correction is performed

For processing individual samples, such as in the case of parallelization, run python calculate\_indel.py --sample <sample name> to produce a file <sample name>\_out.csv. Combine individual sample files by calling python calculate\_indel.py --combine. After running calculate\_indel.py, calculated indels will be in the output file, which also contains counts of reads that matched perfectly, failed to align, or were rejected due to quality, or had miscalled bases/replacements. There will also be three columns corresponding to the MLE corrected indel rate, as well as the upper and lower bounds for the 95% confidence interval of indels.

- 100** *Determine the fold activation for validating an activation screen.* Prepare cells by seeding the validation cells at a density of 60% confluency with 4 bioreps for each validation cell line in a 96-well poly-d-lysine coated tissue culture plate.
- 101** *Reverse transcription to cDNA.* When the cells are confluent approximately 1 d after seeding, prepare the following reagents for each well: Complete RNA lysis buffer:

Component	Amount (µl)	Final concentration
RNA lysis buffer	100	
Proteinase K, 300 U ml <sup>-1</sup>	1	3 U ml <sup>-1</sup>
DNase I, 50 KU ml <sup>-1</sup>	0.6	300 U ml <sup>-1</sup>
Total	101.6	

#### Reverse Transcription Mix:

Component	Amount (µl)	Final concentration
Reaction Buffer, 5×	5	1×
dNTP Mix, 10 mM	1.25	0.5 mM
Random Hexamer Primer, 100 µM	1.09	4.4 µM
Oligo dT, 100 µM	0.88	3.5 µM

RiboLock RNAse Inhibitor, 40 U $\mu\text{l}^{-1}$	0.125	0.2 U $\mu\text{l}^{-1}$
RevertAid Reverse Transcriptase, 200 U $\mu\text{l}^{-1}$	1.25	10 U $\mu\text{l}^{-1}$
UltraPure water	10.41	
Total	20	

Except for Oligo dT, all components can be found in the Thermo RevertAid RT Reverse Transcription kit.

**Critical Step** Make sure all reagents are RNase free and take proper precautions when working with RNA.

**102** Aliquot 20  $\mu\text{l}$  of the Reverse Transcription Mix into each well of a 96-well PCR plate. Thaw the RNA lysis stop solution, prepare cold DPBS, and keep all reagents except for the complete RNA lysis buffer on ice until needed.

**103** Aspirate the media from each well of the 96-well poly-d-lysine tissue culture plate from Step 100, wash with 100  $\mu\text{l}$  of cold DPBS, and add 100  $\mu\text{l}$  of room temperature complete RNA lysis buffer from Step 101. Incubate at room temperature while mixing thoroughly for 6–12 min to lyse the cells.

**Critical Step** It is important to limit the lysis time to less than 12 min to prevent RNA degradation.

**104** Transfer 30  $\mu\text{l}$  of the cell lysate to a new 96-well PCR plate. Add 3  $\mu\text{l}$  of RNA lysis stop solution to terminate lysis and mix thoroughly.

**Pause Point** The cell lysate with RNA lysis stop solution can be stored at  $-20\text{ }^{\circ}\text{C}$  for additional reverse transcription reactions.

**105** Then, add 5  $\mu\text{l}$  of the cell lysate with RNA lysis stop solution to the Reverse Transcription Mix from Step 101 for a total volume of 25  $\mu\text{l}$  and mix thoroughly.

**106** Reverse transcribe the harvested RNA to cDNA with the following cycling conditions:

Cycle number	Condition
1	25 $^{\circ}\text{C}$ , 10 min
2	37 $^{\circ}\text{C}$ , 60 min
3	95 $^{\circ}\text{C}$ , 5 min

**Pause Point** The cDNA can be stably stored at  $-20\text{ }^{\circ}\text{C}$  for 1 year.

**107** *Perform a TaqMan qPCR for fold activation analysis.* Thermo Fisher Scientific provides design ready TaqMan Gene Expression Assays for candidate genes as well as for endogenous control genes such as GAPDH or ACTB. Make sure that the experimental and control gene expression assays have different probe dyes (i.e. VIC and FAM dyes) that allow for running in the same reaction. Prepare the following qPCR master mix per reverse transcription reaction. We recommend pre-mixing the TaqMan Fast Advanced Mastermix and gene expression assays for all samples with the same target gene first.

Component	Amount ( $\mu$ l)	Final concentration
TaqMan Fast Advanced Master Mix, 2 $\times$	12	1 $\times$
TaqMan Gene Expression Assays for candidate gene, 20 $\times$	1.2	1 $\times$
TaqMan Gene Expression Assays for control gene, 20 $\times$	1.2	1 $\times$
cDNA	9.6	
Total	24	

**108** Aliquot  $4 \times 5 \mu$ l of the qPCR master mix into a 384-well optical plate for technical replicates.

Cycle number	Hold	Denature	Anneal/Extend
1	50 °C, 2 min		
2	95 °C, 20 s		
3–42		95 °C, 3 s	60 °C, 30 s

**109** Perform a qPCR with the following cycling conditions:

**110** Once the qPCR is complete, calculate the candidate gene expression fold change relative to control using the ddCt method according to the instrument manufacturer's protocol.

**111** *Additional steps for combining candidate genes from knockout and activation screens using dead sgRNAs (dRNAs).* dRNAs are sgRNAs with 14- or 15-nt spacer sequences, which are truncated versions of the standard sgRNAs with 20-nucleotide spacer sequences, that are still capable of binding DNA<sup>75,76</sup>. dRNAs are considered catalytically 'dead' because they can guide wild-type Cas9 without inducing double-stranded breaks. Adding MS2 binding loops to the dRNA backbone allows for wild-type Cas9 to activate transcription without cleavage. To achieve simultaneous knockout and activation, repeat Steps 44–55 to generate a cell line that stably expresses wild-type Cas9 and MS2-p65-HSF1.

**112** Transduce the Cas9- and MS2-p65-HSF1-expressing cell line with a standard sgRNA for knocking out a candidate gene and a 14-nt dRNA with MS2 binding loops for activating a second candidate gene according to Steps 69–83.

**113** Verify the screening phenotype, indel percentage, and fold activation according to Steps 84–110.

**Troubleshooting**—Troubleshooting advice can be found in Table 6. To clarify potential points of confusion when performing genome-scale screens using CRISPR-Cas9, we have compiled a list of most-frequently asked questions from our web-based CRISPR forum ([discuss.genome-engineering.org](http://discuss.genome-engineering.org)) (Box 6). (Fig. 4)

**Timing**—Steps 1–17, designing and cloning a targeted screen: 3–5 w, 1 w hands-on  
Steps 18–31, amplification of pooled sgRNA library: 2 d

Steps 32–39, next-generation sequencing of the amplified sgRNA library: 3–5 d

Steps 40–53, lentivirus production and titer: 1–2 w

Steps 54–56, lentiviral transduction and selection: 3–6 w

Steps 57–68, harvest genomic DNA for screening analysis: 5–7 d

Steps 69–110, validation of candidate genes: 4–5 w

## ANTICIPATED RESULTS

As a reference for screening results, we provide data from genome-scale knockout and transcriptional activation screens for genes that confer BRAF inhibitor vemurafenib (PLX) resistance in a BRAF<sup>V600E</sup> (A375) cell line<sup>30,37</sup>. After applying vemurafenib selection, the sgRNA library distribution, which is measured by NGS, in the experimental condition is more skewed than the baseline and vehicle control conditions, with some sgRNAs enriched and others depleted (Fig. 4a,b). SgRNAs targeting genes involved in vemurafenib resistance are enriched because they provide a proliferation advantage upon vemurafenib treatment. RIGER analysis of enriched sgRNAs in the vemurafenib condition relative to the control identified several candidate genes responsible for resistance. Each candidate gene has multiple significantly enriched sgRNAs (Fig. 4c, d) and *P* values that are significantly lower than the rest of the genes (Fig. 4e, f).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We would like to thank O. Shalem, D. A. Scott, and P. D. Hsu for helpful discussions and insights; R. Belliveau for overall research support; R. Macrae for critical reading of the manuscript; and the entire Zhang laboratory for support and advice. O.O.A. is supported by a Paul and Daisy Soros Fellowship, a Friends of the McGovern Institute Fellowship, and the Poitras Center for Affective Disorders. J.S.G. is supported by a D.O.E. Computational Science Graduate Fellowship. F.Z. is supported by the NIH through NIMH (5DP1-MH100706 and 1R01-MH110049), the New York Stem Cell, Poitras, Simons, Paul G. Allen Family, and Vallee Foundations; and David R. Cheng, Tom Harriman, and B. Metcalfe. F.Z. is a New York Stem Cell Foundation-Robertson Investigator. Reagents are available through Addgene; support forums and computational tools are available via the Zhang lab website (<http://www.genome-engineering.org>).

## References

1. Boutros M, Ahringer J. The art and design of genetic screens: RNA interference. *Nat Rev Genet.* 2008; 9:554–566. DOI: 10.1038/nrg2364 [PubMed: 18521077]
2. Echeverri CJ, Perrimon N. High-throughput RNAi screening in cultured cells: a user's guide. *Nat Rev Genet.* 2006; 7:373–384. DOI: 10.1038/nrg1836 [PubMed: 16607398]
3. Shalem O, Sanjana NE, Zhang F. High-throughput functional genomics using CRISPR-Cas9. *Nat Rev Genet.* 2015; 16:299–311. DOI: 10.1038/nrg3899 [PubMed: 25854182]
4. Elbashir SM, et al. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature.* 2001; 411:494–498. DOI: 10.1038/35078107 [PubMed: 11373684]
5. Fire A, et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature.* 1998; 391:806–811. DOI: 10.1038/35888 [PubMed: 9486653]

6. McManus MT, Sharp PA. Gene silencing in mammals by small interfering RNAs. *Nat Rev Genet.* 2002; 3:737–747. DOI: 10.1038/nrg908 [PubMed: 12360232]
7. Meister G, Tuschl T. Mechanisms of gene silencing by double-stranded RNA. *Nature.* 2004; 431:343–349. DOI: 10.1038/nature02873 [PubMed: 15372041]
8. Berns K, et al. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature.* 2004; 428:431–437. DOI: 10.1038/nature02371 [PubMed: 15042092]
9. Boutros M, et al. Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science.* 2004; 303:832–835. DOI: 10.1126/science.1091266 [PubMed: 14764878]
10. Moffat J, Sabatini DM. Building mammalian signalling pathways with RNAi screens. *Nat Rev Mol Cell Biol.* 2006; 7:177–187. DOI: 10.1038/nrm1860 [PubMed: 16496020]
11. Paddison PJ, et al. A resource for large-scale RNA-interference-based screens in mammals. *Nature.* 2004; 428:427–431. DOI: 10.1038/nature02370 [PubMed: 15042091]
12. Root DE, Hacohen N, Hahn WC, Lander ES, Sabatini DM. Genome-scale loss-of-function screening with a lentiviral RNAi library. *Nat Methods.* 2006; 3:715–719. DOI: 10.1038/nmeth924 [PubMed: 16929317]
13. Silva JM, et al. Second-generation shRNA libraries covering the mouse and human genomes. *Nat Genet.* 2005; 37:1281–1288. DOI: 10.1038/ng1650 [PubMed: 16200065]
14. Birmingham A, et al. 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat Methods.* 2006; 3:199–204. DOI: 10.1038/nmeth854 [PubMed: 16489337]
15. Jackson AL, et al. Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol.* 2003; 21:635–637. DOI: 10.1038/nbt831 [PubMed: 12754523]
16. Jackson AL, Linsley PS. Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nat Rev Drug Discov.* 2010; 9:57–67. DOI: 10.1038/nrd3010 [PubMed: 20043028]
17. Wright AV, Nunez JK, Doudna JA. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell.* 2016; 164:29–44. DOI: 10.1016/j.cell.2015.12.035 [PubMed: 26771484]
18. Marraffini LA. CRISPR-Cas immunity in prokaryotes. *Nature.* 2015; 526:55–61. DOI: 10.1038/nature15386 [PubMed: 26432244]
19. Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell.* 2014; 157:1262–1278. S0092-8674(14)00604-7[pii]. DOI: 10.1016/j.cell.2014.05.010 [PubMed: 24906146]
20. Garneau JE, et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature.* 2010; 468:67–71. doi:nature09523 [pii] 10.1038/nature09523. [PubMed: 21048762]
21. Deltcheva E, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature.* 2011; 471:602–607. [pii]. DOI: 10.1038/nature09886 [PubMed: 21455174]
22. Jinek M, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* 2012; 337:816–821. [pii]. DOI: 10.1126/science.1225829 [PubMed: 22745249]
23. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A.* 2012; 109:E2579–2586. doi:1208507109 [pii] 10.1073/pnas.1208507109. [PubMed: 22949671]
24. Cong L, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science.* 2013; 339:819–823. DOI: 10.1126/science.1231143 [PubMed: 23287718]
25. Mali P, et al. RNA-guided human genome engineering via Cas9. *Science.* 2013; 339:823–826. DOI: 10.1126/science.1232033 [PubMed: 23287722]
26. Rouet P, Smih F, Jasin M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol.* 1994; 14:8096–8106. [PubMed: 7969147]
27. Chavez A, et al. Highly efficient Cas9-mediated transcriptional programming. *Nat Methods.* 2015; 12:326–328. DOI: 10.1038/nmeth.3312 [PubMed: 25730490]

28. Gilbert LA, et al. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*. 2014; 159:647–661. DOI: 10.1016/j.cell.2014.09.029 [PubMed: 25307932]
29. Gilbert LA, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*. 2013; 154:442–451. DOI: 10.1016/j.cell.2013.06.044 [PubMed: 23849981]
30. Konermann S, et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*. 2015; 517:583–588. DOI: 10.1038/nature14136 [PubMed: 25494202]
31. Maeder ML, et al. CRISPR RNA-guided activation of endogenous human genes. *Nat Methods*. 2013; 10:977–979. DOI: 10.1038/nmeth.2598 [PubMed: 23892898]
32. Mali P, et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol*. 2013; 31:833–838. DOI: 10.1038/nbt.2675 [PubMed: 23907171]
33. Perez-Pinera P, et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods*. 2013; 10:973–976. DOI: 10.1038/nmeth.2600 [PubMed: 23892895]
34. Qi LS, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*. 2013; 152:1173–1183. DOI: 10.1016/j.cell.2013.02.022 [PubMed: 23452860]
35. Konermann S, et al. Optical control of mammalian endogenous transcription and epigenetic states. *Nature*. 2013; 500:472–476. DOI: 10.1038/nature12466 [PubMed: 23877069]
36. Chavez A, et al. Comparison of Cas9 activators in multiple species. *Nat Methods*. 2016
37. Shalem O, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*. 2014; 343:84–87. DOI: 10.1126/science.1247005 [PubMed: 24336571]
38. Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic screens in human cells using the CRISPR-Cas9 system. *Science*. 2014; 343:80–84. DOI: 10.1126/science.1246981 [PubMed: 24336569]
39. Koike-Yusa H, Li Y, Tan EP, Velasco-Herrera Mdel C, Yusa K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol*. 2014; 32:267–273. DOI: 10.1038/nbt.2800 [PubMed: 24535568]
40. Zhou Y, et al. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature*. 2014; 509:487–491. DOI: 10.1038/nature13166 [PubMed: 24717434]
41. Doench JG, et al. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol*. 2014; 32:1262–1267. DOI: 10.1038/nbt.3026 [PubMed: 25184501]
42. Chen S, et al. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell*. 2015; 160:1246–1260. DOI: 10.1016/j.cell.2015.02.038 [PubMed: 25748654]
43. Shi J, et al. Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat Biotechnol*. 2015; 33:661–667. DOI: 10.1038/nbt.3235 [PubMed: 25961408]
44. Parnas O, et al. A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell*. 2015; 162:675–686. DOI: 10.1016/j.cell.2015.06.059 [PubMed: 26189680]
45. Wang T, et al. Identification and characterization of essential genes in the human genome. *Science*. 2015; 350:1096–1101. DOI: 10.1126/science.aac7041 [PubMed: 26472758]
46. Wong AS, et al. Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. *Proc Natl Acad Sci U S A*. 2016; 113:2544–2549. DOI: 10.1073/pnas.1517883113 [PubMed: 26864203]
47. Doench JG, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol*. 2016; 34:184–191. DOI: 10.1038/nbt.3437 [PubMed: 26780180]
48. Jain IH, et al. Hypoxia as a therapy for mitochondrial disease. *Science*. 2016; 352:54–61. DOI: 10.1126/science.aad9642 [PubMed: 26917594]
49. Marceau CD, et al. Genetic dissection of Flaviviridae host factors through genome-scale CRISPR screens. *Nature*. 2016; 535:159–163. DOI: 10.1038/nature18631 [PubMed: 27383987]
50. Zhang R, et al. A CRISPR screen defines a signal peptide processing pathway required by flaviviruses. *Nature*. 2016; 535:164–168. DOI: 10.1038/nature18625 [PubMed: 27383988]
51. Blondel CJ, et al. CRISPR/Cas9 Screens Reveal Requirements for Host Cell Sulfation and Fucosylation in Bacterial Type III Secretion System-Mediated Cytotoxicity. *Cell Host Microbe*. 2016; 20:226–237. DOI: 10.1016/j.chom.2016.06.010 [PubMed: 27453484]
52. Orchard RC, et al. Discovery of a proteinaceous cellular receptor for a norovirus. *Science*. 2016; 353:933–936. DOI: 10.1126/science.aaf1220 [PubMed: 27540007]



53. Canver MC, et al. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature*. 2015; 527:192–197. DOI: 10.1038/nature15521 [PubMed: 26375006]
54. Korkmaz G, et al. Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat Biotechnol*. 2016; 34:192–198. DOI: 10.1038/nbt.3450 [PubMed: 26751173]
55. Diao Y, et al. A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. *Genome research*. 2016; 26:397–405. DOI: 10.1101/gr.197152.115 [PubMed: 26813977]
56. Sanjana NE, et al. High-resolution interrogation of functional elements in the noncoding genome. *Science*. 2016; 353:1545–1549. DOI: 10.1126/science.aaf7613 [PubMed: 27708104]
57. Fulco CP, et al. Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science*. 2016; 354:769–773. DOI: 10.1126/science.aag2445 [PubMed: 27708057]
58. Fu Y, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*. 2013; 31:822–826. DOI: 10.1038/nbt.2623 [PubMed: 23792628]
59. Hsu PD, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*. 2013; 31:827–832. DOI: 10.1038/nbt.2647 [PubMed: 23873081]
60. Pattanayak V, et al. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol*. 2013; 31:839–843. DOI: 10.1038/nbt.2673 [PubMed: 23934178]
61. Evers B, et al. CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nat Biotechnol*. 2016; 34:631–633. DOI: 10.1038/nbt.3536 [PubMed: 27111720]
62. Munoz DM, et al. CRISPR Screens Provide a Comprehensive Assessment of Cancer Vulnerabilities but Generate False-Positive Hits for Highly Amplified Genomic Regions. *Cancer Discov*. 2016; 6:900–913. DOI: 10.1158/2159-8290.CD-16-0178 [PubMed: 27260157]
63. Aguirre AJ, et al. Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer Discov*. 2016; 6:914–929. DOI: 10.1158/2159-8290.CD-16-0154 [PubMed: 27260156]
64. Morgens DW, Deans RM, Li A, Bassik MC. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat Biotechnol*. 2016; 34:634–636. DOI: 10.1038/nbt.3567 [PubMed: 27159373]
65. Hasson SA, et al. High-content genome-wide RNAi screens identify regulators of parkin upstream of mitophagy. *Nature*. 2013; 504:291–295. DOI: 10.1038/nature12748 [PubMed: 24270810]
66. Moffat J, et al. A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell*. 2006; 124:1283–1298. DOI: 10.1016/j.cell.2006.01.040 [PubMed: 16564017]
67. Neumann B, et al. High-throughput RNAi screening by time-lapse imaging of live human cells. *Nat Methods*. 2006; 3:385–390. DOI: 10.1038/nmeth876 [PubMed: 16628209]
68. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods*. 2014; 11:783–784. DOI: 10.1038/nmeth.3047 [PubMed: 25075903]
69. Konig R, et al. A probability-based approach for the analysis of large-scale RNAi screens. *Nat Methods*. 2007; 4:847–849. DOI: 10.1038/nmeth1089 [PubMed: 17828270]
70. Luo B, et al. Highly parallel identification of essential genes in cancer cells. *Proc Natl Acad Sci U S A*. 2008; 105:20380–20385. DOI: 10.1073/pnas.0810485105 [PubMed: 19091943]
71. Li W, et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol*. 2014; 15:554. [PubMed: 25476604]
72. Ran FA, et al. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013; 8:2281–2308. nprot.2013143 [pii]. DOI: 10.1038/nprot.2013.143 [PubMed: 24157548]
73. Ran FA, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*. 2013; 154:1380–1389. DOI: 10.1016/j.cell.2013.08.021 [PubMed: 23992846]
74. Zetsche B, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*. 2015; 163:759–771. DOI: 10.1016/j.cell.2015.09.038 [PubMed: 26422227]
75. Dahlman JE, et al. Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nat Biotechnol*. 2015; 33:1159–1161. DOI: 10.1038/nbt.3390 [PubMed: 26436575]



76. Kiani S, et al. Cas9 gRNA engineering for genome editing, activation and repression. *Nat Methods*. 2015; 12:1051–1054. DOI: 10.1038/nmeth.3580 [PubMed: 26344044]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Box 1****Different types of selection applicable for CRISPR screens**

Before setting up a screen, it is important to determine the type of screening selection based on the phenotype of interest and available selection pressures for the screen, as illustrated below:

Positive selection screens rely on enrichment of sgRNAs for genetic perturbations that produce the screening phenotype as a result of cell proliferation. These typically have the highest signal-to-noise ratio compared to the other types of screens because the number of phenotypically relevant sgRNAs increases relative to the rest of the sgRNAs. On the other hand, negative selection screens involve depletion of sgRNAs that correspond to the phenotype due to cell death. However, for a large number of screens the phenotype of interest will not result in cell proliferation or cell death and thus the phenotypically relevant sgRNAs are not enriched or depleted. For these phenotypes, the screen may be read out by capturing sgRNAs that produce changes in marker gene protein expression using either endogenous-tagged fluorescent proteins or highly specific antibodies followed by FACS.

**Box 2****Considerations for setting screening parameters**

Optimal screening parameters should maximize the difference in sgRNA distribution between the experimental and control conditions. Selection conditions such as drug dosage or FACS bin cutoff should be predetermined, if possible, using positive and negative controls from the literature and set to the level at which the greatest difference is observed. As for determining the duration of the screen, collection of time points throughout the screen helps identify the best time point for harvesting and analyzing the screen. These time points are also informative for assessing whether it is necessary to increase the duration to enhance the difference between experimental and control conditions.

Throughout the screen, it is imperative to maintain sufficient coverage to avoid losing sgRNA representation or bias the screening results. Try to maintain sufficient coverage at >500 cells per sgRNA in the library during library transduction, screening selection, and screening harvest. In addition, we recommend 2–4 infection replicates per screen to account for stochastic noise. Increase the coverage and number of infection replicates if the screening selection is noisy. Finally, consistency of screening conditions such as sgRNA representation and passaging reduces the variability between infection replicates.

**Box 3****Additional considerations for *ex vivo* and *in vivo* pooled screening**

*Ex vivo* screening involves removing a primary cell type of interest from a living animal, culturing them *in vitro* and then performing the screen. For example, Parnas et al. demonstrated this strategy by deriving immune dendritic cells from Cas9 mice, transducing them with a CRISPR knockout library, triggering an immune response with lipopolysaccharide (LPS), and then FACS sorting different populations of cells based on immune response (e.g. TNF expression)<sup>44</sup>. This *ex vivo* screen identified many known as well as novel regulators of LPS response. When performing an *ex vivo* screen, it is necessary to be able to obtain enough cells to maintain library representation, deliver appropriate reagents to the cells, and culture the cells for long enough to perform the screen. In cases where these conditions cannot be met, adapt the screening strategy by, for instance, reducing the library size to capture a subset of genes.

*In vivo* screening is performed with either a) transduction of cells *in vitro* followed by *in vivo* cell transplantation, or b) direct transduction of tissues *in vivo*. The first strategy was demonstrated by Chen et al., whereby a cancer cell line was transduced with a CRISPR knockout library and injected subcutaneously in immunocompromised mice<sup>42</sup>. NGS analysis of harvested tumors identified known and novel tumor suppressors associated with tumor growth and metastasis. The main challenge of this approach is engrafting cells *in vivo*. Special care must be taken to ensure that the library is not only maintained upon infection of cells *in vitro* but also after engraftment of cells *in vivo*. While it is not required to maintain library representation on a per animal basis, a sufficient number of animals should be used such that library representation is maintained for each experimental cohort. Because the engraftment efficiency and time of engraftment can change for each application it is necessary to sequence the library at several time points after injection of cells *in vivo*. The optimal time point is one where engraftment is complete and selection (i.e. proliferation, death, or migration) has not yet occurred. Identifying this time point is critical as it is used as a reference to identify enriched and/or depleted perturbations.

For the second method of *in vivo* screening, special considerations will vary widely depending on the specific animal model, tissue, cell type, developmental time point, or biological question. Thus, each screen should be uniquely designed. In addition to the screening considerations outlined previously, the additional challenge for this strategy is the delivery of reagents in a complex environment while maintaining library representation and also infecting cells at a low MOI. Beyond specific circumstances, it may not be feasible to achieve appropriate cell numbers suitable for a genome-scale library. In these cases it is recommended to design smaller, targeted libraries with a specific hypothesis in mind. The complexity of the *in vivo* environment makes it difficult to meet the critical requirements for performing an informative screen. In assessing whether a direct *in vivo* screening strategy is feasible for any particular application, consider these guiding questions: 1) Is there a delivery strategy for infecting the target cells at low MOI? 2) Can enough of the target population be infected and purified to

maintain library representation? 3) Can a reference population be identified before the guide RNA abundance changes?

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Box 4****Designing and analyzing a saturated mutagenesis screen**

Although most pooled CRISPR screens to date have focused on knockout or activation of protein-coding genes, CRISPR screens can also be used to identify functional elements in noncoding regions of the genome such as enhancers or repressors. These functional elements are often inferred using biochemical hallmarks associated with function (e.g. chromatin accessibility, transcription factor binding sites, or post-translational histone modifications). In contrast, CRISPR screens enable direct testing of how mutagenesis at a specific noncoding site impacts phenotype.

Several strategies can be used to design libraries to target noncoding regions. For understanding regulation of a particular gene, tiling mutagenesis libraries were designed to include many or all possible target sites within a noncoding region near a gene<sup>53,56,57</sup>. This allows unbiased identification of all regulatory elements in regions near a gene that has already been established to be important for the screening phenotype, as illustrated below:

Another approach is to design the library to target all instances across the genome of a specific biochemical hallmark, such as all binding sites of a transcription factor like p53<sup>54</sup>. With this kind of library, it is possible to identify specific binding sites or regulatory elements associated with a phenotype of interest.

As with screening the coding genome, a key factor in assessing the performance of the screen is to find multiple sgRNAs targeting the same element that are enriched or depleted together. In coding regions, this is straightforward, as the library is designed to have multiple sgRNAs that target the same gene. In noncoding regions, the same principle of consistent enrichment or depletion can be applied to multiple sgRNAs that target neighboring regions, as the indels are of variable length. Once a functional element is validated using multiple sgRNAs with adjacent target sites, expression of nearby genes and potential mechanisms such as alterations in transcription factor binding at the site can be used to gain further insight into biological mechanisms.

**Box 5****Lentivirus production with Polyethylenimine**

1. *Prepare polyethylenimine (PEI) transfection reagent.* Dissolve 50 mg PEI Max in 45 ml UltraPure Water. Adjust pH to 7.1 by adding 10 M NaOH dropwise until the pH approaches 6 and then by adding 1 M NaOH dropwise until final pH reaches 7.1. Adjust final volume to 50 ml with UltraPure Water. Sterilize using Millipore's 0.45  $\mu\text{m}$  Steriflip filter. Prepare 50  $\times$  1 ml aliquots and store at  $-20\text{ }^{\circ}\text{C}$  until use. PEI is stable for up to one year and can undergo 5 freeze-thaw cycles without a drop in transfection efficiency.
2. *Lentivirus plasmid transfection.* Prepare HEK293FT cells for lentivirus transfection as described in Steps 42–44.
3. For each lentiviral target, combine the following lentiviral target mix in a 50-ml Falcon tube and scale up accordingly:

Component	Amount per T225 flask	Final Concentration
DMEM (serum free)	651 $\mu\text{l}$	
pMD2.G (lentiviral helper plasmid)	3.4 $\mu\text{g}$	5.2 $\mu\text{g ml}^{-1}$
psPAX2 (lentiviral helper plasmid)	6.8 $\mu\text{g}$	10.4 $\mu\text{g ml}^{-1}$
Lentiviral target plasmid	$\mu\text{g}$	20.9 $\mu\text{g ml}^{-1}$

4. Add 195  $\mu\text{l}$  PEI transfection reagent, vortex, and incubate at room temperature for 10 min.
5. Add 25 ml D10 media to the transfection reagent mixture.
6. Aspirate old media from cells, gently add new media containing the transfection reagent mixture and shake gently to mix. Return the T225 flask to the incubator.
7. 2 d after transfection, harvest and store lentivirus as described in Step 52.



**Box 6****Frequently asked screening questions from the CRISPR Forum**

The following questions are selected from the CRISPR Discussion Forum ([discuss.genome-engineering.org](http://discuss.genome-engineering.org)).

**Q1: Can I use liquid culture amplification of the library rather than solid plates?**

We recommend using plates because liquid culture can generate more bias in the plasmid library. Beta-lactamase, the enzyme responsible for ampicillin resistance, is secreted and eventually in liquid culture the selective pressure on the plasmid is decreased causing bias. Additionally, it is more difficult for certain clones to predominate on solid plates because they are spatially limited in growth and each clone is spatially separated to prevent potential intercolony competition. However, it is important to note that some studies have had success with liquid culture amplification<sup>45</sup>.

**Q2: Is there a difference between using HEK293FT vs HEK293T cells for lentivirus production?**

Yes, HEK293FT cells are generally more ideal for lentivirus production. HEK293T cells are a cell line stably expressing the SV40 large T antigen, which helps to boost protein production off expression constructs containing the SV40 enhancer element. HEK293FT cells are a fast-growing, highly transfectable clonal derivation of HEK293T cells that yield higher lentivirus titer than the HEK293T line.

**Q3: For activation, how do I design guides relative to the TSS of the transcript? Additionally, can I expect these guides to work with transient transfection of dCas9-VP64 and MS2-VP65-HSF1 plasmids?**

The TSS is the first base of the transcript, i.e. beginning of the 5' UTR. The UCSC table browser is a good resource for TSS annotations. We have observed the most robust transcriptional upregulation when sgRNAs are designed to target the 200bp region upstream of the TSS. We have created a web tool using these parameters to simplify activation sgRNA design for human and mouse genes (<http://sam.genome-engineering.org/database/>). SAM is highly robust and should yield significant activation levels even in the case of transient transfection<sup>36</sup>.

**Q4: What are important considerations for NGS PCR amplification?**

When designing primers, it is important to include stagger between the primer binding site and the Illumina adapter sequence such that the sequencing regions of different amplicons are offset, improving the sequence diversity and quality. For genomic DNA amplification, it can be helpful to optimize the DNA input for the sequencing readout PCR step. Generally, it is recommended for any given instance of the screen to titrate the DNA input and use the highest possible input without a decrease in the target band intensity. It is critical to minimize amplification bias. The optimal cycle number should always be determined by doing a series of different cycle numbers (e.g. 5, 10, and 15) and identifying the lowest cycle number that generates a visible band by gel electrophoresis. Avoid conditions that yield additional bands at higher cycle numbers.

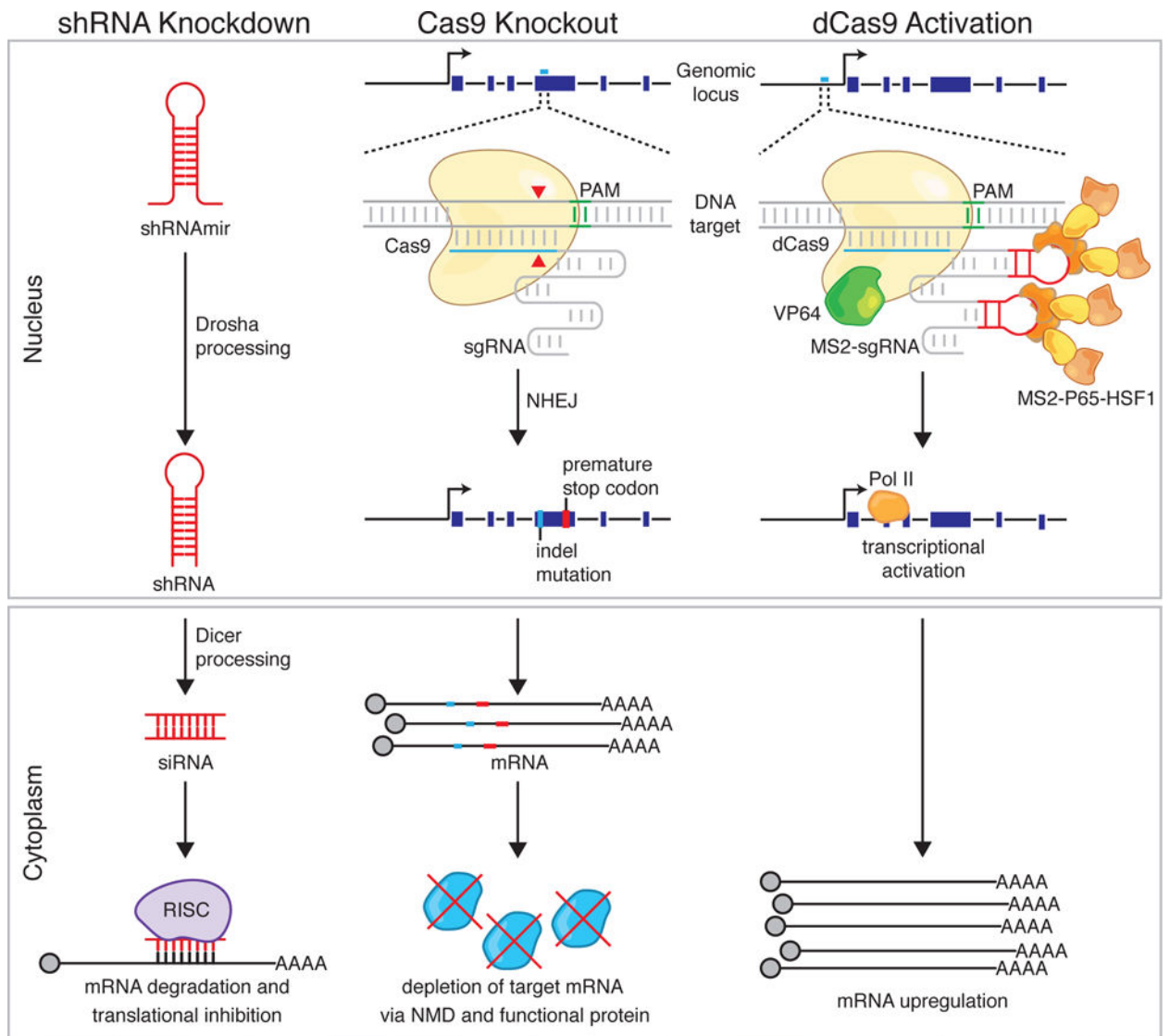
**Q5: My screening design requires too many cells. Can I reduce the coverage?**

We recommend screening at a coverage of >500 cells per sgRNA. Because there is always variability in the copy number of each sgRNA in a given library, it is important to have high coverage to overcome any bias. If it is impossible to screen at this coverage (e.g. insufficient primary cells or cells are difficult to transduce), consider screening with a smaller, targeted library.

**Q6: How do you measure the quality of a cloned plasmid library?**

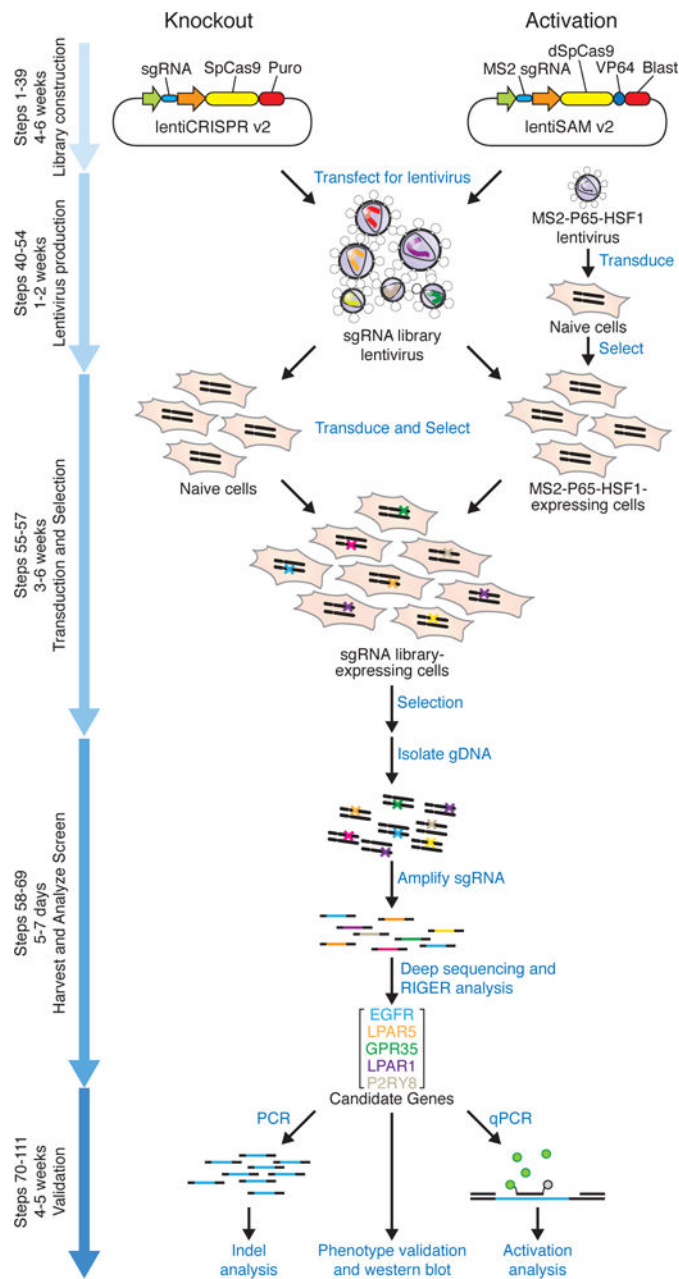
While there are many methods for determining the quality of a library, we typically use the following measures for a sequencing depth of >100 reads per sgRNA:

1. Overall representation: <0.5% of sgRNAs have dropped out with no reads.
2. Library uniformity: <10-fold difference between the 90th percentile and the 10th percentile.



**Figure 1. Approaches to genetic perturbation: shRNA knockdown, Cas9 knockout, and Cas9 transcriptional activation**

Schematic of the mechanisms behind shRNA knockdown, Cas9 knockout, and Cas9 transcriptional activation. ShRNA knockdown begins with processing of the shRNA by Drosha/Dicer machinery and results in degradation of an RNA transcript with a complementary target site by the RNA-induced silencing complex (RISC). Cas9 knockout is accomplished by targeted indel formation at a genomic site complementary to the sgRNA. An indel can result in a frameshift, causing early termination, and either production of non-functional protein or non-sense mediated decay (NMD) of the mRNA transcript. Programmable transcriptional activation can be achieved using dCas9 and activation domains (e.g. VP64/p65/HSF1) to recruit transcriptional machinery to the transcriptional start site of the desired gene target, resulting in upregulation of the target transcript. PAM, protospacer adjacent motif; NHEJ, non-homologous end joining; Pol II, RNA Polymerase II.



**Figure 2. Timeline and overview of experiments**

Genome-scale Cas9 knockout and transcriptional activation screens begin with the construction of a plasmid library encoding the effector protein and sgRNAs. These plasmid libraries are packaged into lentivirus and then transduced into the cell type of interest to generate stably expressing lines for the screen, along with an accessory transcriptional activator complex (MS2-p65-HSF1) lentivirus for the case of activation screening. A selection pressure is applied depending on the nature of the screen and at given timepoints, genomic DNA is harvested. The sgRNA regions (colored bars) are amplified from genomic DNA and then analyzed by next generation sequencing followed by statistical analyses (e.g. RIGER) to identify candidate genes. Candidate genes are then validated by various forms of

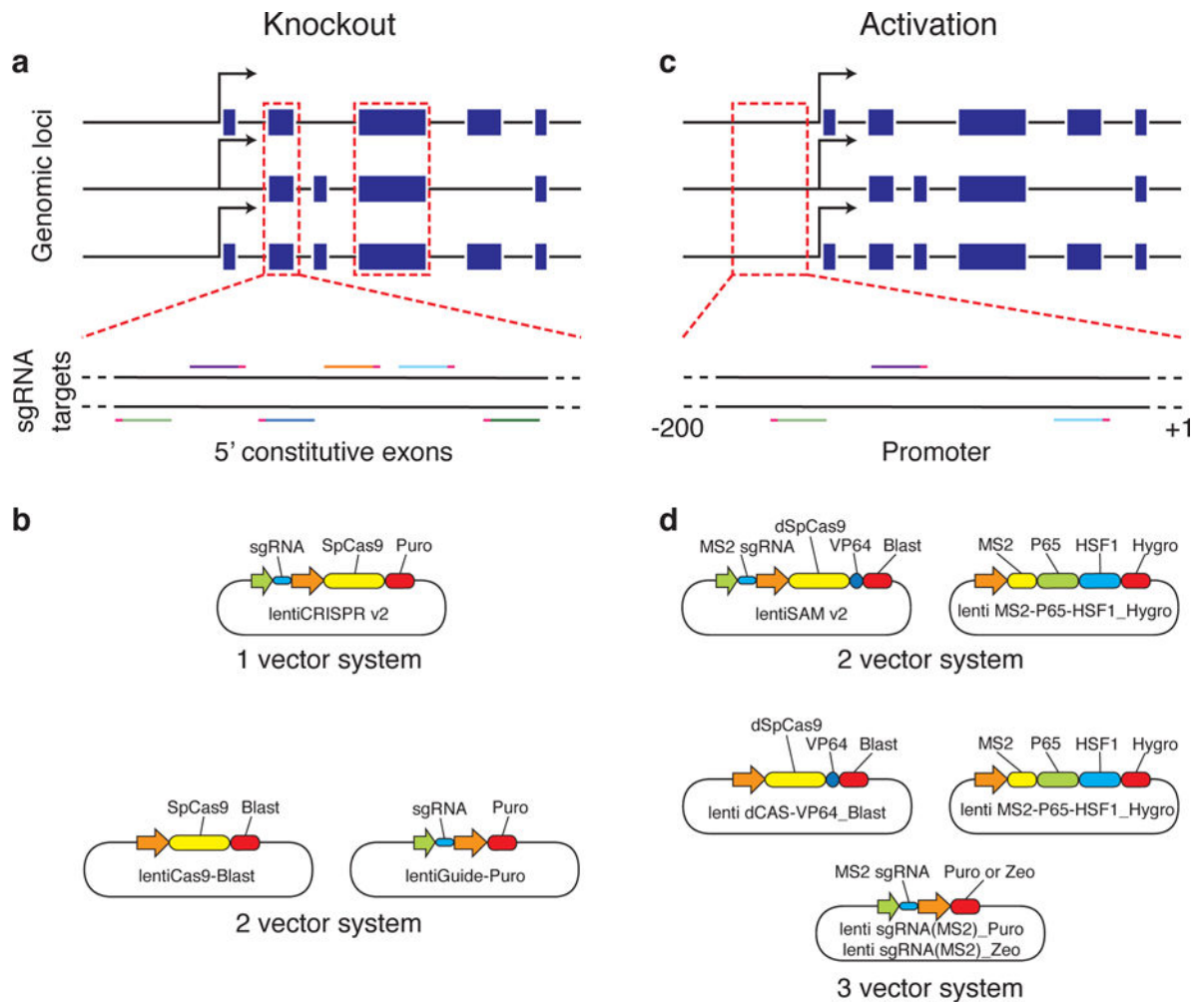
analysis, including testing individual sgRNAs for the screening phenotype, indel formation by targeted sequencing, or transcript upregulation by qPCR.

Author Manuscript

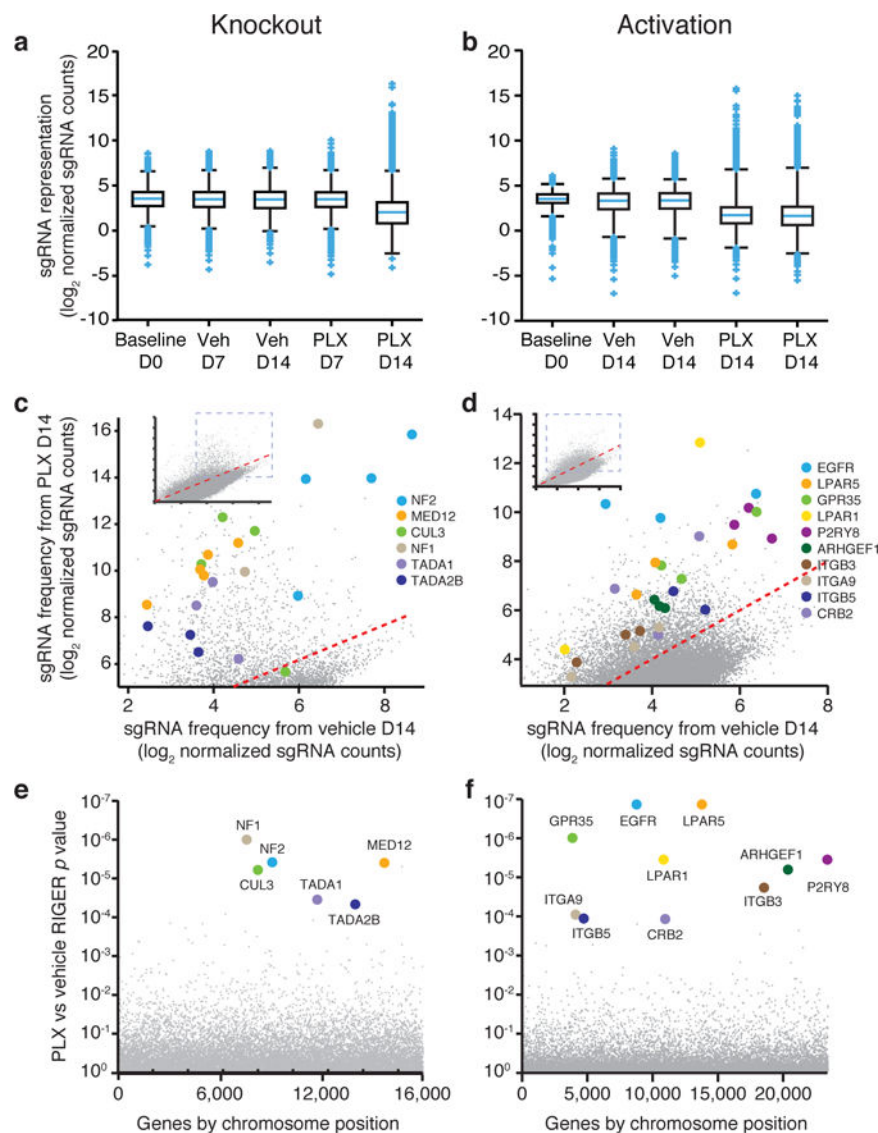
Author Manuscript

Author Manuscript

Author Manuscript



**Figure 3. GeCKO and SAM libraries for genome-scale knockout and activation screens**  
**(a)** For knockout screening, the GeCKO v2 libraries target the 5' conserved coding exons of 19,050 human or 20,611 mouse coding genes with 6 sgRNAs per gene. **(b)** The GeCKO library is available in a 1 vector or 2 vector format. **(c)** For activation screening, the SAM libraries target the 200bp region upstream of the transcriptional start site of 23,430 human or 23,439 mouse RefSeq coding isoforms with 3 sgRNAs per isoform. **(d)** The library has to be combined with additional SAM effectors in a 2 vector or 3 vector format. Both libraries select sgRNAs with minimal off-target activity.



**Figure 4. Anticipated results for genome-scale knockout and activation screens**

We provide data from genome-scale knockout and activation screens for identifying drivers of resistance to the BRAF inhibitor vemurafenib (PLX) in a BRAF<sup>V600E</sup> (A375) melanoma cell line<sup>30,37</sup>. **(a,b)** Box plots showing the distribution of sgRNA frequencies after control (Veh, vehicle) or PLX treatment from  $n = 2$  infection replicates. A significant number of guides are seen enriched and depleted in the PLX day 14 condition, revealing depletion of guides essential for cell growth and enrichment of guides that promote resistance to BRAF inhibitor. Boxes, 25<sup>th</sup> to 75<sup>th</sup> percentile; Whiskers, 1<sup>st</sup> to 99<sup>th</sup> percentile. **(c,d)** Scatterplot showing enrichment of sgRNAs targeting the top candidate genes identified by RIGER (colored dots) compared to other sgRNAs in the library (grey dots) after PLX treatment. Each gene has multiple sgRNAs that are enriched. Inset panel represents the entire dataset. Many of these genes are known tumor suppressors or oncogenes that play a role in PLX4720 resistance. **(e,f)** The top hits of the screen are seen as distributed across the genome, revealing the necessity of genome-scale screens for identifying drivers of resistance. RIGER

*P* values for candidate enriched genes (colored dots) are significantly lower compared to other genes (grey dots) targeted by the sgRNA library.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Table 1**

Previously published screens using Cas9.

Type of screen	Selection	Organism	Cas9 variant	Reference
Knockout	Vemurafenib resistance (positive)	<i>In vitro</i> ; A375 (human melanoma cell line)	Wild-type Cas9	37
Knockout	6-thioguanine/etoposide resistance (positive); gene essentiality (negative)	<i>In vitro</i> ; HL60, KBM70 (human leukemic cell line)	Wild-type Cas9	38
Knockout	6-thioguanine/ <i>Clostridium septicum</i> alpha-toxin resistance (positive)	<i>In vitro</i> ; mouse embryonic stem cells	Wild-type Cas9	39
Knockout	Anthrax/diphtheria toxin resistance (positive)	<i>In vitro</i> ; HeLa (human adenocarcinoma cell line)	Wild-type Cas9	40
Knockout	Surface receptor expression (negative)	<i>In vitro</i> ; EL4 (mouse thymic cell line) and MOLM13/NB4/TF1 (human acute myeloid leukemia cell lines)	Wild-type Cas9	41
Knockout	Metastasis (positive)	<i>In vivo</i> ; mouse	Wild-type Cas9	42
Knockout	Chromatin regulatory domain dependence (negative)	<i>In vitro</i> ; RN2 (murine acute myeloid leukemia cell line)	Wild-type Cas9	43
Knockout	Bacterial lipopolysaccharide response (marker gene)	<i>Ex vivo</i> ; bone-marrow derived dendritic cells (mouse)	Wild-type Cas9	44
Knockout	Fetal hemoglobin regulation by BCL11A enhancer (marker gene)	<i>In vitro</i> ; HUDEP-2 (human erythroid progenitor cell line)	Wild-type Cas9	53
Knockout	Gene dependency for essential genes (negative)	<i>In vitro</i> ; KBM7 (human chronic myelogenous leukemia cell line)	Wild-type Cas9	45
Knockout	p53 binding sites (positive); ESR1 binding sites (negative)	<i>In vitro</i> ; BJ (human fibroblast cell line); MCF-7, T47D, and MDA-MB-231 (human breast cancer cell line)	Wild-type Cas9	54
Knockout	POU5F1 regulation (marker gene)	<i>In vitro</i> ; H1 (human embryonic stem cell line)	Wild-type Cas9	55
Knockout	Combinatorial gene dependency (negative)	<i>In vitro</i> ; OVCAR8-ADR (human ovarian cancer cell line)	Wild-type Cas9	46
Knockout	Vemurafenib resistance (positive); essential genes (negative); 6-thioguanine resistance (positive); interferon survival (positive)	<i>In vitro</i> ; A375 (human melanoma cell line); HT29 (human colorectal adenocarcinoma cell line); HEK293T (human embryonic kidney cell line); BV2 (mouse cell line)	Wild-type Cas9	47
Knockout	Survival under oxidative stress (positive)	<i>In vitro</i> ; K562 (human leukemic cell line)	Wild-type Cas9	48
Knockout	Dengue virus resistance/Hepatitis C virus resistance	<i>In vitro</i> ; Huh7.5.1 (human hepatocyte cell line)	Wild-type Cas9	49
Knockout	West Nile virus resistance	<i>In vitro</i> ; 293T (human embryonic kidney cell line)	Wild-type Cas9	50
Knockout	Type III secretion system resistance	<i>In vitro</i> ; HT29 (human colorectal adenocarcinoma cell line)	Wild-type Cas9	51
Knockout	Norovirus resistance	<i>In vitro</i> ; BV2 (mouse microglial cells)	Wild-type Cas9	52
Knockout	CUL3 regulation and Vemurafenib resistance (positive)	<i>In vitro</i> ; A375 (human melanoma cell line)	Wild-type Cas9	56
Activation	Ricin sensitivity (both); cell growth (both)	<i>In vitro</i> ; K562 (human leukemic cell line)	sunCas9-VP64	28
Activation	Vemurafenib resistance (positive)	<i>In vitro</i> ; A375 (human melanoma cell line)	dCas9-VP64/P65/HSF1	30

Type of screen	Selection	Organism	Cas9 variant	Reference
Knockdown	Ricin resistance (positive)/essential genes (negative)/cholera sensitivity (both)	<i>In vitro</i> ; K562 (human leukemic cell line)	dCas9 or dCas9-KRAB	28
Knockdown	Gene essentiality	<i>In vitro</i> ; K562 (human leukemic cell line)	KRAB-dCas9	57

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

Primers for amplifying the sgRNA oligo library.

<b>Primer</b>	<b>Sequence (5'-3')</b>	<b>Purpose</b>
Oligo-Fwd	GTAACCTGAAAGTATTTTCGATTCTTGGCTTTATATATCTTGTGAAAGGACGAAACACC	Targeted knockout or activation library cloning
Oligo-Knockout-Rev	ACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC	Targeted knockout library cloning
Oligo-Activation-Rev	ATTTTAACTTGCTAGGCCCTGCAGACATGGGTGATCCTCATGTTGGCCTAGCTCTAAAAC	Targeted activation library cloning

Primer sequences for amplifying sgRNA library and NGS. The 10 different NGS-Lib-Fwd primers each contain 1–10 additional nucleotides after the Illumina Read 1 Sequencing primer designed to increase the diversity of the NGS library. All 10 NGS-Lib-Fwd primers are required when amplifying sgRNA library for NGS. The 8 NGS-Lib-Rev primers provide unique barcodes (bold) for distinguishing different sgRNA libraries (e.g. experimental and control conditions for 4-screening bioreps) in a pooled sequencing run, and the number of NGS-Lib-Rev primers needed depends on the number of different sgRNA libraries to be sequenced together in a pooled run. Since the sgRNA backbone is different between the GeCKO and SAM libraries, separate NGS-Lib-KO-Rev and NGS-Lib-SAM-Rev primers have been provided for each library.

**Table 3**

Primer	Sequence (5'–3')	Purpose
NGS-Lib-Fwd-1	AATGATACGGCGACCCAGAGATCTACACTCTTCCCTACAGACGGCTTCCGATCTTAAGTAGAGGCTTTATATAATCTTGTGGAAAGGACGAAACACC	GeCKO or SAM sgRNA library NGS
NGS-Lib-Fwd-2	AATGATACGGCGACCCAGAGATCTACACTCTTCCCTACAGACGGCTTCCGATCTATCATATGCTTAGCTTTATATAFATCTTGTGGAAAGGACGAAACACC	GeCKO or SAM sgRNA library NGS
NGS-Lib-Fwd-3	AATGATACGGCGACCCAGAGATCTACACTCTTCCCTACAGACGGCTTCCGATCTGATGACATCTGCTTTATATAATCTTGTGGAAAGGACGAAACACC	GeCKO or SAM sgRNA library NGS
NGS-Lib-Fwd-4	AATGATACGGCGACCCAGAGATCTACACTCTTCCCTACAGACGGCTTCCGATCTGATGCTCGAGCTTTATATAATCTTGTGGAAAGGACGAAACACC	GeCKO or SAM sgRNA library NGS
NGS-Lib-Fwd-5	AATGATACGGCGACCCAGAGATCTACACTCTTCCCTACAGACGGCTTCCGATCTTCGATFAGCAATTCGGCTTTATATAATCTTGTGGAAAGGACGAAACACC	GeCKO or SAM sgRNA library NGS
NGS-Lib-Fwd-6	AATGATACGGCGACCCAGAGATCTACACTCTTCCCTACAGACGGCTTCCGATCTATCGATGTTGCTTATATAATCTTGTGGAAAGGACGAAACACC	GeCKO or SAM sgRNA library NGS
NGS-Lib-Fwd-7	AATGATACGGCGACCCAGAGATCTACACTCTTCCCTACAGACGGCTTCCGATCTGATCGATCCAGCTTCCGATCTGATGCTTTATATAATCTTGTGGAAAGGACGAAACACC	GeCKO or SAM sgRNA library NGS
NGS-Lib-Fwd-8	AATGATACGGCGACCCAGAGATCTACACTCTTCCCTACAGACGGCTTCCGATCTCGATCCGATCTTATATAATCTTGTGGAAAGGACGAAACACC	GeCKO or SAM sgRNA library NGS
NGS-Lib-Fwd-9	AATGATACGGCGACCCAGAGATCTACACTCTTCCCTACAGACGGCTTCCGATCTACAGCTTCCGATCTGATGATGCTTTATATAATCTTGTGGAAAGGACGAAACACC	GeCKO or SAM sgRNA library NGS
NGS-Lib-Fwd-10	AATGATACGGCGACCCAGAGATCTACACTCTTCCCTACAGACGGCTTCCGATCTTACAGCTTCCGATCTTACAGCTTCCGATCTTATATAATCTTGTGGAAAGGACGAAACACC	GeCKO or SAM sgRNA library NGS
NGS-Lib-KO-Rev-1	CAAGCAGAAAGACGGCATACGAGATTCGGCTTGGTACTGGAGTTTCAGACGTGTGCTCTTCCGATCTCCGACTCCGGTGCACCTTTTTCAA	GeCKO sgRNA library NGS and barcode (bold)
NGS-Lib-KO-Rev-2	CAAGCAGAAAGACGGCATACGAGATTAAGCGTCTGAGTTTCAGACGTGTGCTCTTCCGATCTCCGACTCCGGTGCACCTTTTTCAA	GeCKO sgRNA library NGS and barcode (bold)
NGS-Lib-KO-Rev-3	CAAGCAGAAAGACGGCATACGAGATGAAGAAGTGTGACTGGAGTTTCAGACGTGTGCTCTTCCGATCTCCGACTCCGGTGCACCTTTTTCAA	GeCKO sgRNA library NGS and barcode (bold)
NGS-Lib-KO-Rev-4	CAAGCAGAAAGACGGCATACGAGATATCTAGGGTGTGACTGGAGTTTCAGACGTGTGCTCTTCCGATCTCCGACTCCGGTGCACCTTTTTCAA	GeCKO sgRNA library NGS and barcode (bold)
NGS-Lib-KO-Rev-5	CAAGCAGAAAGACGGCATACGAGATCGTTACCAAGTGTGACTGGAGTTTCAGACGTGTGCTCTTCCGATCTCCGACTCCGGTGCACCTTTTTCAA	GeCKO sgRNA library NGS and barcode (bold)
NGS-Lib-KO-Rev-6	CAAGCAGAAAGACGGCATACGAGATGTTGATGTGACTGGAGTTTCAGACGTGTGCTCTTCCGATCTCCGACTCCGGTGCACCTTTTTCAA	GeCKO sgRNA library NGS and barcode (bold)
NGS-Lib-KO-Rev-7	CAAGCAGAAAGACGGCATACGAGATTTACGCACGTGTGACTGGAGTTTCAGACGTGTGCTCTTCCGATCTCCGACTCCGGTGCACCTTTTTCAA	GeCKO sgRNA library NGS and barcode (bold)
NGS-Lib-KO-Rev-8	CAAGCAGAAAGACGGCATACGAGATTTGAATAGTGTGACTGGAGTTTCAGACGTGTGCTCTTCCGATCTCCGACTCCGGTGCACCTTTTTCAA	GeCKO sgRNA library NGS and barcode (bold)
NGS-Lib-SAM-Rev-1	CAAGCAGAAAGACGGCATACGAGATTAAGCGTGTGACTGGAGTTTCAGACGTGTGCTCTTCCGATCTCCGACTCCGGTGCACCTTTTTCAA	SAM sgRNA library NGS and barcode (bold)
NGS-Lib-SAM-Rev-2	CAAGCAGAAAGACGGCATACGAGATTAAGCGTGTGACTGGAGTTTCAGACGTGTGCTCTTCCGATCTCCGACTCCGGTGCACCTTTTTCAA	SAM sgRNA library NGS and barcode (bold)
NGS-Lib-SAM-Rev-3	CAAGCAGAAAGACGGCATACGAGATGAAGAAGTGTGACTGGAGTTTCAGACGTGTGCTCTTCCGATCTCCGACTCCGGTGCACCTTTTTCAA	SAM sgRNA library NGS and barcode (bold)
NGS-Lib-SAM-Rev-4	CAAGCAGAAAGACGGCATACGAGATATCTAGGGTGTGACTGGAGTTTCAGACGTGTGCTCTTCCGATCTCCGACTCCGGTGCACCTTTTTCAA	SAM sgRNA library NGS and barcode (bold)
NGS-Lib-SAM-Rev-5	CAAGCAGAAAGACGGCATACGAGATCGTTACCAAGTGTGACTGGAGTTTCAGACGTGTGCTCTTCCGATCTCCGACTCCGGTGCACCTTTTTCAA	SAM sgRNA library NGS and barcode (bold)
NGS-Lib-SAM-Rev-6	CAAGCAGAAAGACGGCATACGAGATGTTGATGTGACTGGAGTTTCAGACGTGTGCTCTTCCGATCTCCGACTCCGGTGCACCTTTTTCAA	SAM sgRNA library NGS and barcode (bold)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Primer	Sequence (5'-3')	Purpose
NGS-Lib-SAM-Rev-7	CAAGCAGAAAGACGGGCATACGAGATTTACGGCACGTGACTGGAGTTCAGAGGTTCAGAGGTGTGCTCTTCCGATCTGCCAAGTTGATAACGGGACTAGCCCTT	SAM sgRNA library NGS and barcode (bold)
NGS-Lib-SAM-Rev-8	CAAGCAGAAAGACGGGCATACGAGATTTGAAATAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCCAAGTTGATAACGGGACTAGCCCTT	SAM sgRNA library NGS and barcode (bold)

**Table 4**

Primers for amplifying target sites to determine percentage of indels by NGS. The custom first round primer consists of a universal sequence followed by an optional stagger sequence (1–10 additional random nucleotides for introducing diversity in low-diversity NGS libraries) and the custom priming site designed to amplify the 100–300bp region centered around the sgRNA cut site. Second round universal, barcoded primers amplify the first round PCR products. We have provided 8 Fwd and 12 Rev second round primers, each with unique barcodes (bold) to provide for a total of 96 different combinations when sequencing indels in a pooled NGS run.

Primer	Sequence (5'–3')	Purpose
NGS-indel-R1-Fwd	CTTCCCTACACGACGCTCTCCGATCT(stagger)[priming_site]	Custom first round amplification for NGS
NGS-indel-R1-Rev	GACTGGAGTTCAGACGTGTGCTCTCCGATCT[priming_site]	Custom first round amplification for NGS
NGS-indel-R2-Fwd-1	AATGATACGGCGACCACCGAGATCTACACT <b>TATAGCCT</b> ACACTCTTCCCTACACGACGCTCTTCC	Universal second round amplification for NGS
NGS-indel-R2-Fwd-2	AATGATACGGCGACCACCGAGATCTACAC <b>ATAGAGGC</b> ACACTCTTCCCTACACGACGCTCTTCC	Universal second round amplification for NGS
NGS-indel-R2-Fwd-3	AATGATACGGCGACCACCGAGATCTACACC <b>TATCCT</b> ACACTCTTCCCTACACGACGCTCTTCC	Universal second round amplification for NGS
NGS-indel-R2-Fwd-4	AATGATACGGCGACCACCGAGATCTACAC <b>GGCTCTGA</b> AACTCTTCCCTACACGACGCTCTTCC	Universal second round amplification for NGS
NGS-indel-R2-Fwd-5	AATGATACGGCGACCACCGAGATCTACAC <b>AGGCGAAG</b> AACTCTTCCCTACACGACGCTCTTCC	Universal second round amplification for NGS
NGS-indel-R2-Fwd-6	AATGATACGGCGACCACCGAGATCTACAC <b>TAATCTTA</b> AACTCTTCCCTACACGACGCTCTTCC	Universal second round amplification for NGS
NGS-indel-R2-Fwd-7	AATGATACGGCGACCACCGAGATCTACAC <b>CAGGACGT</b> ACACTCTTCCCTACACGACGCTCTTCC	Universal second round amplification for NGS
NGS-indel-R2-Fwd-8	AATGATACGGCGACCACCGAGATCTACAC <b>GTACTGAC</b> AACTCTTCCCTACACGACGCTCTTCC	Universal second round amplification for NGS
NGS-indel-R2-Rev-1	CAAGCAGAAGACGGCATAACGAGAT <b>CGAGTAAT</b> GTGACTGGAGTTCAGACGTGTGCTCTTC	Universal second round

Primer	Sequence (5'–3')	Purpose
		amplification for NGS
NGS-indel-R2-Rev-2	CAAGCAGAAGACGGCATAACGAGAT <b>TCTCCGGAGT</b> GACTGGAGTTCAGACGTGTGCTCTTC	Universal second round amplification for NGS
NGS-indel-R2-Rev-3	CAAGCAGAAGACGGCATAACGAGATA <b>AATGAGCGGT</b> GACTGGAGTTCAGACGTGTGCTCTTC	Universal second round amplification for NGS
NGS-indel-R2-Rev-4	CAAGCAGAAGACGGCATAACGAGAT <b>GGAATCTCGT</b> GACTGGAGTTCAGACGTGTGCTCTTC	Universal second round amplification for NGS
NGS-indel-R2-Rev-5	CAAGCAGAAGACGGCATAACGAGAT <b>TTCTGAAATGT</b> GACTGGAGTTCAGACGTGTGCTCTTC	Universal second round amplification for NGS
NGS-indel-R2-Rev-6	CAAGCAGAAGACGGCATAACGAGAT <b>ACGAATTCGT</b> GACTGGAGTTCAGACGTGTGCTCTTC	Universal second round amplification for NGS
NGS-indel-R2-Rev-7	CAAGCAGAAGACGGCATAACGAGAT <b>AGCTTCAGGT</b> GACTGGAGTTCAGACGTGTGCTCTTC	Universal second round amplification for NGS
NGS-indel-R2-Rev-8	CAAGCAGAAGACGGCATAACGAGAT <b>GCGCATTA</b> GTGACTGGAGTTCAGACGTGTGCTCTTC	Universal second round amplification for NGS
NGS-indel-R2-Rev-9	CAAGCAGAAGACGGCATAACGAGAT <b>CATAGCCGGT</b> GACTGGAGTTCAGACGTGTGCTCTTC	Universal second round amplification for NGS
NGS-indel-R2-Rev-10	CAAGCAGAAGACGGCATAACGAGAT <b>TTCCGGAGT</b> GACTGGAGTTCAGACGTGTGCTCTTC	Universal second round amplification for NGS
NGS-indel-R2-Rev-11	CAAGCAGAAGACGGCATAACGAGAT <b>GCGCGAG</b> AGTGGAGTTCAGACGTGTGCTCTTC	Universal second round amplification for NGS
NGS-indel-R2-Rev-12	CAAGCAGAAGACGGCATAACGAGAT <b>CTATCGCT</b> GTGACTGGAGTTCAGACGTGTGCTCTTC	Universal second round amplification for NGS

**Table 5**

Primers for sgRNA cloning and validation.

Primer	Sequence (5'-3')	Purpose
sgRNA-top	CACCGNNNNNNNNNNNNNNNNNN	Top strand primer for cloning sgRNA into sgRNA library backbone; appended guanine in lowercase
sgRNA-bottom	AAACNNNNNNNNNNNNNNNNNNc	Bottom strand primer for cloning sgRNA into sgRNA library backbone; appended cytosine in lowercase
NT1-top	CACCGCTGAAAAAGGAAGGAGTTGA	Top strand primer for cloning non-targetting sgRNA into sgRNA library backbone
NT1-bottom	AAACTCAACTCCTTCCTTTTTCAGC	Bottom strand primer for cloning non-targetting sgRNA into sgRNA library backbone
NT2-top	CACCGAAGATGAAAGGAAAGGCGTT	Top strand primer for cloning non-targetting sgRNA into sgRNA library backbone
NT2-bottom	AAACAACGCCTTTCCTTTCATCTTC	Bottom strand primer for cloning non-targetting sgRNA into sgRNA library backbone
U6-Fwd	CGTGACGTAGAAAGTAATAATTCTTGGG	Sanger sequencing of sgRNA cloned into sgRNA library backbone



Table 6

Troubleshooting table.

Step	Problem	Possible reason	Possible solution
26	Insufficient number of colonies	Inefficient electroporation; plates dried out during bacteria spreading; salt contamination or low concentration of DNA for electroporation (custom library only).	Increase number of electroporations if feasible; follow the manufacturer's instructions for optimal efficiency (i.e. ensure cuvettes, tubes and cells are chilled thoroughly and recovery media is added immediately after electroporation); spread bacterial suspension evenly without letting plates dry; limit plate warming to <2h and add 1 volume of additional LB media before spreading as recommended; if the DNA for electroporation has a 260/230 ratio <1.8 or is <50 ng $\mu\text{l}^{-1}$ by NanoDrop, consider re-prepping (custom library only).
27	Colonies growing on negative control plate	Incomplete digestion of library plasmid backbone	Increase the amount of restriction enzyme and phosphatase, ensure added DTT is fresh and stored at $-20\text{ }^{\circ}\text{C}$ as recommended
39	Less than 70% of perfectly matched sgRNAs	NGS reads are low quality; Incorrect flag for presence of 'G' before guide spacer sequence; Low oligo synthesis quality (custom library only)	If <75% of clusters passed filter, consider resequencing the library with more optimal cluster density and cluster passing filter; Check for the proper flag indicating presence of 'G'; if both the negative gibson control and sequencing quality were good, contact your synthesis company or core to discuss your results
39	More than 0.5% of sgRNAs are undetected	Insufficient colonies during amplification process; Not enough cloning reactions during custom sgRNA library cloning	Increase the number of electroporations; scale up the number of reactions throughout cloning
39	Skew ratio of custom sgRNA library is greater than 10	Insufficient representation and potential bias during cloning process; Low oligo synthesis quality (custom library only)	Reduce the number of PCR cycles during oligo amplification; increase the number of reactions throughout cloning, ensure >500 colonies per guide during library amplification; contact your synthesis company or core to discuss your results.
51	Low survival following Lipofectamine transfection	Cell density at transfection <75%; failure to replace media within 6 h post transfection	Ensure even plating and a cell density of 80–90% during transfection. Replace media 4 h after transfection. The cell density should not be <80% at the time of lentivirus harvest.
51	Low Lipofectamine transfection efficiency	Incorrect amount or poor quality DNA used for transfection; poorly or unevenly seeded cells	Use low-passage-number cells (passage number <10) and do not let cells reach >90% confluency; reseed cells evenly at recommended density; prepare new DNA for transfection
53	Significant cell survival in the no virus, antibiotic-selected control	Incorrect antibiotic concentration	Construct a kill curve for the antibiotic and cell line of interest by titrating different concentrations of the antibiotic on cells seeded at 10% confluency. Choose the lowest concentration that kills all cells after 4–7 days. When using zeocin, only 90–95% of cells may be killed at the optimal concentration after 7 days, but this should not interfere with screening when using zeocin for selection of SAM sgRNAs.
53	Low lentivirus titer	Low lentivirus production or transduction efficiency	Concentrate the lentivirus using Amicon Ultra-15 Centrifugal Filter Unit with Ultracel-100 membrane or ultracentrifugation at $88,000 \times g$ for 2 h at $4\text{ }^{\circ}\text{C}$ . Discard the supernatant and resuspend the lentivirus in cold PBS at $4\text{ }^{\circ}\text{C}$ overnight.
58	No amplification of gDNA	Incorrect template or primer; incorrect template or primer concentration	Use appropriate NGS-Rev primer corresponding to the screen; titrate gDNA concentration to 1–3 $\mu\text{g}$ for a 50 $\mu\text{l}$ reaction; titrate primer concentration to a final concentration of 0.1–0.5 $\mu\text{M}$
58	Primer dimer after gDNA amplification	Insufficient number of PCR cycles	Increase the number of cycles for PCR up to 25; lower gDNA input and perform more reactions
76	Very few or no colonies on plates	Incorrect template; Low transformation efficiency	Use an appropriate antibiotic for the resistance marker on the plasmid backbone; verify competency of <i>E. coli</i>
80	Low percentage of correctly cloned sgRNAs	Incorrect template or primer; incorrect Golden Gate Assembly reaction	Check that the primer design is correct and that the top and bottom strands are complementary; verify restriction enzyme site on plasmid backbone; ensure use of the

Step	Problem	Possible reason	Possible solution
94	Multiple bands or no amplification during indel analysis	Inefficient or nonspecific priming; incorrect template concentration	suggested restriction enzyme (Fermentas Fast Digest) and DNA ligase (Enzymatics T7 ligase)  Test different annealing temperatures (58–62 °C); re-design primers to minimize primer dimer formation or nonspecific amplification; test a different high fidelity DNA polymerase; vary the magnesium concentration in the reaction. If none of the above steps solve the issue it may be helpful to clean up the Quickextract gDNA using Ampure beads to provide a cleaner PCR template.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript