

# Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening

Julia Joung<sup>1–4,7</sup>, Silvana Konermann<sup>2–4,7,8</sup>, Jonathan S Gootenberg<sup>2–5</sup>, Omar O Abudayyeh<sup>2–4,6</sup>, Randall J Platt<sup>2–4,8</sup>, Mark D Brigham<sup>2–4</sup>, Neville E Sanjana<sup>2–4,8</sup> & Feng Zhang<sup>1–4</sup>

<sup>1</sup>Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. <sup>2</sup>Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. <sup>3</sup>McGovern Institute for Brain Research at MIT, Cambridge, Massachusetts, USA. <sup>4</sup>Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. <sup>5</sup>Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA. <sup>6</sup>Department of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. <sup>7</sup>These authors contributed equally to this work. <sup>8</sup>Present addresses: Salk Institute for Biological Studies, La Jolla, California, USA (S.K.); Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland (R.J.P.); New York Genome Center, New York, New York, USA (N.E.S.); Department of Biology, New York University, New York, New York, USA (N.E.S.). Correspondence should be addressed to F.Z. (zhang@broadinstitute.org).

Published online 23 March 2017; doi:10.1038/nprot.2017.016

**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 by 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.**

## 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 use 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 a low signal-to-noise ratio 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 immune system has proven 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.

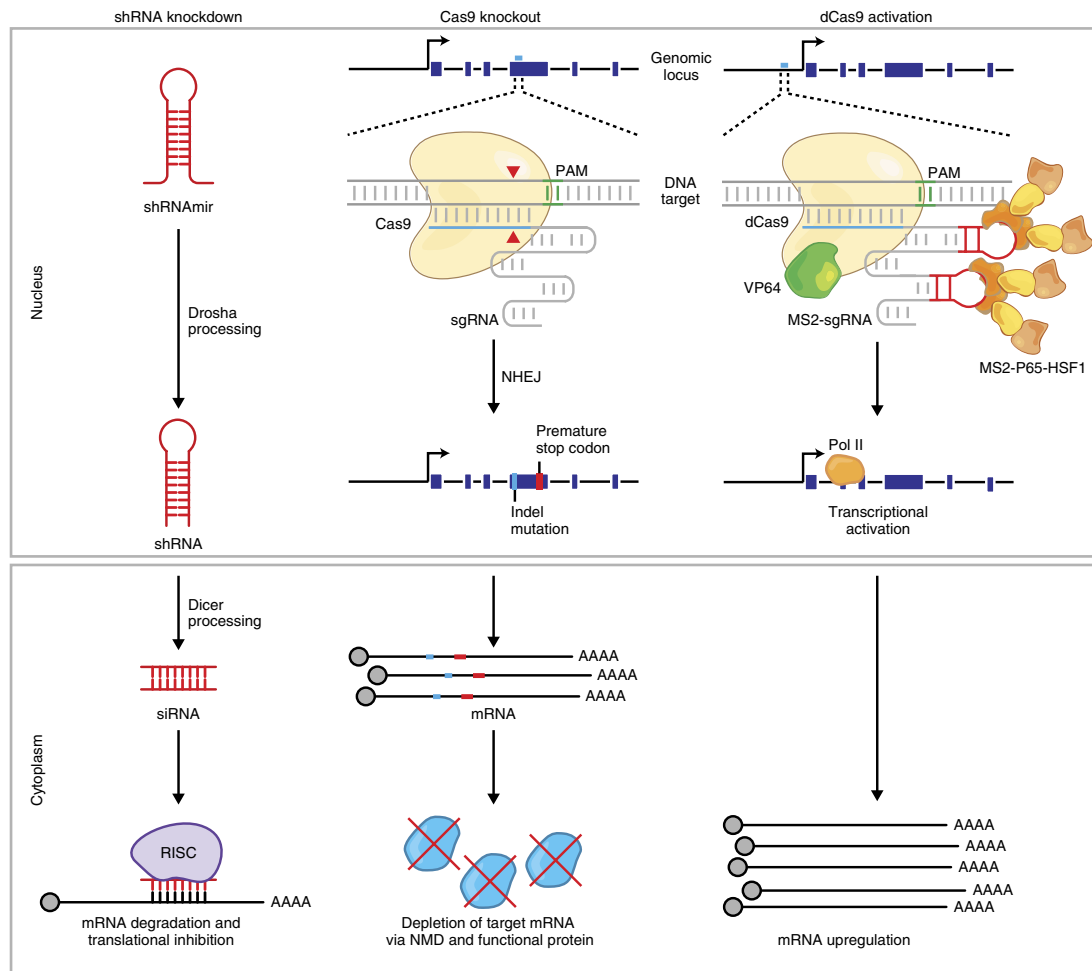
Cas9 generates precise double-strand breaks (DSBs) at target loci that are repaired through either homology-directed repair or, more often, nonhomologous end-joining (NHEJ)<sup>26</sup>. Homology-directed repair precisely repairs the DSB using a homologous

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 nonfunctional 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 substantial 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



**Figure 1** | Approaches to genetic perturbation: shRNA knockdown, Cas9 knockout, and dCas9 transcriptional activation. Schematic of the mechanisms behind shRNA knockdown, Cas9 knockout, and dCas9 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 nonfunctional protein or non-sense-mediated decay (NMD) of the mRNA transcript. Programmable transcriptional activation can be achieved by 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. NHEJ, nonhomologous end joining; PAM, protospacer adjacent motif; Pol II, RNA polymerase II; shRNAmir, microRNA-adapted shRNA.

with single-chain variable fragment antibodies<sup>28</sup>. Another activation method, fuses three activation domains, VP64-p65-Rta (VPR), in tandem to dCas9 to enhance transcriptional activation<sup>27</sup>.

We devised an alternative approach to CRISPRa that involves incorporating MS2 binding loops into 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 investigate 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 nonmetastatic 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>.

**TABLE 1** | Previously published screens using Cas9.

Type of screen	Selection	Organism	Cas9 variant	References
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 <i>BCL11A</i> 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	<i>POU5F1</i> 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 (melanoma cell line); (negative); 6-thioguanine resistance (positive); interferon survival (positive)	<i>In vitro</i> ; A375 (human 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	<i>CUL3</i> 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/HSF 1	30
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

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 by 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).

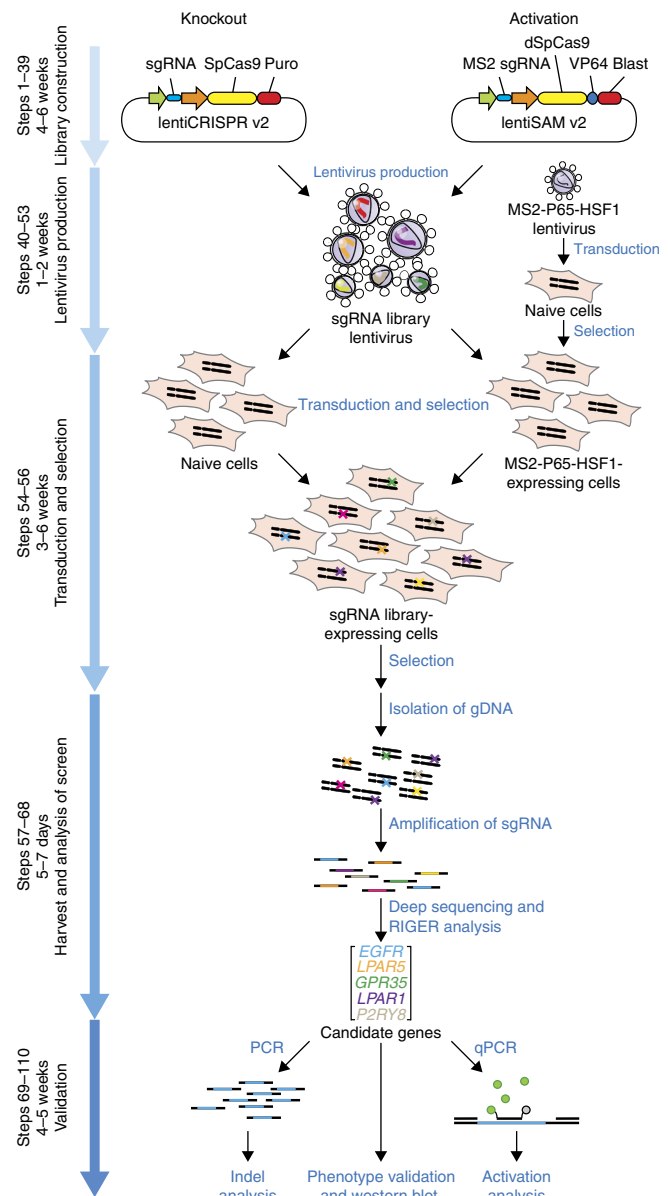
### Comparison with other pooled screening technologies

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 did 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 nonoverlapping gene sets and may have different false-positive or false-negative results<sup>64</sup>. Although Cas9 knockout screening is an enormously valuable LOF screening method, shRNA and CRISPRi screening may complement Cas9 knockout screening to further investigate 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 aliquotted into separate wells in multiwell plates. This format allows for measuring of a diverse range of phenotypes such as fluorescence and luminescence, or even for 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 they 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).

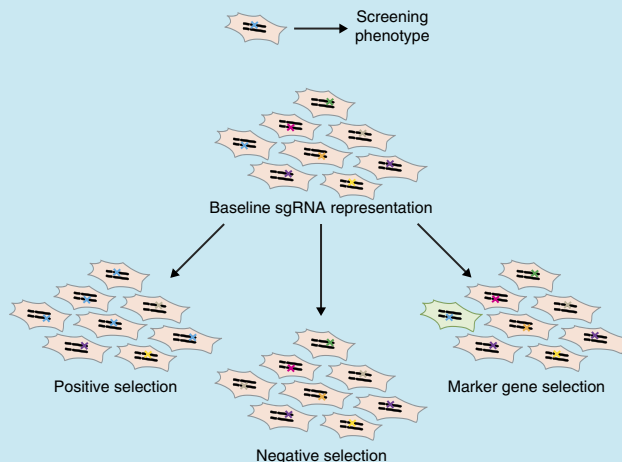


**Figure 2** | Timeline and overview of experiments. Genome-scale Cas9 knockout and transcriptional activation screens begin with the construction of a plasmid library encoding the effector protein and sgRNAs. These plasmid libraries are packaged into lentivirus and then transduced into the cell type 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 time points 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 of individual sgRNAs for the screening phenotype, indel formation by targeted sequencing, and transcript upregulation by qPCR. Blast, blasticidin; Puro, puromycin.

Regardless of the type of screening selection, next-generation sequencing (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.

## Box 1 | Different types of selection applicable to 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:



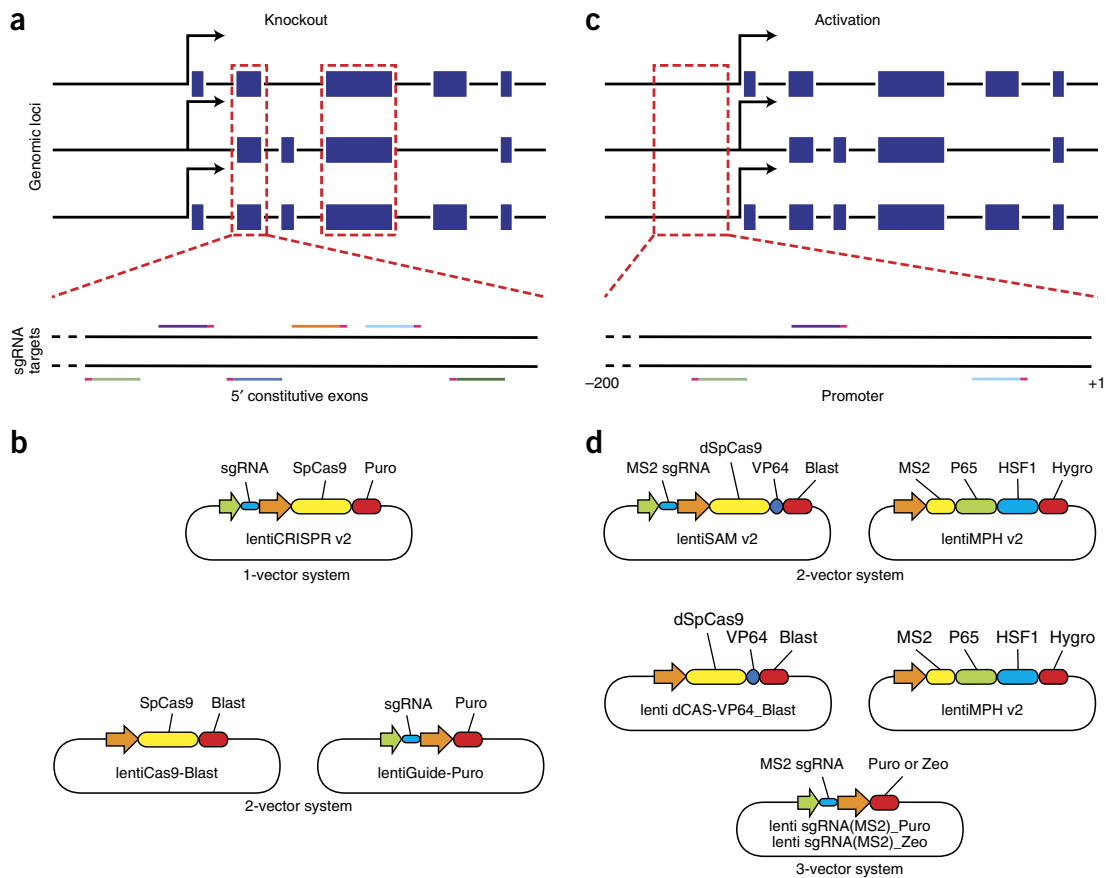
Positive selection screens rely on enrichment of sgRNAs for genetic perturbations that produce the screening phenotype as a result of cell proliferation. These typically have the highest signal-to-noise ratio compared with other types of screens, because the number of phenotypically relevant sgRNAs increases 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.

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 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 probably 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 nontargeting guides that do not target the genome should be included as negative controls. Nontargeting guides are critical to 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 as compared with those in the control condition, whereas the nontargeting guides should remain relatively unchanged between the 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 four sgRNAs per miRNA. Each species-specific library contains 1,000 nontargeting 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 200-bp region upstream of the transcriptional start site of 23,430 human or 23,439 mouse RefSeq coding isoforms with three sgRNAs per isoform (Fig. 3c)<sup>30</sup>. The library must be combined with additional SAM effectors in a 2-vector (lentiSAM v2 and lentiMPH v2) or 3-vector (lenti dCAS-VP64\_Blast, lentiMPH v2, 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 in which a subset of genes is known to be involved in the screening phenotype and/or when the cell number is limited, one



**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 200-bp 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 must be combined with additional SAM effectors in a 2-vector or 3-vector format. Both libraries select sgRNAs with minimal off-target activity.

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. In addition, 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 NGS before 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 nondividing cells. By contrast, retrovirus transduces only dividing cells. In addition, AAV has a smaller insert size capacity as compared with 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.** As 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

## 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 should be 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 biasing the screening results. Try to maintain sufficient coverage (>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.

of the screening selection pressure, at the end of a successful screen, the sgRNA library distribution in the experimental condition should be more skewed as compared with 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 more skewed as compared with the nontargeting 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 toward 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

## 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 the cells *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) (ref. 44). 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 in which these conditions cannot be met, adapt the screening strategy by, for instance, reducing the library size to capture a subset of genes.

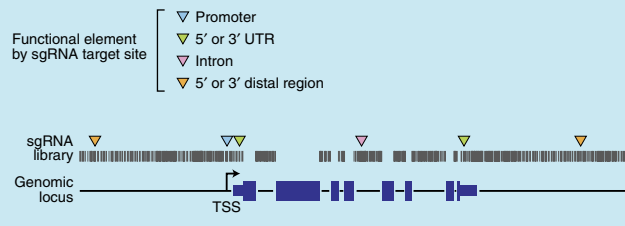
*In vivo* screening is performed with either a) transduction of cells *in vitro* followed by *in vitro* cell transplantation, or b) direct transduction of tissues *in vitro*. The first strategy was demonstrated by Chen *et al.*, whereby a cancer cell line was transduced with a CRISPR knockout library 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 vitro*. 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 vitro*. 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 vitro* 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?

## 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). By contrast, CRISPR screens enable direct testing of how mutagenesis at a specific noncoding site affects phenotype.

Several strategies can be used to design libraries to target noncoding regions. To enable understanding of the 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 such as p53 (ref. 54). 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.

identifying positively and negatively selected genes and pathways using a 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. As 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 by 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 substantially lower than those of the rest of the genes (Fig. 4e,f).

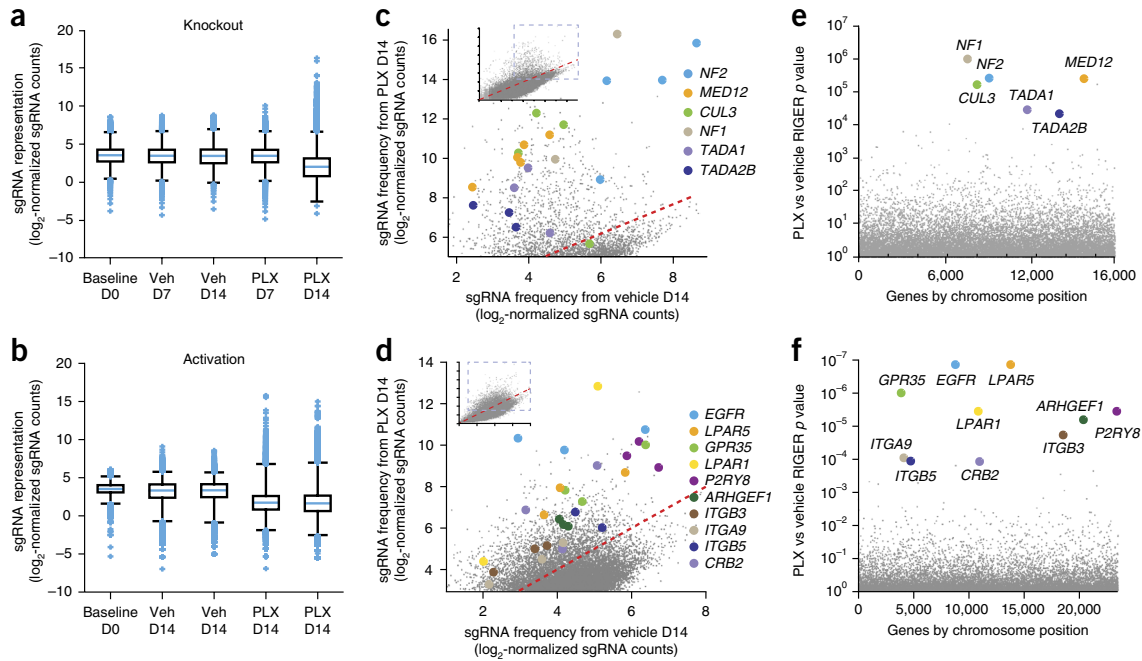
**Validation of candidate genes.** Given that the screening process can be noisy and the analysis produces a ranked list of candidate genes, it is necessary to verify that perturbation of the identified candidate genes confers the phenotype of interest. For validation, each of the sgRNAs that target 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 with SURVEYOR, which we have

described previously<sup>72</sup>, NGS is more suitable to sampling a large number of sgRNA target sites and therefore is 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), which 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, whereas 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.





**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 A375, a BRAF V600E 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. Multiple guides are seen to be enriched in the PLX day 14 (D14) condition, revealing guides that promote resistance to BRAF inhibitor. Boxes, 25th to 75th percentiles; whiskers, 1st to 99th percentiles. **(c,d)** Scatter plot showing enrichment of sgRNAs targeting the top candidate genes identified by RIGER (colored dots) compared with other sgRNAs in the library (gray dots) after PLX treatment. Each gene has multiple sgRNAs that are enriched. Inset panels represent the entire data set. 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 substantially lower as compared with other genes (gray dots) targeted by the sgRNA library.

Measurement of transcriptional activation usually entails isolation of RNA, reverse transcription of the RNA to cDNA, and quantitative PCR (qPCR). Various 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 downregulation or upregulation of protein expression for knockout or transcriptional activation screening, respectively. Immunohistochemistry and western blotting 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 blotting 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 blotting because proteins are not separated by size. Thus, western blotting 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)
- lentiMPH v2 (Addgene, cat. no. 89308)
- 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 (listed in **Table 2**)

TABLE 2 | Primers for amplifying the sgRNA oligo library.

Primer	Sequence (5'–3')	Purpose
Oligo-Fwd	GTAACCTGAAAGTATTTTCGATTCTT GGCTTTATATATCTTGTGAAAGGAC GAAACACC	Targeted knockout or activation library cloning
Oligo-Knockout-Rev	ACTTTTCAAGTTGATAACGGACTAG CCTTATTTAACTTGCTATTCTAGCT CTA AAAAC	Targeted knockout library cloning
Oligo-Activation-Rev	ATTTAACTTGCTAGGCCCTGCAGAC ATGGGTGATCCTCATGTTGGCCTAGC TCTAAAAC	Targeted activation library cloning

Primers longer than 60 bp 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 NEBNext High-Fidelity PCR Master Mix.
- UltraPure DNase/RNase-free dH<sub>2</sub>O (Thermo Fisher Scientific, cat. no. 10977023)
- QIAquick PCR Purification Kit (Qiagen, cat. no. 28104)
- QIAquick Gel Extraction Kit (Qiagen, cat. no. 28704)
- UltraPure TBE Buffer, 10× (Thermo Fisher Scientific, cat. no. 15581028)
- SeaKem LE agarose (Lonza, cat. no. 50004)
- SYBR Safe DNA stain, 10,000× (Thermo Fisher Scientific, cat. no. S33102)
- 1-kb Plus DNA ladder (Thermo Fisher Scientific, cat. no. 10787018)
- 50-bp DNA ladder (Thermo Fisher Scientific, cat. no. 10416014)
- TrackIt Cyan/Orange Loading Buffer (Thermo Fisher Scientific, cat. no. 10482028)
- FastDigest Esp3I (BsmBI; Thermo Fisher Scientific, cat. no. FD0454)
- FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific, cat. no. EF0651)
- DTT, Molecular Grade (Promega, cat. no. P1171)
- Gibson Assembly Master Mix (New England BioLabs, cat. no. E2611L)
- GlycoBlue Coprecipitant (Thermo Fisher Scientific, 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> colony-forming units (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 Maxi Kit, may be used as substitutes.
- One Shot Stbl3 chemically competent *Escherichia coli* (Thermo Fisher Scientific, 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 Midi Kit, may be used as substitutes.

**Next-generation sequencing**

- Primers (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 60 bp 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 Scientific, cat. no. Q32851)

TABLE 3 | Primer sequences for amplifying sgRNA library and NGS.

Primer	Sequence (5'–3')	Purpose
NGS-Lib-Fwd-1	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCTACACGACGCTCTTCC GATCTTAAGTAGAGGCTTTATATATCT TGTGAAAGGACGAAACACC	GeCKO or SAM sgRNA library NGS
NGS-Lib-Fwd-2	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCTACACGACGCTCTTCC GATCTATCATGCTTAGCTTTATATATC TTGTGAAAGGACGAAACACC	GeCKO or SAM sgRNA library NGS
NGS-Lib-Fwd-3	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCTACACGACGCTCTTCC GATCTGATGCATCTGCTTTATATAT CTTGTGAAAGGACGAAACACC	GeCKO or SAM sgRNA library NGS
NGS-Lib-Fwd-4	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCTACACGACGCTCTTCC GATCTCGATTGCTCGACGCTTTATATA TCTTGTGAAAGGACGAAACACC	GeCKO or SAM sgRNA library NGS
NGS-Lib-Fwd-5	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCTACACGACGCTCTTCC GATCTTCGATAGCAATTCGCTTTATAT ATCTTGTGAAAGGACGAAACACC	GeCKO or SAM sgRNA library NGS
NGS-Lib-Fwd-6	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCTACACGACGCTCTTCC GATCTATCGATAGTTGCTTGCTTTATA TATCTTGTGAAAGGACGAAACACC	GeCKO or SAM sgRNA library NGS
NGS-Lib-Fwd-7	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCTACACGACGCTCTTCC GATCTGATCGATCCAGTTAGGCTTTAT ATATCTTGTGAAAGGACGAAACACC	GeCKO or SAM sgRNA library NGS

(continued)

**TABLE 3** | Primer sequences for amplifying sgRNA library and NGS (continued).

Primer	Sequence (5'–3')	Purpose
NGS-Lib-Fwd-8	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCTACACGACGCTCTCC GATCTCGATCGATTTGAGCCTGCTTTA TATATCTTGTGGAAAGGACGAAACAC C	GeCKO or SAM sgRNA library NGS
NGS-Lib-Fwd-9	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCTACACGACGCTCTCC GATCTACGATCGATACACGATCGCTTT ATATATCTTGTGGAAAGGACGAAACA CC	GeCKO or SAM sgRNA library NGS
NGS-Lib-Fwd-10	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCTACACGACGCTCTCC GATCTTACGATCGATGGTCCAGAGCTT TATATATCTTGTGGAAAGGACGAAAC ACC	GeCKO or SAM sgRNA library NGS
NGS-Lib-KO-Rev- 1	CAAGCAGAAGACGGCATAACGAGATTC <b>GCCTTGGTGACTGGAGTTCAGACGTG</b> TGCTCTCCGATCTCCGACTCGGTGCC ACTTTTTCAA	GeCKO sgRNA library NGS and barcode (bold)
NGS-Lib-KO-Rev- 2	CAAGCAGAAGACGGCATAACGAGATAT <b>AGCGTCGTGACTGGAGTTCAGACGTG</b> TGCTCTCCGATCTCCGACTCGGTGCC ACTTTTTCAA	GeCKO sgRNA library NGS and barcode (bold)
NGS-Lib-KO-Rev- 3	CAAGCAGAAGACGGCATAACGAGATGA <b>AGAAGTGTGACTGGAGTTCAGACGTG</b> TGCTCTCCGATCTCCGACTCGGTGCC ACTTTTTCAA	GeCKO sgRNA library NGS and barcode (bold)
NGS-Lib-KO-Rev- 4	CAAGCAGAAGACGGCATAACGAGATAT <b>TCTAGGGTGACTGGAGTTCAGACGTG</b> TGCTCTCCGATCTCCGACTCGGTGCC ACTTTTTCAA	GeCKO sgRNA library NGS and barcode (bold)
NGS-Lib-KO-Rev- 5	CAAGCAGAAGACGGCATAACGAGATCG <b>TTACCA</b> GTGACTGGAGTTCAGACGTG TGCTCTCCGATCTCCGACTCGGTGCC ACTTTTTCAA	GeCKO sgRNA library NGS and barcode (bold)
NGS-Lib-KO-Rev- 6	CAAGCAGAAGACGGCATAACGAGATGT <b>CTGATGGTGACTGGAGTTCAGACGTG</b> TGCTCTCCGATCTCCGACTCGGTGCC ACTTTTTCAA	GeCKO sgRNA library NGS and barcode (bold)
NGS-Lib-KO-Rev- 7	CAAGCAGAAGACGGCATAACGAGATTT <b>ACGCACGTGACTGGAGTTCAGACGTG</b> TGCTCTCCGATCTCCGACTCGGTGCC ACTTTTTCAA	GeCKO sgRNA library NGS and barcode (bold)
NGS-Lib-KO-Rev- 8	CAAGCAGAAGACGGCATAACGAGATTT <b>GAATAGGTGACTGGAGTTCAGACGTG</b> TGCTCTCCGATCTCCGACTCGGTGCC ACTTTTTCAA	GeCKO sgRNA library NGS and barcode (bold)
NGS-Lib-SAM-Rev-1	CAAGCAGAAGACGGCATAACGAGATTC <b>GCCTTGGTGACTGGAGTTCAGACGTG</b> TGCTCTCCGATCTGCCAAGTTGATAA CGGACTAGCCTT	SAM sgRNA library NGS and barcode (bold)
NGS-Lib-SAM-Rev-2	CAAGCAGAAGACGGCATAACGAGATAT <b>AGCGTCGTGACTGGAGTTCAGACGTG</b> TGCTCTCCGATCTGCCAAGTTGATAA CGGACTAGCCTT	SAM sgRNA library NGS and barcode (bold)
NGS-Lib-SAM-Rev-3	CAAGCAGAAGACGGCATAACGAGATGA <b>AGAAGTGTGACTGGAGTTCAGACGTG</b> TGCTCTCCGATCTGCCAAGTTGATAA CGGACTAGCCTT	SAM sgRNA library NGS and barcode (bold)
NGS-Lib-SAM-Rev-4	CAAGCAGAAGACGGCATAACGAGATAT <b>TCTAGGGTGACTGGAGTTCAGACGTG</b> TGCTCTCCGATCTGCCAAGTTGATAA CGGACTAGCCTT	SAM sgRNA library NGS and barcode (bold)
NGS-Lib-SAM-Rev-5	CAAGCAGAAGACGGCATAACGAGATCG <b>TTACCA</b> GTGACTGGAGTTCAGACGTG TGCTCTCCGATCTGCCAAGTTGATAA CGGACTAGCCTT	SAM sgRNA library NGS and barcode (bold)
NGS-Lib-SAM-Rev-6	CAAGCAGAAGACGGCATAACGAGATGT <b>CTGATGGTGACTGGAGTTCAGACGTG</b> TGCTCTCCGATCTGCCAAGTTGATAA CGGACTAGCCTT	SAM sgRNA library NGS and barcode (bold)
NGS-Lib-SAM-Rev-7	CAAGCAGAAGACGGCATAACGAGATTT <b>ACGCACGTGACTGGAGTTCAGACGTG</b> TGCTCTCCGATCTGCCAAGTTGATAA CGGACTAGCCTT	SAM sgRNA library NGS and barcode (bold)
NGS-Lib-SAM-Rev-8	CAAGCAGAAGACGGCATAACGAGATTT <b>GAATAGGTGACTGGAGTTCAGACGTG</b> TGCTCTCCGATCTGCCAAGTTGATAA CGGACTAGCCTT	SAM sgRNA library NGS and barcode (bold)

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 the sgRNA library for NGS. The eight NGS-Lib-Rev primers provide unique barcodes (bold) for distinguishing different sgRNA libraries (e.g., experimental and control conditions for four 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. As 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.

- 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 the cases of

**TABLE 4** | Primers for amplifying target sites to determine percentage of indels by NGS.

Primer	Sequence (5'–3')	Purpose
NGS-indel-R1-Fwd	CTTCCCTACACGACGCTCTCCGATC T(stagger)[priming_site]	Custom first-round amplification for NGS
NGS-indel-R1-Rev	GACTGGAGTTCAGACGTGTGCTCTTC CGATCT[priming_site]	Custom first-round amplification for NGS
NGS-indel-R2-Fwd-1	AATGATACGGCGACCACCGAGATCTA CACT <b>TATAGCCT</b> ACACTCT TCCCTAC ACGACGCTCTTC	Universal second-round amplification for NGS
NGS-indel-R2-Fwd-2	AATGATACGGCGACCACCGAGATCTA CAC <b>ATAGAGGC</b> ACACTCT TCCCTAC ACGACGCTCTTC	Universal second-round amplification for NGS
NGS-indel-R2-Fwd-3	AATGATACGGCGACCACCGAGATCTA CACC <b>TATCCT</b> ACACTCT TCCCTAC ACGACGCTCTTC	Universal second-round amplification for NGS
NGS-indel-R2-Fwd-4	AATGATACGGCGACCACCGAGATCTA CAC <b>GGCTCTGA</b> CAACTCTT TCCCTAC ACGACGCTCTTC	Universal second-round amplification for NGS
NGS-indel-R2-Fwd-5	AATGATACGGCGACCACCGAGATCTA CAC <b>AGGCGAAG</b> ACACTCTT TCCCTAC ACGACGCTCTTC	Universal second-round amplification for NGS
NGS-indel-R2-Fwd-6	AATGATACGGCGACCACCGAGATCTA CACT <b>AATCTT</b> AACACTCTT TCCCTAC ACGACGCTCTTC	Universal second-round amplification for NGS
NGS-indel-R2-Fwd-7	AATGATACGGCGACCACCGAGATCTA CACC <b>AGGACG</b> TACACTCT TCCCTAC ACGACGCTCTTC	Universal second-round amplification for NGS
NGS-indel-R2-Fwd-8	AATGATACGGCGACCACCGAGATCTA CAC <b>GTA</b> CTGACACACTCTT TCCCTAC ACGACGCTCTTC	Universal second-round amplification for NGS
NGS-indel-R2-Rev-1	CAAGCAGAAGACGGCATA <b>CAGATC GAGTA</b> ATGTGACTGGAGT TCAGACG TGTGCTCTTC	Universal second-round amplification for NGS
NGS-indel-R2-Rev-2	CAAGCAGAAGACGGCATA <b>CAGATT CTCCG</b> AGTGACTGGAGTT CAGACG TGTGCTCTTC	Universal second-round amplification for NGS
NGS-indel-R2-Rev-3	CAAGCAGAAGACGGCATA <b>CAGATA ATGAG</b> CGGTGACTGGAG TTCAGACG TGTGCTCTTC	Universal second-round amplification for NGS
NGS-indel-R2-Rev-4	CAAGCAGAAGACGGCATA <b>CAGATG GAAT</b> CTCGTGACTGGAG TTCAGACG TGTGCTCTTC	Universal second-round amplification for NGS
NGS-indel-R2-Rev-5	CAAGCAGAAGACGGCATA <b>CAGATT TCTGA</b> ATGTGACTGGAG TTCAGACG TGTGCTCTTC	Universal second-round amplification for NGS
NGS-indel-R2-Rev-6	CAAGCAGAAGACGGCATA <b>CAGATA CGAAT</b> TCGTGACTGGAG TTCAGACG TGTGCTCTTC	Universal second-round amplification for NGS
NGS-indel-R2-Rev-7	CAAGCAGAAGACGGCATA <b>CAGATA GCTT</b> CAGGTGACTGGAG TTCAGACG TGTGCTCTTC	Universal second-round amplification for NGS
NGS-indel-R2-Rev-8	CAAGCAGAAGACGGCATA <b>CAGATG CGC</b> ATTAGTGACTGGAG TTCAGACG TGTGCTCTTC	Universal second-round amplification for NGS
NGS-indel-R2-Rev-9	CAAGCAGAAGACGGCATA <b>CAGATC ATAG</b> CCGGTGACTGGAG TTCAGACG TGTGCTCTTC	Universal second-round amplification for NGS
NGS-indel-R2-Rev-10	CAAGCAGAAGACGGCATA <b>CAGATT TCG</b> CGAGTGACTGGAG TTCAGACG TGTGCTCTTC	Universal second-round amplification for NGS
NGS-indel-R2-Rev-11	CAAGCAGAAGACGGCATA <b>CAGATG CGG</b> GAGTGACTGGAG TTCAGACG TGTGCTCTTC	Universal second-round amplification for NGS
NGS-indel-R2-Rev-12	CAAGCAGAAGACGGCATA <b>CAGATC TAT</b> CGCTGTGACTGGAG TTCAGACG TGTGCTCTTC	Universal second-round amplification for 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- to 300-bp region centered around the sgRNA cut site. Second-round universal, barcoded primers amplify the first-round PCR products. We have provided 8 forward (Fwd) and 12 reverse (Rev) second-round primers, each with unique barcodes (bold) to provide a total of 96 different combinations when sequencing indels in a pooled NGS run.

**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	CACCGCTGAAAAGGAAGGAGTTGA	Top-strand primer for cloning nontargeting sgRNA into sgRNA library backbone
NT1-bottom	AAACTCAACTCCTCTCTTTTCAGC	Bottom-strand primer for cloning nontargeting sgRNA into sgRNA library backbone
NT2-top	CACCGAAGATGAAAGGAAAGGCGTT	Top-strand primer for cloning nontargeting sgRNA into sgRNA library backbone
NT2-bottom	AAACAACGCCCTTCTTTCATCTTC	Bottom-strand primer for cloning nontargeting sgRNA into sgRNA library backbone
U6-Fwd	CGTGACGTAGAAAGTAATAATTCTTGGG	Sanger sequencing of sgRNA cloned into sgRNA library backbone

contact with skin, contact with eyes, ingestion, and inhalation. Wear protective clothing and avoid contact.

- Tris, pH 7.0 (Thermo Fisher Scientific, cat. no. AM9850G)

**Mammalian cell culture**

- HEK293FT (Thermo Fisher Scientific, cat. no. R70007) **! CAUTION** Check cell lines regularly to ensure that they are authentic and are not infected with mycoplasma.
- DMEM, high glucose, GlutaMAX supplement, pyruvate (Thermo Fisher Scientific, cat. no. 10569010)
- Penicillin–streptomycin, 100× (Thermo Fisher Scientific, cat. no. 15140122)
- FBS, premium grade (VWR, cat. no. 97068-085)
- TrypLE Express, no phenol red (Thermo Fisher Scientific, cat. no. 12604021)
- HUES 66 cell line (Harvard Stem Cell Science) **! CAUTION** Check cell lines regularly to ensure that they are authentic and that they are not infected with mycoplasma.
- Geltrex LDEV-free reduced growth factor basement membrane matrix (Thermo Fisher Scientific, 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 Scientific, cat. no. 14190250)

**Lentivirus production and titer**

- Opti-MEM I reduced serum medium (Thermo Fisher Scientific, 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 Scientific, cat. no. 11668019)
- PLUS Reagent (Thermo Fisher Scientific, 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 Scientific, cat. no. A1113903)
- Puromycin dihydrochloride (Thermo Fisher Scientific, cat. no. A1113803)
- Hygromycin B (Thermo Fisher Scientific, cat. no. 10687010)
- Zeocin (Thermo Fisher Scientific, 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 (Integrated DNA Technologies; these primers are listed in **Table 5**)
- 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 Scientific, cat. no. AM9855G)
- Deoxyribonuclease I bovine (Sigma-Aldrich, cat. no. D2821-50KU)
- UltraPure 1 M Tris–HCl Buffer, pH 7.5 (Thermo Fisher Scientific, 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 Scientific, cat. no. AM9530G)
- Triton X-114 (Sigma-Aldrich, cat. no. X114-100ML)
- Proteinase K Inhibitor (EMD Millipore, cat. no. 539470-10MG)
- DMSO (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 Scientific, cat. no. K1691)
- Oligo dT (TTTTTTTTTTTTTTTTTTTTTNN; Integrated DNA Technologies)
- TaqMan target probes, FAM dye (Thermo Fisher Scientific)
- Taqman endogenous control probe, VIC dye (e.g., Human GAPD, GAPDH, Endogenous Control VIC/MGB probe, primer-limited; Thermo Fisher Scientific, cat. no. 4326317E)
- TaqMan Fast Advanced Master Mix, 2× (Thermo Fisher Scientific, 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)
- 245-mm 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 Fisher Scientific, cat. no. 156367)
- Nunc EasYFlask 75 cm<sup>2</sup>, filter cap, 25-ml working volume, (T75 flask; Thermo Fisher Scientific, cat. no. 156499)
- Nunc EasYFlask 225 cm<sup>2</sup>, filter cap, 70-ml working volume (T225 flask; Thermo Fisher 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, 100 mm (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 (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 Fisher Scientific)
- Plate spectrophotometer (e.g., Synergy H4 Hybrid Multi-Mode Microplate Reader; BioTek)
- Qubit Assay Tubes (Thermo Fisher Scientific, cat. no. Q32856)
- Qubit Fluorometer (Thermo Fisher Scientific, 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 Scientific, 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 dH<sub>2</sub>O to a 1× working concentration, 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>) proteinase K. Store it 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 the solution at –20 °C for up to 2 years.

Component	Amount (µl)	Final concentration
Tris–HCl, pH 7.5, 1 M	250	50 mM
CaCl <sub>2</sub> , 1 M	50	10 mM
Glycerol	2,500	50% (vol/vol)
UltraPure water	2,200	
Total	5,000	

**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 ~7.8. Store it at 4 °C for up to 1 year. **▲ CRITICAL** Prepare the solution under RNase-free conditions.

Component	Amount (ml)	Final concentration
Tris, pH 8.0, 1 M	1.2	4.8 mM
Tris, pH 7.5, 1 M	1.2	4.8 mM
MgCl <sub>2</sub> , 1 M	0.125	0.5 mM
CaCl <sub>2</sub> , 1 M	0.110	0.44 mM
DTT, 0.1 M	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 the solution under RNase-free conditions. Take aliquots to measure the pH in order to keep the main stock from becoming RNase-contaminated by the pH probe.

Component	Amount	Final concentration
EGTA	9.5 g	0.5 M
Tris, pH 8.0, 1 M	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 the solution 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. Divide the solution into aliquots in 8-strip PCR tubes to avoid freeze–thaw cycles and facilitate sample processing with multichannel pipettes. Store aliquots at –20 °C for up to 1 year. **▲ CRITICAL** Prepare the 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 µM
UltraPure water	12.14	
Total	15	

**Oligo dT, 100 µM** Resuspend oligo dT to 100 µM in UltraPure water. Prepare aliquots and store them 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 g liter<sup>-1</sup> in deionized water and swirl to mix. Autoclave the mixture to sterilize. Allow the LB agar to cool to 55 °C before adding ampicillin to a final concentration of 100 µg ml<sup>-1</sup> and swirl to mix. On a sterile bench area, pour ~300 ml of LB

## PROTOCOL

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 them sit for several more hours or overnight. Agar plates can be stored in plastic bags or sealed with Parafilm and stored 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 that for large LB agar plates, except ~20 ml of LB agar should be poured per 100-mm Petri dish. Store the plates at 4 °C for up to 3 months.

### PROCEDURE

#### Designing a custom sgRNA library ● **TIMING** 3–5 weeks; 5 d 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 the command `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
<i>EGFR</i>	chr7	55086525	55086725
<i>LPAR5</i>	chr12	6745297	6745497
<i>GPR35</i>	chr2	241544625	241544825

- (iii) *Design custom library.* Download the genome 2bit file that the target gene coordinates correspond 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 <code>.csv</code> 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	<code>final_guides.csv</code>
'-i'	Prefix of input genome 2-bit file	<code>hg19</code>
'-g'	Target-gene <code>.csv</code> file with the gene name, chromosome, start of the targeted region, and end of the targeted region in columns from left to right	<code>genes.csv</code>
'-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-gene <code>.csv</code> file	3
'-db'	Used to specify an existing off-target database constructed from a previous custom library design for a new library	False
'-gecko' or '-sam'	Used to 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 (>50 kb), we recommend splitting the target-gene `.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 dramatically 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 targets 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 the 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 corresponds to the target genes by running `design_targeted_library.py` with the following optional parameters:

Flag	Description	Default
'-o'	Output .csv file with names for target genes, corresponding spacer sequences, and oligo library sequences in columns from left to right	oligos.csv
'-l'	Annotated library .csv file with names in the first column and corresponding spacer sequences in the second column	annotated_library.csv
'-g'	Target-gene .csv file with names of target genes	target_genes.csv
'-gecko' or '-sam'	Used to 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. Wrap the pooled oligos in Parafilm and store them at -20 °C.

**Cloning a custom sgRNA library ● TIMING 2 d**

3| *PCR amplification of pooled oligo library.* Throughout the sgRNA library cloning process, refer to the following table for 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 (µl)	Final concentration
NEBNext High Fidelity PCR Master Mix, 2×	12.5	1×
Pooled oligo library template from Step 2	1	0.04 ng µl <sup>-1</sup>
Oligo-Fwd primer (universal)	1.25	0.5 µM
Oligo-Knockout-Rev or Oligo-Activation-Rev primer	1.25	0.5 µM
UltraPure water	9	
Total	25	

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



## PROTOCOL

4| Divide the PCR master mix into 25- $\mu$ l reactions and perform a PCR using the following cycling conditions:

Cycle number	Denature	Anneal	Extend
1	98 °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. Use a NanoDrop UV spectrophotometer to quantify the product.

6| Run the PCR-purified oligo library from Step 5 on a gel along with a 50-bp 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  $\text{cm}^{-1}$  for 45 min.

▲ **CRITICAL STEP** Run on a 2% (wt/vol) agarose gel for long enough to separate the target library (140 bp) from a possible primer-dimer of ~120 bp. 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 use a NanoDrop UV spectrophotometer to quantify the final product.

8| *Restriction digest of plasmid backbone.* Digest the desired library plasmid backbone with the restriction enzyme Esp3I (BsmBI), which cuts around the sgRNA target region. Refer to the master mix setup below for the reaction ratios:

Component	Amount per reaction ( $\mu$ l)	Final concentration
FastDigest Buffer, 10 $\times$	2	1 $\times$
Library Plasmid Backbone	1	50 ng $\mu$ l $^{-1}$
FastDigest Esp3I (BsmBI)	1	
FastAP Thermosensitive Alkaline phosphatase	1	
DTT, 100 mM	0.2	1 mM
UltraPure water	14.8	
Total	20	

9| Prepare aliquots of 20- $\mu$ l 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  $\text{cm}^{-1}$  for 30 min. Gel-extract the library plasmid backbone using the QIAquick Gel Extraction Kit according to the manufacturer's protocol and use a NanoDrop UV spectrophotometer to quantify the product. Note that the GeCKO library backbones contain a 1,880-bp filler sequence, which should be visible as a dropout. The SAM library backbones do not contain a filler sequence, and the expected dropout of 20 bp 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 $\times$	10 $\mu$ l	1 $\times$
Digested library plasmid Backbone from Step 10	330 ng	16.5 ng $\mu$ l $^{-1}$
SgRNA library insert from Step 7 or UltraPure water control	50 ng	2.5 ng $\mu$ l $^{-1}$
UltraPure water	To 20 $\mu$ l	
Total	20 $\mu$ l	

12| Prepare 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 the mixture at room temperature for 15 min and centrifuge at >15,000g for 15 min at room temperature 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. Incubate at 55 °C for 10 min and use a NanoDrop UV spectrophotometer to quantify the custom sgRNA library. Isopropanol-purified sgRNA libraries can be stored at -20 °C for several months.

#### Amplification of pooled sgRNA library ● **TIMING 2 d**

18| *Pooled sgRNA library transformation.* Electroporate the library at 50–100 ng  $\mu$ l<sup>-1</sup> using Endura ElectroCompetent cells according to the manufacturer's directions. If you are amplifying a ready-made genome-scale library from Addgene, repeat for a total of 1 electroporation per 10,000 sgRNAs in the library. If you are 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| Prewarm 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. Prewarm 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 the transformation efficiency. To prepare the dilution mix, add 10  $\mu$ l of the pooled electroporated cells to 990  $\mu$ l of LB medium for a 100-fold dilution and mix well. Then add 100  $\mu$ l of the 100-fold dilution to 900  $\mu$ l of LB medium for a 1,000-fold dilution and mix well.

22| Plate 100  $\mu$ l of the 1,000-fold dilution on a prewarmed 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 you are amplifying a custom sgRNA library, repeat Steps 21 and 22 for the control Gibson reaction.

## PROTOCOL

**24|** To plate pooled electroporated cells, add 1 volume of LB medium to the pooled electroporated cells from Step 20, mix well, and plate the mixture 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 prewarmed 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 that the liquid culture does not completely dry out, as this will lead to poor survival rates.

### **(B) Plating on standard LB agar plates**

- (i) Alternatively, plate 200  $\mu$ l of electroporated cells on each of the prewarmed standard LB agar plates from Step 19 using the same technique as described in Step 24A(i).

▲ **CRITICAL STEP** Plating of 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 you are 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 you are 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 critical to 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 as compared with 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 to 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 OD<sub>600</sub> value of the harvested bacterial suspension as follows: number of maxipreps = OD<sub>600</sub> value · (total volume of suspension)/1,200. Perform maxipreps of 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 OD<sub>600</sub> value within the linear range of the spectrophotometer, typically ~0.1–0.5, and measure the OD<sub>600</sub> value of the dilution. Then multiply the OD<sub>600</sub> value by the dilution factor to obtain the OD<sub>600</sub> value of the bacterial suspension. Approximately one maxiprep is needed for two densely plated large LB agar plates.

**31|** Pool the resulting plasmid DNA and use a NanoDrop UV spectrophotometer to quantify the product. Maxiprep sgRNA library can be aliquotted and stored at –20 °C for at least 1 year.

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

**32|** *Library PCR for NGS.* We have provided NGS primers that amplify the sgRNA target region with Illumina adaptor 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 (μl)	Final concentration
NEBNext High Fidelity PCR Master Mix, 2×	25	1×
Pooled sgRNA library template from Step 31	1	0.4 ng μl <sup>-1</sup>
NGS-Lib-Fwd primer (unique)	1.25	0.25 μM
NGS-Lib-KO-Rev or NGS-Lib-SAM-Rev primer (barcode)	1.25	0.25 μM
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 such as NEBNext. 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	98 °C, 3 min		
2–23	98 °C, 10 s	63 °C, 10 s	72 °C, 25 s
24			72 °C, 2 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 μg of the product on a 2% (wt/vol) agarose gel. Successful reactions should yield a ~260- to 270-bp product for the knockout library and a ~270- to 280-bp product for the activation library. Perform gel extraction using the QIAquick Gel Extraction Kit according to the manufacturer’s directions.

■ **PAUSE POINT** Gel-extracted samples can be stored at –20 °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 a 5% PhiX control on the MiSeq or a 20% PhiX control on the NextSeq to improve library diversity; we recommend 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’	Indicates the absence of a guanine before the guide spacer sequence	Guanine is present

## PROTOCOL

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, nonperfect guide matches, sequencing reads without key, 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 >70% perfectly matching guides, <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 ● TIMING 1–2 weeks

40| *Perform an antibiotic kill curve.* Before 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 the media with antibiotic every 3 d. After 4–7 d, choose the lowest concentration of antibiotic sufficient to kill all cells.

▲ **CRITICAL STEP** It is important to use the lowest concentration of antibiotic in order 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 then 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 the cells to reach >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 in 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 of 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	2,250 µl
pMD2.G (lentiviral helper plasmid)	15.3 µg
psPAX (lentiviral helper plasmid)	23.4 µg
Lentiviral target plasmid	30.6 µg

▲ **CRITICAL STEP** Transfecting at the recommended cell density is critical to maximizing transfection efficiency. Lower densities can result in Lipofectamine 2000 toxicity for cells, whereas higher densities can reduce transfection efficiency.

## Box 5 | Lentivirus production with polyethylenimine ● TIMING 3 d

### Procedure

1. *Prepare polyethylenimine (PEI) transfection reagent.* Dissolve 50 mg of PEI Max in 45 ml of UltraPure water. Adjust the pH to 7.1 by adding 10 M NaOH dropwise until the pH approaches 6 and then by adding 1 M NaOH dropwise until the pH reaches 7.1. Adjust the final volume to 50 ml with UltraPure water. Sterilize using Millipore's 0.45- $\mu$ m Steriflip filter. Prepare 50  $\times$  1-ml aliquots and store them at  $-20$  °C until use. PEI is stable for up to 1 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 $\mu$ l	
pMD2.G (lentiviral helper plasmid)	3.4 $\mu$ g	5.2 $\mu$ g ml <sup>-1</sup>
psPAX2 (lentiviral helper plasmid)	6.8 $\mu$ g	10.4 $\mu$ g ml <sup>-1</sup>
Lentiviral target plasmid	13.6 $\mu$ g	20.9 $\mu$ g ml <sup>-1</sup>

4. Add 195  $\mu$ l of PEI transfection reagent, vortex, and incubate the mixture at room temperature for 10 min.
5. Add 25 ml of D10 medium to the transfection reagent mixture.
6. Aspirate the old medium from the cells, gently add new medium 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 the lentivirus as described in Step 52.

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

Component	Amount per T225 flask
Opti-MEM	2,250 $\mu$ l
PLUS reagent	297 $\mu$ l

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	4,500 $\mu$ l
Lipofectamine 2000	270 $\mu$ l

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 the flasks gently to mix. Return the T225 flasks to the incubator.

51| After 4 h, replace the medium with 45 ml of prewarmed 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 aliquotted and stored at  $-80$  °C for up to 1 year. Avoid freezing and thawing the 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 per cell line. Lentiviral titer should be determined using the relevant cell line for the screen.

## PROTOCOL

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 and 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 of D10 medium per well with  $8 \mu\text{g ml}^{-1}$  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 1,000g 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  $\mu\text{l}$  of 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 53A(iv)), based on the cell count determined in Step 53A(v) 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 two wells and 100  $\mu\text{l}$  of regular D10 medium to the other two 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 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 cell maintenance.* We routinely maintain HUES66 cells (an 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 before plating. During passaging and plating, mTeSR1 medium should be supplemented further with 10  $\mu\text{M}$  ROCK inhibitor. Refresh the mTeSR1 medium daily.
- (ii) *Passaging HUES66 cells.* 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 200g for 5 min. Remove the supernatant, resuspend the cells in mTeSR1 medium with 10  $\mu\text{M}$  ROCK inhibitor, and replate the cells on 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 of mTeSR1 medium per well. To 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, fill to a total volume of 3 ml with DPBS, and supplement with 10  $\mu\text{M}$  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  $\mu\text{l}$  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** ● **TIMING 3–6 weeks**

**CRITICAL:** Skip to Step 56 if you are screening using a 1-vector format.

**54|** *Generation of a cell line with stably expressed Cas9 components.* Before 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 of < 0.7. If two additional Cas9 components are required—for instance, dCas9-VP64 and MS2-p65-HSF1—the two 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 of D10 medium per well with  $8 \mu\text{g ml}^{-1}$  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 1,000g for 2 h at 33 °C. Return the plates to the incubator after spinfection.
- (iii) 24 h after the end of spinfection, remove the medium, gently wash with 400  $\mu\text{l}$  of TrypLE per well, add 100  $\mu\text{l}$  of TrypLE, and incubate at 37 °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 in 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 to a T75 flask and add 13 ml of D10 medium with selection antibiotic.
- (v) 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 2 ml of 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  $\mu\text{M}$  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, 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 in order 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 you are culturing for more than 7 d after selection or after freezing cells, reselect 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 and 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 d, the cells are ready for screening. For knockout screening, we have found that maximal knockout efficiency is achieved 7 d after sgRNA transduction and therefore recommend selecting for 7 d before starting the screen selection. By contrast, maximal SAM activation is achieved as early as 4 d after sgRNA transduction. If selection is complete based on the no-virus control, gain-of-function screening can be started 5 d after transduction. We generally recommend performing 4 independent screening bioreps (i.e., 4 separate sgRNA library infections followed by separate screening selection). Multiple screening bioreps are critical to determining screening hits with a high rate of validation. As the parameters of each screen depend on the screening phenotype of interest, we provide guidelines and technical considerations for the screening selection process (**Boxes 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 you are 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 MOIs may confound screening results.



## PROTOCOL

### Harvest of genomic DNA for screening analysis ● 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  $\mu$ 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 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 a 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 the mixture at 500g for 5 min at room temperature. Discard the flow-through.

**61|** Add 2 ml of DNA Wash Buffer and centrifuge at 500g 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 (>12,000g) 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 (>12,000g) 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 RIGER.** Before 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' is a list of the genes that the sgRNAs target, and the 'biorep' columns show 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 data set. Determine the top candidate genes based on either the overlap or the average ranking between the screening bioreps.

**Validation of candidate genes for screening phenotype** ● **TIMING 4–5 weeks**

**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 nontargeting sgRNAs (NT1 and NT2) for controls are also provided in **Table 5**.

**70** Resuspend the top- and bottom-strand primers to a final concentration of 100 μM. Prepare the following mixture for phosphorylating and annealing the top- and bottom-strand primers for each validation sgRNA:

Component	Amount (μl)	Final concentration
sgRNA-top, 100 μM	1	10 μM
sgRNA-bottom, 100 μM	1	10 μM
T4 ligation buffer, 10×	1	1×
T4 PNK	0.5	
UltraPure water	6.5	
Total	10	

**71** Phosphorylate and anneal the primers in a thermocycler using the following conditions: 37 °C for 30 min; 95 °C for 5 min; and ramp down to 25 °C at 5 °C min<sup>-1</sup>.

**72** After the annealing reaction is complete, dilute the phosphorylated and annealed oligos 1:10 by adding 90 μl of UltraPure water.

■ **PAUSE POINT** Annealed oligos can be stored at -20 °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 (μl)	Final concentration
Rapid Ligase Buffer, 2×	12.5	1×
FastDigest Esp3I (BsmBI)	1	
DTT	0.25	1 mM
BSA, 20 mg ml <sup>-1</sup>	0.125	0.1 mg ml <sup>-1</sup>
T7 ligase	0.125	
Diluted oligo duplex from Step 72	1	0.04 μM
sgRNA library backbone	1	1 ng μl <sup>-1</sup>
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.

## PROTOCOL

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 Stbl3 strain for quick transformation. Thaw the chemically competent Stbl3 cells on ice, add 2 µl of the product from Step 74 to ice-cold Stbl3 cells, and incubate the mixture on ice for 5 min. Heat-shock the mixture at 42 °C for 30 s and return it to ice immediately for 2 min. Add 100 µl of SOC medium and plate the mixture on 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 two separate 25-ml cultures of LB medium with 100 µg ml<sup>-1</sup> ampicillin in 50-ml Falcon tubes. 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 midpreps have higher purity and produce higher lentiviral titers.

■ **PAUSE POINT** Midiprep validation sgRNA constructs can be stored at –20 °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 with 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-µm syringe filters.

82| Titer the lentivirus according to Step 53. If you are preparing multiple validation sgRNA lentiviruses in the same plasmid backbone at the same time, titer lentiviruses from 2–3 different sgRNAs and extend the average titer to the rest of the lentiviruses.

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 and 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 to the validation cell lines and assessing cell proliferation, death, or fluorescence for positive, negative, or 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- to 300-bp 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 in order to allow for the detection of longer indels. Aim for an annealing temperature of 60 °C and check for potential off-target sites using Primer-BLAST. If necessary, include a 1- to 10-bp 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 three bioreps in a 96-well clear-bottom black tissue culture plate.

**87|** 1 d after seeding, when the cells have reached confluency, aspirate the medium and add 50 µl 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 (µl)	Final concentration
KAPA HiFi HotStart ReadyMix, 2×	10	1×
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	

▲ **CRITICAL STEP** To minimize error in amplifying sgRNAs, it is important to use a high-fidelity polymerase such as KAPA. 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	95 °C, 5 min		
2–13	98 °C, 20 s	60 °C, 15 s	72 °C, 15 s
14			72 °C, 1 min

## PROTOCOL

**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 ( $\mu\text{l}$ )	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 $\mu\text{M}$
NGS-indel-R2-Rev	1	0.5 $\mu\text{M}$
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  $\mu\text{l}$  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 in TBE buffer with SYBR Safe dye and run the gel at 15 V  $\text{cm}^{-1}$  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** Because of 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, using a NanoDrop UV spectrophotometer to quantify the products, and normalizing them to the same concentration before 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  $-20\text{ }^{\circ}\text{C}$  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 sample or control in columns from left to right. The last column is required only when performing MLF correction. When MLE is performed, control samples that 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:

Example 1	example1.fastq	GCCCGAT CGCTATA TCCACG	TGTATATACCTCGCGCCTAACTGC CAGCTGACCACGCCGTACAGGTTCTGTTGTC TACTGATGCATTACATC TCCTTAGGGTACTTCCGCTGAGAT CATTGCCCGATCG CTATATCCACG CGTTTGGCTCGTTCAACCAAGTCCA ACCGACTCGTTGGTCTGG TAATGT GCCCAAGTTAAGGTGAGTATGGA CATGGCGGG	Experimental
Example 2	example2.fastq	CGAGATA AGTCAGC AGGGGC	CTCTTCTGCTCAAGCGAGTTCCTCA GAGGTCCTTGCCGAGGGGGTTATA TCGATCCA CGAAACATAGTATGTA ATACGAAAGTCATCGGCGCCTATG CCCTCGAGATAAGTC AGCAGGGG CTTTTCTGACATTTTCCAAGATTTCG GGATTGACGTTGCATCGCAAGCTA ATTGGTTACCATTAGACCCAGTC CTCAGCCC	Experimental
Example 3	example3.fastq	CACCCAC ACCAACC GCAGAA	CTGGGTTTAAACCGAGCTAGTCCTG AAGATCTTGAGTAACTGCACATGT AAATAATC CGAGGCGTTTTTCGGCG CCTAGCATGGCGTGAAGCTCTCC GCAGCACCCACACCA ACCGCAGA AGACAGCCCTGGGACGTGTAATT TAGCCATGTTGGTCGTTTCGGGA GCAACAGTGGTGGCCGTAGCG GACACCACCA	Control

**99** | If you are processing all files with a single command, run `python calculate_indel.py` with the following optional parameters:

Flag	Description	Default
'-f'	Indicates that input file is in .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, respectively, 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, use the command `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 because of 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 four 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, ~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 μl <sup>-1</sup>	0.125	0.2 U μl <sup>-1</sup>
RevertAid Reverse Transcriptase, 200 U μl <sup>-1</sup>	1.25	10 U μl <sup>-1</sup>
UltraPure water	10.41	
Total	20	

## PROTOCOL

Except for Oligo dT, all components in the reverse transcription mix can be found in the Thermo RevertAid RT Reverse Transcription Kit. **▲ CRITICAL STEP** Make sure that all reagents are RNase free, and take proper precautions when working with RNA.

**102|** Add 20- $\mu$ l aliquots of the Reverse Transcription Mix to 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 medium from each well of the 96-well poly-D-lysine tissue culture plate from Step 100, wash with 100  $\mu$ l of cold DPBS, and add 100  $\mu$ 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  $\mu$ l of the cell lysate to a new 96-well PCR plate. Add 3  $\mu$ 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  $-20^{\circ}\text{C}$  for additional reverse transcription reactions.

**105|** Add 5  $\mu$ l of the cell lysate with RNA lysis stop solution to the Reverse Transcription Mix from Step 101 for a total volume of 25  $\mu$ 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  $-20^{\circ}\text{C}$  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 premixing 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|** Transfer 4  $\times$  5  $\mu$ l of the qPCR master mix to a 384-well optical plate for technical replicates.

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

Cycle number	Hold	Denature	Anneal/Extend
1	50 $^{\circ}\text{C}$ , 2 min		
2	95 $^{\circ}\text{C}$ , 20 s		
3–42		95 $^{\circ}\text{C}$ , 3 s	60 $^{\circ}\text{C}$ , 30 s

**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.

**Additional steps for combining candidate genes from knockout and activation screens using dead sgRNAs (dRNAs)**

● **TIMING 4–5 weeks**

**111** | dRNAs are sgRNAs with 14- or 15-nt spacer sequences that 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 (<https://groups.google.com/forum/#!forum/crispr>) (**Box 6; Fig. 4**).

**TABLE 6** | Troubleshooting table.

Step	Problem	Possible reason	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 the number of electroporations if feasible; follow the manufacturer's instructions for optimal efficiency (i.e., ensure that cuvettes, tubes, and cells are chilled thoroughly and that recovery medium is added immediately after electroporation); spread the bacterial suspension evenly without letting plates dry; limit plate warming to <2 h and add 1 volume of additional LB medium before spreading as recommended; if the DNA for electroporation has a 260/230 ratio <1.8 or is <50 ng μl <sup>-1</sup> as assessed by NanoDrop UV spectrophotometer, 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, and ensure that added DTT is fresh and stored at -20 °C as recommended
39	<70% of clusters represent perfectly matched sgRNAs	NGS reads are of low quality; incorrect flag for the presence of 'G' before guide spacer sequence; low oligo synthesis quality (custom library only)	If <75% of clusters passed the filter, consider resequencing the library with a more optimal cluster density and cluster-passing filter; check for the proper flag, indicating the 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 are present 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, and ensure there are >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 was <75%; failure to replace medium within 6 h post transfection	Ensure even plating and a cell density of 80–90% during transfection. Replace the medium 4 h after transfection. The cell density should not be <80% at the time of lentivirus harvest

(continued)



## PROTOCOL

**TABLE 6** | Troubleshooting table (continued).

Step	Problem	Possible reason	Solution
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 the recommended density; prepare new DNA for transfection
53	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 an Amicon Ultra-15 Centrifugal Filter Unit with an Ultracel-100 membrane or ultracentrifugation at 88,000 <i>g</i> for 2 h at 4 °C. Discard the supernatant and resuspend the lentivirus in cold PBS at 4 °C overnight
58	No amplification of gDNA	Incorrect template or primer; incorrect template or primer concentration	Use an appropriate NGS-Rev primer corresponding to the screen; titrate the gDNA concentration to 1–3 µg for a 50-µl reaction; titrate the primer concentration to a final concentration of 0.1–0.5 µM
58	Primer–dimers present after gDNA amplification	Insufficient number of PCR cycles	Increase the number of cycles for PCR up to 25; lower the 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 the use of the suggested restriction enzyme (Fermentas Fast Digest) and DNA ligase (Enzymatics T7 ligase)
94	Multiple bands or no amplification during indel analysis	Inefficient or nonspecific priming; incorrect template concentration	Test different annealing temperatures (58–62 °C); redesign 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

### ● TIMING

Steps 1–17, designing and cloning a targeted screen: 3–5 weeks, 1 week 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 weeks

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

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

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

Steps 111–113, additional steps for combining candidate genes from knockout and activation screens using dead sgRNAs: 4–5 weeks

**Box 5**, lentivirus production with polyethylenimine: 3 d

## Box 6 | Frequently asked screening questions from the CRISPR Forum

The following questions are selected from the CRISPR Discussion Forum (<https://groups.google.com/forum/#!forum/crispr>).

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

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

The transcriptional start site (TSS) is the beginning of the 5' UTR. The UCSC Table Browser (<https://genome.ucsc.edu/cgi-bin/hgTables>) is a good resource for TSS annotations. We have observed the most robust transcriptional upregulation when sgRNAs are designed to target the 200-bp 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>30</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 adaptor 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 performing PCR with 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 present 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.

### 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>FV600E</sup> (A375) cell line<sup>30,37</sup>. After applying vemurafenib selection, the sgRNA library distribution in the experimental condition, which is measured by NGS, 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 substantially lower than those for the rest of the genes (**Fig. 4e,f**).

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

**ACKNOWLEDGMENTS** We 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. was 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. was supported by a DOE Computational Science Graduate Fellowship. F.Z. was supported by the NIH through the National Institute of Mental Health (NIMH; grants 5DP1-MH100706 and 1R01-MH110049),

the National Science Foundation (NSF), the Howard Hughes Medical Institute (HHMI), the New York Stem Cell Foundation, the Simons Foundation, the Paul G. Allen Family Foundation, and the Vallee Foundation, and James and Patricia Poitras, Robert Metcalfe, and David Cheng. 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 laboratory website (<http://www.genome-engineering.org>).

**AUTHOR CONTRIBUTIONS** J.J., S.K., J.S.G., O.O.A., R.J.P., M.D.B., N.E.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: details are available in the [online version of the paper](#).

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Boutos, M. & Ahringer, J. The art and design of genetic screens: RNA interference. *Nat. Rev. Genet.* **9**, 554–566 (2008).
2. Echeverri, C.J. & Perrimon, N. High-throughput RNAi screening in cultured cells: a user's guide. *Nat. Rev. Genet.* **7**, 373–384 (2006).
3. Shalem, O., Sanjana, N.E. & Zhang, F. High-throughput functional genomics using CRISPR-Cas9. *Nat. Rev. Genet.* **16**, 299–311 (2015).
4. Elbashir, S.M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).
5. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
6. McManus, M.T. & Sharp, P.A. Gene silencing in mammals by small interfering RNAs. *Nat. Rev. Genet.* **3**, 737–747 (2002).
7. Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343–349 (2004).
8. Berns, K. *et al.* A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431–437 (2004).
9. Boutos, M. *et al.* Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science* **303**, 832–835 (2004).
10. Moffat, J. & Sabatini, D.M. Building mammalian signalling pathways with RNAi screens. *Nat. Rev. Mol. Cell Biol.* **7**, 177–187 (2006).
11. Paddison, P.J. *et al.* A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**, 427–431 (2004).
12. Root, D.E., Hacohen, N., Hahn, W.C., Lander, E.S. & Sabatini, D.M. Genome-scale loss-of-function screening with a lentiviral RNAi library. *Nat. Methods* **3**, 715–719 (2006).
13. Silva, J.M. *et al.* Second-generation shRNA libraries covering the mouse and human genomes. *Nat. Genet.* **37**, 1281–1288 (2005).
14. Birmingham, A. *et al.* 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat. Methods* **3**, 199–204 (2006).
15. Jackson, A.L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.* **21**, 635–637 (2003).
16. Jackson, A.L. & Linsley, P.S. Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nat. Rev. Drug Discov.* **9**, 57–67 (2010).
17. Wright, A.V., Nunez, J.K. & Doudna, J.A. Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* **164**, 29–44 (2016).
18. Marraffini, L.A. CRISPR-Cas immunity in prokaryotes. *Nature* **526**, 55–61 (2015).
19. Hsu, P.D., Lander, E.S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).
20. Garneau, J.E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
21. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
22. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
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. USA* **109**, E2579–E2586 (2012).
24. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
25. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
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.* **14**, 8096–8106 (1994).
27. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12**, 326–328 (2015).
28. Gilbert, L.A. *et al.* Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).
29. Gilbert, L.A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
30. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).
31. Maeder, M.L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* **10**, 977–979 (2013).
32. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838 (2013).
33. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* **10**, 973–976 (2013).
34. Qi, L.S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
35. Konermann, S. *et al.* Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472–476 (2013).
36. Chavez, A. *et al.* Comparison of Cas9 activators in multiple species. *Nat. Methods* **13**, 563–567 (2016).
37. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
38. Wang, T., Wei, J.J., Sabatini, D.M. & Lander, E.S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
39. Koike-Yusa, H., Li, Y., Tan, E.P., Velasco-Herrera Mdel, C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267–273 (2014).
40. Zhou, Y. *et al.* High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509**, 487–491 (2014).
41. Doench, J.G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* **32**, 1262–1267 (2014).
42. Chen, S. *et al.* Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* **160**, 1246–1260 (2015).
43. Shi, J. *et al.* Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat. Biotechnol.* **33**, 661–667 (2015).
44. Parnas, O. *et al.* A genome-wide CRISPR screen in primary immune cells to dissect regulatory networks. *Cell* **162**, 675–686 (2015).
45. Wang, T. *et al.* Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
46. Wong, A.S. *et al.* Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. *Proc. Natl. Acad. Sci. USA* **113**, 2544–2549 (2016).
47. Doench, J.G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
48. Jain, I.H. *et al.* Hypoxia as a therapy for mitochondrial disease. *Science* **352**, 54–61 (2016).
49. Marceau, C.D. *et al.* Genetic dissection of *Flaviviridae* host factors through genome-scale CRISPR screens. *Nature* **535**, 159–163 (2016).
50. Zhang, R. *et al.* A CRISPR screen defines a signal peptide processing pathway required by flaviviruses. *Nature* **535**, 164–168 (2016).
51. Blondel, C.J. *et al.* CRISPR/Cas9 screens reveal requirements for host cell sulfation and fucosylation in bacterial type III secretion system-mediated cytotoxicity. *Cell Host Microbe* **20**, 226–237 (2016).
52. Orchard, R.C. *et al.* Discovery of a proteinaceous cellular receptor for anorovirus. *Science* **353**, 933–936 (2016).
53. Canver, M.C. *et al.* BCL11A enhancer dissection by Cas9-mediated *in situ* saturating mutagenesis. *Nature* **527**, 192–197 (2015).
54. Korkmaz, G. *et al.* Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat. Biotechnol.* **34**, 192–198 (2016).
55. Diao, Y. *et al.* A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. *Genome Res.* **26**, 397–405 (2016).
56. Sanjana, N.E. *et al.* High-resolution interrogation of functional elements in the noncoding genome. *Science* **353**, 1545–1549 (2016).
57. Fulco, C.P. *et al.* Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016).
58. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–826 (2013).
59. Hsu, P.D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
60. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31**, 839–843 (2013).
61. Evers, B. *et al.* CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nat. Biotechnol.* **34**, 631–633 (2016).
62. Munoz, D.M. *et al.* CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. *Cancer Discov.* **6**, 900–913 (2016).
63. Aguirre, A.J. *et al.* Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. *Cancer Discov.* **6**, 914–929 (2016).
64. Morgens, D.W., Deans, R.M., Li, A. & Bassik, M.C. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat. Biotechnol.* **34**, 634–636 (2016).

65. Hasson, S.A. *et al.* High-content genome-wide RNAi screens identify regulators of parkin upstream of mitophagy. *Nature* **504**, 291–295 (2013).
66. Moffat, J. *et al.* A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**, 1283–1298 (2006).
67. Neumann, B. *et al.* High-throughput RNAi screening by time-lapse imaging of live human cells. *Nat. Methods* **3**, 385–390 (2006).
68. Sanjana, N.E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
69. König, R. *et al.* A probability-based approach for the analysis of large-scale RNAi screens. *Nat. Methods* **4**, 847–849 (2007).
70. Luo, B. *et al.* Highly parallel identification of essential genes in cancer cells. *Proc. Natl. Acad. Sci. USA* **105**, 20380–20385 (2008).
71. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
72. Ran, F.A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
73. Ran, F.A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).
74. Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
75. Dahlman, J.E. *et al.* Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nat. Biotechnol.* **33**, 1159–1161 (2015).
76. Kiani, S. *et al.* Cas9 gRNA engineering for genome editing, activation and repression. *Nat. Methods* **12**, 1051–1054 (2015).