



Research review paper

# Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases

Julia Zischewski<sup>a</sup>, Rainer Fischer<sup>a,b</sup>, Luisa Bortesi<sup>a,\*</sup><sup>a</sup> Institute for Molecular Biotechnology, RWTH Aachen University, Worringerweg 1, 52074 Aachen, Germany<sup>b</sup> Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Forckenbeckstraße 6, 52074 Aachen, Germany

## ARTICLE INFO

## Article history:

Received 23 August 2016

Received in revised form 18 November 2016

Accepted 19 December 2016

Available online 21 December 2016

## Keywords:

CRISPR

Cas9

Genome editing

Engineered nucleases

Indel detection

Unbiased off-target analysis

Mismatch detection assays

## ABSTRACT

The development of customizable sequence-specific nucleases such as TALENs, ZFNs and the powerful CRISPR/Cas9 system has revolutionized the field of genome editing. The CRISPR/Cas9 system is particularly versatile and has been applied in numerous species representing all branches of life. Regardless of the target organism, all researchers using sequence-specific nucleases face similar challenges: confirmation of the desired on-target mutation and the detection of off-target events. Here, we evaluate the most widely-used methods for the detection of on-target and off-target mutations in terms of workflow, sensitivity, strengths and weaknesses.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Contents

1.	Introduction . . . . .	96
2.	Detection of on-target mutations . . . . .	96
2.1.	Mismatch detection assays . . . . .	97
2.1.1.	The mismatch cleavage assay . . . . .	97
2.1.2.	High-resolution melting analysis . . . . .	97
2.1.3.	Heteroduplex mobility assay . . . . .	98
2.2.	Analysis of cleaved amplified polymorphic sequences . . . . .	98
2.3.	Loss of a primer binding site . . . . .	98
2.4.	Sequencing . . . . .	98
2.5.	Amplified fragment length polymorphisms . . . . .	99
2.6.	Fluorescent PCR capillary gel electrophoresis . . . . .	99
3.	Prevention, detection and quantification of off-target mutations caused by SSNs . . . . .	99
3.1.	Upstream measures to reduce the likelihood of off-target mutations . . . . .	99
3.1.1.	Engineered nucleases with reduced off-target activity . . . . .	100
3.2.	Off-target detection using biased and unbiased screening methods . . . . .	100
3.2.1.	Amplification and sequencing of pre-selected off-target sites . . . . .	100
3.2.2.	Whole exome sequencing . . . . .	100
3.2.3.	Whole genome sequencing . . . . .	100
3.2.4.	BLESS . . . . .	100
3.2.5.	GUIDE-seq . . . . .	101
3.2.6.	LAM-HTGTS . . . . .	102
3.2.7.	Digenome-seq . . . . .	102
4.	Conclusions . . . . .	102

\* Corresponding author at: Institute for Molecular Biotechnology, RWTH Aachen University, Worringerweg 1, 52074 Aachen, Germany.

E-mail address: [luisa.bortesi@molbiotech.rwth-aachen.de](mailto:luisa.bortesi@molbiotech.rwth-aachen.de) (L. Bortesi).

Acknowledgements . . . . .	102
References . . . . .	102

## 1. Introduction

The use of sequence-specific nucleases (SSNs) for genome editing has become routine in many laboratories. Genome editing tools such as zinc finger nucleases (ZFNs) (Kim, Cha, & Chandrasegaran, 1996), transcription activator-like effector nucleases (TALENs) (Christian et al., 2010) and especially the more recent clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated protein 9 (Cas9) system (Jinek et al., 2012), have provided researchers with the ability to create double-strand breaks (DSBs) at any desired position in the genome. In higher eukaryotes, DSBs are usually resolved by the endogenous DNA repair mechanism of non-homologous end-joining (NHEJ) which is intrinsically error-prone, typically resulting in small insertions and/or deletions (indels) at the site of the break. If the indels cause a frameshift mutation, they can knock out the function of the gene due to the production of truncated polypeptides and/or nonsense-mediated mRNA decay (Perez et al., 2008; Ramlee, Yan, Cheung, Chuah, & Li, 2015; Santiago et al., 2008; Sung et al., 2013).

The target sequence of the CRISPR/Cas9 system can be changed simply by altering the 20-nt sequence of the single guide RNA (gRNA), so the generation and testing of multiple targeting constructs has become straightforward. However, once the components of the system have been introduced into the host organism, the next major challenge is to confirm and characterize the resulting mutations. In the relatively simple case of targeting a single diploid cell, there are four potential outcomes: no mutation, a heterozygous mutation (only one allele is mutated), a biallelic mutation (both alleles are mutated but the sequence of each allele is distinct) or a homozygous mutation (both alleles carry the same mutation). The latter can also occur if one allele is used as a template to repair the break in the other allele. More complex outcomes are possible in polyploid host species, when the mutated organism is a chimera, or when pools of samples are screened. Off-target mutations can further complicate the analysis, but specific methods

have been developed to identify such events as discussed later in this article. All methods for the analysis of on-target and off-target mutations have pros and cons and the ideal method in any situation depends on a number of factors, including the type of sample, the anticipated size and frequency of the mutations, and the cost of the method.

## 2. Detection of on-target mutations

The most widely used methods for the detection of targeted mutations are summarized in Table 1. These are all based on the polymerase chain reaction (PCR) and therefore tend to underestimate the frequency of on-target activity because large deletions that extend beyond the boundaries of the PCR amplicon are not detected, and large insertions are amplified less efficiently than small mutations (if at all) and are therefore less likely to be identified. This tends not to be a critical issue when a single gRNA is used because small indels are much more common than large deletions or insertions, but larger indels arise at higher frequency when two gRNAs are designed to target sites on the same chromosome. In the case of mutants present at a very low frequency in an otherwise wild-type background (such as chimeras or pooled clones), the PCR step is often biased towards the more abundant template, and the small number of mutated sequences may not be detected. One way to reduce this problem is to pre-digest the genomic DNA with a restriction enzyme recognizing the wild-type sequence, thus eliminating the wild-type template before the amplification step, although this depends on the availability of restriction sites overlapping the nuclease target sequence. An alternative is “co-amplification at lower denaturation temperature” (ice-COLD-PCR), which improves the detection of rare mutant sequences in chimeric clones because it does not favor the amplification of the proportionally dominant wild-type sequence (Milbury, Li, & Makrigiorgos, 2011). Of course, if the mutated sequences are intentionally enriched, the results cannot be considered quantitative. Regardless of the detection method, it is always

**Table 1**  
Overview of methods for the detection of on-target mutations induced by SSNs.

Methods	Type of mutations preferentially detected	Reported sensitivity	Determination of mutation type?	Cost <sup>a</sup>	Throughput	Limitations	References
Mismatch cleavage assay	Small indels	0.5–3%	No	\$	Moderate	T7E1 can overlook single nucleotide changes; Surveyor less sensitive than T7E1	Kim et al., 2013; Qiu et al., 2004; Ran et al., 2013; Vouillot, Th��ie, & Pollet, 2015; Zhu et al., 2014
HRMA	Small indels	2%	If insertion or deletion	\$ (+ equipment)	High	Misses large indels	Dahlem et al., 2012; Wang et al., 2015
Heteroduplex mobility assay by PAGE	Small indels	0.5%	No	\$	Moderate	Misses large indels	Zhu et al., 2014
CAPS	All		No	\$	Moderate	Availability of restriction site	Ran et al., 2013
Loss of primer binding site	Indels	10%	Yes	\$	High	Misses substitutions	Yu, Zhang, Yao, & Wei, 2014
Sanger sequencing	All	1–2%	Yes	\$\$/\$\$\$ <sup>b</sup>	Low	Costly, labor intensive	Brinkman, Chen, Amendola, & van Steensel, 2014; Liu et al., 2015
NGS	All	0.01%	If insertion or deletion	\$\$\$\$	High	Misses large indels	G��ell, Yang, & Church, 2014
AFLP	Large indels, also Mb		If insertion or deletion	\$	Moderate	Misses small indels	Bauer, Canver, & Orkin, 2015
Fluorescent PCR-capillary gel electrophoresis	Small indels	1%	Number of bp	\$\$	High	Misses substitutions	Ramlee et al., 2015; Yang et al., 2015

<sup>a</sup> Estimated cost per assay. \$: <1 US\$; \$\$: <5 US\$, \$\$\$: >100 US\$, \$\$\$\$ : >500 US\$.

<sup>b</sup> Sequencing of bulk/cloned PCR products.

necessary to sequence the target region to confirm and ultimately determine the outcome of genome editing.

2.1. Mismatch detection assays

Assays that detect mismatches typically consist of three steps: (1) amplification of the target site and its flanking region by PCR; (2) denaturing and reannealing the DNA to allow the mutant and wild-type strands to form heteroduplex DNA; and (3) detection of the heteroduplex using a method that is selective for the difference in structure or melting temperature (Fig. 1).

The general advantage of mismatch detection assays is that they are simple, rapid and cost-effective. They can be used to genotype single clones or analyze pooled samples and populations. Although they detect mutations, they do not reveal any details of the mutation structure. Furthermore, if the targeted locus is highly polymorphic then the results can be difficult to interpret because different wild-type alleles can also form heteroduplex DNA (Kim, Kim, Kim, & Kim, 2014a). Mismatch detection is often used in a semi-quantitative manner, e.g. to compare the efficiency of several gRNAs, to evaluate experimental conditions that affect genome editing, or as a preliminary screening approach to identify lines for further analysis using more accurate sequencing-based methods.

2.1.1. The mismatch cleavage assay

The mismatch cleavage assay is a simple and cost-effective method for the detection of indels and is therefore the most widely used procedure to detect mutations induced by genome editing. The assay uses enzymes that cleave heteroduplex DNA at mismatches and extrahelical loops formed by multiple nucleotides, yielding two or more smaller fragments. A PCR product of ~300–1000 bp is generated with the predicted nuclease cleavage site off-center so that the resulting fragments are dissimilar in size and can easily be resolved by conventional gel electrophoresis or high-performance liquid chromatography (HPLC). End-labeled digestion products can also be analyzed by automated gel or capillary electrophoresis (Qiu et al., 2004). The frequency of indels at the locus can be estimated by measuring the integrated intensities of the PCR amplicon and cleaved DNA bands (Ran et al., 2013). The digestion step takes 15–60 min, and when the DNA preparation and PCR steps are added the entire assays can be completed in <3 h.

Two alternative enzymes are recommended for this assay. T7 endonuclease 1 (T7E1) is a resolvase that recognizes and cleaves imperfectly matched DNA at the first, second or third phosphodiester bond upstream of the mismatch. The sensitivity of a T7E1-based assay is 0.5–5% (Kim et al., 2013; Zhu et al., 2014). In contrast, Surveyor™ nuclease (Transgenomic Inc., Omaha, NE, USA) is a member of the CEL family of

mismatch-specific nucleases derived from celery. It recognizes and cleaves mismatches due to the presence of single nucleotide polymorphisms (SNPs) or small indels, cleaving both DNA strands downstream of the mismatch. It can detect indels of up to 12 nt and is sensitive to mutations present at frequencies as low as ~3%, i.e. 1 in 32 copies (Qiu et al., 2004).

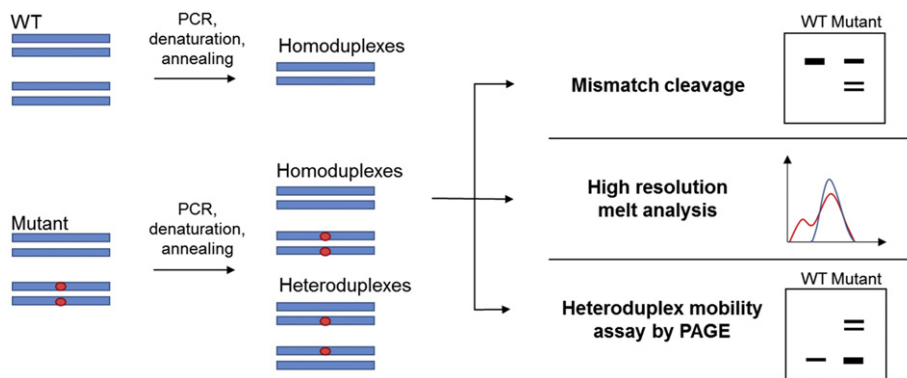
T7E1 outperforms Surveyor nuclease in terms of sensitivity when the substrates carry indels, but completely ignores SNPs and also tends to miss small indels. Surveyor nuclease is less sensitive but is better suited for the detection of SNPs and small indels (Vouillot et al., 2015). Therefore the choice depends on which types of mutations are anticipated or need to be detected. The Surveyor™ kit is more expensive, but it is also more robust and comes with a standardized protocol, whereas T7E1 is sensitive to the reaction conditions (e.g. incubation time, temperature, DNA/enzyme ratio, salt concentration) and the assay may therefore require optimization. The mismatch cleavage assay usually underestimates the mutation frequency due to the preferential cleavage properties of each enzyme.

There are two further caveats associated with the mismatch cleavage assay: (1) homozygous mutations can only be detected by adding wild-type DNA to the PCR step in order to allow the formation of heteroduplexes; and (2) if the mutation frequency is high enough the mutant sequences form homoduplexes that cannot be detected, so the number of mutations will be under-reported (Kim et al., 2011).

2.1.2. High-resolution melting analysis

High-resolution melting analysis (HRMA) involves the amplification of a DNA sequence spanning the genomic target (90–200 bp) by real-time PCR with the incorporation of a fluorescent dye, followed by melt curve analysis of the amplicons (Dahlem et al., 2012; Wang et al., 2015). HRMA is based on the loss of fluorescence when intercalating dyes are released from double-stranded DNA during thermal denaturation. It records the temperature-dependent denaturation profile of amplicons and detects whether the melting process involves one or more molecular species.

Unlike the melt curves analyzed in typical quantitative PCR (qPCR) experiments, the data are typically collected over narrower temperature increments of 0.2 °C, followed by signal normalization and analysis. Melting temperature shifts and the shape of the melting curves can both provide useful information: homozygous allelic variants may cause a temperature shift in the melt curve compared to the wild-type homoduplex, whereas heteroduplexes representing heterozygous mutations change the shape of the melt curve due to the presence of mismatches (Taylor et al., 2010). Unlike the mismatch cleavage assay, HRMA can therefore distinguish among different mutant alleles and can also distinguish homozygous wild-type and homozygous mutant



**Fig. 1.** Mismatch-based assays for the detection of on-target mutations. The first steps of the workflow are identical for the three mismatch-based assays and rely on amplification of the target sequence, denaturation of the dsDNA and reannealing to form homoduplexes and heteroduplexes. Mismatch cleavage with T7 or Surveyor endonuclease relies on the recognition and digestion of DNA bulges in heteroduplexes, and the digested fragments can be detected in an agarose gel. For HRMA, the different melting temperatures of homoduplexes and heteroduplexes are exploited and measured using a fluorescence-based readout. Homoduplexes and heteroduplexes differ in their mobility and can be separated by PAGE. Figure modified from (Zhu et al., 2014) (Creative Commons License).

sequences due to the shift in melting temperatures caused by the different nucleotide composition (Thomas, Percival, Yoder, & Parant, 2014).

HRMA can detect mutations at any target site. However, when selecting a gRNA it is important to consider the suitability of the primers flanking the target site, which should be tested in advance and the PCR conditions optimized to ensure that the predicted amplicon gives a smooth melt curve with only one melt peak (Talbot & Amacher, 2014). The shorter the amplicon, the greater the difference in melting temperature caused by a mutation and therefore the greater the resolution. But if the amplicon is too small it becomes impossible to detect larger indels. The ideal amplicon size to maximize the resolution of HRM is ~100 bp, although a 50-bp amplicon is best for discriminating among sequences that differ at only one nucleotide position (Thomas et al., 2014).

HRMA is a simple and highly sensitive method that is also compatible with a high-throughput screening format (96-well microtiter plates) so direct sample handling following the PCR step is unnecessary. The entire procedure from the preparation of genomic DNA to the identification of mutations takes <2 h. Because HRMA is nondestructive, the amplicons can be analyzed further by other methods such as gel electrophoresis and sequencing. Its sensitivity depends on the amplicon size and the type of mutation. For indels larger than 4 bp, the estimated detection limit in a ~100-bp amplicon is at least 2%, i.e. one mutant among 50 wild-type genomes (Dahlem et al., 2012).

One limitation of HRMA is that the target fragments are relatively short so larger indels cannot be detected. The setup costs are also high, although this can be mitigated by pairing an existing qPCR machine with online HRMA software (e.g. <https://dna.utah.edu/uv/uanalyze.html>). Once the equipment is in place the cost per analysis is low (Talbot & Amacher, 2014).

### 2.1.3. Heteroduplex mobility assay

Mutations can also be detected by analyzing re-hybridized PCR fragments directly by native polyacrylamide gel electrophoresis (PAGE). This method takes advantage of the differential migration of heteroduplex and homoduplex DNA in polyacrylamide gels. The angle between matched and mismatched DNA strands caused by an indel means that heteroduplex DNA migrates at a significantly slower rate than homoduplex DNA under native conditions, and they can easily be distinguished based on their mobility. Fragments of 140–170 bp can be separated in a 15% polyacrylamide gel. The sensitivity of such assays can approach 0.5% under optimal conditions, which is similar to T7E1 (Zhu et al., 2014). After reannealing the PCR products, the electrophoresis component of the assay takes ~2 h.

The advantage of this one-step method is that it does not involve time-consuming enzyme reactions and eliminates the false negative results caused by the incomplete digestion of mismatched DNA fragments. However, only small amplicons can be resolved so the assay can only detect SNPs and small indels, and the sensitivity of PAGE across the spectrum of indels is unclear (Shui, Hernandez Matias, Guo, & Peng, 2016). A variation of the DNA mobility assay uses slower-migrating single-stranded DNA to detect mobility differences between wild-type and mutated DNA strands differing by as little as 1 nt (Zheng et al., 2016).

### 2.2. Analysis of cleaved amplified polymorphic sequences

The position of mutations induced by SSNs is generally predictable because ZFNs and TALENs induce genomic DSBs in the spacer region between their DNA recognition sites, and Cas9 induces DSBs 3 bp upstream of the protospacer adjacent motif (PAM). If it is possible to design an experiment in such a way that the nuclease cuts within a restriction enzyme recognition site or <5 bp away from it, the combination of PCR and restriction enzymes is a straightforward and cost-effective method for the detection of indels, which appear as cleaved amplified polymorphic sequences (CAPS). This approach involves the amplification of a target site and ~300–1000 bp of flanking material so

that the cleavage site is offset from the center of the amplicon. Digestion with the appropriate restriction enzyme is then followed by analysis of the fragment sizes by gel electrophoresis. The entire process from genomic DNA preparation to mutant detection takes a few hours. Like the mismatch cleavage assay, digestion products can be analyzed by conventional 2% agarose gel electrophoresis or HPLC, or end-labeled digestion products can be analyzed by automated gel or capillary electrophoresis (Qiu et al., 2004). By measuring the integrated intensities of the PCR amplicon and cleaved DNA bands, the frequency of indels can be estimated (Ran et al., 2013).

CAPS analysis is one of the most widely used mutation detection methods together with the T7E1 and Surveyor assay. In contrast to other widely-used mismatch-detection assays, CAPS analysis can detect homozygous mutants and, provided that the nuclease target sequence near the nuclease target sites. It can detect all kinds of mutations (SNPs, and both small and large indels) as long as they disrupt the restriction site, and is therefore highly sensitive and convenient. Nevertheless, CAPS analysis is limited by the availability of restriction sites covering the mismatch. Recently, an interesting application of CRISPR/Cas9 was proposed in which the system is used *in vitro* just like a conventional restriction enzyme to overcome this limitation because it can target any genomic sequence as long as a PAM is present (Kim et al., 2014a).

### 2.3. Loss of a primer binding site

When genomic DNA is amplified with two pairs of primers, one spanning the target region but annealing outside it, and another that includes a primer overlapping the putative indel site, mutations at the target site will prevent the latter primer annealing and only the larger amplicon spanning the entire target site will be produced (Yu et al., 2014). If a qPCR approach is used, the mutation frequency can be estimated, and the larger amplicon spanning the entire target site can be sequenced to characterize the mutations. The PCR products are resolved by electrophoresis so the method is rapid and inexpensive, and – as long as the sequence pairing with the 5' end of the primer is conserved – natural polymorphisms in the genome should not interfere with the results. The main limitation is that point mutations can be overlooked because primers can anneal to mismatched templates, and extension remains possible as long as the terminal nucleotide is paired correctly. This method also has a sensitivity of only ~10% so it is not suitable if the SSN has a low targeting efficiency (Yu et al., 2014).

### 2.4. Sequencing

Mutations induced by SSNs can be characterized in detail by sequencing amplicons that span the entire target site. Suitable approaches include the Sanger sequencing of individual cloned fragments or the bulk amplicon mixture, and next generation sequencing (NGS). The great advantage of sequencing-based detection methods is the direct and detailed information about the nature and diversity of mutations. The gold standard for the identification of induced mutations at on-target sites is the cloning of amplicons from independent targeting events at each site followed by Sanger sequencing of the cloned PCR products (50–100 events, depending on the efficiency of the SSN). This reveals both the frequency and type of mutations at the target locus, but it is laborious, time-consuming and expensive when many samples are processed. An alternative is to sequence the PCR products directly. Unless the mutation is homozygous, direct Sanger sequencing generates multiple traces with overlapping peaks. In the case of diploid organisms with heterozygous or biallelic mutations, two overlapping traces are obtained starting at the mutation site. In polyploid organisms or when sequencing pooled clones, even more traces can be found in a single chromatogram. The automatic decoding of superimposed chromatograms derived from PCR amplicons containing various types of

mutations can then be achieved using *ad hoc* bioinformatics tools such as DSDecode (<http://dsdecode.scgene.com/>) and TIDE (<http://tide.nki.nl>). DSDecode can genotype various types of biallelic and heterozygous mutations in diploid organisms (Liu et al., 2015) whereas the more versatile TIDE can also identify indels by decomposition of the quantitative sequence trace data originating from pooled samples, accurately quantifying the editing efficiency and simultaneously determining the predominant type of indels in the sample (Brinkman et al., 2014). If necessary, the decoded sequences can then be verified by cloning the PCR products and sequencing a few clones representing each sample. The reliability of TIDE depends on the purity of the PCR products and the quality of the sequence reads. Highly repetitive sequences around the target site can hamper the decomposition process. TIDE can detect indels with a sensitivity of ~1–2% across various target regions in a pool of cells (Brinkman et al., 2014). This entire workflow takes up to 2 days from genomic DNA extraction to the sequencing of the target locus, but the hands-on time is limited to preparing the PCR step.

Bulk PCR products can also be sequenced using a NGS approach followed by software analysis, e.g. CRISPR-GA (Güell et al., 2014). This method is highly informative and powerful because it can detect mutation frequencies reliably with a sensitivity of 0.01% (Hendel, Fine, Bao, & Porteus, 2015). It is especially useful when large numbers of samples are multiplexed and, although quite expensive, it is commonly used to assess indel formation. NGS produces relatively short reads (~300–700 bp, depending on the platform) and therefore cannot detect larger indels. This issue has been addressed with a new method known as single molecule real-time (SMRT) DNA sequencing, which provides average read lengths of 8.5 kb (Hendel et al., 2014). Although the sensitivity of SMRT sequencing is lower than that of other sequencing platforms, this is currently the only NGS method that can identify large indels.

### 2.5. Amplified fragment length polymorphisms

When large deletions are anticipated, e.g. when targeting a gene with multiple gRNAs separated by a few hundred base pairs, a simple PCR followed by product resolution on a standard agarose gel to detect the different sizes of amplicons provides a rapid and cost-effective solution. Such differences are known as amplified fragment length polymorphisms (AFLPs). If nested primer pairs are used, this method can also detect very large chromosome deletions of several million base pairs (Bauer et al., 2015). In the latter case, the presence/absence of an amplification product rather than differences in product sizes would confirm whether or not a deletion had occurred in the target genome. However, the resolution of conventional AFLPs with agarose gel electrophoresis is limited, and more sophisticated techniques are necessary to detect size differences with the resolution of few base pairs.

### 2.6. Fluorescent PCR capillary gel electrophoresis

Indels caused by SSNs can also be characterized by fluorescent PCR coupled with capillary gel electrophoresis. The genomic region containing the anticipated indel site is amplified by PCR using fluorophore-labeled primers. The resulting labeled amplicons (<500 bp) are then resolved by capillary gel electrophoresis and any mutations are revealed by the differences in mobility compared to the wild-type amplicon (Ramlee et al., 2015; Yang et al., 2015). This technique has single-nucleotide resolution and can thus identify frameshift mutations, and has the sensitivity to detect mutations with a frequency of ~1% (Yang et al., 2015). It also facilitates the multiplexed genotyping of different events in a single sample, which makes it suitable for high-throughput screening. Sanger sequencing is currently the most widely used genotyping method, but fluorescent capillary electrophoresis is less expensive and more readily scalable (Ramlee et al., 2015). However, it does not reveal which nucleotides have been inserted or deleted and it is not able to detect SNPs, which can generate missense or nonsense mutations. The

limited size of the amplicons that can be analyzed also prevents the detection of larger indels. Therefore, conventional PCR followed by agarose gel electrophoresis should be combined with fluorescent PCR capillary gel electrophoresis to analyze samples in which both large indels (e.g. deletions between multiple gRNAs) and small indels (mutations at individual gRNA targets) are expected. Furthermore, this technique does not report large indels accurately, because it tends to overestimate mutations longer than 30 base pairs (Ramlee et al., 2015). The equipment and analytical software required are rather expensive, but analysis can be outsourced to companies that provide standard sequencing services.

## 3. Prevention, detection and quantification of off-target mutations caused by SSNs

In some cases it is absolutely necessary that no off-target mutations are induced in the genome, but the detection of off-target mutations is more challenging than the detection of on-target mutations because the number and position of off-target mutations is unknown. Even so, the likelihood of an off-target mutation at a given site can be predicted to some extent, and as for many other scientific endeavors, the saying “An ounce of prevention is worth a pound of cure” is a good maxim to follow when designing experiments with SSNs. The CRISPR/Cas9 system is more prone to off-target effects than ZFNs and TALENs because Cas9 works as a monomer in contrast to the dimeric ZFN and TALEN assemblies, and thus recognizes a shorter target sequence. Furthermore, the gRNA can tolerate a certain number of mismatches. The first reports of significant Cas9-induced off-target effects were reported in human cancer cell lines (Cradick, Fine, Antico, & Bao, 2013; Fu et al., 2013a; Fu, Reyon, & Keith, 2013b), although the frequency was unusually high because DNA repair pathways do not function correctly in tumor cells. Efforts to increase the accuracy of targeting stepped up when the crystal structure of Cas9 was solved, revealing that the apoenzyme is inactive and only gains endonuclease activity when the gRNA binds, resulting in a conformational change (Anders, Niewoehner, Duerst, & Jinek, 2014; Jinek et al., 2014; Nishimasu et al., 2014). This showed that the gRNA determines the likelihood of off-target activity and three major off-target categories were identified. Generally, off-target sites are similar in sequence to the desired target sites but they may feature: (i) up to seven mismatches (Tsai et al., 2015); (ii) small indels that cause DNA or RNA bulges (Lin et al., 2014); or (iii) a different PAM, e.g. NAG often acts as a PAM in addition to NGG, although the interaction with Cas9 is weaker (Hsu et al., 2013; Jiang, Bikard, Cox, Zhang, & Marraffini, 2013). The potential for off-target effects should be kept in mind when designing the gRNA, but their overall impact should be considered in a wider context because the frequency of off-target mutations is much lower than on-target mutations when the DNA repair machinery is intact (Hruscha et al., 2013; Veres et al., 2014; Yang et al., 2013). The extent of off-target activity is highly dependent on the gRNA, and the number of off-targets varies from 0 to >150 (Frock et al., 2015; Kim, Kim, Kim, Park, & Kim, 2016; Tsai et al., 2015). These findings demonstrate that the CRISPR/Cas9 system can be highly specific, but robust methods are needed to experimentally evaluate candidate gRNAs. The purpose of this article is to provide a comprehensive overview of the available detection/characterization techniques for both on-target and off-target mutations, without going into details about the mechanisms underlying off-target mutations. We refer readers interested in learning how to predict and mitigate off-target effects to some excellent reviews covering that topic (Tsai & Joung, 2016; Tycko, Myer, & Hsu, 2016; Yee, 2016). The ever-growing body of knowledge should eventually make it possible to design accurate gRNAs lacking off-target activity, but until then the methods presented in this section will remain relevant.

### 3.1. Upstream measures to reduce the likelihood of off-target mutations

For the CRISPR/Cas9 system, the careful selection of gRNAs and an appropriate nuclease variant can substantially reduce the risk of off-

target mutations. Numerous bioinformatics tools have been developed to identify optimal gRNAs, and several groups have published strategies that help to reduce the off-target activity of the CRISPR/Cas9 system.

One of the simplest measures is the use of a truncated gRNA, which is 17–18 rather than 20 nt in length (Fu et al., 2013b). Although shortening the specificity-determining region of the gRNA to improve specificity may sound counterintuitive, this approach makes sense when the binding energy between RNA and DNA is considered. The truncated gRNAs work because the binding energy is reduced to an extent that is just sufficient to bind a perfect target, but not targets containing mismatches (Fu et al., 2013b).

Whatever the length of the selected gRNA, it is always advisable to use one of the many available bioinformatics tools to identify potential off-target sites in the genome. These tools exploit the greatest advantage of the CRISPR/Cas9 system, *i.e.* that target recognition is dependent on the Watson-Crick pairing between the gRNA and target DNA. CRISPR/Cas9 off-target sites are therefore much easier to predict compared to ZFNs and TALENs, where specificity is based on protein-DNA binding. Many different online and offline tools are now available and it is beyond the scope of this article to discuss them, so the reader is referred to Cameron MacPherson's CRISPR Software Matchmaker post on Addgene for further information (<http://goo.gl/8Yse8H>).

The method used to deliver the *cas9* and *gRNA* genes can also affect the frequency of off-target mutations: lower off-target activity has been observed following the transient expression of Cas9 and gRNA (by mRNA delivery) or when using Cas9 protein and gRNA produced *in vitro* in comparison to Cas9 expression from a plasmid (Kim, Kim, Cho, Kim, & Kim, 2014b; Liang et al., 2015). If cells are transfected with Cas9 protein and gRNA, the amount of each component can be titrated to find the optimal conditions for efficient on-target activity and low off-target activity (Hsu et al., 2013). The ribonucleoprotein consisting of Cas9 and the gRNA is subject to the same endogenous degradation processes as host cell molecules, and thus has a limited window of opportunity to introduce DSBs in the genome (Kim et al., 2014b).

### 3.1.1. Engineered nucleases with reduced off-target activity

Inspired by the dimeric nature of TALENs and ZFNs, two different strategies have been used to reduce the off-target activity of Cas9 > 50-fold by ensuring that the enzyme introduces two cuts instead of one. In one approach, two closely-spaced gRNAs are combined with a Cas9 nickase mutant to introduce staggered double-strand breaks by nicking the DNA in two positions (Ran et al., 2013). In the other approach, two groups have independently developed fusions comprising the endonuclease domain of *FokI* and a catalytically dead Cas9 that remains able to recognize the PAM (Guilinger, Thompson, & Liu, 2014; Tsai et al., 2014). The endonuclease domain of *FokI* (also used in the ZFN and TALEN systems) acts as a dimer and is not sequence specific. Therefore, two Cas9-*FokI* monomers must bind to neighboring target sites simultaneously to induce the DSB. More recently, engineered versions of Cas9 have been developed with different PAM specificities (Kleinstiver et al., 2015), as well an enhanced specificity Cas9 (eSpCas9) (Slaymaker et al., 2016) and the high-fidelity variant SpCas9-HF1 (Kleinstiver et al., 2016). Both eSpCas9 and SpCas9-HF1 were generated by mutating different sets of amino acid residues involved in non-specific DNA contacts, thus reducing the binding energy and making the Cas9/gRNA complex less tolerant of mismatches. It is possible that a combination of these mutations may further improve the specificity of Cas9.

### 3.2. Off-target detection using biased and unbiased screening methods

Even if careful measures are taken to reduce the probability of off-target activity, it is still necessary to screen the genome for unintended changes after the confirmation of on-target mutations. All the methods listed above for the detection of on-target mutations can also be used for the analysis of off-target sites provided the sites are known. However,

other methods have been developed specifically to detect off-target mutations, most of which rely on some form of sequencing to detect mutations either in pre-selected regions or on a genome-wide scale (Table 2).

These methods can be described as biased or unbiased, the former referring to methods that confirm off-target mutations at predicted sites and the latter referring to methods that identify off-target mutations anywhere in the genome.

#### 3.2.1. Amplification and sequencing of pre-selected off-target sites

The easiest detection method is the amplification of pre-selected potential off-target sites, followed by the sequencing of the PCR products using Sanger or NGS procedures. Many of the CRISPR/Cas9 design tools include information about potential off-target sites in the genome of interest, but it is important to keep in mind that not every algorithm searches for every kind of off-target (*e.g.* DNA or RNA bulges). Importantly, the predictions are not always correct because the CRISPR/Cas9 system is not completely understood, so some predicted off-target sites may be ignored by the enzyme while DSBs may be introduced elsewhere (Anderson et al., 2015; Sander & Joung, 2014). However, this simple method is available to most molecular biology laboratories, whereas the more sophisticated unbiased detection methods require special equipment and knowhow. The choice of sequencing method depends on the number of off-target sites and the nature of the sample (genomic DNA from a cell pool or individual clones). The larger the number of sites and samples, the more likely that Sanger sequencing will become too expensive and difficult to manage, and NGS then becomes more attractive because of the ability to process many amplicons in parallel. The greatest drawback of screening pre-selected sites is its biased nature and thus the inherent risk of overlooking mutations at other loci.

#### 3.2.2. Whole exome sequencing

Whole exome sequencing was originally developed as a compromise between targeted sequencing and the expense of whole genome sequencing to study variants in human genes. The targeted sequencing of all protein-coding regions in the genome allows the identification of relevant variants in the exome but costs only a fraction of whole genome sequencing (Ng et al., 2009). The presence of on-target and off-target mutations in the exome can be detected using this method, as demonstrated for a set of modified human K326 cell lines (although no off-target mutations were observed) (Cho et al., 2014). Depending on the organism, only a small percentage of the genome needs to be covered in this approach, but mutations in regulatory or non-coding regions such as introns are not detected. Exome sequencing is thus limited by its high false-negative rate and many off-target mutations may be overlooked (Cho et al., 2014; Karakoc et al., 2012).

#### 3.2.3. Whole genome sequencing

The unbiased detection of off-target mutations requires whole-genome sequencing but this is expensive and can only be applied to a relatively small number of clones. This approach has been used to screen for off-target mutations induced by TALENs or CRISPR/Cas9 in a range of organisms, including human inducible pluripotent stem cells (Smith et al., 2014; Veres et al., 2014; Yang et al., 2014), mice (Iyer et al., 2015), nematodes (Paix et al., 2014), the malaria parasite *Plasmodium falciparum* (Ghorbal et al., 2014) and several species of plants (Feng et al., 2014; Zhang et al., 2014). By sequencing the whole genome, it is possible to identify not only small indels and SNPs but also structural variants such as inversions, rearrangements, duplications and major deletions (Veres et al., 2014). The restriction of whole genome sequencing to a small number of clones means that most low-frequency off-target events are missed (Wu, Kriz, & Sharp, 2014).

#### 3.2.4. BLESS

Off-target mutations occur when a nuclease-induced DSB is repaired by error-prone NHEJ. The most direct way to detect and quantify the off-

**Table 2**  
Overview of methods for the detection of off-target mutations caused by SSNs.

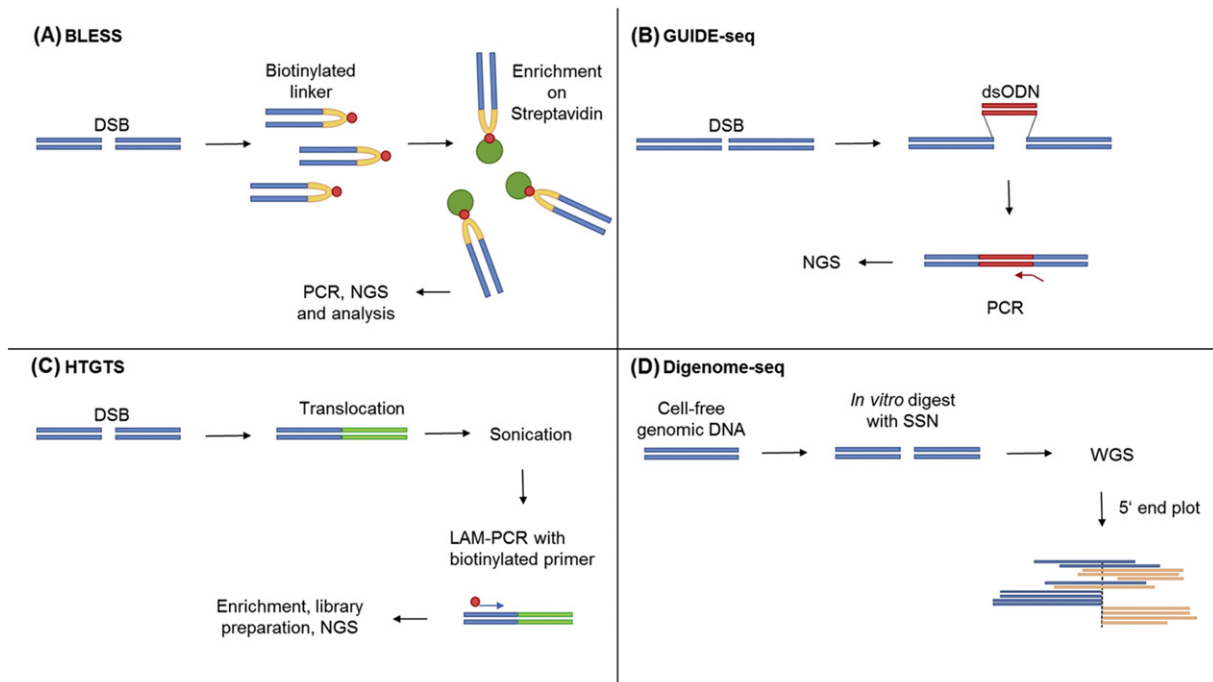
Method	Description	Advantages	Disadvantages
Targeted sequencing	Amplification and Sanger or next-generation sequencing of pre-selected off-target sites	Easy, fast and widely available	Biased, no detection of unexpected mutation sites, gets more expensive and time-consuming when many sites are screened
Exome sequencing	Exome capture and targeted sequencing of all protein-coding regions in a genome	Unbiased detection of mutations in coding regions, less expensive than whole-genome sequencing	No detection of mutations in non-coding regions, reference genome required
Whole genome sequencing	Next-generation sequencing of library prepared from genomic DNA	Unbiased, comprehensive analysis, detects SNPs, indels and structural variants	Expensive, reference genome required
BLESS	Direct <i>in situ</i> labeling of breaks in fixed cells, and next-generation sequencing of enriched and amplified fragments	Unbiased, genome-wide direct labeling of DSBs	Only detection of DSBs present at the time of labeling, reference genome required
GUIDE-seq	Integration of dsODNs into DSBs by NHEJ, amplification and next-generation sequencing	Unbiased, sensitive (0.03% reported)	The dsODNs integrate only in ~30–50% of DSBs, only detection of DSBs present at the time of analysis, reference genome required
HTGTS	Induction of “bait” DSB that can capture DNA ends from other DSBs, amplification of the translocated sequence and next-generation sequencing	Unbiased, detection of large chromosomal rearrangements	Relies on concurrence of DSBs, reference genome required
Digenome-seq	<i>In vitro</i> digestion of genomic DNA with Cas9 and gRNA(s) of interest, fragmentation and whole genome next-generation sequencing	Unbiased, sensitive (0.1% reported), multiplexing possible, does not capture random DSBs, no repair of DSBs by cell machinery, no amplification step	Expensive when testing one gRNA, sequencing depth can be challenging when testing several gRNAs, reference genome required

target activity of a given nuclease is to track these breaks in the genome. This led to the development of BLESS (direct *in situ* breaks labeling, enrichment on streptavidin, and next-generation sequencing) (Crosetto et al., 2013). The genome-wide mapping of DSBs is achieved by ligating the break ends to a biotinylated linker, capturing the biotinylated fragments with streptavidin, ligating a second barcoded linker, and then identifying the products by PCR amplification and sequencing (Fig. 2). The advantage of BLESS compared to other break-labeling methods is that the DSB itself is labeled rather than proteins associated with DSBs (Crosetto et al., 2013). Several groups using BLESS to identify DSBs introduced by different Cas9 variants in mice (Ran et al., 2015) and human cells (Slaymaker et al., 2016) observed low frequencies of off-

target activity. Although BLESS allows the unbiased and genome-wide identification of DSBs, it can only detect breaks present at the time of labeling, and not earlier breaks that have already been repaired (Tsai & Joung, 2016).

3.2.5. *GUIDE-seq*

Another approach to detect DSBs caused by nuclease activity is *GUIDE-seq* (genome-wide, unbiased identification of DSBs enabled by sequencing), which is based on the integration of double-stranded oligodeoxynucleotides (dsODNs) into DSBs by NHEJ, followed by the amplification and sequencing of the tagged DNA fragments (Fig. 2). Because two primers are used to bind both strands of the dsODN, the



**Fig. 2.** Overview of methods for the detection of off-target mutations caused by site-specific nucleases (SSNs). (A) In the BLESS method, a double-strand break (DSB) is captured with a biotinylated linker, enriched on streptavidin and analyzed by PCR and next-generation sequencing (NGS). (B) In the *GUIDE-seq* method, a known double-stranded oligodeoxynucleotide (dsODN) is incorporated into the DSB, the known sequence is used as a primer binding site and the resulting products are analyzed by NGS. (C) *HTGTS* relies on a translocation involving a bait DSB and other DSBs, followed by sonication of the genomic DNA to produce fragments of suitable size, amplification of the known bait sequence with biotinylated primers, enrichment and subsequent NGS analysis. (D) *Digenome-seq* uses cell-free genomic DNA digested *in vitro* by a SSN, followed by whole genome sequencing (WGS) and 5' end plotting of the resulting sequences to identify all target sites cleaved by the SSN.

genomic sequences flanking the former DSB can be sequenced and the break site can be mapped at the single-nucleotide level. This approach was first applied in human cells, and all previously known off-target sites were identified as well as many new off-target sites with indel frequencies as low as 0.03% that were not predicted using bioinformatics (Tsai et al., 2015). Although GUIDE-seq relies on the incorporation of dsODNs into break sites, which happens in only 30–50% of the DSBs (Tsai et al., 2015), it is a powerful method to detect off-target effects (Lee, Cradick, Fine, & Bao, 2016). GUIDE-seq has therefore been used in several systematic studies to determine the underlying rules of CRISPR/Cas9 off-target activity (Doench et al., 2016; Kleinstiver et al., 2016; Kleinstiver et al., 2015). Like BLESS, GUIDE-seq can only detect DSBs present at the time of labeling.

### 3.2.6. LAM-HTGTS

Linear amplification-mediated high-throughput genome-wide translocation sequencing (LAM-HTGTS) is a method developed to track genomic translocations caused by end-joining between genomic DSBs (Frock et al., 2015; Hu et al., 2016). The method was designed to detect DSBs generated by SSNs such as TALENs and Cas9 in a sensitive, unbiased and robust manner by translocation to a known so-called bait DSB (Hu et al., 2016). The introduced nuclease cuts the bait sequence and the break is repaired by fusion with another DSB, which can cause chromosomal translocations if the breaks are located on different chromosomes or chromosome regions. Because the bait sequence is known, linear amplification PCR with a bait-specific primer can be used to amplify the translocated sequence from bulk genomic DNA. Following the addition of barcodes and adapters, the amplicons are then used for NGS and subsequent analysis (Fig. 2) (Hu et al., 2016). Repaired DSBs that are not translocated carry a restriction enzyme site that can be used for selective digestion so that these sequences are not amplified and sequenced. Standard LAM-HTGTS including the optional digestion step will therefore not identify small indels or SNPs, but could be modified accordingly. It would then require greater sequencing depth to compensate for the higher number of mutated but not translocated and wild-type bait sequence reads. LAM-HTGTS is a useful method to screen cells for large genomic rearrangements caused by nuclease-induced on-target and off-target DSBs that tend to be missed by other methods, but it relies on the simultaneous presence of the bait DSB and another DSB.

### 3.2.7. Digenome-seq

Cas9 is not only a genome editing tool, but it can also be used as an *in vitro* nuclease to generate an unbiased profile of the off-target effects of selected gRNAs. This is the basis of Digenome-seq (*in vitro* Cas9-digested whole genome sequencing), where cell-free genomic DNA is digested by Cas9 *in vitro* and the resulting fragments are sequenced by NGS (Fig. 2). For Cas9-induced breaks, many sequence reads with identical ends should be produced, which can be identified by alignment. Digenome-seq involves the sequencing of many individual genomes, so it becomes possible to identify off-target sites that are mutated at frequencies lower than 0.1% (Kim et al., 2015). Digenome-seq is also amenable to multiplexing, and can be used to analyze up to 10 gRNAs in one sequencing run, effectively saving time and money. The biggest strengths of this method include the fact that it does not capture DSBs introduced randomly in the cell, DSBs introduced *in vitro* are not processed by the DNA repair machinery, and it does not require a PCR amplification step (Kim et al., 2016). However, the artificial environment offered by cell-free genomic DNA also gives rise to the potential disadvantage of Digenome-seq: differences between *in vitro* and *in vivo* activity and specificity of Cas9 (Fu, St. Onge, Fire, & Smith, 2016) could lead to the false positive or false negative results.

## 4. Conclusions

Given the growing body of literature concerning the practice of genome editing, numerous strategies have been developed to enhance the efficiency of on-target mutations while reducing or even eliminating the off-target activity of SSNs such as Cas9. All of the methods discussed above have strengths and weaknesses, and careful selection is necessary to find the best method for a particular genome editing experiment. Algorithms for gRNA design will be improved with the growing body of knowledge concerning the mechanisms of on-target and off-target activity, and the combination of high-fidelity nucleases with optimally-designed gRNAs will further enhance the accuracy of the CRISPR/Cas9 system. At some point, the existing methods will not be sensitive enough to quantify further improvements (Nelson & Gersbach, 2016). Therefore, more sensitive off-target detection methods are required, especially for applications such as gene therapy that require absolute fidelity. Further research efforts aiming to identify or engineer high-fidelity nuclease variants, optimize gRNA design, and develop highly sensitive detection methods, will eventually make the CRISPR/Cas9 system suitable for application in humans. In addition, technical limitations concerning the small size of NGS amplicons should be overcome while maintaining high coverage, to allow the detection of even rare larger indels. Other site-specific nucleases differing from the CRISPR/Cas9 system but with comparable efficiency and higher specificity may also be discovered in the near future. Even so, the risk of off-target effects should be considered in a wider context: the few off-target mutations that occur in an experiment with a carefully selected gRNA are far less abundant than the spontaneous mutations that occur during the clonal expansion of cell cultures.

## Acknowledgements

This work was funded by the European Research Council [Grant Number 269110]. We would like to thank Dr. Richard M. Twyman for his assistance with editing the manuscript. The authors have no conflict of interest to declare.

## References

- Anders, C., Niewoehner, O., Duerst, A., Jinek, M., 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 513 (7519), 569–573.
- Anderson, E.M., Haupt, A., Schiel, J.A., Chou, E., Machado, H.B., Strezoska, Z., Lenger, S., McClelland, S., Birmingham, A., Vermeulen, A., Smith, A.V.B., 2015. Systematic analysis of CRISPR–Cas9 mismatch tolerance reveals low levels of off-target activity. *J. Biotechnol.* 211, 56–65.
- Bauer, D.E., Canver, M.C., Orkin, S.H., 2015. Generation of genomic deletions in mammalian cell lines via CRISPR/Cas9. *J. Vis. Exp.* 95.
- Brinkman, E.K., Chen, T., Amendola, M., van Steensel, B., 2014. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* 42 (22), e168.
- Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., Kim, J.S., 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* 24 (1), 132–141.
- Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J., Voytas, D.F., 2010. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186 (2), 757–761.
- Cradick, T.J., Fine, E.J., Antico, C.J., Bao, G., 2013. CRISPR/Cas9 systems targeting  $\beta$ -globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res.* 41 (20), 9584–9592.
- Crosetto, N., Mitra, A., Silva, M.J., Bienko, M., Dojer, N., Wang, Q., Karaca, E., Chiarle, R., Skrzypczak, M., Ginalski, K., Pasero, P., Rowicka, M., Dikic, I., 2013. Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat. Methods* 10 (4), 361–365.
- Dahlem, T.J., Hoshijima, K., Juryneć, M.J., Gunther, D., Starker, C.G., Locke, A.S., Weis, A.M., Voytas, D.F., Grunwald, D.J., 2012. Simple methods for generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome. *PLoS Genet.* 8 (8), e1002861.
- Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H.W., Listgarten, J., Root, D.E., 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR–Cas9. *Nat. Biotechnol.* 34 (2), 184–191.
- Feng, Z., Mao, Y., Xu, N., Zhang, B., Wei, P., Yang, D.L., Wang, Z., Zhang, Z., Zheng, R., Yang, L., Zeng, L., Liu, X., Zhu, J.K., 2014. Multigeneration analysis reveals the inheritance,



- specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 111 (12), 4632–4637.
- Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.-J., Kii, E., Alt, F.W., 2015. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol.* 33 (2), 179–186.
- Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., Sander, J.D., 2013a. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 31 (9), 822–826.
- Fu, Y., Reyon, D., Keith, J.J., 2013b. Targeted genome editing in human cells using CRISPR/Cas nucleases and truncated guide RNAs. *Methods Enzymol.* 546, 21–45.
- Fu, B.X.H., St. Onge, R.P., Fire, A.Z., Smith, J.D., 2016. Distinct patterns of Cas9 mismatch tolerance in vitro and in vivo. *Nucleic Acids Res.* 44 (11), 5365–5377.
- Ghorbal, M., Gorman, M., Macpherson, C.R., Martins, R.M., Scherf, A., Lopez-Rubio, J.-J., 2014. Genome editing in the human malaria parasite *Plasmodium falciparum* using the CRISPR-Cas9 system. *Nat. Biotechnol.* 32 (8), 819–821.
- Güell, M., Yang, L., Church, G.M., 2014. Genome editing assessment using CRISPR Genome Analyzer (CRISPR-GA). *Bioinformatics* 30 (20), 2968–2970.
- Guilinger, J.P., Thompson, D.B., Liu, D.R., 2014. Fusion of catalytically inactive Cas9 to *FokI* nuclease improves the specificity of genome modification. *Nat. Biotechnol.* 32 (6), 577–582.
- Hendel, A., Kildebeck, E.J., Fine, E.J., Clark, J.T., Punjya, N., Sebastiano, V., Bao, G., Porteus, M.H., 2014. Quantifying genome-editing outcomes at endogenous loci with SMRT sequencing. *Cell Rep.* 7 (1), 293–305.
- Hendel, A., Fine, E.J., Bao, G., Porteus, M.H., 2015. Quantifying on- and off-target genome editing. *Trends Biotechnol.* 33 (2), 132–140.
- Hruscha, A., Krawitz, P., Rechenberg, A., Heinrich, V., Hecht, J., Haass, C., Schmid, B., 2013. Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development* 140 (24), 4982–4987.
- Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., Cradick, T.J., Marraffini, L.A., Bao, G., Zhang, F., 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31 (9), 827–832.
- Hu, J., Meyers, R.M., Dong, J., Panchakshari, R.A., Alt, F.W., Frock, R.L., 2016. Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nat. Protoc.* 11 (5), 853–871.
- Iyer, V., Shen, B., Zhang, W., Hodgkins, A., Keane, T., Huang, X., Skarnes, W.C., 2015. Off-target mutations are rare in Cas9-modified mice. *Nat. Methods* 12 (6), 479.
- Jiang, W., Bikard, D., Cox, D., Zhang, F., Marraffini, L.A., 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* 31 (3), 233–239.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337 (6096), 816–821.
- Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., Kaplan, M., Iavarone, A.T., Charpentier, E., Nogales, E., Doudna, J.A., 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 343 (6176), 1247–1249.
- Karakoc, E., Alkan, C., O’Roak, B.J., Dennis, M., Vives, L., Mark, K., Rieder, M.J., Nickerson, D.A., Eichler, E.E., 2012. Detection of structural variants and indels within exome data. *Nat. Methods* 9 (2), 176–178.
- Kim, Y.G., Cha, J., Chandrasegaran, S., 1996. Hybrid restriction enzymes: zinc finger fusions to *FokI* cleavage domain. *Proc. Natl. Acad. Sci. U. S. A.* 93 (3), 1156–1160.
- Kim, H., Um, E., Cho, S.-R., Jung, C., Kim, H., Kim, J.-S., 2011. Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat. Methods* 8 (11), 941–943.
- Kim, Y., Kweon, J., Kim, A., Chon, J.K., Yoo, J.Y., Kim, H.J., Kim, S., Lee, C., Jeong, E., Chung, E., 2013. A library of TAL effector nucleases spanning the human genome. *Nat. Biotechnol.* 31 (3), 251–258.
- Kim, J.M., Kim, D., Kim, S., Kim, J.-S., 2014a. Genotyping with CRISPR-Cas-derived RNA-guided endonucleases. *Nat. Commun.* 5, 3157.
- Kim, S., Kim, D., Cho, S.W., Kim, J., Kim, J.-S., 2014b. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* 24 (6), 1012–1019.
- Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H.R., Hwang, J., Kim, J.-I., Kim, J.-S., 2015. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* 12 (3), 237–243.
- Kim, D., Kim, S., Kim, S., Park, J., Kim, J.S., 2016. Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq. *Genome Res.* 26 (3), 406–415.
- Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., Nguyen, N.T., Zheng, Z., Gonzales, A.P., Li, Z., Peterson, R.T., Yeh, J.J., Aryee, M.J., Joung, J.K., 2015. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 523 (7561), 481–485.
- Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., Joung, J.K., 2016. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529 (7587), 490–495.
- Lee, C.M., Cradick, T.J., Fine, E.J., Bao, G., 2016. Nuclease target site selection for maximizing on-target activity and minimizing off-target effects in genome editing. *Mol. Ther.* 24 (3), 475–487.
- Liang, X., Potter, J., Kumar, S., Zou, Y., Quintanilla, R., Sridharan, M., Carte, J., Chen, W., Roark, N., Ranganathan, S., Ravinder, N., Chesnut, J.D., 2015. Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J. Biotechnol.* 208, 44–53.
- Lin, Y., Cradick, T.J., Brown, M.T., Deshmukh, H., Ranjan, P., Sarode, N., Wile, B.M., Vertino, P.M., Stewart, F.J., Bao, G., 2014. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res.* 42 (11), 7473–7485.
- Liu, W., Xie, X., Ma, X., Li, J., Chen, J., Liu, Y.-G., 2015. DSDecode: a web-based tool for decoding of sequencing chromatograms for genotyping of targeted mutations. *Mol. Plant* 8 (9), 1431–1433.
- Milbury, C.A., Li, J., Makrigiorgos, G.M., 2011. Ice-COLD-PCR enables rapid amplification and robust enrichment for low-abundance unknown DNA mutations. *Nucleic Acids Res.* 39 (1), 1–10, e2.
- Nelson, C.E., Gersbach, C.A., 2016. Cas9 loosens its grip on off-target sites. *Nat. Biotechnol.* 34 (3), 298–299.
- Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E., Bamshad, M., Nickerson, D.A., Shendure, J., 2009. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461 (7261), 272–276.
- Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O., 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156 (5), 935–949.
- Paix, A., Wang, Y., Smith, H.E., Lee, C.-Y.S., Calidas, D., Lu, T., Smith, J., Schmidt, H., Krause, M.W., Seydoux, G., 2014. Scalable and versatile genome editing using linear DNAs with microhomology to Cas9 sites in *Caenorhabditis elegans*. *Genetics* 198 (4), 1347–1356.
- Perez, E.E., Wang, J., Miller, J.C., Jouvenot, Y., Kim, K.A., Liu, O., Wang, N., Lee, G., Bartsevich, V.V., Lee, Y.-L., Guschin, D.Y., Rupniewski, I., Waite, A.J., Carpenito, C., Carroll, R.G., Orange, J.S., Urnov, F.D., Rebar, E.J., Ando, D., Gregory, P.D., Riley, J.L., Holmes, M.C., June, C.H., 2008. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat. Biotechnol.* 26 (7), 808–816.
- Qiu, P., Shandilya, H., D’Alessio, J.M., O’Connor, K., Durocher, J., Gerard, G.F., 2004. Mutation detection using Surveyor™ nuclease. *Biotechniques* 36 (4), 702–707.
- Ramlee, M.K., Yan, T., Cheung, A.M., Chuah, C.T., Li, S., 2015. High-throughput genotyping of CRISPR/Cas9-mediated mutants using fluorescent PCR-capillary gel electrophoresis. *Sci. Rep.* 5, 15587.
- Ran, F., Hsu, P.D., Lin, C.-Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154 (6), 1380–1389.
- Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., Koonin, E.V., Sharp, P.A., Zhang, F., 2015. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* 520 (7546), 186–191.
- Sander, J.D., Joung, J.K., 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32 (4), 347–355.
- Santiago, Y., Chan, E., Liu, P.-Q., Orlando, S., Zhang, L., Urnov, F.D., Holmes, M.C., Guschin, D., Waite, A., Miller, J.C., Rebar, E.J., Gregory, P.D., Klug, A., Collingwood, T.N., 2008. Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc. Natl. Acad. Sci. U. S. A.* 105 (15), 5809–5814.
- Shui, B., Hernandez Matias, L., Guo, Y., Peng, Y., 2016. The rise of CRISPR/Cas for genome editing in stem cells. *Stem Cells Int.* 2016, 17.
- Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., Zhang, F., 2016. Rationally engineered Cas9 nucleases with improved specificity. *Science* 351 (6268), 84–88.
- Smith, C., Gore, A., Yan, W., Abalde-Atristain, L., Li, Z., He, C., Wang, Y., Brodsky, R.A., Zhang, K., Cheng, L., Ye, Z., 2014. 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.
- Sung, Y.H., Baek, I.-J., Kim, D.H., Jeon, J., Lee, J., Lee, K., Jeong, D., Kim, J.-S., Lee, H.-W., 2013. Knockout mice created by TALEN-mediated gene targeting. *Nat. Biotechnol.* 31 (1), 23–24.
- Talbot, J.C., Amacher, S.L., 2014. A streamlined CRISPR pipeline to reliably generate zebrafish frameshifting alleles. *Zebrafish* 11 (6), 583–585.
- Taylor, S., Scott, R., Kurtz, R., Fisher, C., Patel, V., Bizouarn, F., 2010. A Practical Guide to High Resolution Melt Analysis Genotyping. Bio-Rad Laboratories, Inc., Hercules, CA 94547.
- Thomas, H.R., Percival, S.M., Yoder, B.K., Parant, J.M., 2014. High-throughput genome editing and phenotyping facilitated by high resolution melting curve analysis. *PLoS ONE* 9 (12), e114632.
- Tsai, S.Q., Joung, J.K., 2016. Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nat. Rev. Genet.* 17 (5), 300–312.
- Tsai, S.Q., Wyvekens, N., Khayter, C., Foden, J.A., Thapar, V., Reyon, D., Goodwin, M.J., Aryee, M.J., Joung, J.K., 2014. Dimeric CRISPR RNA-guided *FokI* nucleases for highly specific genome editing. *Nat. Biotechnol.* 32 (6), 569–576.
- Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafate, A.J., Le, L.P., Aryee, M.J., Joung, J.K., 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* 33 (2), 187–197.
- Tycko, J., Myer, V.E., Hsu, P.D., 2016. Methods for optimizing CRISPR-Cas9 genome editing specificity. *Mol. Cell* 63 (3), 355–370.
- Veres, A., Gosis, B.S., Ding, Q., Collins, R., Ragavendran, A., Brand, H., Erdin, S., Cowan, C.A., Talkowski, M.E., Musunuru, K., 2014. Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell Stem Cell* 15 (1), 27–30.
- Vouillot, L., Thélie, A., Pollet, N., 2015. Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3 (Bethesda, Md.)* 5 (3), 407–415.
- Wang, K., Mei, D.Y., Liu, Q.N., Qiao, X.H., Ruan, W.M., Huang, T., Cao, G.S., 2015. Research of methods to detect genomic mutations induced by CRISPR/Cas systems. *J. Biotechnol.* 214, 128–132.
- Wu, X., Kriz, A.J., Sharp, P.A., 2014. Target specificity of the CRISPR-Cas9 system. *Quant. Biol.* 2 (2), 59–70.
- Yang, H., Wang, H., Shivalila, C.S., Cheng, A.W., Shi, L., Jaenisch, R., 2013. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* 154 (6), 1370–1379.
- Yang, L., Grishin, D., Wang, G., Aach, J., Zhang, C.-Z., Chari, R., Homsy, J., Cai, X., Zhao, Y., Fan, J.-B., Seidman, C., Seidman, J., Pu, W., Church, G., 2014. Targeted and genome-wide sequencing reveal single nucleotide variations impacting specificity of Cas9 in human stem cells. *Nat. Commun.* 5, 5507.

- Yang, Z., Steentoft, C., Hauge, C., Hansen, L., Thomsen, A.L., Niola, F., Vester-Christensen, M.B., Frödin, M., Clausen, H., Wandall, H.H., Bennett, E.P., 2015. Fast and sensitive detection of indels induced by precise gene targeting. *Nucleic Acids Res.* 43 (9), e59.
- Yee, J.-K., 2016. Off-target effects of engineered nucleases. *FEBS J.* 283 (17), 3239–3248.
- Yu, C., Zhang, Y., Yao, S., Wei, Y., 2014. A PCR based protocol for detecting indel mutations induced by TALENs and CRISPR/Cas9 in zebrafish. *PLoS ONE* 9 (6), e98282.
- Zhang, H., Zhang, J., Wei, P., Zhang, B., Gou, F., Feng, Z., Mao, Y., Yang, L., Zhang, H., Xu, N., Zhu, J.-K., 2014. The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnol. J.* 12 (6), 797–807.
- Zheng, X., Yang, S., Zhang, D., Zhong, Z., Tang, X., Deng, K., Zhou, J., Qi, Y., Zhang, Y., 2016. Effective screen of CRISPR/Cas9-induced mutants in rice by single-strand conformation polymorphism. *Plant Cell Rep.* 35 (7), 1545–1554.
- Zhu, X., Xu, Y., Yu, S., Lu, L., Ding, M., Cheng, J., Song, G., Gao, X., Yao, L., Fan, D., Meng, S., Zhang, X., Hu, S., Tian, Y., 2014. An efficient genotyping method for genome-modified animals and human cells generated with CRISPR/Cas9 system. *Sci. Rep.* 4, 6420.