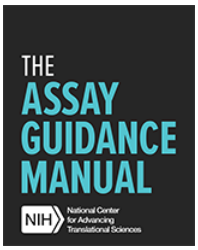
U.S. National Library of Medicine
National Center for Biotechnology Information

NLM Citation: Costa JR, Bejcek BE, McGee JE, et al. Genome Editing Using Engineered Nucleases and Their Use in Genomic Screening. 2017 Nov 20. In: Markossian S, Grossman A, Arkin M, et al., editors. Assay Guidance Manual [Internet]. Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences; 2004-.
Bookshelf URL: https://www.ncbi.nlm.nih.gov/books/



Genome Editing Using Engineered Nucleases and Their Use in Genomic Screening

Joana R. Costa, Ph.D.,[1] Bruce E. Bejcek, Ph.D.,[2] James E. McGee, Ph.D.,[3] Adam I. Fogel, Ph.D.,[4] Kyle R. Brimacombe, M.S.,[5] and Robin Ketteler, Ph.D.[1]

Created: November 20, 2017.

Abstract

Understanding gene function is critical for developing therapeutic strategies to target disease. Common approaches to understanding gene function in a systematic and unbiased way include loss-of-function and gain-of-function genomic screening. Some of these rely on artificially increasing the copy number of gene transcripts using cDNA expression libraries. Others interrogate endogenous protein expression through genetic loss-of-function approaches such as siRNA screening. Over the past two decades, targeted approaches that reduce the endogenous expression level of genes or proteins have been developed that facilitate a much better understanding of genes in the context of the living cell. Technologies have been developed that enable precise modification of the genome rather than reduction at the transcript level. These include zinc finger nucleases, TALENs and the CRISPR/Cas9 technology. The use of these systems in large-scale high-throughput screening is an emerging field and we herein highlight recommendations for such applications.

Introduction

The ability to modify the expression of single genes and proteins has become one of the most important tools in molecular and cellular biology. Several methodologies have been developed to allow for specific gene manipulation in tissue culture cells, which have become colloquially known as "genome-editing". These rely on nucleases that are engineered to cut specific genomic target sequences, including *Zinc Finger Nucleases* (ZFN), *Transcription Activator-like Effector Nucleases* (TALEN) and *Clustered Regularly Interspaced Short Palindromic Repeats* (CRISPR) nucleases (1). Homing meganucleases have also been used for these purposes but because they have not achieved widespread use they will not be discussed further.

The ZFNs, TALENs, and CRISPR/Cas9 (CRISPR associated) enzymes create double stranded breaks in the target DNA sequence which the cell will then repair using one of two pathways (Figure 1, Figure 2). The first of these processes is non-homologous end joining (NHEJ), which occurs without the use of a repair template. NHEJ results in a deletion or insertion (indel), the resulting sequence of which is essentially random and impossible to control. Practically, this pathway is useful in generating loss-of-function (knockout) of the gene of interest. The

Author Affiliations: 1 MRC Laboratory for Molecular Cell Biology, University College London; Email: j.costa@ucl.ac.uk; Email: r.ketteler@ucl.ac.uk. 2 Western Michigan University; Email: Bruce.bejcek@wmich.edu. 3 Eli Lilly; Email: Mcgee_james_e@lilly.com. 4 Biogen Inc.; Email: adam.fogel@biogen.com. 5 NIH/NCATS; Email: kyle.brimacombe@nih.gov.

Corresponding author.

second pathway is homology-directed repair (HDR), which utilizes a repair template that contains homology to the sequences proximal to the DNA break (the endogenous template in cells is the sister chromatid). This pathway allows for precise control of the resulting insertion, but requires additional DNA sequences to be transfected, and is often less efficient than the NHEJ pathway.

In addition to these methods, *recombinant Adeno-Associated Viral vectors* (rAAVs) have been used to modify specific genomic DNA targets as well (2) (reviewed in (3) and (4)). The rAAV methodology for genomic modification does not rely on the activity of an exogenously derived nuclease but relies on the replacement of sequences carried on a rAAV for the endogenous cellular homologue. rAAV methods will not be discussed further in this chapter.

All of these methods can be used in a wide variety of cell types to modify specific DNA sequences. In this chapter, we discuss the general application of different genome editing techniques for cell line generation and the specific use of CRISPR as a scalable platform for genetic screening.

A. Zinc Finger Nucleases (ZFNs)

Typically, genome editing platforms require two key features: an endonuclease that can cut a target location in the genome; and a sequence-specific adaptor that targets the endonuclease to a specified region on the DNA.

Zinc finger nucleases are chimeric proteins comprised of a DNA binding domain composed of zinc fingers (based on zinc finger transcription factors) and an endonuclease (5) (Figure 1, reviewed in (6)). Zinc fingers are ~30 amino acids that can bind to a limited combination of ~3 nucleotides. By using a combination of different zinc fingers, a unique DNA sequence within the genome can be targeted. Similarly, it was discovered that the *Fok I* type IIS restriction endonuclease had distinct and separable DNA binding and DNA cleavage domains, and the cleavage domain only has activity when there is dimerization (7). Making a chimera of the zinc finger binding domain with the *Fok I* digestion domain resulted in an artificial nuclease with a specificity that can theoretically be tailored to any sequence. When used in whole live cells it has become a powerful tool for the creation of cells that have single, insertion/deletion (indel) mutations in a region of interest in the genome.

The concept of use for zinc finger nucleases is relatively straightforward. Two chimeras are produced. Each chimera recognizes a specific and unique DNA sequence where the two sequences are relatively close to each other. When these chimeras are bound to the DNA they cut the DNA to create a fragment that is released, resulting in the activation of the cellular repair machinery in the form of NHEJ. NHEJ results in indels of various sizes and thus can disrupt open reading frames and result in a nonfunctional protein or a lack of an exon. Alternatively, a fragment of DNA can be also supplied to the cell, resulting in replacement of the sequence between the two cut sites by homologous recombination. The replacement fragment can be a normal or mutant form of the gene that is to be replaced.

There are some drawbacks with this technique:

Cost and time to engineer. ZFNs were the original platform used for genome editing and date back to the 1980s (8). Their use has always been hampered by the need to computationally predict and then iteratively engineer and test for sequence specificity. In comparison with the modular nature of TALE repeat binding, or the direct Watson-Crick affinity pairing of the CRISPR/Cas system, this makes the ZFNs much slower to develop and much costlier to pursue for a given target. This is the major reason why this technique has fallen out of favor in recent years.

Targeting specificity. Although the sequence that is targeted for digestion may be unique within the genome and the ZFN may have a high affinity for this sequence, off-target binding and cleavage of other DNA sequences can occur and result in undesired genome modification at other sites. For this interaction to result in digestion of DNA that will be repaired by NHEJ both halves of the ZFN must recognize sequences around the target site.

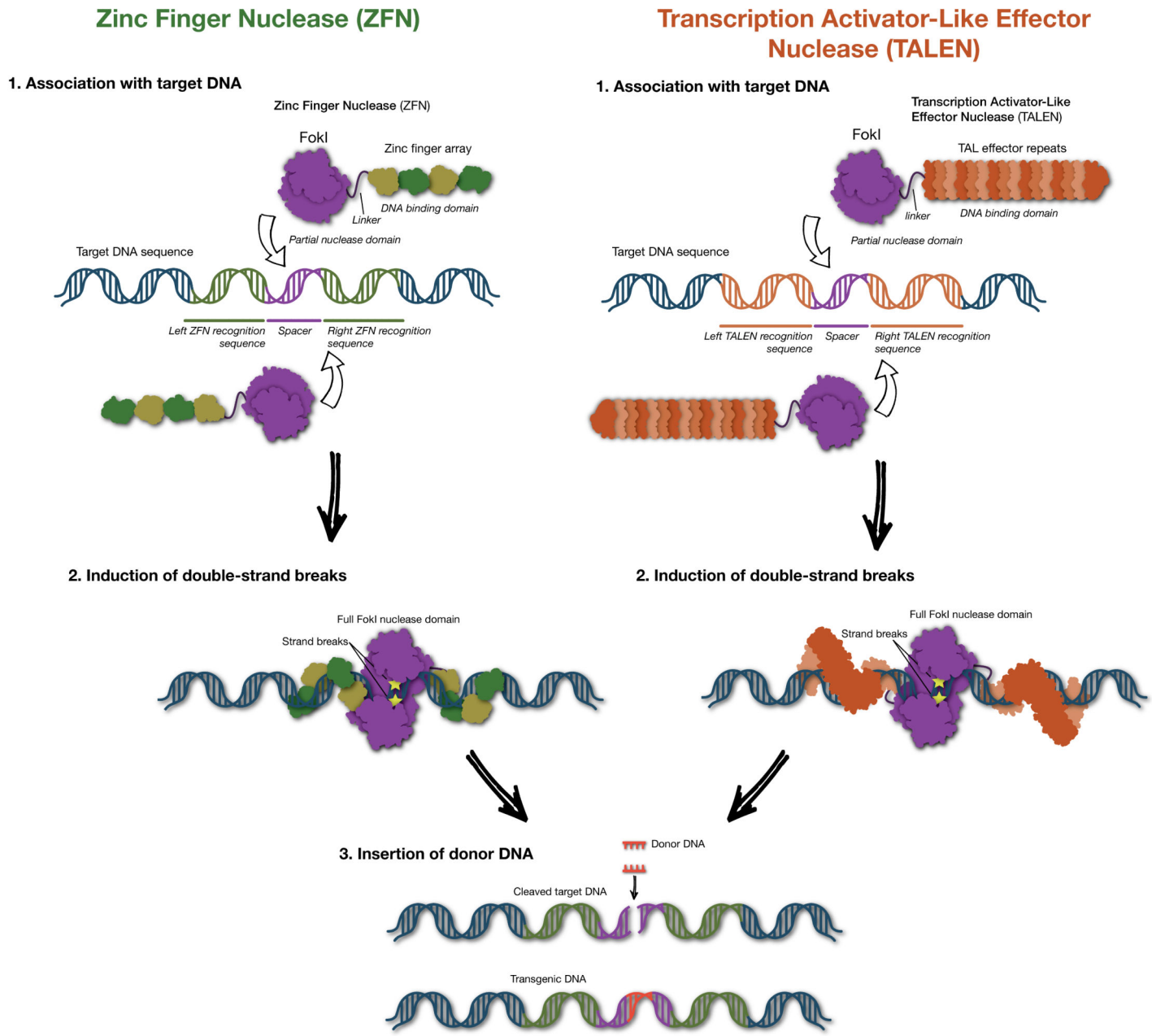


Figure 1: Overview of genome editing by zinc finger nucleases (ZFNs) and Transcription Activator-Like Effector Nucleases (TALENs). Zinc finger nucleases are composed of the Fok I endonuclease and an array of zinc finger binding domains that recognize the target DNA sequence. The action of two ZFNs on both strands of the DNA will result in a double strand break. A donor DNA provided in trans can then be integrated at the site of the break, resulting in a transgenic DNA sequence. In the absence of a donor template, the double strand break will be repaired by the host machinery, often resulting in insertions/deletions (indels) that disrupt the open reading frame. For TALENs, the zinc finger array is replaced by TAL effector repeats that guide targeting to the DNA.

Although this probability is small there are known zinc fingers that have off-site digestion patterns that will result in disruption of more than one site. Such off-target activities can be determined by sequencing the entire genome, but this is time consuming and costly. The best way to determine whether off-target activities might be influencing the biology of interest is to create more than one ZFN to digest the DNA at different sites (and presumably different off-target sites). When assessing knockout function, it is also imperative that the phenotype is rescued with a reintroduction of a functional gene in order to demonstrate the phenotype achieved is due to disruption of the gene of interest.

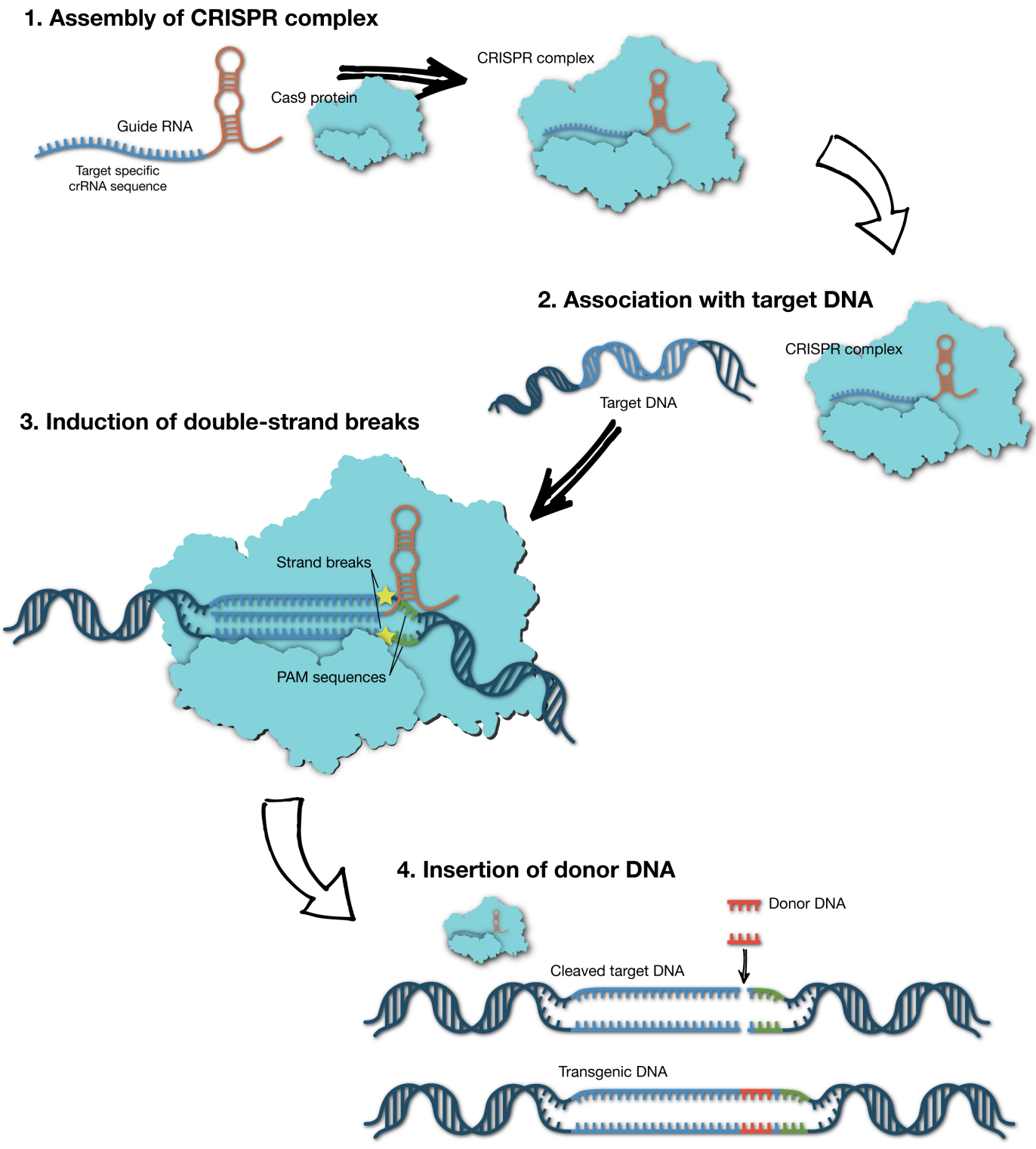


Figure 2: Mechanism of CRISPR/Cas9 genome editing. CRISPR/Cas9 genome editing requires a single guide (sg) RNA that directs the Cas9 endonuclease to a specific region of the genomic DNA, resulting in a double strand break. By providing a donor DNA in trans, a transgenic DNA can be created, whereas in the absence of a donor DNA, the double strand break will be repaired by the host cell, resulting in an insertion or deletion, thus potentially disrupting the open reading frame of a gene.

B. TALENs

TALE-Nucleases (TALENs) are based on the transcription activator-like effector (TALE) proteins from *Xanthomonas* bacterial species (reviewed in (9,10)) that have a DNA binding domain and an effector domain (Figure 1). The DNA binding domain consists of several subdomains of slightly variable length (~34 amino acids) but have a nearly identical amino acid sequence. There is a region at the 12^{th} to 13^{th} amino acids that is highly variable (the repeat variable dinucleotide, RVD) and known to mediate nucleotide specificity (see Table 1). The combination of these ~34 amino acid regions and the RVD dictate which DNA sequence is bound. Therefore, the engineering of a sequence specific binding domain is more easily accomplished than with ZFNs as this is a more modular approach. It requires only varying the RVD nucleotides and not the rest of the repeat backbone to achieve specificity with reasonable avidity. As with ZFNs, this DNA recognition domain can then be fused with a variety of proteins including the *Fok I* catalytic domain.

Like ZFNs, TALENs can be used to make a site-specific cut in genomic DNA, which results in either indel mutations or allows for replacement specific pieces of DNA if a suitable HDR template is delivered. Since the TALE domains allow for a more highly predictable, modular creation of binding domains, they gained immediate popularity when compared with ZFNs because they obviated the need for expensive and time-consuming engineering. Cloning TALEN constructs could be done in 1-2 weeks and the highly modular nature of the binding domains made binding fidelity much more likely. They have been used in cells from a variety of species both in tissue culture as well as *in vivo*. The advantages of TALENs over ZFNs are increased predictability of binding of the TALE domain to the desired target sequence and cost.

Like ZFNs, TALEN construction requires verification *in vitro* to demonstrate an adequate level of cleavage efficiency before they can be employed in experiments. Due to the modular nature of the TALE repeats, the probability of rationally designing a TALEN that will be specific for a particular sequence appears to be greater than that for ZFNs, however it is not possible to rule out off-target cleavage in a given cell line without extensive testing.

Table 1. Nucleotide binding by amino acids of the RVD.

Amino acids in the repeat variable dinucleotide (RVD) at aa 12 and 13	Nucleotide(s) recognized
HG	T, A, C
NG	T
N*	C, T
HD	C
NN	G, A
NS	A, C, G, T
NI	A

C. CRISPR

The identification of clustered regularly interspaced short palindromic repeats (CRISPR) along with the CRISPR associated (Cas) protein is the most recent development and is even more rapid and modular than the TALEN platform. The CRISPR/Cas9 technology is a three-component system consisting of an endonuclease (Cas9), a sequence-specific targeting element (the crRNA), and another RNA that links Cas9 with the crRNA, the tracrRNA (Figure 3) (11). Both crRNA and tracrRNA can be combined in a single molecule termed single guide (sg) RNA (12,13). Since an RNA molecule acts as the guide in this system, and the technology for cloning DNA oligonucleotides has been well developed, this system results in a methodology that is simpler and more rapid to

develop than TALENs and ZFNs. CRISPR/Cas9 is now widely used in multiple organisms, both *in vitro* and *in vivo*.

The mechanism of action for the CRISPR/Cas system relies on the expression of at least two components: an effector enzyme and a guide RNA (Figure 2). The effector enzyme most commonly used to date is Cas9 from the type II CRISPR/Cas system from *Streptococcus pyogenes*. Although there are three different types of CRISPR/Cas systems currently known, the type II enzymes are preferable as they are single polypeptides. To be used in this system the CAS9 protein must be expressed and translocated to the nucleus. Several laboratories have created codon-optimized versions with nuclear translocation signal sequences (NLS).

The crRNA recognizes and pairs with a sequence of 20 nucleotides in length within the targeted genome (Figure 3). Targeting relies on sequences at the 3' end of the desired site for digestion. These sequences, known as protospacer adjacent motifs (PAMs), are critical and unique to the CAS9 proteins isolated from different species. Alternative endonucleases such as the type V CPF1 exist that utilize different PAM motifs for targeting (14), thus expanding the repertoire of potential target sites (see Table 2). While the requirement for a PAM motif in the genome is a major restriction in this technique, it is estimated that every gene harbors multiple sites, allowing disruption of virtually any gene.

One advantage of the CRISPR/Cas9 technology is its great versatility. In effect, the sgRNA directs the endonuclease to a specific site in the genome. By attaching functional domains to Cas9, it is possible to target any functionality to a specific genomic location, including promoter sites and introns. For instance, using a translational fusion of catalytically inactive Cas9 to transcriptional activators, such as multimers of the VP16 peptide, and targeting a promoter region, it is possible to activate gene expression (15). Similarly, systems for gene repression, histone modification and epigenetic alterations have been created (16,17,18). In addition, the system can be used to specifically paint genomic sequences or tag sequence regions with a fluorescent protein or protein tag (19,20). It is also possible to generate an inducible CRISPR/Cas9 system either using classical repressor systems or small molecule inhibitors of Cas9 (21). The inventiveness of researchers to create novel applications of this technology seems to be infinite.

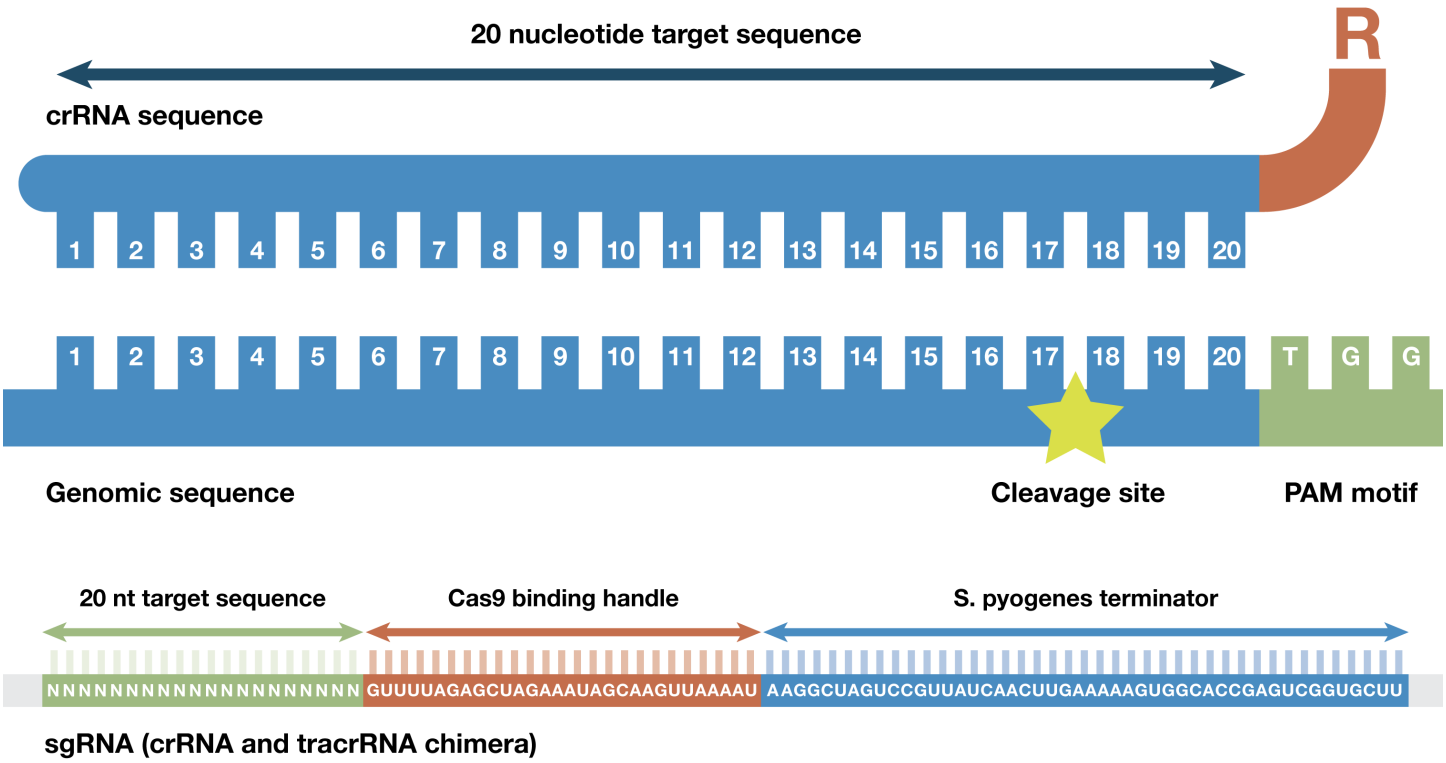


Figure 3: Anatomy of a single guide (sg) RNA. The sgRNA is composed of a crRNA sequence that is fused to the tracrRNA. The crRNA contains a 20 nt sequence that is identical to the genomic target DNA sequence. A pre-requisite for binding of the sgRNA to the target site is the presence of a PAM motif following the 20 nt recognition sequence. "R" denotes the remaining part of the sgRNA, consisting of a Cas9 binding handle and the *S. pyogenes* terminator sequence. Cleavage occurs at the -3 position upstream of the PAM motif.

Table 2: Endonucleases with different PAM specificity.

Name	Organism	PAM	Type	Ref
SpCas9	*Streptococcus pyogenes*	5'-NGG-3'	II	(11)
SpCas9 VQR	*Streptococcus pyogenes*	5'-NGAN-3'	II	(105,106)
SpCas9 EQR	*Streptococcus pyogenes*	5'-NGNG-3'	II	(105,106)
SpCas9 VRER	*Streptococcus pyogenes*	5'-NGCG-3'	II	(105,106)
SaCas9	*Staphylococcus aureus*	5'-NNGRRT-3'	II	(105,107)
NmCas9	*Neisseria meningitides*	5'-NNNNGATT-3'	II	(108)
St1Cas9	*Streptococcus thermophiles*	5'-NNAGAAW-3'	II	(109)
St3Cas9	*Streptococcus thermophiles*	5'-NNGGNG-3'	II	(109)
BlCas9	*Brevibacillus laterosporus*	5'-NNNCND-3'	II	(110)
AsCpf1 LbCpf1	*Acidaminococcus* *Lachnospiraceae*	5'-TTTC-3'	V	(14)
C2c1	*A. acidoterrestris*	5'-TTC-3'	V	(111)

Comparison of the Various Genome Editing Technologies

While commercial availability and technical aspects of design, assembly, and delivery are significant criteria when choosing a nuclease, performance is of equal importance. Factors that govern performance are influenced by the model system of choice (cell line, model species, etc.), the efficiency of nucleic acid delivery, and the presence of polymorphisms in the target region. Indeed, there are examples for each of the major genome editing technologies where very high frequency of modification is observed. Targeting efficiencies of 1%-50% have been reported for TALENS and ZFNs (22,23,24) and efficiencies of up to 70% have been reported for CRISPRs (13,25,26).

Despite the growing amount of data using these systems, there are very few examples of side-by-side comparisons, making it difficult to evaluate which system will work most effectively for an unvalidated target gene or sequence. Recently, the efficiency of TALENs was directly compared to that of CRISPR/Cas9. Site-specific editing of an EGFP transgene in mammalian cells showed that CRISPR/Cas9 is more efficient and precise than TALENs in the absence of a homology-repair DNA template. Contrarily, when supplied with a repair template, TALENs performed more efficiently (27).

Another study has focused on the signature of TALENs and ZFNs (insertions vs deletions). TALENs introduced mostly deletions whereas ZFNs introduced both insertions and deletions (28). However, a more systematic comparison of these techniques is required to draw any conclusions.

Very recently, a direct comparison of CRISPR, CRISPR-interference (CRISPRi) and shRNA on gene modulation was established by using a lethality screen (essential genes). Targeting of 46 essential and 47 nonessential genes showed that CRISPR performed better than the shRNA and CRISPRi methodologies as it allowed a better distinction of the two groups of genes. In this experiment, shRNA-mediated knockdown resulted in more noise (29).

A key aspect of genome editing technologies is their potential use in high-throughput screening applications for understanding gene function and drug target identification. While in principle all mentioned genome editing techniques are suitable for production of gene targeting libraries, there are some practicalities that make the use of the CRISPR/Cas9 the primary choice for such applications. In the next section, we will discuss the advantages of CRISPR/Cas9 in this context and its use in screening.

Screening

Cell-based high-throughput screening in the pharmaceutical industry serves two main purposes: a) the identification of genes underlying phenotypes that correlate with disease (identification of drug targets); and b) the identification of chemical or biological agents that bind to or modulate the activity of such drug targets in cell-based assays (phenotypic screening). Of particular interest in recent years has been high-content screening, due to the ease of use, the potential for multiplexing several readouts, and the high amount of information that can be recorded from relatively simple experiments (30). Nonetheless, enzymatic reporter assays and general viability assays are also very powerful tools even in the age of HCS (31).

Traditionally, modulation of gene expression using genomic libraries has been achieved with siRNA/shRNA libraries (see book-part://[cbrnai]). Although successful in many applications, the use of these libraries presents some challenges. Notably, off-target activities of siRNAs or shRNAs and incomplete knockdown represent major sources of false positives and false negatives, respectively. Nonetheless, the technology has been widely used and been successful for the identification of gene function in multiple organisms. Further, siRNA-based approaches (but also CRISPR/Cas9 activation and repression screens; see below) are preferable when identifying the gene function of essential genes, since in this case complete knockout will obscure any functional readout by disruption of cellular function. The two key reasons why genome editing technologies may overtake the

widespread use of siRNA technology in the near future are: *1) the possibility to create a "clean" knockout of a gene, where residual activity is not present and can be neglected; and 2) the reduced confusion created by off-target activities.*

The use of genome editing technologies in high-throughput approaches has not yet reached the same ease of automation and throughput as other technologies, but protocols are being developed by industry and academic laboratories that will address the current limitations (32).

There is general consensus that the applicability of both TALENs and ZFNs methodologies in screening approaches is more reduced than that of CRISPR. To date, there has been very little use of TALENs, ZFNs and AAV-related techniques for functional genomic screening. This is due to the design complexity and time-consuming generation of TALENs and ZFNs, relative to CRISPR. Nevertheless, an innovative cloning system has recently allowed the generation of a TALEN library comprising TALEN plasmids for 18,740 protein-coding genes (33,34). In a pilot study targeting 126 of those genes, the efficiency was around 90%. Moreover, TALEN-mediated knockout of genes involved in the NF-κB pathway performed better than siRNA knockdown (35). However, no large-scale screens or using TALENs have been published to date. Therefore, we will focus below mainly on the use of CRISPR/Cas9 in screening applications.

The main advantage of the CRISPR/Cas9 technology for screening approaches is that the DNA-targeting molecule is a short RNA. Single guide RNA and crRNA libraries can be produced easily and cost-effectively, thus putting this technology at the forefront of functional genomic screening.

When using genomic libraries to modulate genome sequences such as those mentioned above, two types of screening approaches can be designed: 1. Pooled screening, and 2. Arrayed screening (32). Both have very different requirements in terms of library design, assay design, screen optimization, quality control, hit selection and post-screen validation that is summarized in Table 3 and discussed below.

Table 3: Advantages and disadvantages of pooled vs arrayed screening.

	Pooled screening	Arrayed screening
Advantages	• simple setup • requires no automation and specialized equipment • requires less experimental manipulation • suitable for cell viability/proliferation assays identification of drug resistance genes • comprehensive profiling of genome-wide library • cost-effective	• easy genotype-phenotype correlation • suitable for high-content screening assays • multi-parametric readouts possible • suitable for primary cells and neurons • custom-made libraries are available
Disadvantages	• requires high cell number • not suitable for primary cells and neurons • not suitable for high-content screening • biosafety considerations for virus libraries • requires a selection step • analysis depends on the availability of next-generation sequencing (NGS) • rare gene transcripts may be more difficult to identify • limited number of readouts	• requires specialized equipment (lab automation) • costly and time-consuming • quality control difficult (internal controls and statistical adjustments may be needed) • genome editing efficiency and transfection efficiency may be low • plate and well variability high

Pooled Library Screening

CRISPR/Cas9 screens can be performed in two different formats (arrayed vs. pooled libraries). Most CRISPR-based screens performed to date utilize pooled libraries. In short, a pooled library is a single preparation of many different sgRNA plasmids. The use of pooled libraries requires a positive or negative selection step, and thus has limited use for high-content screening approaches. However, there is great potential for pooled library screening

as the setup is relatively simple and does not require extensive automation and high-throughput screening capabilities. Usual readouts include cell viability and/or proliferation measurements, and the identification of drug resistance genes (36). Several studies reported in the literature comprise investigations in several fields, such as for the identification of cancer tumour suppressor genes (37), viral infection (38), cell cycle regulation and DNA replication. CRISPR/Cas9 screens have been successfully performed not only in mammalian cells but also in Drosophila cells (39) and Zebrafish (40). Recently, a resource summarizing all high-throughput CRISPR/Cas9 screens has been published and is frequently updated (41).

In general, pooled lentiviral screening with CRISPR/sgRNA libraries uses several of the same principles as pooled shRNA screening. This is likely the main reason why pooled screening approaches have been developed very quickly, with essentially all the technology for library production, screening workflow and hit identification already established in key laboratories. Many of these principles can be found in book-part://[cbrnai] and will not be discussed in detail here. However, there are key differences in terms of library design, assay optimization, hit identification and validation that will be outlined below.

Library Design

Pooled Libraries consist of a pool of sgRNAs in a lentiviral vector backbone that is used to transfect or transduce in bulk the cell line of interest. Several tools exist for identification of guide RNA target sequences in the genome. The two key features in producing a high-quality library is a) identification of unique target sites in the genome as to limit potential off-target activities; and b) to predict highest efficiency for genome editing. The identification of unique target sites is usually done by a search for PAM motifs combined with a BLAST algorithm (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to remove similar regions. Deleterious consequences of off-target activity can be mitigated by strategies described above, and improvements to Cas9-based platforms have been occurring at a rapid pace in the last couple of years resulting in increased fidelity (42,43,44).

The rules for predicting the highest efficiency genome editing sequence targets are not well understood and are lagging behind our understanding of how to select good siRNA molecules. Nonetheless, recent progress has made it clear that there are key determinants that improve design of sgRNA sequence. A recent study has investigated the cleavage activity of sgRNAs on single nucleotide mismatched targets (45). The study found a 4 nucleotide sequence located at +4 to -7 upstream of the PAM motif that can help in the design of gene specific sgRNAs. Another study found nucleotides both at PAM distal and proximal sites to be important for on-target activity (46). Another feature that has been recognized to interfere with genome editing efficiency is PAM density, reasoning that sites harboring multiple PAMs are less amenable to editing (47). Further determinants for improved design have been made (48), suggesting that G is preferred at the -1 and -2 positions proximal to the PAM sequence, a preference for C at the -3 position, a preference for A in positions from -5 to -12 and for G in positions from -14 to -17. However, this prediction is based on a limited set of data and the overall design rules for efficient guide RNAs need to be further investigated.

Several online CRISPR design tools now exist, and these enable the prediction of off-target sites, along with the probability of those events based on the individual sgRNA designs. Some of these are summarized in recent reviews (49,50). A list of online tools and websites for sgRNA design can be found in the appendix.

A very interesting experimental approach to generate sgRNA libraries from any gene without the need for bioinformatics prediction of target sequences has recently been described as an alternative (51). This method relies on the use of a semi-random primer containing a PAM complementary sequence and standard molecular biology tools to generate a sgRNA library. This method is particularly useful for generating sgRNA libraries for species where limited sequence information is available.

Several companies have developed their own sgRNA design tools, which make it easy to design and order the reagents with a very simple two-click system. Often, the problem is that too many sequences are identified and

selection becomes a matter of being spoiled for choice. Given that most sgRNA sequences work reasonably well, it might be best to opt for location in the gene sequence rather than predicted efficiency. For instance, sequences close to the 5'-end of a gene might be favorable, but one should keep in mind that alternative start sites might result in expression of a downstream gene fragment. Also, ribosomal frameshifting (52) and "illegitimate translation" (53) can result in expression of a gene from out-of-frame alleles. An alternative is to target within a functional or catalytic domain. Often, even an in-frame insertion can result in disruption of protein function, as we have observed for the ATG4B protease (Robin Ketteler, unpublished observation).

Currently available libraries are listed in the appendix.

Assay Optimization

A pooled CRISPR screen typically requires three components: the sgRNA library, Cas9 endonuclease, a cell model of interest and a sequence analyzer for hit identification.

Cells: When choosing the right cell line, the first choice should always be that the most relevant cell type is used to address the specific biological question. However, one might want to consider the ploidy of cells. For instance, in diploid cells, one would need to edit two alleles in order to get a complete loss of function phenotype. In polyploid cells, this might be more complicated. To date there is no evidence that the number of alleles has an effect on genome editing efficiency, although it has been suggested that the efficiency may be lower with increasing number of alleles (54). Another consideration is that 1 in 3 modifications can result in an in-frame indel, thus potentially not leading to a loss of gene function. Identified phenotypes can therefore be a consequence of mosaicism in gene function in a certain cell. Furthermore, interfering with the endogenous DNA repair mechanism may interfere with genome editing efficiencies. All of these caveats will require addressing when validating individual hits from pooled library screening.

Delivery of library: Often, the pooled library is supplied as a glycerol stock where each plasmid needs to be amplified before use. This is a critical step, as during amplification one has to ensure that the representation of each sgRNA is maintained. In a next step, the plasmids will be delivered to producer cells that produce the lentivirus particles for transduction of the target cells. This step is critical and assessment of virus titer is necessary. During transduction of target cells, the key requirement is that each cell receives only one virus on average. This can be best achieved by transducing cells with a mixture of lentiviral particles at a low multiplicity of infection that allows a single integration per cell. On the other hand, one has to ensure that each plasmid is represented in the infected cell population. Therefore, an excess of cells is usually used, at least 100-fold higher than the number of plasmids. This also means that a large portion of cells will not be transduced. Another consideration in pooled lentiviral screening that affects the number of cells required for hit identification is that each transcript is represented at high enough numbers so that even rare transcripts can be identified. For this reason, the number of cells in the beginning of the screen should be sufficiently high to enable such representation. Typically, 100-200 million cells are transduced and screened.

Delivery of Cas9 endonuclease: The CAS9 endouclease can be delivered in conjunction with the sgRNA (using the same vector) or provided in trans, either in the form of a stable Cas9 expressing cell line or through transient delivery of the Cas9 gene or protein into cells. Some studies have suggested that the transient delivery of Cas9 protein can be advantageous to reduce off-target effects when compared to a stable Cas9 expressing cell line. However, for practical reasons and to reduce variability between batches, a stable Cas9 cell line is often used.

Selection: In the next step, a positive or negative selection will be applied to allow the identification of sgRNA-mediated editing events that lead to a desired phenotype. In a positive selection screen, a positive selective pressure will eliminate all cells that do not survive the selection (e.g. treatment with puromycin). The few cells that survive can then be sequenced and the underlying sgRNA identified. Positive selection screens are typically very robust, if the selection pressure is very strong. Obviously, the right amount of selective pressure has to be

optimized in a first experiment on a control population. In a negative selection screen, most cells will survive. Therefore, one has to perform next-generation sequencing on the starting population, apply selection, and perform another round of next-generation sequencing on the surviving cells. By comparing the results from the initial population with the selected population, one can identify the list of gRNAs that are underrepresented during negative selection. For both, positive and negative selection screens, access to next-generation sequencing platforms is essential. An alternative to positive or negative selection is the use of a fluorescent marker protein and cell sorting.

Pooled library screening is not feasible for cell types that have limited proliferative potential or a limited passage number such as neurons or primary cells. Also, pooled screening is limited for identification of morphological phenotypes, such as those typically observed in high-content screening assays.

Hit Identification

Pooled screens require a selection step in order to enrich for cells displaying the desired phenotype. Ultimately, this will result in the differential representation of transcripts in the selected cell population that can be identified by next-generation sequencing. The identity of each sgRNA in the final population can be determined by one of two approaches: the sgRNA sequence can be determined by designing primers for PCR amplification that recognize the vector backbone and are flanking the sgRNA, thus effectively directly sequencing the sgRNA. An alternative is to use sgRNA-specific barcodes in the lentiviral vector backbone that are sequenced and then infer the corresponding sgRNA sequence. The advantage of a barcode approach is that probe sequences can be designed so that optimal hybridization settings can be obtained for each plasmid, resulting in equal PCR amplification and representation of transcripts. Identification of sequences can be done by microarrays (55) or deep sequencing. Recently, a multiplexed barcoded library approach has been developed that allows hit identification in a high-throughput manner (56). In order to capture underrepresented sgRNAs, it is important to reach a high level of transcript representation, and – as mentioned above – to start with a sufficiently high cell number. There are a few things worth noting:

Sequencing Depth: A key consideration is sequencing depth. It has been estimated that more than $1x10^7$ reads are required for a complex library. While normalization is usually not required, it is important to assess the representation of each sgRNA in the initial library before selection. Therefore, a cumulative distribution normalized to the total number of reads may help in this. This can also be important for quality control, as deviations from this curve may indicate a loss of diversity in the starting population.

Representation/Reads: After determination of the representation of each sgRNA, it is important to apply statistical measures to determine the significance of the results. Several user-friendly programs have been designed to aid in statistical analysis, including RIGER (57), RSA (58), HitSelect (59) and MaGeCK (60).

Recently, an R package for data analysis and documentation of pooled CRISPR library screens has been developed (caRpools (61)).

Validation

Post-screen validation of hits from pooled and arrayed library screening use common principles and will be discussed below.

Arrayed Library Screening

The use of arrayed libraries is still in its early days and only recently have they become available. Arrayed libraries present a broader applicability, allowing direct analysis of image-based phenotypes as well as biochemical readouts based on intensity (colorimetric/fluorescence/luminescence). In this sense, the use of arrayed libraries is more direct. In some cases a selection step for the cells that incurred a genetic modification

may be required, whereas for robust biochemical readouts (such as luminescence), this step might be unnecessary. The same is true when the targeting efficiency is very high.

Library Design

Guide RNA design follows very similar rules for arrayed libraries as for pooled libraries. Tools for identifying target sequences are described above and listed in the appendix. Here, we will only focus on particular requirements for arrayed libraries.

An arrayed library usually consists of guide or crRNAs targeting a single gene within each well in a multi-well microplate format. In some cases, there is more than one sgRNA/crRNA per well, but the key point is that each well contains agents that target the same single gene. The main advantage is that the gene-function correlation can be much more easily established than by selection of over- or underrepresented sequences in a mixed pool.

Unlike pooled libraries, arrayed libraries can occur in many different formats.

Lentiviral arrayed libraries are available, and these often contain the same vectors and sequences as the corresponding pooled libraries. The use of lentiviral libraries provide an efficient delivery to almost any cell type, but may require a selection step and culture of the cells for extended time (>7 days) before analysis.

Synthetic crRNA libraries: Synthetic libraries of small RNA molecules have been successfully delivered to cells at high efficiency and with similar protocols developed for siRNA oligonucleotides. In this case, high transfection efficiency is essential as there is no ability to select cells. Experiments are on a shorter time frame (within a few days). When using crRNAs, the tracrRNA has to be provided in trans alongside the Cas9 endonuclease.

In vitro transcribed sgRNA libraries: Another format that is being offered, although usually at lower scale due to the cost in manufacturing, is an *in vitro* transcribed sgRNA library. These are being offered as custom libraries and it is unlikely that they will be provided as an off-the-shelf solution for large libraries.

It is worth noting that chemically synthesized crRNAs have been suggested to be more versatile due to the possibility of having incorporated chemical modifications that increase gene editing efficiency (see review (62)). It can be expected that further improvements in this area may result in alternative library design in the future.

The key advantages of arrayed libraries are that the gene-function correlation is much easier to establish, the handling can be automated using standard liquid handlers, and an arrayed format is easily adaptable to high-content screening technologies. Also, it is possible to deliver the Cas9-sgRNA as a ribonucleoprotein complex, which has been suggested to reduce off-target activities (63). A key disadvantage is the cost of arrayed libraries, plus uncertainties associated with the genome editing efficiency, which may be below the detection limit of conventional high-content screening assays (see Table 3).

The libraries currently available are designed to target the whole genome, individual classes of genes (e.g. tumour suppressors, hormone receptors) and specific pathways (e.g. angiogenesis, apoptosis, cell cycle). Additionally, custom-made libraries that match the need of the research community are also available. Resources for pooled and arrayed libraries are listed in the appendix.

Assay Optimization

An optimal assay is one that delivers the maximum number of reproducible positive hits and a minimum of false negatives. In arrayed CRISPR library screening, very similar measures can be applied as in siRNA or small molecule library based approaches. Thus, the optimization of an assay starts with evaluation of a positive (i.e. exerting the maximum known phenotypic effect) and a negative (i.e. neutral) control. A good signal-to-noise window is necessary and statistical measures for effect size such as a Z'-factor can be used to evaluate the reagents (64). One complication in CRISPR based approaches is that penetrance of a phenotype is typically

much lower than for siRNA or small molecule screening approaches. For instance, if genome editing efficiency is below 20%, it becomes hard to identify a phenotype in image-based approaches (unless the phenotype is an off-on switch of a fluorescently labeled reporter gene). Assays with a low background noise and a high signal-to-noise window obviously perform better than assays that record morphological changes in organelle structure or number.

The main issue is to ensure that the observed phenotype is indeed linked to the desired mutation or edited locus.

The parameters that will affect this are:

- transfection efficiency (the phenotype should only be observed in cells that have been transfected)
- genome editing efficiency (the phenotype should only be observed in cells that have been successfully edited)
- off-target activity (the phenotype should only be observed in cells that have been edited at the correct location)

Accordingly, controls should be deployed to test each of these steps.

- For transfection/transduction efficiency, one could include a control such as a plasmid encoding for a fluorescent protein that can be easily quantified. Alternatively, a reagent that results in cell death can be used, such as sgRNAs targeting the Plk1 gene, which is often used as readout for the setup of RNAi screens. As a negative control, one could use a scrambled sgRNA or a sgRNA targeting an exogenous gene such as firefly luciferase or GFP.
- Assessing genome editing efficiency during a screen is often difficult. Ideally, a positive control could be used here, targeting a gene that is known to exert a strong phenotypic effect. Genes often used are HPRT1, TK or DHFR, which can be used as selection markers.
- Off-target activity is most often assessed post-screening, when secondary assays are used to validate hits. However, off-target effects can be potentially reduced by selecting the right library in the beginning, as discussed above.

Depending on the method of delivery, the optimal conditions of transfection/transduction should be determined beforehand. For lentiviral particles, the multiplicity of infection (MOI) that gives the best transduction and genome editing efficiency must be empirically determined. Unlike pooled screens where it is crucial to have one integrant per cell, in arrayed screening it is less important to restrict the number of integrations per cell as all viral particles within one well will target the same gene. Therefore much higher MOIs are typically used. We recommend the use of an MOI of 2-5 for the sgRNA particles.

Assessing genome editing efficiency is usually done for identified hits post-screening. There are possibilities for investigating genome editing in a high-throughput manner, though, and if costs for such methods decrease in the coming years, it is in principle feasible to do this important quality control step while screening. The technologies for assessment of genome editing efficiency will be discussed below in the "validation" section.

Hit Identification

Hits in arrayed CRISPR library screening can be identified by the commonly used Z score measures that incorporate sample size and deviations from normal distribution to identify genes that enhance or reduce a phenotype. When edge effects might interfere with the assay, a B score can be applied post-screening to normalize for these effects. Often, a simple signal to background calculation is sufficient to identify the strongest phenotypes.

Validation

For high-throughput screening using genome editing technologies, it is very important to attribute the phenotypic effect to a successfully edited genome modification. There are multiple methods available, from a very simple setup requiring only a PCR instrument to very complex technologies that rely on next-generation sequencing (Table 4). Accordingly, the cost, but most importantly the potential of each technique is very different and requires careful assessment before choosing the appropriate method.

Table 4: Overview of validation methods.

Method	Cost	Resolution	High-throughput potential	Equipment	Comments
Sanger sequencing	low	high	yes	Sequencer	needs sub-cloning, time-consuming; can use deconvolution software
Mismatch Cleavage Detection Assay	low	low	no	Standard lab equipment	not sensitive, not quantitative
Poly-acrylamid gel electrophoresis	low	low	no	Standard lab equipment	simple, cheap method, but non-quantitative
Indel Detection by Amplicon Analysis (IDAA)	moderate	moderate	yes	PCR machine	simple method with good resolution
High Resolution Melt analysis (HRM)	low	high	yes	RT-PCR machine	simple, cheap method with high resolution
Competitive PCR	low	moderate	yes	PCR machine	simple and cheap method, but non-quantitative
Digital Droplet PCR	high	high	yes	Digital Droplet PCR machine	simple method with high resolution, amenable to 96-well format
Next generation sequencing	high	high	yes	Next generation sequencer	most comprehensive analysis of genome editing; has potential to detect off-target effects

PCR based assays

Mismatch Cleavage Detection Assay: Technologies with relatively low throughput include mismatch cleavage assay detection (65) such as the Surveyor assay (Invitrogen) and GeneArt Genomic Cleavage Detection Kit (ThermoFisher) that utilize the properties of an error-prone polymerase for amplification of fragments that harbor mismatch mutations.

PCR-PAGE: An alternative is the use of poly-acrylamide gel electrophoresis, where PCR fragments are generated from the genomic region, subjected to denaturation and re-annealing, and analysis of differential patterns of migration by homo- or heteroduplexes (66).

High Resolution Melting Curve Analysis: An alternative with higher throughput is high resolution melting (HRM) curve analysis (67). HRM makes use of different melting curve patterns for two individual PCR products and is efficient to detect indels with a resolution close to 1 nucleotide.

Indel Detection by Amplicon Analysis (IDAA): Another PCR-based method termed Indel Detection by Amplicon Analysis (IDAA) relies on tri-primer labeling of a PCR product and detection by gel electrophoresis (68).

Competitive PCR: A simple and cheap semiquantitative method to identify mutated cells based on competition-based PCR has been developed, where a mixture of three primers with one primer overlapping the Cas9 cleavage site can be used (69).

Digital Droplet PCR: Recently, Digital Droplet PCR has been used to detect genome editing events in a high-throughput manner (70). This method is based on partitioning of PCR reactions on single targets into droplets and using a fluorescence-based readout for each individual droplet PCR to visualize changes in the DNA sequence. Quantification of edited and wild-type alleles can be done in the same sample. More recent versions of this approach have enabled the simultaneous detection of NHEJ and HDR events in a single reaction mixture (71).

Sequencing based assays

Sanger Sequencing: Traditionally, genome editing has been assessed for single genes by sub-cloning the affected locus followed by Sanger sequencing. However, this method is time-consuming and not amenable to large-scale interrogation of genome editing events.

Next-Generation Sequencing: One caveat of the PCR based assays is that they only detect that a mutation has occurred but they do not detect the specific sequence alterations exerted by genome editing. Next-generation sequencing technologies can overcome this limitation. For instance, a variant of next generation sequencing, CRISPR Genome Analyzer (CRISPR-GA), can be used to map the specific alterations detected (72). CRISPR-GA provides information on size and location of indels, as well as the efficiency of NHEJ and HDR events. This is a huge advantage in terms of validation, as one could use this to determine if the phenotype corresponds to an edited site. An alternative is BATCH-GE (73), which provides batch analysis features, thus reducing the time required for analysis.

Additional validation steps

After a cell is identified as carrying the appropriate mutation additional steps are necessary for validation that the phenotype is directly the result of the known mutation and not due to some other effect. This can be done using secondary phenotypic assays which can include one or more of the following:

1. Rescue of the phenotype by expression of an exogenous normal copy of the mutated gene. This can be achieved by transfection or transduction of a plasmid or retrovirus particle, respectively. One caveat is that overexpression may result in a higher copy number, thus not representing the original endogenous levels.
2. Reversion of the genome modification back to wild-type by genome editing. In this approach it is important to include controls to demonstrate that the mutated cell line has been rescued and that the initial unmutated cell line has not been "reisolated" due to contamination. This can be easily accomplished by including insertion of silent mutations into the reversion step.

Conclusion

In summary, genome editing technologies have greatly advanced over the last decade and provided new insights into the function of genes and proteins. Recently, these technologies have been applied to high-throughput screening applications, with the CRISPR/Cas9 system at the forefront of this development. It can be expected that further improvements to this technology will be made in this fast-moving field and some of the challenges such as delivery and efficiency will be overcome in the not too distant future.

Acknowledgements

This work was supported by the Medical Research Council, grant number MC_U12266B.

Appendix

Commercial availability of ZFNs, TALENs, CRISPR/Cas9

ZFNs,TALENs and rAAVs may be custom synthesized or created. They may be ordered either by user design or by having the company design and synthesize and clone the relevant protein. There is a consortium for the development of ZFNs.

As of 2017, the following companies are known to provide ZFN synthesis services and/or reagents to create ZFNs:

- Sigma-Aldrich (www.sigma-aldrich.com)
- Available through Addgene (www.addgene.com)
  - Oligomerized Pool Engineering
  - Context dependent assembly
  - Modular assembly (Barbas kit, Joung kit, Wolfe system)

As of 2017, the following companies offer Tal synthesis services and or reagents to create TALENs:

- Cellectis Bioresearch (http://www.cellectis.com/)
- Life Technologies (www.lifetechnologies.com)
- Available through Addgene (www.addgene.com)
  - FLASH assembly
  - Modular assembly (Voytas kit, Joung kit, Zhang kit)

As of 2017, the following companies offer CRISPR/Cas9 development services:

- Agilent
- Dharmacon/GE
- Genecopeia
- Horizon Genomics
- IDT
- MolDiag Solutions
- Sigma/Merck
- Synthego
- ThermoFisher

Online Resources for Guide and crRNA Design

When choosing the best design algorithm for sgRNA, there are multiple choices (see below). All have in common a basic search algorithm for a PAM motif sequence and some blast search against potential off-target effects, i.e. similar sequences. The most useful websites are the ones that provide clear information and visualization of results, for example when displaying the location of the cut site, providing a suggestion of potential sequencing primers and scores about similar sequences that are easy to understand.

Design tools	Considerations	Reference
sgRNA Scorer 2.0	Identify putative guide sequences and assign a predicted activity. Considers off-target effects. Uses CasFinder algorithm.	(74)
CRISPR-DO (CRISPR Design and Optimization)	Predicts sgRNA efficiency and off-target scoring; applied to several genomes	(75)

Table continued from previous page.

Design tools	Considerations	Reference
Drosophila RNAi Screening Center (DRSC) Find CRISPRs	Very good graphical visualization in JBrowse. Considers off-target effects.	(76)
ge-CRISPR	Evaluates potential off-target sequences; very user-friendly tool	(77)
CT-Finder	Graphical visualization on JBrowse. Considers off-target effects. Applied to several genomes	(78)
CHOPCHOP	Considers off-target effects. Very informative table. Cas9 nuclease, Cas9 nickase, Cpf1 and TALEN	(79)
Breaking-Cas	Interactive design	(80)
Cas-Database	Graphical visualization in JBrowse. Both Cas9 and Cpf1	(81)
CLD (CRISPR Library Designer)	Design of CRISPR/Cas9 libraries. Also suitable for targeting non coding regions.	(82)
WU-CRISPR	Human or mouse genomes. No graphical visualization.	(83)
CRISPR-GA (Genome Analyzer)		(72)
CRISPRscan	Predictions are available as tracks that can be uploaded in the UCSC genome browser. Applies to several species	(84)
CRISPR-ERA	Editing, repression and activation; Cas9 nuclease and Cas9 nickase. Predictions are available as tracks that can be uploaded in the UCSC genome browser.	(85)
Protospacer Workbench		(86)
WGE | Sanger Institute	Along with Gibson assembly PCR oligo designer. Genoverse browser	(87)
E-CRISP		(88)
Phyto-CRISP-Ex	Protist genomes, 2 sequential filters are applied.	(89)
CRISPR-P	Applicable to Plants	(90)
Off-Spotter	Interactive way of selecting the seed sequence	(91)
CRISPR Direct		(92)
CRISPRseek	Bioconductor package	(93)
sgRNAcas9	Software package	(94)
GPP Web portal sgRNA designer for CRISPRa and CRISPRi	Gene activation and gene repression. No graphical visualization	(95)

Available Libraries for Screening

As of this date (2017), the following libraries for screening using CRISPR/Cas9 are available:

Commercial	Sanger Whole Genome & Sigma CRISPR arrayed libraries Thermo Fisher (LentiArray CRISPR Libraries, TrueGuide CRISPR Libraries) Dharmacon (synthetic guide RNA and lentiviral guide RNA) Genecopeia GenScript Horizon Adgene Agilent

Table continued from previous page.

Academic	Pooled lentiviral libraries: Mouse genome-wide lentiviral CRISPR-gRNA library (96) Human and mouse genome-wide lentivrial CRISPR/-gRNA library (GECKO) (97,98) Human genome-wide lentiviral CRISPR/gRNA library (99) Drosophila genome-wide CRISPR library (39) Human genome-wide lentiviral library (Toronto KnockOut TKO) (100) Human and mouse CRISPRi and CRISPRa libraries (101) Human lncRNA targeting library (102) Arrayed libraries: Human genome-wide, sequence-verified, arrayed lentiviral CRISPR library (103) Human and mouse lentiviral arrayed library (104)

References

1. Gaj T, Gersbach CA, Barbas CF 3rd. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends in biotechnology. 2013;31:397–405. PubMed PMID: 23664777.
2. Hirata R, Chamberlain J, Dong R, Russell DW. Targeted transgene insertion into human chromosomes by adeno-associated virus vectors. Nature biotechnology. 2002;20:735–738. PubMed PMID: 12089561.
3. Daya S, Berns KI. Gene therapy using adeno-associated virus vectors. Clinical microbiology reviews. 2008;21:583–593. PubMed PMID: 18854481.
4. Russell DW. AAV vectors, insertional mutagenesis, and cancer. *Molecular therapy : the journal of the American Society of Gene Therapy* **15**, 1740-1743 (2007).
5. Kim YG, Cha J, Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proceedings of the National Academy of Sciences of the United States of America. 1996;93:1156–1160. PubMed PMID: 8577732.
6. Durai S, Mani M, Kandavelou K, Wu J, Porteus MH, Chandrasegaran S. Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. Nucleic acids research. 2005;33:5978–5990. PubMed PMID: 16251401.
7. Li L, Wu LP, Chandrasegaran S. Functional domains in Fok I restriction endonuclease. Proceedings of the National Academy of Sciences of the United States of America. 1992;89:4275–4279. PubMed PMID: 1584761.
8. Miller J, McLachlan AD, Klug A. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. The EMBO journal. 1985;4:1609–1614. PubMed PMID: 4040853.
9. Mussolino C, Cathomen T. TALE nucleases: tailored genome engineering made easy. Current opinion in biotechnology. 2012;23:644–650. PubMed PMID: 22342754.
10. Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing. Nature reviews Molecular cell biology. 2013;14:49–55. PubMed PMID: 23169466.
11. Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. RNA-programmed genome editing in human cells. eLife. 2013;2:e00471. PubMed PMID: 23386978.
12. Cong L, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339:819–823. PubMed PMID: 23287718.
13. Mali P, et al. RNA-guided human genome engineering via Cas9. Science. 2013;339:823–826. PubMed PMID: 23287722.
14. Zetsche B, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell. 2015;163:759–771. PubMed PMID: 26422227.
15. Gilbert LA, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. 2013;154:442–451. PubMed PMID: 23849981.
16. Qi LS, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. 2013;152:1173–1183. PubMed PMID: 23452860.

17. Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nature protocols. 2013;8:2180–2196. PubMed PMID: 24136345.
18. Hilton IB, Gersbach CA. Enabling functional genomics with genome engineering. Genome research. 2015;25:1442–1455. PubMed PMID: 26430154.
19. Chen B, et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell. 2013;155:1479–1491. PubMed PMID: 24360272.
20. Ma H, Naseri A, Reyes-Gutierrez P, Wolfe SA, Zhang S, Pederson T. Multicolor CRISPR labeling of chromosomal loci in human cells. Proceedings of the National Academy of Sciences of the United States of America. 2015;112:3002–3007. PubMed PMID: 25713381.
21. Rauch BJ, *et al.* Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. *Cell* **168**, 150-158 e110 (2017).
22. Miller JC, et al. A TALE nuclease architecture for efficient genome editing. Nature biotechnology. 2011;29:143–148. PubMed PMID: 21179091.
23. Mussolino C, Morbitzer R, Lutge F, Dannemann N, Lahaye T, Cathomen T. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic acids research. 2011;39:9283–9293. PubMed PMID: 21813459.
24. Maeder ML, et al. Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Molecular cell. 2008;31:294–301. PubMed PMID: 18657511.
25. Hwang WY, et al. Heritable and precise zebrafish genome editing using a CRISPR-Cas system. PloS one. 2013;8:e68708. PubMed PMID: 23874735.
26. Feng Z, et al. Efficient genome editing in plants using a CRISPR/Cas system. Cell research. 2013;23:1229–1232. PubMed PMID: 23958582.
27. He Z, Proudfoot C, Whitelaw CB, Lillico SG. Comparison of CRISPR/Cas9 and TALENs on editing an integrated EGFP gene in the genome of HEK293FT cells. SpringerPlus. 2016;5:814. PubMed PMID: 27390654.
28. Kim Y, Kweon J, Kim JS. TALENs and ZFNs are associated with different mutation signatures. Nature methods. 2013;10:185. PubMed PMID: 23396284.
29. Evers B, Jastrzebski K, Heijmans JP, Grernrum W, Beijersbergen RL, Bernards R. CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. Nature biotechnology. 2016;34:631–633. PubMed PMID: 27111720.
30. Ketteler R, Kriston-Vizi J. High-Content Screening in Cell Biology. *Waltham, MA: Academic Press* **4**, 234-244 (2016).
31. Ketteler R. The Feynman trajectories: determining the path of a protein using fixed-endpoint assays. Journal of biomolecular screening. 2010;15:321–326. PubMed PMID: 20130209.
32. Agrotis A, Ketteler R. A new age in functional genomics using CRISPR/Cas9 in arrayed library screening. Frontiers in genetics. 2015;6:300. PubMed PMID: 26442115.
33. Reyon D, Tsai SQ, Khayter C, Foden JA, Sander JD, Joung JK. FLASH assembly of TALENs for high-throughput genome editing. Nature biotechnology. 2012;30:460–465. PubMed PMID: 22484455.
34. Reyon D, *et al.* Engineering customized TALE nucleases (TALENs) and TALE transcription factors by fast ligation-based automatable solid-phase high-throughput (FLASH) assembly. *Current protocols in molecular biology / edited by Frederick M Ausubel [et al]* **Chapter 12**, Unit 12 16 (2013).
35. Kim Y, et al. A library of TAL effector nucleases spanning the human genome. Nature biotechnology. 2013;31:251–258. PubMed PMID: 23417094.
36. Kurata M, et al. Using genome-wide CRISPR library screening with library resistant DCK to find new sources of Ara-C drug resistance in AML. Scientific reports. 2016;6:36199. PubMed PMID: 27808171.
37. Katigbak A, Cencic R, Robert F, Senecha P, Scuoppo C, Pelletier J. A. CRISPR/Cas9 Functional Screen Identifies Rare Tumor Suppressors. Scientific reports. 2016;6:38968. PubMed PMID: 27982060.
38. Park RJ, et al. A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors. Nature genetics. 2017;49:193–203. PubMed PMID: 27992415.

39. Bassett AR, Kong L, Liu JL. A Genome-Wide CRISPR Library for High-Throughput Genetic Screening in Drosophila Cells. *Journal of genetics and genomics = Yi chuan xue bao* **42**, 301-309 (2015).
40. Shah AN, Davey CF, Whitebirch AC, Miller AC, Moens CB. Rapid reverse genetic screening using CRISPR in zebrafish. Nature methods. 2015;12:535–540. PubMed PMID: 25867848.
41. Rauscher B, Heigwer F, Breinig M, Winter J, Boutros M. GenomeCRISPR - a database for high-throughput CRISPR/Cas9 screens. Nucleic acids research. 2017;45:D679–D686. PubMed PMID: 27789686.
42. Cho SW, et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome research. 2014;24:132–141. PubMed PMID: 24253446.
43. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nature biotechnology. 2014;32:279–284. PubMed PMID: 24463574.
44. Ran FA, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell. 2013;154:1380–1389. PubMed PMID: 23992846.
45. Zheng T, et al. Profiling single-guide RNA specificity reveals a mismatch sensitive core sequence. Scientific reports. 2017;7:40638. PubMed PMID: 28098181.
46. Liu X, Homma A, Sayadi J, Yang S, Ohashi J, Takumi T. Sequence features associated with the cleavage efficiency of CRISPR/Cas9 system. Scientific reports. 2016;6:19675. PubMed PMID: 26813419.
47. Malina A, Cameron CJ, Robert F, Blanchette M, Dostie J, Pelletier J. PAM multiplicity marks genomic target sites as inhibitory to CRISPR-Cas9 editing. Nature communications. 2015;6:10124. PubMed PMID: 26644285.
48. Xu H, et al. Sequence determinants of improved CRISPR sgRNA design. Genome research. 2015;25:1147–1157. PubMed PMID: 26063738.
49. Chuai GH, Wang QL, Liu Q. In Silico Meets In Vivo: Towards Computational CRISPR-Based sgRNA Design. Trends in biotechnology. 2017;35:12–21. PubMed PMID: 27418421.
50. Mohr SE, Hu Y, Ewen-Campen B, Housden BE, Viswanatha R, Perrimon N. CRISPR guide RNA design for research applications. The FEBS journal. 2016;283:3232–3238. PubMed PMID: 27276584.
51. Arakawa H. A method to convert mRNA into a gRNA library for CRISPR/Cas9 editing of any organism. Science advances. 2016;2:e1600699. PubMed PMID: 27574704.
52. Ketteler R. On programmed ribosomal frameshifting: the alternative proteomes. Frontiers in genetics. 2012;3:242. PubMed PMID: 23181069.
53. Makino S, Fukumura R, Gondo Y. Illegitimate translation causes unexpected gene expression from on-target out-of-frame alleles created by CRISPR-Cas9. Scientific reports. 2016;6:39608. PubMed PMID: 28000783.
54. Miles LA, Garippa RJ, Poirier JT. Design, execution, and analysis of pooled in vitro CRISPR/Cas9 screens. The FEBS journal. 2016;283:3170–3180. PubMed PMID: 27250066.
55. Brummelkamp TR, et al. An shRNA barcode screen provides insight into cancer cell vulnerability to MDM2 inhibitors. Nature chemical biology. 2006;2:202–206. PubMed PMID: 16474381.
56. Wong AS, et al. Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. Proceedings of the National Academy of Sciences of the United States of America. 2016;113:2544–2549. PubMed PMID: 26864203.
57. Luo B, et al. Highly parallel identification of essential genes in cancer cells. Proceedings of the National Academy of Sciences of the United States of America. 2008;105:20380–20385. PubMed PMID: 19091943.
58. Konig R, et al. A probability-based approach for the analysis of large-scale RNAi screens. Nature methods. 2007;4:847–849. PubMed PMID: 17828270.
59. Diaz AA, Qin H, Ramalho-Santos M, Song JS. HiTSelect: a comprehensive tool for high-complexity-pooled screen analysis. Nucleic acids research. 2015;43:e16. PubMed PMID: 25428347.
60. Li W, et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome biology. 2014;15:554. PubMed PMID: 25476604.
61. Winter J, et al. caRpools: an R package for exploratory data analysis and documentation of pooled CRISPR/Cas9 screens. Bioinformatics. 2016;32:632–634. PubMed PMID: 26508755.

62. Kelley ML, Strezoska Z, He K, Vermeulen A, Smith A. Versatility of chemically synthesized guide RNAs for CRISPR-Cas9 genome editing. Journal of biotechnology. 2016;233:74–83. PubMed PMID: 27374403.
63. Liang X, et al. Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. Journal of biotechnology. 2015. PubMed PMID: 26003884.
64. Zhang JH, Chung TD, Oldenburg KR. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. Journal of biomolecular screening. 1999;4:67–73. PubMed PMID: 10838414.
65. Langhans MT, Palladino MJ. Cleavage of mispaired heteroduplex DNA substrates by numerous restriction enzymes. Current issues in molecular biology. 2009;11:1–12. PubMed PMID: 18541926.
66. Zhu X, et al. An efficient genotyping method for genome-modified animals and human cells generated with CRISPR/Cas9 system. Scientific reports. 2014;4:6420. PubMed PMID: 25236476.
67. Thomas DK, et al. Comparative variation within the genome of Campylobacter jejuni NCTC 11168 in human and murine hosts. PloS one. 2014;9:e88229. PubMed PMID: 24516617.
68. Yang Z, et al. Fast and sensitive detection of indels induced by precise gene targeting. Nucleic acids research. 2015;43:e59. PubMed PMID: 25753669.
69. Harayama T, Riezman H. Detection of genome-edited mutant clones by a simple competition-based PCR method. PloS one. 2017;12:e0179165. PubMed PMID: 28586390.
70. Nelson CE, et al. In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. Science. 2016;351:403–407. PubMed PMID: 26721684.
71. Miyaoka Y, et al. Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. Scientific reports. 2016;6:23549. PubMed PMID: 27030102.
72. Guell M, Yang L, Church GM. Genome editing assessment using CRISPR Genome Analyzer (CRISPR-GA). Bioinformatics. 2014;30:2968–2970. PubMed PMID: 24990609.
73. Boel A, et al. BATCH-GE: Batch analysis of Next-Generation Sequencing data for genome editing assessment. Scientific reports. 2016;6:30330. PubMed PMID: 27461955.
74. Chari R, Mali P, Moosburner M, Church GM. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. Nature methods. 2015. PubMed PMID: 26167643.
75. Ma J, et al. CRISPR-DO for genome-wide CRISPR design and optimization. Bioinformatics. 2016;32:3336–3338. PubMed PMID: 27402906.
76. Housden BE, Lin S, Perrimon N. Cas9-based genome editing in Drosophila. Methods in enzymology. 2014;546:415–439. PubMed PMID: 25398351.
77. Kaur K, Gupta AK, Rajput A, Kumar M. ge-CRISPR - An integrated pipeline for the prediction and analysis of sgRNAs genome editing efficiency for CRISPR/Cas system. Scientific reports. 2016;6:30870. PubMed PMID: 27581337.
78. Zhu H, Misel L, Graham M, Robinson ML, Liang C. CT-Finder: A Web Service for CRISPR Optimal Target Prediction and Visualization. Scientific reports. 2016;6:25516. PubMed PMID: 27210050.
79. Labun K, Montague TG, Gagnon JA, Thyme SB, Valen E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. Nucleic acids research. 2016;44:W272-276. PubMed PMID: 27185894.
80. Oliveros JC, et al. Breaking-Cas-interactive design of guide RNAs for CRISPR-Cas experiments for ENSEMBL genomes. Nucleic acids research. 2016;44:W267-271. PubMed PMID: 27166368.
81. Park J, Kim JS, Bae S. Cas-Database: web-based genome-wide guide RNA library design for gene knockout screens using CRISPR-Cas9. Bioinformatics. 2016;32:2017–2023. PubMed PMID: 27153724.
82. Heigwer F, et al. CRISPR library designer (CLD): software for multispecies design of single guide RNA libraries. Genome biology. 2016;17:55. PubMed PMID: 27013184.
83. Wong N, Liu W, Wang X. WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. Genome biology. 2015;16:218. PubMed PMID: 26521937.
84. Moreno-Mateos MA, et al. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nature methods. 2015;12:982–988. PubMed PMID: 26322839.

85. Liu H, Wei Z, Dominguez A, Li Y, Wang X, Qi LS. CRISPR-ERA: a comprehensive design tool for CRISPR-mediated gene editing, repression and activation. Bioinformatics. 2015;31:3676–3678. PubMed PMID: 26209430.
86. MacPherson CR, Scherf A. Flexible guide-RNA design for CRISPR applications using Protospacer Workbench. Nature biotechnology. 2015;33:805–806. PubMed PMID: 26121414.
87. Hodgkins A, et al. WGE: A CRISPR database for genome engineering. Bioinformatics. 2015. PubMed PMID: 25979474.
88. Heigwer F, Kerr G, Boutros M. E-CRISP: fast CRISPR target site identification. Nature methods. 2014;11:122–123. PubMed PMID: 24481216.
89. Rastogi A, Murik O, Bowler C, Tirichine L. PhytoCRISP-Ex: a web-based and stand-alone application to find specific target sequences for CRISPR/CAS editing. BMC bioinformatics. 2016;17:261. PubMed PMID: 27363443.
90. Lei Y, Lu L, Liu HY, Li S, Xing F, Chen LL. CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. Molecular plant. 2014;7:1494–1496. PubMed PMID: 24719468.
91. Pliatsika V, Rigoutsos I. "Off-Spotter": very fast and exhaustive enumeration of genomic lookalikes for designing CRISPR/Cas guide RNAs. Biology direct. 2015;10:4. PubMed PMID: 25630343.
92. Naito Y, Hino K, Bono H, Ui-Tei K. CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. Bioinformatics. 2015;31:1120–1123. PubMed PMID: 25414360.
93. Zhu LJ, Holmes BR, Aronin N, Brodsky MH. CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. PloS one. 2014;9:e108424. PubMed PMID: 25247697.
94. Xie S, Shen B, Zhang C, Huang X, Zhang Y. sgRNAcas9: a software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites. PloS one. 2014;9:e100448. PubMed PMID: 24956386.
95. Doench JG, et al. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nature biotechnology. 2014;32:1262–1267. PubMed PMID: 25184501.
96. Koike-Yusa H, Li Y, Tan EP, Velasco-Herrera Mdel C, Yusa K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nature biotechnology. 2014;32:267–273. PubMed PMID: 24535568.
97. Shalem O, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science. 2014;343:84–87. PubMed PMID: 24336571.
98. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. Nature methods. 2014;11:783–784. PubMed PMID: 25075903.
99. Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic screens in human cells using the CRISPR-Cas9 system. Science. 2014;343:80–84. PubMed PMID: 24336569.
100. Hart T, et al. High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. Cell. 2015;163:1515–1526. PubMed PMID: 26627737.
101. Horlbeck MA, et al. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. eLife. 2016;5 PubMed PMID: 27661255.
102. Liu SJ, et al. CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. Science. 2017;355 PubMed PMID: 27980086.
103. Erard N, Knott SRV, Hannon GJ. A CRISPR Resource for Individual, Combinatorial, or Multiplexed Gene Knockout. *Molecular cell* **67**, 348-354 e343 (2017).
104. Metzakopian E, et al. Enhancing the genome editing toolbox: genome wide CRISPR arrayed libraries. Scientific reports. 2017;7:2244. PubMed PMID: 28533524.
105. Kleinstiver BP, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature. 2015. PubMed PMID: 26098369.
106. Anders C, Bargsten K, Jinek M. Structural Plasticity of PAM Recognition by Engineered Variants of the RNA-Guided Endonuclease Cas9. Molecular cell. 2016;61:895–902. PubMed PMID: 26990992.
107. Ran FA, et al. In vivo genome editing using Staphylococcus aureus Cas9. Nature. 2015;520:186–191. PubMed PMID: 25830891.

108. Hou Z, et al. Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proceedings of the National Academy of Sciences of the United States of America. 2013;110:15644–15649. PubMed PMID: 23940360.
109. Muller M, *et al.* Streptococcus thermophilus CRISPR-Cas9 Systems Enable Specific Editing of the Human Genome. *Molecular therapy : the journal of the American Society of Gene Therapy* **24**, 636-644 (2016).
110. Karvelis T, et al. Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. Genome biology. 2015;16:253. PubMed PMID: 26585795.
111. Yang H, Gao P, Rajashankar KR, Patel DJ. PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. *Cell* **167**, 1814-1828 e1812 (2016).

Suggested Reading

Iiizumi et al, Simple one-week method to construct gene-targeting vectors: application to production of human knockout cell lines, Biotechniques 41:311-316 2006

Jantz and Berg, Probing the DNA-binding affinity and specificity of designed zinc finger proteins Biophys J 98:852-860, 2010

Klug, Towards therapeutic applications of engineered zinc finger proteins FEBS Letters 579:892-894 2005

Perez-Pinera et al, Advances in targeted genome editing, Curr Opin Chem Biol 16:268-277, 2012

Suggested Websites

www.addgene.com

www.agilent.com

www.cellectis.com

www.genecopoeia.com

www.horizondiscovery.com

www.idtdna.com

www.lifetechnologies.com

www.moldiag.in

www.sigma-aldrich.com

www.synthego.com

License