

Guidelines for optimized gene knockout using CRISPR/Cas9

Claude Van Campenhout¹, Pauline Cabochette², Anne-Clémence Veillard¹, Miklos Laczik¹, Agnieszka Zelisko-Schmidt¹, Céline Sabatel¹, Maxime Dhainaut³, Benoît Vanhollebeke^{2,4}, Cyril Gueydan^{3,5} & Véronique Kruys^{*,1,5}

ABSTRACT

CRISPR/Cas9 technology has evolved as the most powerful approach to generate genetic models both for fundamental and preclinical research. Despite its apparent simplicity, the outcome of a genome-editing experiment can be substantially impacted by technical parameters and biological considerations. Here, we present guidelines and tools to optimize CRISPR/Cas9 genome-targeting efficiency and specificity. The nature of the target locus, the design of the single guide RNA and the choice of the delivery method should all be carefully considered prior to a genome-editing experiment. Different methods can also be used to detect off-target cleavages and decrease the risk of unwanted mutations. Together, these optimized tools and proper controls are essential to the assessment of CRISPR/Cas9 genome-editing experiments.

KEYWORDS:

gene targeting • methodology • quality control

¹Diagenode, SA, Liège Science Park, 4102 Seraing, Belgium; ²Laboratoire de Signalisation Neurovasculaire, Faculté des Sciences, Université libre de Bruxelles (ULB), 12 rue des Profs. Jeener et Brachet, 6041 Gosselies, Belgium; ³Precision Immunology Institute, Department of Genetics & Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; ⁴Walloon Excellence in Life Sciences & Biotechnology (WELBIO), Belgium; ⁵Laboratoire de Biologie Moléculaire du Gène, Faculté des Sciences, Université libre de Bruxelles (ULB), 12 rue des Profs. Jeener et Brachet, 6041 Gosselies, Belgium; *Author for correspondence: vkruys@ulb.ac.be; [†]These authors contributed equally

BioTechniques: 295-302 (June 2019)
10.2144/btn-2018-0187

Engineered nucleases, from zinc-finger nucleases to TALENs and CRISPRs, form a powerful class of genome-editing tools [1]. Among these, the CRISPR/Cas9 system has become the most popular, owing to its ease of use and rapidity. The CRISPR/Cas system was discovered in prokaryotes where it provides adaptive immunity against foreign elements [2]. In 2013, the CRISPR/Cas9 system from *Streptococcus pyogenes* (spCas9, further indicated in the text as Cas9) was successfully adapted for genome editing in eukaryotic cells [3]. Since then, the technique has become extremely popular as it can modify the genome of a large variety of organisms from animals to plants with unprecedented ease.

However, there is still limited predictability of whether the CRISPR system will be able to effectively target a given region of interest. This aspect is of particular importance in the context of CRISPR/Cas9-based screens in model organisms and is related to the definition of the target site and the sequence of small guide RNA (sgRNA). Another major hurdle common to all engineered nucleases is the risk of unwanted mutations at sites other than the intended on-target site (off-target effects). The off-target mutations are the consequence of sgRNA binding to DNA sites with less than perfect complementarity [4]. Current strategies to increase targeting specificity notably include: refinements in guide RNA selection, enzyme and guide engineering, and improvements in the delivery method. Here, we describe a series of guidelines to optimize CRISPR/Cas9 efficiency and specificity.

ANALYSIS OF THE TARGET LOCUS

Careful determination of target sites is essential. For many applications, a loss of function may be desirable or even required. Targeting of functional protein domains was recently demonstrated to result in higher

proportions of loss-of-function mutations [5]. A common strategy is to select sgRNAs that will target Cas9 nuclease to the N-terminal-coding exons of protein-coding genes. After the action of Cas9 nuclease, the introduction of indels by the error-prone non-homologous end joining repair of double-strand breaks (DSBs) introduces frame-shift mutations and subsequent premature stop codons, leading to mRNA elimination by nonsense-mediated mRNA decay. Genome-editing experiments to generate knockouts should be designed to disrupt exons that are shared by all transcript variants of a given gene. This strategy can also be applied to whole gene families using a sgRNA against exons that are conserved between all family members [6]. The CRISPy algorithm aims to design the optimal sequence to target multiple members of a gene family [7].

The high frequency with which CRISPR-induced mutations can be directed to target genes enables easy isolation of homozygous gene knockouts. Paradoxically, a potential caveat is found in this high efficiency. This holds particularly true in cell lines upon targeting genes essential for cell viability and fitness. In this regard, two distinct genome-wide CRISPR-Cas9-based screens have identified ~2000 essential genes in the human genome [8]. More recently, Lenoir and colleagues published a database of pooled *in vitro* CRISPR knockout library essentiality screens that can be searched to identify genes that are essential across different human tissues [9].

Genetic screens in zebrafish and mouse have estimated that as many as 30% of genes are embryonic lethal [10,11]. The functional characterization of such essential genes requires the generation of heterozygous knockouts. The generation of hypomorphic alleles with the CRISPR system has been reported by different groups [12,13], but the method is not, at the moment, commonly used. ▶

► RNAi or CRISPRi [14] are efficient alternative loss-of-function methods and their effects can be directly evaluated at the transcriptome level. In addition, the development of inducible CRISPR tools provides a solution for genome editing with tight temporal control [15]. They additionally circumvent the mechanisms of genetic compensation that not unfrequently mask the phenotypes of knockout but not knockdown models [16].

Genetic polymorphism in the target region should be carefully assessed as it might have a profound influence on CRISPR/Cas9 efficacy. Although base mismatches (up to five) may be tolerated between the sgRNA and targeted sequences, the PAM and its proximal sequence have a stricter adherence to the consensus [17]. When a sgRNA is selected, the potential presence of a SNP in the PAM and the sgRNA-binding site should be verified as it can abolish Cas9 binding and cleavage. Of note, commonly used laboratory cell lines such as HeLa cells present a variant spectrum that slightly differs from the one found in the human population [18]. In general, sequences found in genomic databases may not exactly correspond to the DNA sequences of the model used for the genome-editing experiment. Sequencing of the target locus prior to sgRNA design will solve this potential pitfall. On the other hand, this PAM constraint can be exploited to target and disrupt heterozygous single-nucleotide mutations in certain dominant autosomal disorders, while leaving the wild-type allele intact [19]. Cell line ploidy is an additional consideration to take into account. Many common laboratory cancer cell lines carry four or more copies of a chromosome. Full knockouts would then require the introduction of mutations in all copies of the target gene. In practice, it is strongly advised to sequence the target loci to verify homozygous knockout when generating mutant clonal cell lines.

Besides the influence of the sequence features, chromatin states also strongly impact Cas9 binding and nuclease activity in vertebrates. Nucleosomes constitute fundamental units of chromatin and their positioning directly impedes Cas9 binding and cleavage *in vitro* and *in vivo*. Highly active sgRNAs for Cas9 are found almost exclusively in regions of low nucleosome occupancy [20]. Higher order chromatin

structure (i.e., organization beyond the level of the linear array of nucleosomes) also influences Cas9 binding and enzymatic activity. Several authors showed that Cas9 cleavage efficiency positively correlates with open chromatin based on DNase I hypersensitivity. Along the same line, the activity of Cas9 can be significantly hindered by compact heterochromatin in cells [21]. Interestingly, the engineered Cas9 variants designed to improve specificity, Cas9-HF1 [22] and eSpCas9(1.1) [23], might be even more impacted than Cas9 by the chromatin-related factors [24]. While some gene editing applications have the option to select easy-to-cleave targets, such practice may not be feasible for gene corrections and other potential therapeutic applications. Many CRISPR genome-editing experiments focus on gene targeting and the study of the phenotypic consequences. In these applications, the gene of interest is usually transcriptionally active and the associated chromatin is relatively accessible to Cas9. Nevertheless, chromatin compactness can vary considerably between different genomic sites and from one cell type to another. Gene targeting in model organisms presents an additional challenge as the chromatin landscape is under constant change to ensure coordinated growth and differentiation during early development. Atlases of transcriptional activity (RNA-Seq) and of chromatin accessibility (e.g., ATAC-Seq and ChIP-Seq) are valuable information resources (of note see the ENCODE project: www.encodeproject.org) to predict sgRNA efficiency [25] and have been used to elaborate a predictive algorithm for zebrafish sgRNA selection taking into account chromatin accessibility [26]. Gene editing in mouse and human cells has been greatly facilitated by the publication of the genome-wide Brie and Brunello libraries [27]. These optimized sgRNA libraries respectively target the mouse and human genomes and provide 3–4 sgRNA sequences per gene with predicted high on-target efficiency and low off-target effects.

For more challenging applications such as the editing of heterochromatin-embedded sequences, chromatin manipulation might enhance the CRISPR targeting efficiency. While treatment with chromatin-disrupting drugs does not appear sufficient, transient overexpression of a targeted transcriptional activator might be an effective method to

enhance Cas9 editing at closed chromatin regions [21].

DELIVERY METHODS

Introduction of the CRISPR/Cas9 components into cultured cells is often achieved by DNA-based delivery systems such as transfection of plasmids encoding nuclear targeted Cas9 and sgRNA. Transduction with viral particles is also commonly used and is typically more efficient compared with plasmid transfection and is applicable to many cell types including primary cells. Plasmid transfection and viral transduction methods lead to a prolonged or a permanent expression of Cas9, respectively. Extended expression of Cas9 in cells can lead to accumulation of off-targeting events [28]. Indeed, constitutive expression of lentiviral-based Cas9 and sgRNAs leads to an enrichment of predicted off-target sites over time. Reducing the concentration of delivered plasmid during transfection was shown to decrease off-targeting [4]. These data support the idea that controlling the expression of Cas9 and the sgRNAs in order to limit the time of action can reduce genome-wide off-targeting. A doxycycline-inducible promoter allows for transient Cas9 expression and is compatible with lentiviral delivery of the nuclease [29]. Because gene editing results in a permanent change in the genome, CRISPR-mediated editing can be achieved using Cas9 protein/sgRNA ribonucleoprotein (RNP) complexes. Both sgRNAs and nuclear-targeted Cas9 protein can be produced in-house [30] or obtained from different commercial suppliers. RNP complexes can be delivered by a variety of techniques such as lipid-mediated transfection [30], electroporation [28], induced transduction by osmocytosis and propanebetaine (iTOP) [31], microinjection [32] or cell-penetrating peptide-mediated delivery [33]. Uncoupling administration of the sgRNA and Cas9 protein (e.g., in the context of genome-scale screens) can lead to successful gene editing in human primary cells [34] and appears to be more efficient upon delivery of Cas9 protein complexed with a nontargeting gRNA [35]. Finally, biolistic transfer of Cas9/sgRNA RNP complexes or of Cas9- and sgRNA-encoding plasmids appears to be an attractive alternative for cells resistant to other delivery methods, such as plant cells [36].

The sgRNA-Cas9 RNPs were shown to cleave the target chromosomal DNA between 12 and 24 h after delivery and the frequency of gene editing reaches a plateau after 1 day. For plasmid expression of Cas9 and sgRNA, equivalent gene editing levels were only achieved at 3 days after delivery. Furthermore, the Cas9 protein has been shown to be degraded rapidly in cells, within 24–48 h after delivery, compared with several days when continuously expressed from a plasmid [28]. Moreover, the ratio of the indel frequency at the on-target site to off-target sites strongly increases when RNPs are transfected in comparison with plasmid [33]. While off-target effects may be less of a concern in screening applications since any identified ‘hits’ will be confirmed through follow-up experiments, constitutive expression or high stability of Cas9 nuclease and/or sgRNA may be undesirable for many applications, such as generation of clonal cell lines for a phenotypical study of a specific gene knockout. In addition to the increased potential for off-target effects due to prolonged or constitutive expression of components of the CRISPR-Cas9 system, unwanted incorporation of the plasmid DNA into the cell genome is not uncommon. When the DNA repair pathways are activated after Cas9-mediated double-strand breaks, the risk of foreign DNA integration is increased. The absence of transgene eliminates the risk of unintended DNA integration.

The delivery of RNP complexes also has the major advantage of being easily applicable to a wide range of model organisms and cell types. In *in vivo* contexts, the functionality of the Cas9–sgRNA RNP complexes has been reported as being superior to other delivery methods. In zebrafish, mutagenesis can be performed through microinjection of Cas9-encoding mRNA or of Cas9 protein together with sgRNA into fertilized embryos. Contrary to Cas9-encoding mRNA, RNPs are immediately active upon microinjection and are generally more effective in the fast dividing blastomeres of the zebrafish and *Xenopus* embryos (Figure 1) [37,38], in which mutagenesis occurring after the first cleavages leads to increased mosaicism. In plants, this DNA-free approach allows generation of marker-free ‘cisgenic’ variants that could be exempted from current GMO regulations and thereby supersede all the technologies of gene editing in plants [39].

The fact that sgRNAs can be easily synthesized *in vitro* makes it possible to use multiple sgRNAs simultaneously to achieve multi-genetic targeting [40]. The DNA-free system also suppresses the variability that can arise from the choice of promoter used to drive expression from vector-based CRISPR-Cas9 systems. It is well known that not all promoters are functional in every cell type or cell line, so delivery of Cas9 protein or Cas9 mRNA avoids incompatibilities of certain promoters in specific cells. Codon usage patterns also vary between species and Cas9 derived from DNA or mRNA expression may not yield the expected result as every organism has its own codon bias. Optimization of codon usage is a routine process but can be relatively time-consuming. Codon optimization becomes unnecessary when using Cas9 protein instead of a DNA- or mRNA-based delivery method. Independent of the delivery mode, specific anti-Cas9 antibodies can be used to measure Cas9 expression level and to confirm Cas9 presence in the nucleus (Figure 2).

Lentiviral or plasmid delivery of Cas9 and sgRNA often utilizes a selection gene encoding either a drug-selectable marker (hygromycin, blasticidin or puromycin) or a reporter protein (GFP and NGFR) to isolate cells that are successfully transduced or transfected. When RNPs are transfected or electroporated, alternative strategies can be used, such as surrogate reporters [44]. However, these methods are inefficient for assessing sgRNA efficiency at a large scale because it is both time- and labour-consuming to construct a specific reporter for each individual sgRNA. To avoid specific cloning, the transfection efficiency can also be indirectly evaluated with a dTomato reporter assay [31]. Fluorescent versions of Cas9 such as Cas9–GFP [45] or Cas9–Cy3 [46] can also be used to sort RNP-transfected cells. These latter methods focus on the physical separation of edited cells from unedited cells. An important aspect to consider is that CRISPR experiments lead to a genetic heterogeneity due to the random nature of DNA repair by the NHEJ pathway. As this genetic heterogeneity could yield phenotypic heterogeneity, monoclonal populations should be isolated prior to phenotypic analysis. The first step is to determine the editing efficiency of the entire cell population. This information can

indicate how many individual clones should be isolated and checked for editing. If limited dilutions are used to isolate individual cells, it should be realized as soon as possible after termination of the edition process, as nonedited cells could potentially outgrow edited cells.

Gene editing in *in vivo* mouse models was greatly facilitated by the generation of a knock-in (KI) transgenic mouse in which a Cre-inducible Cas9-P2A-GFP cassette was inserted in the Rosa26 locus [47]. Cre-mediated recombination leads to cell- or tissue-specific Cas9 expression, as evidenced by GFP expression. Apart from allowing for gene editing following *in vivo* delivery of sgRNAs, this model can also be used to efficiently edit the genome of primary cells *ex vivo*.

GUIDE RNA EFFICIENCY & SPECIFICITY

The performance of sgRNAs targeting the same gene can vary dramatically. This was recently highlighted in a novel approach to CRISPR genomics where expression of sgRNAs was coupled with specific protein barcodes, allowing for simultaneous multi-dimensional phenotypic analysis of several dozen knockouts at a single-cell resolution [48]. In a pooled parallel analysis of gene editing efficiency for ten genes (3–4 sgRNAs per gene), the authors demonstrated that the gene knockout at the protein level was highly variable depending on the sgRNAs used. There are many bioinformatic tools available for sgRNA design and some of these tools also apply filters or show ‘scores’ related to predicted effectiveness. sgRNAs with potential for weak secondary structures are likely to be more efficient than alternatives with strong secondary structures [49]. Nevertheless, no computational tool can guarantee the efficacy of a sgRNA and, when possible, several sgRNAs should be tested. Endonuclease cleavage assays can be used to characterize the *in vitro* efficacy of a particular sgRNA. Experimental validation of sgRNAs before practical application is particularly important to minimize wasted experiments on sgRNAs with poor activity. In these *in vitro* assays, the target DNA site, including its PAM motif, is either inserted into a plasmid or provided in the form of a PCR product. The Cas9 recombinant ►

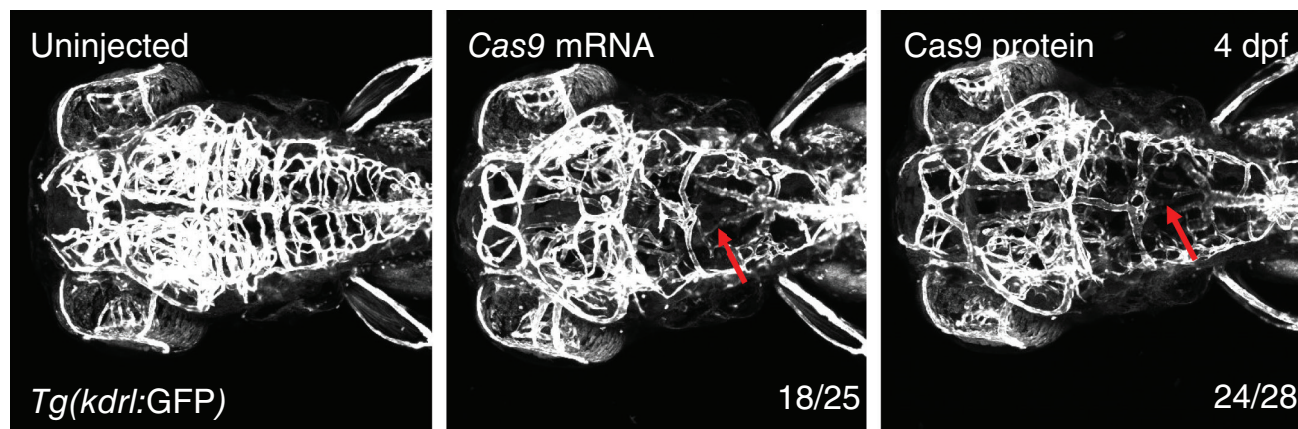


Figure 1. Cas9 Nuclease Protein NLS and Cas9 mRNA injections achieve highly efficient bi-allelic somatic gene disruptions in zebrafish. Tg(kdrl:GFP)^{s843} embryos were injected at the one-cell stage with 300 pg of Cas9 protein (Cas9 Nuclease Protein NLS from Diagenode, cat n° C29010001) or 150 pg of cas9 mRNA *in vitro* transcribed from the XbaI linearized pT3TS-nls-zcas9-nls vector (Addgene #46757) and 30 pg of two sgRNAs targeting *reck*, an essential regulator of CNS vascularization [41]. Right panel: maximal intensity projections of confocal z-stacks of the cranial vasculature of Tg(kdrl:GFP) larvae at 4 dpf. Red arrows point to the avascular hindbrains. The fraction of injected larvae showing cerebrovascular defects is indicated (18/25 and 24/28).
dpf: Days post-fertilization.

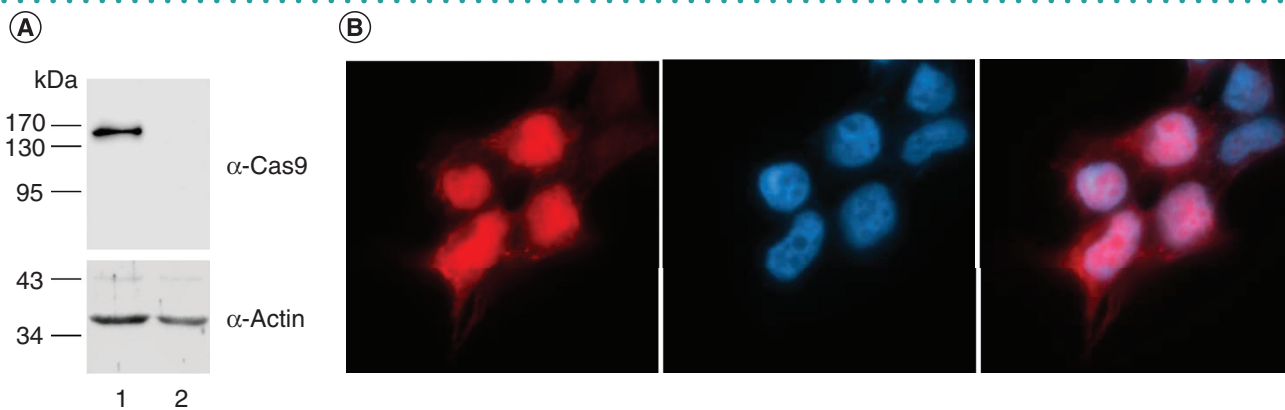


Figure 2. Cas9 expression can be detected by western blot and immunofluorescence. (A) Western blot analysis of HEK293T cells transfected with pSpCas9(BB)-2A-Puro (PX459) V2.0 plasmid (Addgene #62988) [42] encoding Cas9 using Lipofectamine 2000 (Life Technologies). Cell lysates from transfected (lane 1) or untransfected cells (lane 2) were analyzed by immunoblotting with mouse anti-Cas9 (Diagenode Cat No. C15200216). (B) HEK293T cells were transfected with dCas9 plasmid (Addgene #100091) [43]. dCas9 localization was analyzed by immunostaining using anti-Cas9 mouse monoclonal antibody (Diagenode Cat No. C15200216) and Alexa-Fluor 594-conjugated donkey anti-mouse antibody. Nuclei were counterstained with DAPI.

► protein and the sgRNA are pre-incubated in a 1:1 molar ratio in the cleavage buffer to reconstitute the Cas9–sgRNA complex prior to the addition of target DNA. Cleavage of plasmid or PCR substrates is monitored by agarose gel electrophoresis with an intercalating dye (Figure 3). The reaction rate can strongly vary in function of DNA source and length (PCR product versus plasmid, circular plasmid versus linear plasmid), optimal enzyme and substrate concentrations, and also reaction time points need to be determined empirically [50]. This *in vitro* test validates sgRNA intrinsic capacity to

form cleavage-competent complexes; however, it does not guarantee *in vivo* effectiveness, which also greatly depends on chromatin accessibility, as previously mentioned.

The targeting specificity of Cas9 is believed to be tightly controlled by the 20-nucleotide guide sequence of the sgRNA and the presence of a PAM adjacent to the target sequence in the genome. Nevertheless, potential off-target cleavage activity can still occur on DNA sequence with even 3–5 basepair mismatches in the PAM-distal part of the sgRNA-guiding sequence [4]. Of note,

shortening of the sgRNA guide sequence to 17 nucleotides was shown to improve target specificity [51]. Numerous online tools are available to assist in sgRNA design but the correlation between the predictions and the actual measurements varies considerably since sequence homology alone is not fully predictive of off-target sites [52]. These tools also suggest probable off-target sites but the appropriate number of potential sites to experimentally assay remains unclear. Moreover, there are still contradictory conclusions as to the prevalence of off-target effects, from low [53] to high levels of off-targeting [54].

Cleavage at on- and off-target sites can be assessed using various methods, which include mismatch-sensitive enzymes (Surveyor or T7 Endonuclease I assay), restriction fragment length polymorphism (RFLP) analysis, high-resolution melting curve analysis (HRMA) or PCR amplification of the locus of interest followed by sequencing. Surveyor and T7 Endonuclease I specifically cleave heteroduplex DNA mismatch. The T7 Endonuclease I assay outperforms the Surveyor nuclease in terms of sensitivity with deletion substrates, whereas Surveyor is better for detecting single nucleotide changes. The limit of sensitivity for the T7 Endonuclease I assay is around 5% [55]. HRMA utilizes the difference in melting curve of the heteroduplex and mutant homoduplex. A recent report demonstrates that techniques such as targeted NGS, tracking indels by decomposition (TIDE) and indel detection by amplicon analysis (IDAA) outperform nuclease-based methods to detect Cas9-mediated edition in pools of cells [56]. Ultimately, Sanger sequencing of DNA from individual clones is the gold standard for confirming the presence of indels at on-target site but is not easily applicable to off-target detection. Overall, these indels detection methods are relatively straightforward but are low throughput and interrogate one locus at a time.

Unbiased off-target analysis requires the detection of mutations generated in the target cells by the CRISPR/Cas9 system outside their target locus. In theory, WGS of cells before and after editing could be used to study CRISPR/Cas9 specificity. In a clonal population, off-target sites can be determined by the analysis of the new mutations that have been generated outside the intended locus. However, WGS faces its own challenges and might not be easily applicable to the detection of off-target mutations. While sequencing costs continue to drop, a certain degree of bioinformatic expertise is necessary to detect small indels and separate signal from noise. In fact, many spontaneous new mutations may appear during clonal expansion and it might not be possible to distinguish them from off-target effects. WGS of individual induced pluripotent stem cell clones reveals a large number of indels in the genome that are not the result of

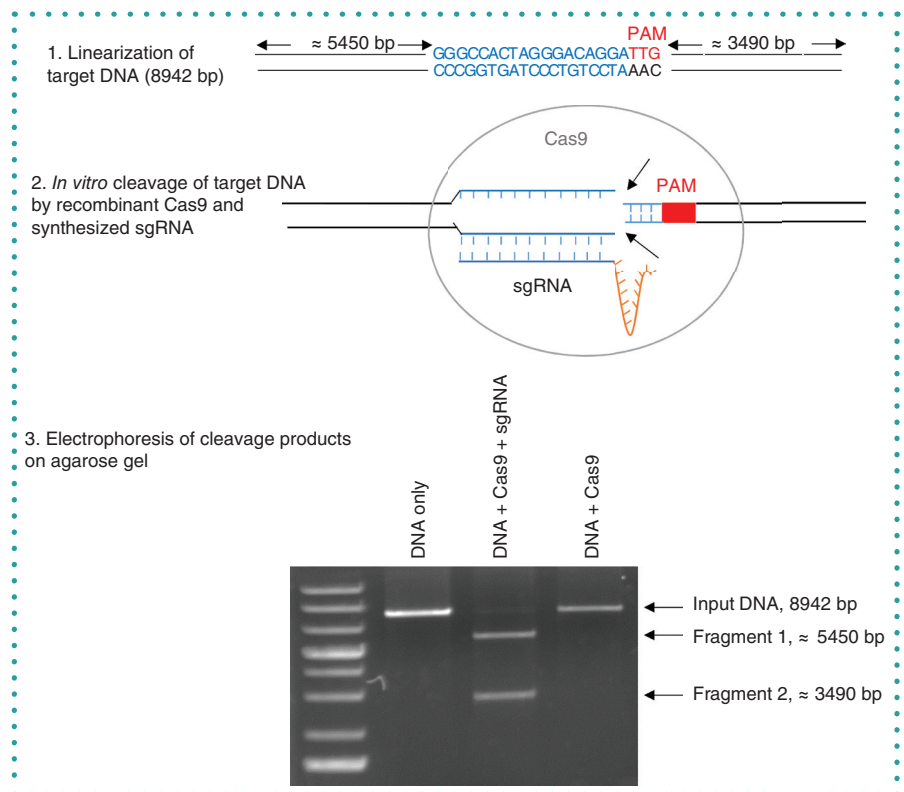


Figure 3. *In vitro* validation of sgRNA by Cas9 cleavage assay. Experimental design of the assay. DNA target is PvuI-linearized CRISPR-SP-Cas9 reporter plasmid (Addgene #62733) [31]. Corresponding sgRNA (GGGCCACUAGGGACAGGAU) was synthesized by *in vitro* transcription. Target DNA was incubated with (lanes 2 and 3) or without (lane 1) Cas9 recombinant protein (Cas9 Nuclease protein NLS from Diagenode, Cat No. C29010001). Reactions were set-up with a ratio of 20:20:1 (Cas9:sgRNA:DNA target) and incubated for 1 h at 37°C. The products were resolved on 0.8% TAE agarose gel stained with ethidium bromide.

Cas9 activity, but rather a consequence of clonal variation or technical artefacts [57]. To circumvent these limitations, several methods have recently been developed to measure Cas9 off-target activity across the genome such as BLESS (labeling of DSBs followed by enrichment and sequencing), HTGTS (high-throughput genome-wide translocation sequencing), GUIDE-Seq (genome-wide unbiased identification of DSBs enabled by sequencing), Digenome-Seq (*in vitro* Cas9-digested WGS) [58], IDLV (detection of off-targets using integrase-deficient lentiviral vectors) [59] and most recently, SITE-Seq (a biochemical method that identifies DNA cut sites) [60] and CIRCLE-Seq (an *in vitro* method for identifying off-target mutations) [61]. Overall, these unbiased methods tend to be less sensitive and have a lower throughput than biased targeted sequencing, in addition to typically requiring higher sequencing coverage and much more complex protocols. These techniques also require

manipulation of the genome and might be difficult to apply on some samples (e.g., primary cells, or *in vivo*).

Chromatin immunoprecipitation followed by next-generation sequencing (ChIP-Seq) is a technique of choice for studying protein–DNA interactions. ChIP has been used to pull down the Cas9 nuclease protein together with the DNA fragments to which the nuclease was bound. The immunoprecipitation of Cas9 bound to the genome is technically challenging due to the nuclease activity of Cas9. However, the introduction of two amino-acid changes (D10A and H840A) in Cas9-coding sequence results in a nuclease-inactive DNA-binding protein named ‘dead Cas9’ (dCas9). Specific enrichment of dCas9 at on-target regions can be evaluated by ChIP-qPCR using ChIP-grade Cas9 antibodies (Figure 4A). Moreover, this approach can be extended to the unbiased analysis of off-target sites by ChIP-Seq (Figure 4B). dCas9-based ChIP-PCR/Seq is thus a powerful approach to score several sgRNAs at once thanks to ▶

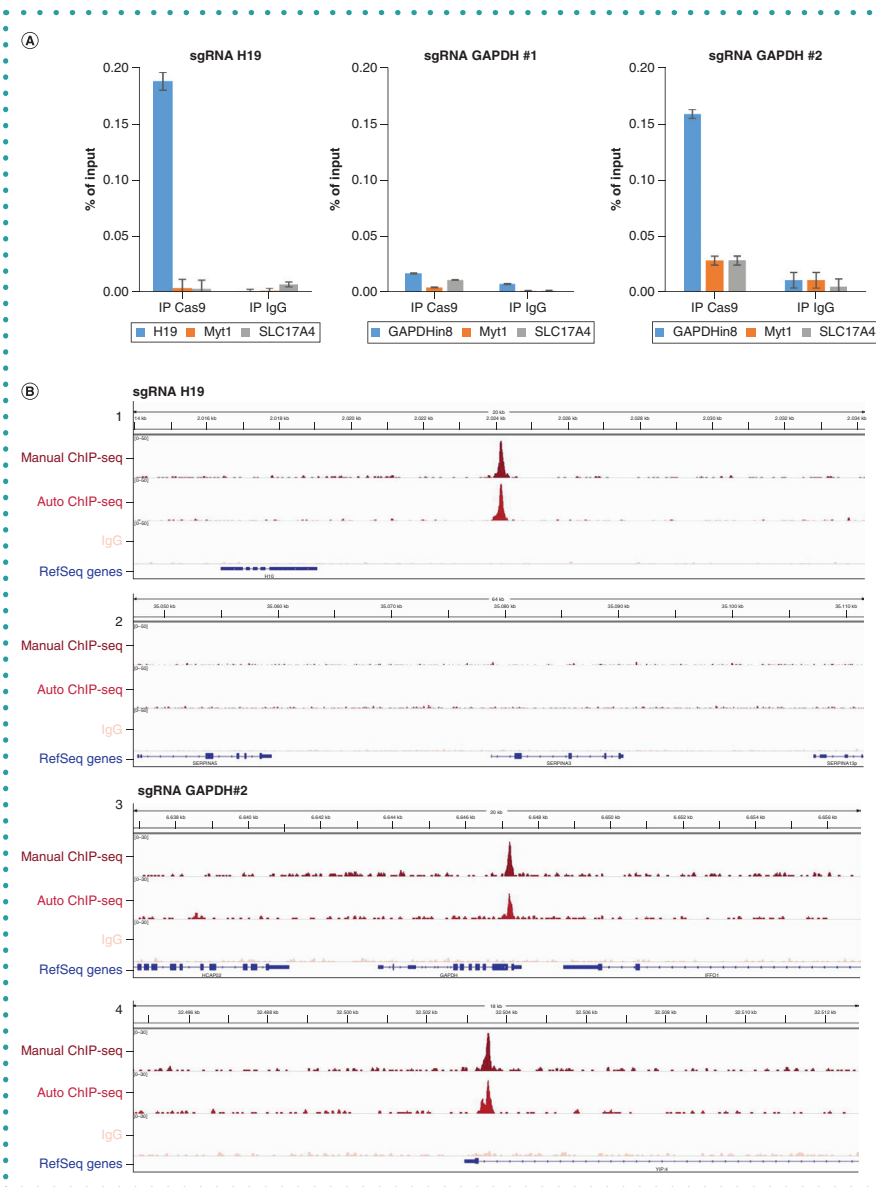


Figure 4. Evaluation of sgRNA specificity by dCas9 ChIP-PCR and ChIP-Seq. Sequences of sgRNA targeting the human H19 (GTCTATCTCTGACAACCCTC) or GAPDH sgRNA#1 GTCTGGCGCCTCTGTGGC and sgRNA#2 AAAGACTCGGTGGTGGTCT) loci were designed using an online selection tool (<http://crispr.mit.edu/>) and cloned in the pLenti-dCas9-2xAM plasmid (Addgene 92220) [63]. ChIP was performed on sheared chromatin from HEK293T cells using the iDeal ChIP-seq Kit for TF (Diagenode Cat No. C01010170), 5 μ l of the polyclonal Cas9 antibody (Diagenode Cat No. C15310258) and 5 μ l of control IgG (Diagenode Cat No. C15410206). (A) dCas9 ChIP-PCR analysis of one representative out of three independent experiments is shown (error bars show SD of replicates). Primers specific for the human H19, GAPDH, Myt1 and SLC17A4 were used for the qPCR. The figure shows the recovery, expressed as % of input. (B) dCas9 ChIP-Seq analysis. ChIP was performed manually or with the IP-Star[®] Compact Automated System (Diagenode Cat. No. B03000002) as described in the manual. Libraries were prepared using the kit MicroPlex (Diagenode Cat. No. C05010012) and sequenced on HiSeq 3000 using SE 50bp reads. The figure shows the read distribution for the manual IP (top), the automated IP (middle) and IgG (bottom) samples. Peaks show distribution of the three datasets in the region surrounding H19 (1) and a representative region of the genome (2) for HEK293T cells expressing a sgRNA for H19. Peaks show distribution of the three datasets in the region surrounding GAPDH (3) and YIPF4 (4) for HEK293T cells expressing a sgRNA targeting GAPDH (sgRNA GAPDH #2).

► its rapidity, reduced sequencing cost and high coverage. Moreover, it is of predictive value for sgRNA performance upon association with catalytically active Cas9, although Cas9 DNA-binding and cleavage activities are sometimes uncoupled [62]. As no single method guarantees a complete coverage of off-target sites, multiple approaches should ideally be combined. Therefore, sequence-based *in silico* prediction combined with genome-wide ChIP-Seq dCas9-binding analysis can efficiently identify off-target sites.

Variants of dCas9 have recently been generated that allow repurposing of the system to a variety of applications [64]. Fusing dCas9 to various transcriptional activating or repressing modules proved to be a potent way of regulating gene expression. Moreover, dCas9 can be fused to domains that regulate the epigenetic landscape at endogenous loci. It can also be used to label endogenous loci for live visualization [65] or to edit a single base in the genome [66]. In those applications, the binding specificity of dCas9 fused to various effectors could be tested by dCas9 ChIP-Seq as we describe here.

FUTURE PERSPECTIVE

From the first description of Cas9 derived from *S. pyogenes* for gene editing in 2013, incredible progress has been made to optimize and adapt its use in a wide range of applications. Structural studies of Cas9 led to the generation of several variants such as enhanced specificity Cas9 (eSpCas9), high-fidelity Cas9 (Cas9-HF1) and hyper-accurate Cas9 (HypaCas9), which display increased specificity due to reduced DNA-binding affinity (eSpCas9 and Cas9-HF1) [22,23] or locking of the nuclease domain upon guide/target mismatches (HypaCas9) [67]. In addition, Cas9 nickases (Cas9n) were developed by inactivating the cleavage activity on target or nontarget DNA and have been demonstrated to nick only one DNA strand instead of generating a DSB. DSBs are generated only upon recruitment of a Cas9n pair with two sgRNAs that target opposite strands in close proximity [4,68], thereby increasing specificity by double selection. A similar strategy was used to develop a fusion of dCas9 with the catalytic domain of FokI nuclease (fCas9), which induces DSBs only upon dimerization of the FokI domains by sgRNA pairing to comple-

mentary strands. With the same aim of reducing Cas9 off-target activity, several Cas9 variants whose editing activity can be irreversibly or reversibly programmed are now also available. Finally, engineered Cas9 variants with novel PAM specificities enlarge the edition spectrum to previously inaccessible sites [69,70].

Several limitations have yet to be addressed to promote Cas9 use in gene therapy. First, the source of Cas9 nucleases, that is, *S. pyogenes* and *Staphylococcus aureus*, are common human pathogens. Recent reports have highlighted pre-existing immunity towards both SpCas9 and SaCas9 in the human population, with a high prevalence of both Cas9-reactive T cells and antibodies [71]. Although it is still unclear whether adeno-associated virus delivery of Cas9 leads to the immune rejection of transduced cells *in vivo*, strategies to control the anti-Cas9 T cell responses, such as transient immunosuppression or engineering Cas9 proteins with mutated T cell epitopes, are being considered [72]. Another limitation of Cas9 for its use in gene therapy resides in its rather large size, which is incompatible for efficient packaging into adeno-associated virus vectors, the most commonly used delivery systems in gene therapy. Although this hurdle can be overcome by the separation of the recognition lobe from the nuclease lobe into two separate vectors, the emergence of Cas9 orthologs of smaller size might provide more efficient alternatives [69]. Beside solving the delivery problem, CRISPR effectors from other bacterial and archaeal species offer different substrate specificities or operate according to different mechanisms. This is notably the case of Cas12a (Cpf1), which is structurally different from Cas9, has no requirement for trailer RNA, recognizes a T-rich (TTTN) PAM sequence lying 5' of the target sequence, and uses a different mechanism for target recognition and cleavage [69]. Cpf1 also possesses the ability to cleave RNA and generate multiple crRNAs from a single pre-crRNA array. This capacity has been harnessed to achieve multiplex gene editing using a

single pre-crRNA array, which can both increase knockout efficiency (when using multiple crRNAs targeting the same locus) or easily knockout multiple genes with a single construct [73]. Moreover, gene editing by Cpf1 results in lower off-target effects than Cas9, as evidenced by genome-wide analysis of edited cells [74]. Finally, the discovery of Cas13a and CasRx as RNA-guided nucleases targeting RNA paves the way to new therapeutic approaches based on RNA editing [69,75].

While several solutions and guidelines to harness CRISPR-Cas9-based gene targeting have been provided, it is expected that therapeutic, industrial and research applications will still place high demand on improving the specificity and efficiency of the CRISPR/Cas9 system. As CRISPR-based gene targeting technology continues to become more sophisticated and diverse, optimized procedures and quality control guidelines should be established.

AUTHOR CONTRIBUTIONS

All the authors contributed to the content of the manuscript; CVC, MD, BV, CG and VK wrote the manuscript. CG and VK share senior authorship.

ACKNOWLEDGMENTS

We thank Romuald Soïn and Nadège Delacourt for technical support.

FINANCIAL & COMPETING INTERESTS DISCLOSURE

CVC was funded by a FIRST Enterprise grant from DGO6 from the Walloon Region, Belgium. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

OPEN ACCESS

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/> ►

COMPACT TISSUE HOMOGENIZATION



The 1200 GenoLyte® is the ideal solution for labs that want a compact yet powerful tissue homogenizer and cell lyser.

Equipped with interchangeable sample vial holders allowing a variety of vial types from 2 mL to 12 mL.

Designed for fast and efficient extraction of nucleic acids, proteins and other molecules of interest.

SPEX® SamplePrep®

(+1) 732-623-0465

learnmore@spex.com

SPEXSAMPLEPREP.COM/GENOLYTE

REFERENCES

- Gaj T, Gersbach CA, Barbas III CF. ZFN, TALEN and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 31(7), 397–405 (2013).
- Barrangou R, Fremaux C, Deveau H *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712 (2007).
- Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346(6213), 1258096 (2014).
- Hsu PD, Scott DA, Weinstein JA *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31(9), 827–832 (2013).
- Shi J, Wang E, Milazzo JP *et al.* Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat. Biotechnol.* 33, 661–667 (2015).
- Endo M, Mikami M, Toki S. Multigene knockout utilizing off-target mutations of the CRISPR/Cas9 system in rice. *Plant Cell. Physiol.* 56(1), 41–47 (2015).
- Hyams G, Abadi S, Lahav S *et al.* CRISPRs: optimal sgRNA design for editing multiple members of a gene family using the CRISPR system. *J. Mol. Biol.* 430(15), 2184–2195 (2018).
- Wang T, Birsoy K, Hughes NW *et al.* Identification and characterization of essential genes in the human genome. *Science* 350(6264), 1096–1101 (2015).
- Lenoir WF, Lim TL, Hart T. PICKLES: the database of pooled *in vitro* CRISPR knockout library essentiality screens. *Nucleic Acids Res.* 46(D1), D776–D780 (2018).
- Haffter P, Granato M, Brand M *et al.* The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* 123, 1–36 (1996).
- Ayadi A, Birling MC, Bottomley J *et al.* Mouse large-scale phenotyping initiatives: overview of the European Mouse Disease Clinic (EUMODIC) and of the Wellcome Trust Sanger Institute Mouse Genetics Project. *Mamm. Genome* 23(9–10), 600–610 (2012).
- Challa AK, Boitet ER, Turner AN *et al.* Novel hypomorphic alleles of the mouse tyrosinase gene induced by CRISPR-Cas9 nucleases cause non-albino pigmentation phenotypes. *PLoS One* 25, 11(5), e0155812 (2016).
- Goto T, Hara H, Nakauchi H, Hoshi S, Hirabayashi M. Hypomorphic phenotype of *Foxn1* gene-modified rats by CRISPR/Cas9 system. *Transgenic Res.* 25, 533–544 (2016).
- Qi LS, Larson MH, Gilbert LA *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152(5), 1173–1183 (2013).
- Cao J, Wu L, Zhang M *et al.* An easy and efficient inducible CRISPR/Cas9 platform with improved specificity for multiple gene targeting. *Nucleic Acid Res.* 44(19), e149 (2016).
- El-Brolosy MA, Stainier DFR. Genetic compensation: a phenomenon in search of mechanisms. *PLoS Genet.* 13(7), e1006780 (2017).
- Zheng T, Hou Y, Zhang P *et al.* Profiling single-guide RNA specificity reveals a mismatch sensitive core sequence. *Sci. Rep.* 7, 40638 (2017).
- Landry JJ, Pyl PT, Rausch T *et al.* The genomic and transcriptomic landscape of a HeLa cell line. *G3 (Bethesda)* 3(8), 1213–1224 (2013).
- Li Y, Mendiratta S, Ehrhardt K *et al.* Exploiting the CRISPR/Cas9 PAM constraint for single-nucleotide resolution interventions. *PLoS One* 11(1), e0144970 (2016).
- Horlbeck MA, Witkowski LB, Guglielmi B *et al.* Nucleosomes impede Cas9 access to DNA *in vivo* and *in vitro*. *Elife* 5, e12677 (2016).
- Daer RM, Cutts JP, Brannan DA, Haynes KA. The impact of chromatin dynamics on Cas9-mediated genome editing in human cells. *ACS Synth Biol.* 6(3), 428–438 (2017).
- Kleistner BP, Pattanayak V, Prew MS *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529(7587), 490–495 (2016).
- Slaymaker IM, Gao L, Zetsche B *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* 351(6268), 84–88 (2016).
- Chen X, Liu J, Janssen JM, Gonçalves MAFV. The chromatin structure differentially impacts high-specificity CRISPR-Cas9 nuclease strategies. *Mol. Ther. Nucleic Acids* 8, 558–563 (2017).
- Uusi-Mäkelä MIE, Barker HR, Bäuerlein CA *et al.* Chromatin accessibility is associated with CRISPR-Cas9 efficiency in the zebrafish (*Danio rerio*). *PLoS One* 13(4), e0196238 (2018).
- Chen Y, Zeng S, Hu R *et al.* Using local chromatin structure to improve CRISPR/Cas9 efficiency in zebrafish. *PLoS One* 12(8), e0182528 (2017).
- Doench JG, Fusi N, Sullender M *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 34(2), 184–191 (2016).
- Kim S, Kim D, Cho SW, Kim J, Kim JS. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* 24(6), 1012–1019 (2014).
- Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343(6166), 80–84 (2014).
- Zuris JA, Thompson DB, Shu Y *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing *in vitro* and *in vivo*. *Nat. Biotechnol.* 33(1), 73–80 (2015).
- D'Astolfo DS, Pagliaro RJ, Pras A *et al.* Efficient intracellular delivery of native proteins. *Cell* 161(3), 674–690 (2015).
- Gagnon JA, Valen E, Thyme SB *et al.* Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS One* 9(5), e98186 (2014).
- Ramakrishna S, Kwaku Dad AB, Beloor J *et al.* Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. *Genome Res.* 24(6), 1020–1027 (2014).
- Shifrut E, Carnevale J, Tobin V *et al.* Genome-wide CRISPR screens in primary human T cells reveal key regulators of immune function. *Cell* 175(7), 1958–1971 (2018).
- Ting PY, Parker AE, Lee JS *et al.* Guide Swap enables genome-scale pooled CRISPR-Cas9 screening in human primary cells. *Nat. Methods* 15(11), 941–946 (2018).
- Hamada H, Liu Y, Nagira Y *et al.* Biolistic-delivery-based transient CRISPR/Cas9 expression enables in planta genome editing in wheat. *Sci. Rep.* 8(1), 14422 (2018).
- Burger A, Lindsay H, Felker A *et al.* Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes. *Development* 143(11), 2025–37 (2016).
- Aslan Y, Tadjuidje E, Zorn AM, Cha SW. High-efficiency non-mosaic CRISPR-mediated knock-in and indel mutation in *F0 Xenopus*. *Development* 144, 2852–2858 (2017).
- Soda N, Verma L, Giri J. CRISPR-Cas9 based plant genome editing: significance, opportunities and recent advances. *Plant Physiol. Biochem.* 131, 2–11 (2018).
- Song J, Yang D, Ruan J *et al.* Production of immunodeficient rabbits by multiplex embryo transfer and multiplex gene targeting. *Sci. Rep.* 7(1), 12202 (2017).
- Eubelen M, Bostaille N, Cabochette P *et al.* A molecular mechanism for Wnt ligand-specific signaling. *Science* 361(6403), eaat1178 (2018).
- Ran FA, Hsu PD, Wright J *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8(11), 2281–308 (2013).
- O'Geen H, Ren C, Nicolet CM *et al.* dCas9-based epigenome editing suggests acquisition of histone methylation is not sufficient for target gene repression. *Nucleic Acids Res.* 45(17), 9901–9916 (2017).
- He Z, Shi X, Liu M *et al.* Comparison of surrogate reporter systems for enrichment of cells with mutations induced by genome editors. *J. Biotechnol.* 221, 49–54 (2016).
- Mircetic J, Steinebrunner I, Ding L *et al.* Purified Cas9 fusion proteins for advanced genome manipulation. *Small Methods* 1(4), 1600052 (2017).
- Kim K, Park SW, Kim JH *et al.* Genome surgery using Cas9 ribonucleoproteins for the treatment of age-related macular degeneration. *Genome Res.* 27(3), 419–426 (2017).
- Platt RJ, Chen S, Zhou Y *et al.* CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* 159(2), 440–455 (2014).
- Wroblewska A, Dhainaut M, Ben-Zvi B *et al.* Protein barcodes enable high-dimensional single-cell CRISPR screens. *Cell* 175(4), 1141–1155.e16 (2018).
- Thyme SB, Akhmetova L, Montague TG, Valen E, Schier AF. Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nat. Commun.* 7, 11750 (2016).
- Anders C, Jinek M. *In vitro* enzymology of Cas9. *Methods Enzymol.* 546, 1–20 (2014).
- Zhang JP, Li XL, Neises A *et al.* Different effects of sgRNA length on CRISPR-mediated gene knockout efficiency. *Sci. Rep.* 6, 28566 (2016).
- Haessler M, Schöning K, Eckert H *et al.* Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* 17(1), 148 (2016).
- Kim D, Bae S, Park J *et al.* Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* 12(3), 237–243 (2015).
- Tsai SQ, Zheng Z, Nguyen NT *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* 33(2), 187–197 (2015).
- Vouillot L, Thélie A, Pollet N. Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3 (Bethesda)* 5(3), 407–415 (2015).
- Sentmanat MF, Peters ST, Florian CP, Connelly JP, Pruett-Miller SM. A survey of validation strategies for CRISPR-Cas9 editing. *Sci. Rep.* 8(1), 888 (2018).
- Smith C, Gore A, Yan W *et al.* Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell* 15(1), 12–13 (2014).
- Zischewski J, Fischer R, Bortesi L. Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnol. Adv.* 35(1), 95–104 (2017).
- Wang X, Wang Y, Wu X *et al.* Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotechnol.* 33(2), 175–178 (2015).
- Cameron P, Fuller CK, Donohoue PD *et al.* Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nat. Methods* 14(6), 600–606 (2017).
- Tsai SQ, Nguyen NT, Malagon-Lopez J *et al.* CIRCLE-seq: a highly sensitive *in vitro* screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat. Methods* 14(6), 607–614 (2017).
- Wu X, Scott DA, Kriz AJ *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* 32(7), 670–676 (2014).
- Fujita T, Yuno M, Fujii H. enChIP systems using different CRISPR orthologues and epitope tags. *BMC Res. Notes* 11(1), 154 (2018).
- Xu X, Qi LS. A CRISPR-dCas toolbox for genetic engineering and synthetic biology. *J. Mol. Biol.* 431(1), 34–47 (2019).
- Negembor MV, Sebastian-Perez R, Alicino F *et al.* (Po)STAC (Polycistronic SunTag modified CRISPR) enables live-cell and fixed-cell super-resolution imaging of multiple genes. *Nucleic Acids Res.* 46(5), e30 (2018).
- Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533(7603), 420–424 (2016).
- Chen JS, Dagdas YS, Kleinstiver BP *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* 550(7676), 407–410 (2017).
- Cho SW, Kim S, Kim Y *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* 24(1), 132–141 (2014).
- Cebrian-Serrano A, Davies B. CRISPR-Cas orthologues and variants: optimizing the repertoire, specificity and delivery of genome engineering tools. *Mamm. Genome* 28(7–8), 247–261 (2017).
- Adli M. The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* 9(1), 1911 (2018).
- Wagner DL, Amini L, Wender DJ *et al.* High prevalence of *Streptococcus pyogenes* Cas9-reactive T cells within the adult human population. *Nat. Med.* doi: 10.1038/s41591-018-0204-6 (2018).
- Crudele JM, Chamberlain JS. Cas9 immunity creates challenges for CRISPR gene editing therapies. *Nat. Commun.* 9(1), 3497 (2018).
- Zetsche B, Heidenreich M, Mohanraju P *et al.* Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat. Biotechnol.* 35(1), 31–34 (2017).
- Kim D, Kim J, Hur JK *et al.* Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* 34(8), 863–868 (2016).
- Konermann S, Lotfy P, Briedau NJ *et al.* Transcriptome engineering with RNA-targeting Type VI-D CRISPR effectors. *Cell* 173(3), 665–676.e14 (2018).