BioEssays

www.bioessays-journal.com

Cas9 Cuts and Consequences; Detecting, Predicting, and Mitigating CRISPR/Cas9 On- and Off-Target Damage

Techniques for Detecting, Predicting, and Mitigating the On- and off-target Effects of Cas9 Editing

*Anthony Newman, Lora Starrs, and Gaetan Burgio**

Large deletions and genomic re-arrangements are increasingly recognized as common products of double-strand break repair at Clustered Regularly Interspaced, Short Palindromic Repeats - CRISPR associated protein 9 (CRISPR/Cas9) on-target sites. Together with well-known off-target editing products from Cas9 target misrecognition, these are important limitations, that need to be addressed. Rigorous assessment of Cas9-editing is necessary to ensure validity of observed phenotypes in Cas9-edited cell-lines and model organisms. Here the mechanisms of Cas9 specificity, and strategies to assess and mitigate unwanted effects of Cas9 editing are reviewed; covering guide-RNA design, RNA modifications, Cas9 modifications, control of Cas9 activity; computational prediction for off-targets, and experimental methods for detecting Cas9 cleavage. Although recognition of the prevalence of on- and off-target effects of Cas9 editing has increased in recent years, broader uptake across the gene editing community will be important in determining the specificity of Cas9 across diverse applications and organisms.

1. Introduction: Cas9—A Programmable Nuclease

In the last 50 years, the discovery and development of sequence-specific nucleases including restriction-modification enzymes, meganucleases, zinc-finger nucleases, transcription activator-like effector nucleases, and now clustered regularly interspaced, short palindromic repeats/CRISPR associated proteins (CRISPR/Cas), have empowered researchers to examine genotype–phenotype interactions across many classes of organisms and various diseases.[1] While each editing system has strengths and weaknesses, CRISPR/Cas editing systems are renowned for their versatility, efficacy, and their cost effectiveness compared with other techniques.[2]

CRISPR/Cas functions by recognizing and cleaving a target sequence adjacent to a "protospacer-adjacent motif" (PAM).[3] An RNA "guide" sequence complexed with the Cas effector recognizes target sites through Watson–Crick base pairing, then depending on the nuclease, the resultant activation of the effector nuclease can cleave single stranded DNA (ssDNA), double stranded DNA (dsDNA), or RNA.[4] There are currently a small number of commercially available CRISPR/Cas effectors that are well characterized and used for gene editing (*As*Cas12a, *Sp*Cas9, *Sa*Cas9).

These effectors from Class 2 CRISPR systems have single-unit effector proteins (unlike Class 1), and thus are relatively simple to repurpose for gene editing.[5] Of the many Class 2 effectors, the type II Cas9 nuclease from *Streptococcus pyogenes* (*Sp*Cas9) has risen to prominence. *Sp*Cas9 efficiently cleaves dsDNA, has robust catalytic activity in eukaryotic cells, and has a short —and thus frequently occurring— PAM motif of 5′-NGG-3′.[6,7]

The ability of Cas9 to recognize DNA or RNA makes it a highly versatile platform for genetic engineering. Cas9 specifically cleaves double-stranded DNA, resulting in deletion[6,8,9] — or insertion of a sequence if a repair template is used.[6,9] Furthermore, Cas9 can cleave RNA transcripts for gene knockdown if the effector is complexed with a PAM-simulating oligonucleotide.[10] When fused with another functional protein, "dead" or catalytically inactive Cas9 (dCas9) or Cas9 nickase (nCas9) can be used to visualize loci,[11] regulate transcription,[12] or edit single bases.[13]

This range of capabilities has directly enabled basic and translational work in the molecular life sciences,[6,8,9] industrial microorganisms,[14,15] horticulture,[16,17] agriculture,[18] and medicine.[19,20] Currently over 20 phase 1/2 clinical trials using CRISPR/Cas for gene or cell therapy are underway.[21–24]

Although CRISPR/Cas9 is leagues ahead of earlier techniques in specificity and simplicity, there are notable drawbacks. Since the discovery of CRISPR/Cas9 editing, "off-target" cleavage activity has been observed.[25–27] Recently, other unwanted effects have been described at the target site "on-target effects" such as large

A. Newman, L. Starrs, Dr. G. Burgio
Department of Immunology and Infectious Diseases
The John Curtin School of Medical Research
The Australian National University
131 Garran Road
Canberra, ACT 2601, Australia
E-mail: Gaetan.burgio@anu.edu.au

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/bies.202000047

DOI: 10.1002/bies.202000047

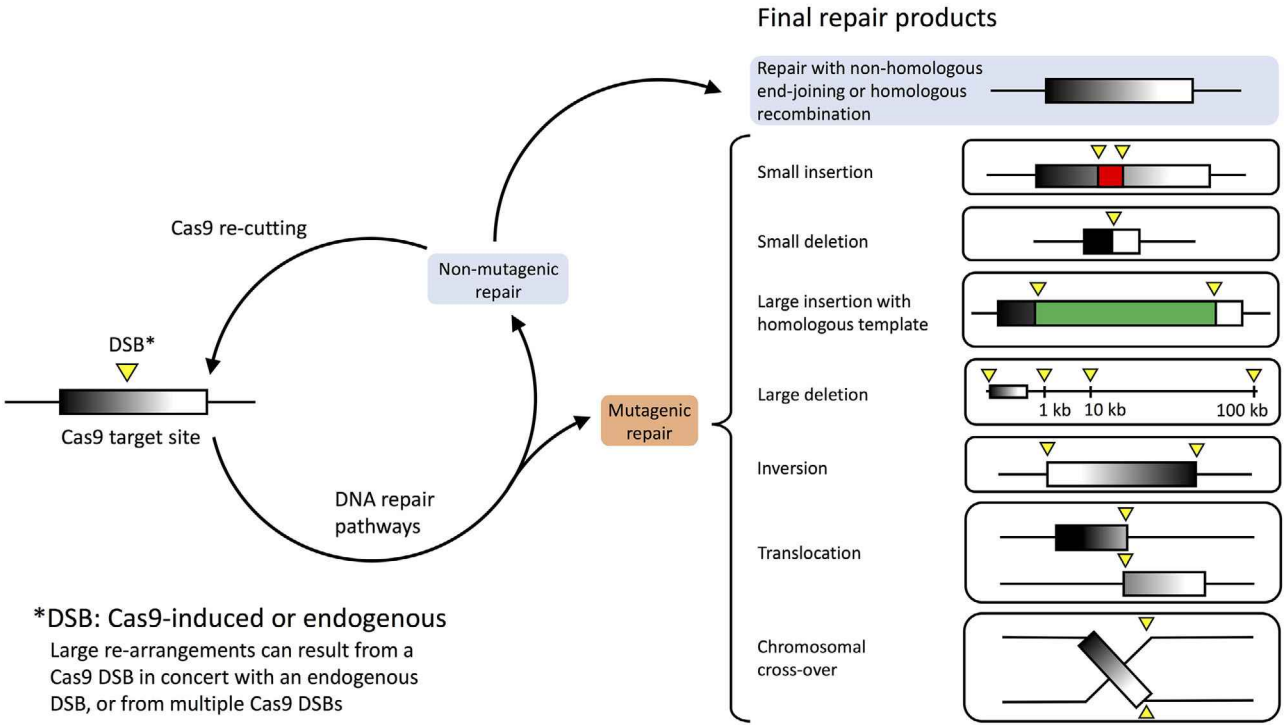

Figure 1. Consequences of a double-stranded break (DSB). DNA repair pathways determine the products of DSBs induced by Cas9 or endogenous. Non-mutagenic repair of a Cas9 cleavage site will result in a cycle of Cas9 cleavage, repair, and re-cutting. Final editing products result when DNA repair mutates a target site beyond possible recognition by Cas9, or when Cas9 is no longer available (degradation of gRNA or Cas9, arrested expression of Cas9). Repair of DSBs can result in a variety of products, from small insertions or deletions, to large-scale genomic rearrangements.

deletions,[28,29] inversions,[29–31] translocations,[30] chromosomal crossovers,[30] and p53 activation.[32,33]

In light of the wide use of CRISPR/Cas9 across the life sciences and in its first human clinical trials, we review the mechanisms of Cas9 on- and off-target activity, methods to detect on- and off-target editing products, and techniques to improve the specificity and efficiency in editing with *Sp*Cas9.

2. Cas9: Editing through Double Stranded Breaks

Cas9 effects genome editing through double-stranded DNA breaks (DSBs). DSBs are a strong insult to genomic integrity—lethal if unrepaired— and provoke a DNA damage repair response from the host cell.[32,33]

Cas9 induces blunt or 1-2 base-pair overhanging DSBs,[31,34] which can be religated by the host's endogenous repair pathways.[35] The mechanisms and factors influencing choice of repair pathway is beyond the scope of this review, please refer to Corn et al.[36] for a review of these mechanisms.

These post-DSB repair mechanisms are responsible for the products of Cas9 editing, the so-called "on-targets" and "off-targets."[37] Most Cas9-induced DSBs result in small insertions or deletions (indels), but like any DSB, aberrant religation of loose DNA ends can result in large deletions, translocations, and other genomic rearrangements at the target site—"on-target effects"—[28,30,38] (**Figure 1**). Cas9 recognizes sequences at the target site and similar, but mismatching, "off-target" sequences (see Section 3), resulting in cleavage and editing products.[25–27] However, undesired editing products following Cas9 cleavage can occur at the target site (on-target effects), and/or away from the target site (off-target effects). Semantic separation of these terms is generally useful to separate "wanted" and "unwanted" effects, but presents a false distinction regarding their cause—all arise from double-stranded DNA breaks. Compared to off-target sites, on-target sites have a higher probability of Cas9 recognition and cleavage. Furthermore, repair of a DSB can result in an intact (or mostly intact) target site, which can be recognized and recut by Cas9 until mutagenic repair edits the target site beyond recognition.[31] This results in a greater cumulative number of DSBs at on-target sites and increases the chance of mutagenic repair products, including large rearrangements.[30] Such on-target effects are thus a consequence of efficient[39] Cas9 on-target activity.

Cas9-induced DSBs are immensely powerful—multiple Cas9 DSBs can reliably cause large genomic deletions[40,41] or inversions;[29,40] as well as truncation,[38] deletion,[42,43] or splicing of entire chromosomes.[44] As such, the specificity of Cas9 target recognition is of paramount importance in gene editing.

3. Mechanisms of Cas9 Target (mis)Recognition

Cas9 target sequence recognition takes place over three steps: PAM recognition; target strand annealing; and nuclease

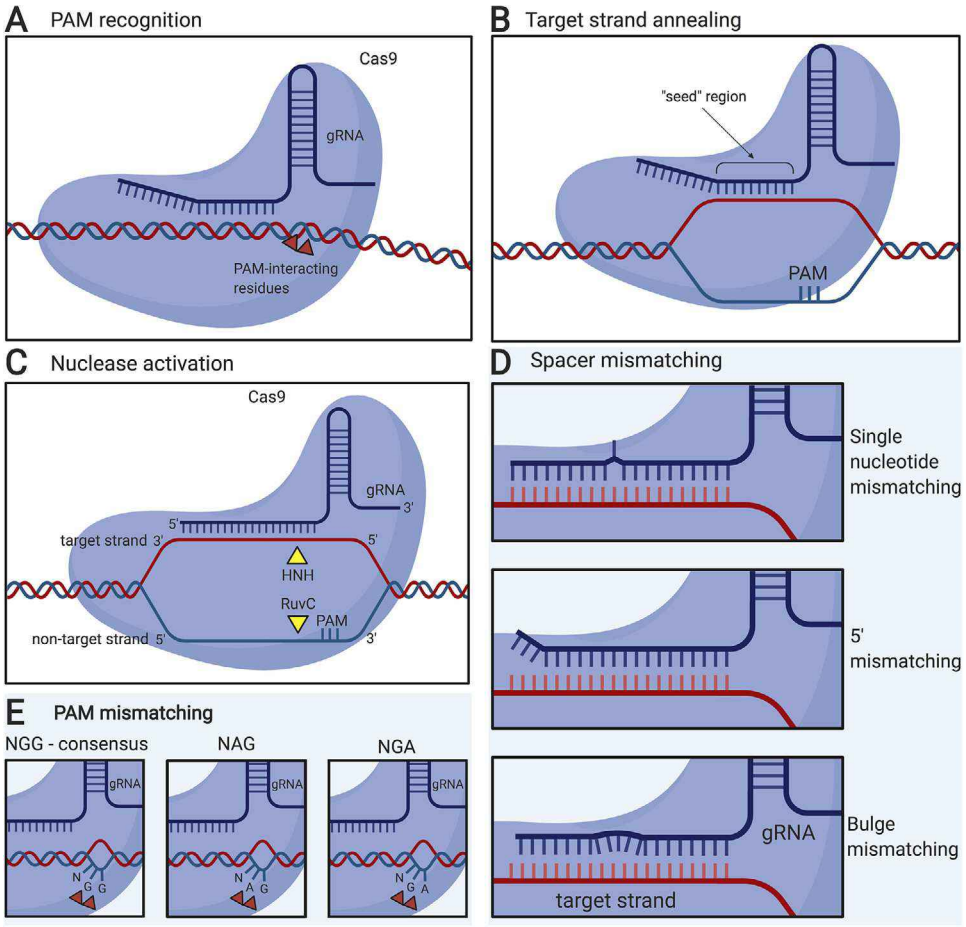

Figure 2. The mechanism of Cas9 target recognition. A) Recognition of a 5′-NGG-3′ PAM sequence in dsDNA, B) dsDNA unwinding and target strand annealing to the "seed" region, C) full R-loop formation and activation of nuclease domains to effect double-strand DNA cleavage. D) Spacer mismatching through single nucleotide mismatching, 5′ mismatching, and bulge mismatching. Note that mismatches can be present in either the RNA or DNA strand. E) PAM mismatching, Cas9 has a consensus NGG PAM, but also recognizes NAG, NGA motifs.

activation. As such, Cas9 can misrecognize a target at any of these three steps.

3.1. PAM Recognition

To efficiently sample the genome for a target, Cas9 undergoes random 3D diffusion[45] and 1D diffusion along dsDNA.[31] The PAM-interacting residues of Cas9 protrude into the major groove to bind to PAM motifs and commence DNA unwinding[3] (**Figure 2**A). *Sp*Cas9 has a consensus 5'-NGG-3' PAM, but will bind to similar PAMs to a lesser extent[46] (NGA, NAG, and so on). Given a similar spacer sequence, this can result in off-target cleavage[46] (Figure 2E).

3.2. Target Strand Annealing

PAM motif binding induces dsDNA bending, breaking the dsDNA duplex and allowing R-loop formation.[3] The target DNA strand can rapidly anneal to the 10–12 bases PAM-proximal of the guide RNA (gRNA; the "seed" region; Figure 2B), which are held in an A-helical conformation.[47] Perfect sequence complementarity drives PAM-distal DNA unwinding, full gRNA:target DNA R-loop formation, and nuclease activation (Figure 2C). To varying degrees, mismatching in the PAM-proximal heteroduplex stops unwinding and causes Cas9 dissociation.[48]

However, Cas9 can stably bind and cleave mismatching sequences depending on the number, type, and spatial distribution of the mismatches. The PAM-proximal 3′ end of the spacer sequence is most sensitive to mismatches, and the PAM-distal 5′ end the least. For example, a single nucleotide mismatch 7 nucleotides from the PAM can completely abolish target cleavage, while two or three nucleotide mismatches at the 5′ end may still permit cleavage[48] (Figure 2D). The type of nucleotide mismatch can be important, and some evidence shows that rG:dT mismatches are more tolerated than rC:dC.[49] Notably, the position of a wobble or transition mismatch can also affect cleavage rates.[49] Furthermore, Cas9 may be able to accommodate "bulge" mismatches in either the gRNA or target DNA strand[50] (Figure 2D), though the extent and importance of this type of mismatching is not well established.[49,51]

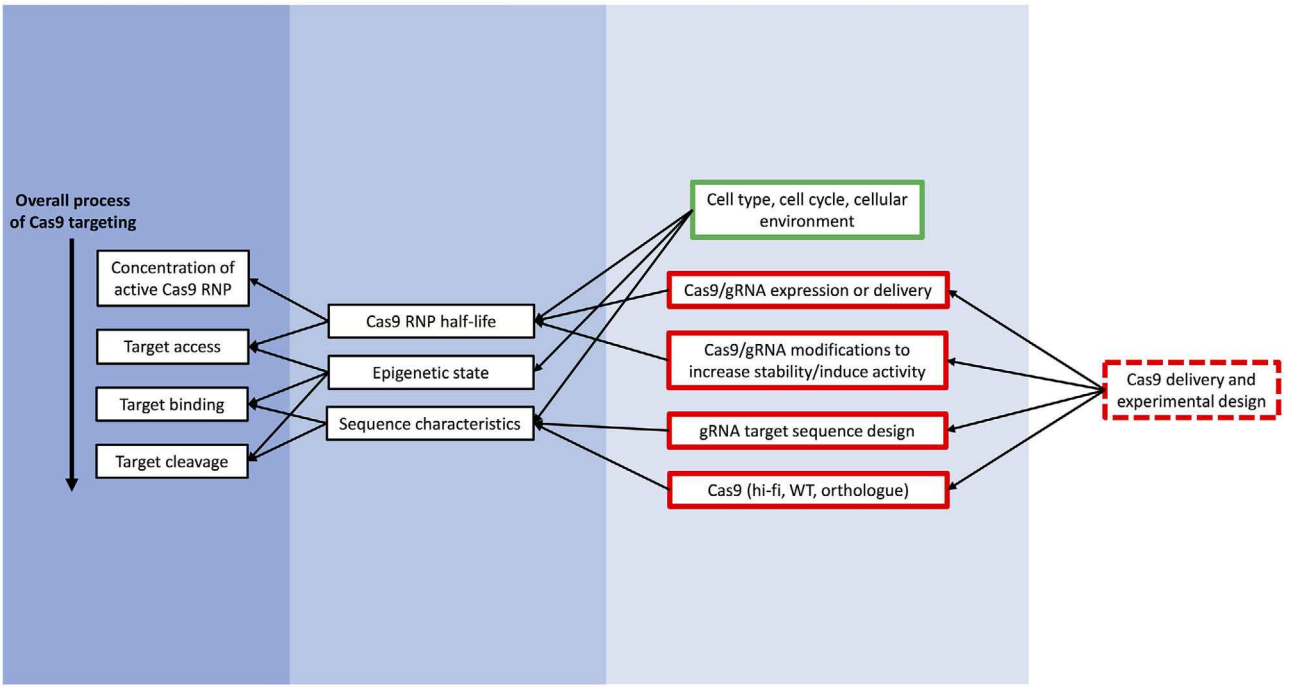

Figure 3. A non-exhaustive schematic of factors that can influence Cas9 targeting. Epigenetic state: encompasses chromatin state, nucleosome occupancy, active transcription/replication. Sequence characteristics: encompasses PAM/spacer sequence mismatches, G/C content, single nucleotide polymorphisms (SNPs), DNA structure distortion. Red boxes indicate factors dependent on experimental design (discussed in Section 5). Green box indicates cell-specific and thus invariable factors. Black boxes are downstream factors influenced by the above.

These examples show how features in the DNA sequence can greatly influence Cas9 cleavage, but these are not universally consistent. The target strand binding is a process that occurs in a context broader than sequence complementarity, and is influenced by numerous factors such as DNA tertiary structure and epigenetic state (see Section 3.4, and **Figure 3**).

3.3. Nuclease Activation

Finally, the HNH and RuvC nuclease domains are activated to cleave the target and non-target strands respectively[7] (Figure 2C). This occurs through a sensing-locking mechanism, which does not reliably sense mismatches at the 5' end of the gRNA;[52] however engineered Cas9 variants have improved mismatch sensing (see Section 5).

3.4. Cas9-Independent Influences on Target (mis)Recognition

In genome editing, Cas9 must contend with the epigenetic and transcription state of the genome before any sequence-based target recognition occurs. Genomic inaccessibility limits the number of possible Cas9 recognition sites,[53,54] preventing a large number of potential off-target cuts. Similarly, Cas9 competes with other DNA-binding proteins such as polymerases,[55] transcription factors,[12] histones,[53] and histone chaperones.[54,56]

Although these cellular processes make much of the genome off-limits to Cas9 cleavage, there are indications they may also directly contribute to off-targets. Logical PAM/gRNA mismatches explain many off-targets, but unbiased off-target detection techniques reveal that many more are not easily explicable.[46] Tsai and colleagues identified DSBs at sites with up to four or even six mismatches, as well as Cas9-independent "breakpoint hotspots" that act in concert with Cas9 DSBs for large-scale genomic rearrangements.[46] These endogenous DSBs may be due to super-helical stress (DNA distortion) caused by transcription and replication.[35] However, in vitro studies have demonstrated Cas9 stably binding and cleaving distorted DNA substrates, such as those generated in transcription and replication.[57,58] Against these DNA "bubbles", Cas9 cleaved sites with up to ten mismatches.[57] This may explain the higher than expected off-target cleavage in frequently transcribed regions.[46]

Given the incompletely predictable nature of Cas9's mechanism of target recognition (and subsequent on-/off-target cleavage), myriad methods have been developed to detect the products of Cas9 editing.

4. Methods for Detecting On- and Off-Target Editing

For researchers using CRISPR/Cas9 editing to examine genotype–phenotype relationships, a variety of methods exist to determine if editing occurred at the intended site, and if there are potentially confounding on- or off-target editing products.

However, analysis is complicated by the complex and highly variable nature of repair products from DSBs. It is well established that two or more gRNAs spanning a region can efficiently

Table 1. Advantages and limitations of five common targeted detection methods.

Name	Method	Advantages	Limitations
Sanger sequencing based detection or tracking of indels by decomposition (TIDE)[67] 2014	PCR amplification followed by Sanger sequencing (and TIDE analysis)	Simple to use, can detect as low as 1 bp indels	Low read depth, low throughput
Standard PCR, long-range PCR[62]	PCR amplification (normal or long-range), analysis of indel products by fragment length polymorphism	Very sensitive, primer design can capture large indels	Low throughput
High throughput amplicon sequencing	PCR amplification, adaptor ligation, followed by high-throughput sequencing and analysis (e.g., CRIS.py[68] or CRISPResso2[69])	High read depth	Cannot detect large indels, not cost effective for smaller sample sizes[70]
Long-read sequencing	SMRT,[28] Nanopore[71] sequencing techniques	Long reads can capture large indels	Not cost effective for smaller sample sizes[70]

induce a large deletion between the target sites.[59] This has been used across bacteria, crop plants, yeast, zebrafish, xenopus, as well as in mammalian cell lines, primary cells, zygotes, and embryos, in order to generate large genomic structural variants.[40] On-target genotyping revealed that not only deletions, but inversion and duplications also occur.[29,40,60] Growing recognition of such outcomes has led to more thorough assessment of such "on-target effects," and the findings that translocations,[61] chromosomal crossovers,[30] and chromosomal truncations[38] can result from editing with a single gRNA.

Importantly, these large rearrangements are not detectable by standard on-target genotyping methods,[28,30,38,41,62] leading to the unfounded conclusion that these outcomes are either not present or occur at low frequency. Generation of structural variants has shown kb to Mb scale deletions and inversions occur in 10–40% of clones,[40] and that larger-than-expected deletions (greater than 250 bp) occur in ≈20% of cases.[30,37,62] This a substantial probability that a primer binding site will be moved or deleted, leading to allelic drop-out, and non-detection of on-target Cas9 editing products.[41]

While the detection of off-target Cas9 editing has rightly been the focus of much study and engineering effort, on-target mutagenesis is arguably of greater consequence. Genomic structural variants remove or create coding sequences, and change gene copy number, resulting in a variety of normal and pathogenic phenotypes.[63]

To detect the products of Cas9 editing, one must know where DSBs have occurred, and have a sequencing strategy to capture the many possible repair outcomes that can result (see Figure 1). These methods can be either "targeted" or "unbiased" in the genomic regions they analyze. In general, the targeted methods are faster and cheaper, and the unbiased are more resource and time intensive.

4.1. Targeted Detection Methods

Targeted methods analyse user-defined amplicons with high sensitivity, but are inherently biased to only detect editing at sites predicted a priori, and products that can be amplified within the parameters of the protocol.

These methods are most commonly used to assess on-target editing, by sequencing a small (<1 kb) amplicon spanning the target site (excepting long-range/long-read methods). However, any deletion larger than the pre-defined amplicon is not detectable due to loss of a primer binding site, and large insertions and rearrangements produce large amplicons that cannot be amplified.[64] Thus, high-throughput sequencing techniques are particularly unsuited to detect on-target deletions, due to their short read lengths. This means studies such as SPROUT[65] and DeepSpCas9[66] are blind to as many as 20% of Cas9 editing events.

Methods to detect on-target deletions include standard PCR, long-range PCR, and long-read sequencing (see **Table 1**). By sequencing larger spans around a Cas9 target site, insertions, deletions, and inversions within this region can be detected.[28,40] Fragment-length polymorphism of PCR amplicons inexpensively and rapidly shows the presence of deletions and insertions, while amplicon sequencing also detects inversions. However, targeted sequencing cannot detect on-target rearrangements such as translocations, where a Cas9 cleavage site is adjacent to an unknown genomic location, and thus a primer site is indeterminate. To assess such repair outcomes, a variety of unbiased methods of DSB detection are available (discussed in Section 4.2).

In the case of cell-lines with known chromosomal instability, karyotyping of edited cells with appropriate controls is recommended. A study of Cas9 editing in colorectal cancer cell lines demonstrated chromosomal truncations and complex karyotypes resulting from Cas9 editing,[72] as well as from selection during cell passaging. Fluorescence in situ hybridization methods are of use in these circumstances,[30,73] capable of precisely locating genomic sequences nucleus-wide, to elucidate the presence of duplications, translocations, and chromosomal abnormalities.

In assessing off-target editing products, targeted methods have similar bias. Targeted methods will miss unpredicted off-target effects; the off-target detection is only as good as the off-target prediction (see Section 5). GUIDE-seq and other methods have shown that off-target DSBs can participate in deletions and translocations, further complicating strategies to predict genomic locations to sequence. Some common methods for targeted off-target detection are listed in Table 1.

4.2. Unbiased Detection Methods

Unbiased off-target detection methods do not presume to predict where Cas9 will cleave and use various methods to

determine the locations of Cas9-induced DSBs. Whole genome sequencing compares edited versus unedited genomes, while other methods use physical capture of DSBs in combination with high-throughput sequencing methods to determine their location.[46,61,74–83] With genome-wide detection, these unbiased methods are thus well suited to the characterization of both on- and off-target editing products, from small off-target indels to large on-target rearrangements.

Whole-genome sequencing (WGS) can detect genome-wide off-target editing products generated from DSB and DSB-free editing (e.g., base editing), although the quality of the experimental controls and the computational pipeline are critical to distinguish genuine off-targets from background genetic variation, private variants, or misalignment.[84] Although providing a "complete picture" of indels and single nucleotide polymorphisms (SNPs) of the edited organism, WGS is expensive, bioinformatically intensive, and has low read depth and sensitivity compared to amplicon sequencing. Despite the expense, commercial availability of WGS makes it accessible to many researchers, unlike other methods that require specific expertise in the necessary techniques, instruments, and analysis.

To increase the detection accuracy, other unbiased methods use a variety of techniques to detect a DSB and identify its location (see **Table 2**). These methods use high-throughput sequencing of DNA fragments "captured" at DSBs, and although less sensitive than targeted sequencing, they accurately detect Cas9 DSBs genome-wide. They differ in their methods of DSB capture, and whether Cas9 cleavage is performed on isolated genomic DNA (in vitro), in cell culture or primary cells, or in a model organism (in vivo). However in vitro cleavage and detection of DSBs tend to overestimate off-targets, as the conditions allow physiologically unlikely cleavage events to occur.[76] This discrepancy is most likely explained by factors that may be different or non-existent outside living cells, such as chromatin state[85] and histone chaperones.[56] Such methods recommend in vivo validation for off-target cleavage using targeted sequencing.[77] DSB detection in living cells has the drawbacks of toxicity in some cell-types (such as GUIDE-seq in induced-pluripotent stem cells (iPSCs)), and a limited time-window of cross-linking to capture DSBs.[61] Though increasing the cost and difficulty of the procedures, a large enough sample size and high read depth enables the majority of DSBs to be detected. Importantly, when detecting DSBs from Cas9 cleavage in cell-lines, it is vital that genomic DNA is extracted from the whole cell population and not just live adherent cells. Doing otherwise (like Kleinstiver and colleagues[86]), heavily biases against detection of on-and off-target edits that result in loss of function in cellular attachment or fatal genomic damage.

However, an evaluation of three studies[46,61,87] measuring off-target editing products for the same gRNAs with different DSB detection methods, in different cell types, found significantly non-overlapping off-target sites[89] (at EMX1, not one off-target was found by all three methods; at VEGFA site 1, only 7 of 37) . This heterogeneity in outcome reflects variable sensitivity in DSB detection, and the fact that Cas9 cleavage has many influencing factors dependent on experimental design and cell-type (see Section 5 and Figure 3). As such, interpretations of the on- and off-target activity of Cas9 must take these factors into account (see Section 5.4.3).

4.3. Validation of Cas9 Editing

With targeted methods cheaper and more sensitive, and unbiased methods more accurate but more difficult, there is no single method for routine assessment of Cas9 editing products. Targeted detection of editing products at the most probable Cas9 cleavage sites—the on-targets—provide a sensitive screen for successfully edited clones; and unbiased detection methods provide accurate a genome-wide assessment of Cas9 on- and off-target editing.

While validation of Cas9-editing by multiple methods of on- and off-target detection is the ideal case, real-life limitations on time, money, and equipment often dictate otherwise. This places the field of Cas9 editing in an awkward double-bind, where researchers know that Cas9 can and will induce unwanted on- and off-target editing products, but are not able to assess the extent to which this occurs.

When it comes to assessing the genotype of a Cas9-edited organism, targeted analysis of the on-target offers quick and cheap screening for successfully edited clones. It is important to note that even if an on-target mutation is detected, this does not necessarily translate to successful editing. Even Cas9-induced frameshift mutations can have protein production rescued by exon skipping[90] and translation reinitiation,[91] though this is easily assessed by Western blotting for the target protein. Similarly, targeted sequencing or Western blotting to analyze off-target editing of downstream genes or proteins may also be used.

Validation of a Cas9-knockout can be achieved by a genetic rescue experiment, in which reversal of a phenotype by supplementing the edited organism with the deleted gene would strengthen the assumption of a genotype–phenotype relationship.[92]

Targeted sequencing of high-scoring off-target sites (see Section 5.4) may be indicative of the extent of collateral damage in a Cas9 editing protocol, although dependent on the model system used, cell-death post-transfection can be a sign of toxicity due to widespread off-target mutagenesis. Unbiased genome-wide detection of DSBs in a physiologically relevant setting (i.e., with GUIDE-seq or DISCOVER-seq) are more informative, but more resource intensive. For applications such as somatic gene therapy, this higher standard of genotype assessment is necessary.

In order to advance the use of Cas9 for gene editing, more accessible methods of unbiased assessment of editing products are needed, as well as routine assessment of on-target editing outcomes. Thorough characterization of the frequency and extent of on-target effects is necessary to determine its importance. This very tractable goal is attainable with collaboration across the community of researchers using Cas9 for gene editing in diverse cell-types.

On the whole, on- and off-target detection methods have greatly advanced our understanding of the target specificity of Cas9. These datasets provide the foundation of the next big challenge in the gene editing field—accurately and reproducibly predicting the target specificity of a Cas effector/gRNA complex.

5. Techniques to Mitigate Off-Target Cleavage

The overall process of Cas9 genome editing is influenced by numerous factors, affecting both activity and specificity.

Table 2. Advantages and limitations of common unbiased DSB detection methods.

Name	Method	Advantages	Limitations
Whole genome sequencing	Whole-genome or whole-exon sequencing with shotgun high-throughput sequencing and genome assembly	Easy access to a sequencing provider	Low read depth compared to amplicon sequencing, requires good sample sizes, including controls to eliminate private variants[84]
GUIDE-seq[46] 2014	Cas9 cleavage in living cells, integration of transfected double-stranded oligo (dsODN) at DSB, DNA isolation, shearing, adaptor ligation, and high throughput sequencing. Primer specific to dsODN used to locate DSBs	Unbiased capture of DSBs in living cells	dsODN incorporation 2/3× lower than DSB rate, limited to blunt-end DSBs.[46] Procedure is toxic to iPSCs[83]
IDLV[74] 2015	Integration-deficient lentiviral vector, non-homologous end-joining repair integrates vector at Cas9-induced DSBs in living cells, sequence from primers specific to lentivirus vector	Moderate integration efficiency	Does not integrate at exact DSB location
dCas9 ChIP-seq[75] 2015	Cas9 cleavage in living cells, formaldehyde treatment for protein-DNA cross-linking, immunoprecipitation of HA-tagged dCas9, DNA purification, adaptor ligation, and high-throughput sequencing	Applicable for dCas9 fusions	Detects Cas9 binding sites, not representative of Cas9 cleavage sites, so overestimates potential off-targets
LAM-HTGTS[61] 2016	Cas9 and I-SceI cleavage in living cells, DNA isolation and shearing, LAM-PCR. Detects Cas9 DSBs that translocate to a "bait" DSBs induced by I-SceI, adaptor ligation and high-throughput sequencing	Can estimate frequency of deletions/translocations between DSBs. High sensitivity as it captures DSBs across broad time period	Only detects DSBs that translocate to "bait" DSBs
SITE-seq[76] 2017	DNA isolation, in vitro cleavage by Cas9, adaptor ligation, and high-throughput sequencing	Can titrate Cas9 concentrations	In vitro cleavage overestimates off-targets
CIRCLE-seq[78] 2017	DNA isolation, shearing, and circularization, in vitro Cas9 cleavage, adaptor ligation to linear fragments, high-throughput sequencing.	Highly sensitive detection technique greatly enriches for Cas9-cleaved genomic DNA. In vivo verification provides physiological relevance	Total process is very time consuming, loses true-positives in noise of many hits/false positives[83]
VIVO[77] 2018	VIVO: CIRCLE-seq, confirmation with targeted deep sequencing from in vivo samples		
DIG-seq[79] 2018	Later iteration of DIGENOME-seq[87] (2015). Isolates cell-free chromatin-associated DNA, in vitro Cas9 cleavage, high-throughput whole-genome sequencing, bioinformatics pipeline for indel detection	Informs how chromatin state affects Cas9 targeting	Requires high read depth to capture all events[76]
qDSB-seq[80] 2019	Later version of BLISS[81] and BLESS.[82] Quantitative PCR method for DSBs captured in fixed cells (by BLISS/BLESS), DSB processing and adaptor ligation, high-throughput sequencing	Very high sensitivity, captures in situ DSBs over long-time span (unlike IP methods)	In fixed cells, which compromises nuclear architecture[88]
DISCOVER-seq[83] 2019	Cas9 cleavage in living cells, formaldehyde treatment for protein-DNA cross-linking, followed by immunoprecipitation of DNA-repair associated protein Mre11, DNA purification, adaptor ligation, high-throughput sequencing	Highly sensitive off-target detection in primary cells (patient-derived iPSCs)	Requires high read depth to capture all events

5.1. Control the Cas9 Concentration

High concentrations of Cas9 are associated with an increase in off-target editing and higher cytotoxicity as a direct result.[26,76] As such, temporal and tropic control of Cas9 concentration can decrease off-target activity with negligible difference in on-target efficiency.

Cas9 delivered as a Ribonucleoprotein RNP or mRNA (e.g., by lipofection, electroporation, micro-injection, nanoparticle) has a limited half-life determined by host cell processes.[93,94] Modifying Cas9 with a proteasome degradation signal (ubiquitin) shortens the half-life, limiting the time-span for Cas9 editing and improving specificity.[95] In contrast, modification of the gRNA with polyadenylation or a G-quadruplex improves RNP half-life in the cell, for a longer editing time-span.[96,97]

Delivery of Cas9 in nucleotide form have a longer half-life (excluding mRNA, including plasmids, recombinant lentivirus, or adeno-associated viruses) and require transcriptional control of Cas9 expression. Cell-cycle specific expression of Cas9 can favor different DNA repair pathways,[98] and in vivo

editing is less toxic and highly tropic with tissue specific promoters.[99,100]

Cas9 can only assume a catalytically-competent conformation in complex with its gRNA, so RNA aptamers have been developed that control gRNA folding, and thus Cas9 activity.[101,102] Furthermore, Cas9/gRNA modifications allow for chemical-induced entry to the nucleus, light-switchable enzyme activation, and Cre-switchable gRNAs, for strict temporal control of editing.[103–106]

In a similar vein, the evolutionary arms race between prokaryotes and their viruses have evolved a multitude of anti-CRISPR proteins, which suppress CRISPR effector activity.[107,108] These can inactivate Cas9 in numerous ways, from preventing gRNA loading, target DNA binding, or allosterically preventing nuclease activity.[109] Thus anti-CRISPR diversity provides a rich resource to control the activity of Cas9 at various stages,[110] though at a cost of increasing the complexity of an editing system.

5.2. gRNA Modifications

The seeming indifference of WT *Sp*Cas9 to the PAM-distal mismatches (see Section 3) has led to development of truncated 17/18 nt gRNAs, and extended gRNAs with a PAM-distal hairpin structure. Truncated gRNAs obviate the possibility of PAM-distal mismatches. Although there are fewer unique on-target sites for 17/18 nt versus 20 nt sequence, the shorter sequence is more sensitive to mismatches, resulting in overall greater specificity with similar on-target efficiency.[46,111]

Hairpin-RNAs take another approach, by adding a hairpin-RNA secondary structure to the PAM-distal end of the gRNA, targeting specificity is markedly improved for Cas9 and orthologues.[112] The exact mechanism of this specificity is not clear, but the steric presence of the hairpin appears to strongly decrease tolerance to PAM-distal mismatching. Similarly, modification of the gRNA backbone and DNA–RNA hybrid gRNAs have been shown to improve Cas9's mismatch tolerance.[113,114]

5.3. Protein Modifications

The specificity of a ribonucleoprotein complex of Cas9 and a gRNA (Cas9 RNP) complex is not solely determined by the gRNA, there are complex and not fully understood protein dynamics underlying mismatch-sensing.[52] High-fidelity (HiFi) variants of Cas9 demonstrate this strikingly, producing greatly reduced off-target cleavage even when complexed with a promiscuous gRNA.[115–118] Although significantly reducing off-target DSBs, this comes at the cost of lower on-target efficiency.[119]

To reduce the consequences of off-target cleavage, deletion of a gene can also be achieved via creation of a staggered DSB with two Cas9 nickases.[120] Off-target single-strand DNA "nicks" are readily repaired, while the dual on-target nicks form a staggered DSB, that is thought to undergo successive rounds of microhomology driven repair until the target sites are destroyed.[62] Due to the low probability of multiple off-target nicks in close proximity, dual-nickase editing is more specific than wild-type Cas9;[120] and furthermore nickase-induced DSBs have no detectable translocations.[121] Although dual-nicking is generally thought to be less efficient, D10A nCas9 (retaining an active HNH domain) is reported to have similar to wild-type on-target efficiency.[122] However, nickases are also reported to have high-off target mutagenesis at certain sites,[111] and can result in larger than intended deletions.[123]

5.4. gRNA Design with Computational Tools

Central to specificity of gene editing is the uniqueness of a target sequence in the genome. Knowing the tolerance of Cas9 for PAM and spacer sequence mismatching (see Section 3), any target site in the human genome could plausibly have up to 300 off-target sites across the genome.[124] Under the collective influence of the factors shown in (Figure 3), only a small fraction of potential off-target sites are cleaved.

As such, numerous studies have assessed on-/off-target cleavage by Cas9, and generated datasets used to derive tools to predict on-/off-target activity of a given gRNA. These range from simpler rule-based off-target cleavage prediction, to machine learning models that integrate a multitude of features.

In **Table 3** we present a short list of common and influential computational methods for gRNA design with off-target prediction. These methods have freely available online interfaces or standalone programs available for download, allow user-input of a target genome, and have been cross-evaluated in relevant literature.

5.4.1. Rule-Based Off-Target Prediction

This family of tools identifies potential Cas9 cleavage sites by searching for sequences that are homologous or have high similarity to a PAM + spacer sequence. These vary in alignment methods, permissible PAMs, number of tolerated mismatches, and gRNA scoring/filtering.

Scoring rules can rank gRNAs or remove sequences considered likely to be cleaved by Cas9;, for example, a single PAM-distal mismatch is more likely to be cleaved than a sequence with multiple mismatches in the seed region.[48] These rules can be simple heuristics weighting for factors like mismatch position (closer to PAM more penalized), mismatch number (the more mismatches, the more penalized), and mismatch type (e.g., wobble base pairing, transition, and transversion), such as used in CCTop[126] and the now defunct CRISPR.mit.[131] Further filtering can then remove gRNAs to account for sequence features known to reduce gRNA expression/on-target efficiency (high/low %GC, sequence secondary structure).[137]

However, predicting off-target cleavage solely on the basis of sequence similarity misses some off-targets detected with unbiased genome-wide off-target detection.[48] Experimental studies of on-/off-target activity have provided datasets from which on-/off-target scoring methods have been derived—in CRISPRscan[133] (Moreno–Mateos score), the GPP sgRNA designer[49] (CFD score), WU-CRISPR[134] (Wong score), and the Farboud-GG rule.[138] These are all aggregated into the CRISPOR tool,[89] which recommends choosing a scoring method to match the gRNA expression. One example of this is CFD/Wong score for gRNAs under a U6 promotor, and the Moreno–Mateos/Farboud-GG rule for T7 in vitro transcribed gRNAs.

Table 3. Common and influential gRNA design tools with rule-based off-target prediction.

Name	Method	Advantages	Limitations
GT-Scan[125] (2014)	Simple off-target alignment with Bowtie. User can define versatile "target rule" of high/low/no specificity, applicable for Cas9 targeting (and other CRISPR, TALENs, ZFNs)	Allows any PAM input, spacer length, and target pattern "rule"	Only allows up to 3 mismatches, ranks off-targets based solely on similarity
CCTop[126] (2015)	Uses Bowtie for off-target alignment, off-target scoring with simple weighting against PAM-proximal mismatches	Simple interface shows matching sequences in exons, introns, intragenic regions	Simplistic off-target ranking gives many off-targets, many false positives
Cas-Designer[127] (2015)	Inputs up to 1000 bp seq to be targeted, uses CasOFFinder[128] to search for off-targets, plus accounting for bulge mismatching. Also attempts prediction of microhomology repair outcomes to give an "out-of-frame" score for knockouts	Can filter proposed gRNAs for number of mismatches, GC content	Online tool only allows 2 mismatches
Guide-Scan[129] (2017)	Uses "retrieval tree" to find all possible targets, rather than alignment, for exhaustive off-target searching. Adopts rule set 2[49] for on-target scoring, and CFD[49] for off-target scoring	Also allows paired gRNA design for double-nicking	Does not allow bulge mismatching
CRISPOR[89,130] (2016–2018)	BWA alignment to find off-targets and aggregates on-/off-target scoring from MIT,[131] CFD,[49] CROP-IT,[132] CCTop,[126] CRISPRscan,[133] Azimuth,[49] WU-CRISPR[134] Continuously integrates new databases, enzymes, genomes. Batch primer design for predicted off-targets	Allows user to choose appropriate scoring method	Does not allow bulge mismatching
CHOP-CHOPv3[135] (2019)	Bowtie alignment for off-targets, user choice of on-target scoring, ranks gRNA by predicted efficiency and potential off-targets. Integrates with inDelphi[136] for repair outcome prediction, shows possible translation re-initiation sites	Can pre-filter guides on GC content, secondary structure, 5′ transcription requirements	Online tool only allows three mismatches

5.4.2. Machine Learning Models

Existing experimental off-target detection methods are unlikely to be widely used for detection due to the expertise required to perform these techniques and the prohibitive cost associated with routinely using them,[139] so machine learning techniques aim to use existing datasets to train models for off-target prediction. These models, outlined in **Table 4**, attempt to discern the underlying features responsible for on-/off-target cleavage in the training data and extrapolate this to unseen data.[140]

Feature learning on heterogeneous datasets can improve off-target predictions by accounting for more factors influencing than simple rule-based predictions. However, the style of feature learning greatly affects which features are more or less weighted. Taking only sequence features into account necessarily biases prediction algorithms against other influences, like energetic or epigenetic factors.[51] Providing long lists of sequence/genome/organism feature information allows more flexibility and such analyses indicate that sequence similarity may only account for about 30% of gRNA activity.[51] Furthermore features are not evenly spread across loci, for example, TUSCAN finds that only 12/63 features were shared across on-target sites tested, whilst the remainder were variable.[141] This flexibility of machine-learning is an advantage over "one rule fits all" off-target predictors.

Ironically, the limitations of experimental on-/off-target datasets that machine-learning methods attempt to curtail, are in fact the source of many of their problems. Insufficiently large datasets, imbalanced datasets, over-fitting, and data from heterogeneous experimental systems all place constraints on assembling quality training sets for machine learning models.[142]

For instance, the CRISTA algorithm[51] is criticized for inappropriate data balancing, and found to have a much lower predictive power than claimed.[142] The same authors found that the elevation off-target prediction model could be improved with two different methods of sampling the imbalanced dataset.[142]

Nonetheless, machine learning methods can evaluate the importance of a variety of factors that influence Cas9 cleavage. The prototype algorithm DeepCRISPR[143] uses automated feature annotation from ENCODE combined with feature learning models; aiming to provide a nuanced learning of on-/off-target activity across different sequence and cellular contexts. Such methods show promise, like the DeepSpCas9 trained on a dataset of over 12 000 lentivirally integrated target sequences,[66] however such methods require caution in interpretation and must be optimized and trained on relevant data sets. Although lacking off-target prediction, the on-target prediction from DeepSpCas9's convolutional neural network learning model claimed to be more generalizable—across datasets using U6 and T7 gRNA transcription—than any previously published method.[66]

5.4.3. Overview of gRNA Design Tools

From simple enumeration of possible genomic off-target sites to machine-learning based prediction of on-/off-target cleavage activity, there are dozens of available programs for use—but with important caveats. Benchmarking studies have found that the

Table 4. gRNA design tools with machine-learning off-target prediction.

Name	Method	Advantages	Limitations
WU-CRISPR[134] (2015)	Supervised support vector machine learning algorithm on best and worst performing gRNAs (top/bottom 20%) from Doench et al. (2014)[49] dataset, to determine characteristics of more active gRNAs	Accounts for gRNA secondary structure, translation termination motifs, GC content, and position	Rejects genome inputs with "N" Trained and tested on gRNAs under u6 promotor (limited generalizability[89])
GPP sgRNA designer[49] (2016)	Gradient-boosted regression tree trained on phenotypic output of gRNA screens with sequence features, for on-target scoring system named "rule set 2". Uses CasOFFfinder[128] to find off-targets, and experimentally derived mismatch-type cutting activity to create an aggregate cutting-frequency determination (CFD), to rank likely off-targets	Works well for gRNAs under u6 promotor	Trained on phenotype rather genotype data (limited generalizability[141])
CRISTA[51] (2017)	Supervised random forest regression algorithm with over 30 features accounting for attributes of genomic context, RNA/DNA thermodynamics, and sequence similarity	Lists "cleavage propensity", user can adjust for gRNA filtering	Poorly balanced training dataset predictive power is likely lower than claimed (Gao et al.[142])
Elevation[139] (2018)	(Azimuth[49] for on-target prediction) Supervised learning model that uses two-layers of boosted regression trees, where the first layer predicts off-target activity for single mismatches present, and in case of multiple mismatches, the second layer combines the multiple first layer predictions	Cloud-based results for human exome (GRCh38), for fast searching	Ignores bulge mismatches. Sub-optimal data balancing (Gao et al.[142])
DeepCRISPR[143] (2018)	Prototype model using a hybrid neural network, with a deep convolutional de-noising neural network for unsupervised feature learning of gRNA-targeted regions from ENCODE, and fully convolutional neural network to predict gRNA on-/off-target activity	Circos plot for off-target profile of gRNA	Trained on limited dataset, low generalizability[66]

consensus across tools for recommended gRNAs is generally low, but gRNAs recommended by multiple tools are more likely to be functional.[144]

Of central concern is the observation that most gRNA design tools work best on datasets that are similar to the experimental protocol of the training dataset, and thus generalize poorly.[89] This can be due to differences in gRNA transcription, cell-type, and method of quantifying Cas9 activity, which are not consistently shared across datasets.[145]

For example, experimental datasets using a U6 promoter/poly-T terminator for gRNA expression will favor gRNAs lacking TT motifs, and favor G/C over T at the 3′ end –as this favors efficient transcription termination and higher gRNA expression.[134] Unlike the position or number of mismatches between gRNA and target DNA, TT motifs do not generalize across different expression systems. Similarly, datasets in which the Cas9 ribonucleoprotein was delivered (rather than expressed) favor gRNAs with higher stability, as they persist longer in vivo.[133] This emphasizes the importance of choosing a gRNA design tool to match the experimental system being used.

In summary, organism or experimental system-specific tools are more likely to predict functional and specific gRNAs, so choosing the appropriate tool for a planned editing experiment is vital. Cross-validating gRNAs with more than one such tool is ideal, but options are limited outside common cell types and experimental systems.

6. Conclusion and Prospects

Although gene editing with Cas9 has been transformative to basic research in the life sciences, and translational work in horticulture, agriculture, industrial microorganisms, and medicine, it is important to recognize and work within the limitations of the technology.

It is established that Cas9 DSBs can result in large-scale genomic disruptions at on-target sites and unwanted editing products at off-target sites (see Section 4). Less well established is rigorous validation of Cas9 editing products. Standard genotyping is often insufficient to detect large on-target deletions and rearrangements, computational prediction is currently not accurate enough for targeted sequencing for off-targets, and experimental methods for unbiased off-target detection are impractical for routine use. Nonetheless, our understanding of on- and off-target effects has considerably improved, and wider community uptake in assessing the unwanted effects of Cas9 editing will yield useful data of Cas9 specificity across diverse applications and organisms.

Limitations notwithstanding, Cas9 editing offers a powerful framework for genetic engineering, and one that is improving constantly. The body of literature around the basic science of Cas9 and translational editing techniques is ever-increasing, and is trending toward higher efficiency, greater specificity, and diverse functionalities. Due to the unpredictable nature of

double-stranded break repair, there will always be a risk/benefit trade-off in editing with Cas9. However, with more thorough understanding of Cas9's specificity across different cell-types, and development of improved gene editing techniques, the risk of unwanted editing products can be mitigated.

In recognition of widespread use of Cas9 editing, and the limitations of the technology, we present a range of techniques available for detecting, predicting, and mitigating the on- and off-target effects of Cas9 editing.

Acknowledgements

The authors would like to acknowledge Arash Hajizadeh Dastjerdi, Jovita D'Silva, Aidan O'Brien, and Alexander McKay for fruitful discussions and critical reading of the manuscript. The graphical abstract and Figure 2 were created with BioRender.com. This work was partly supported from the National Health and Medical Research Council of Australia (APP1143008), the Australian Research Council (DP180101494), the National Natural Science Foundation of China (81772214), The Australian Rese the National Collaborative Research Infrastructure Strategy (NCRIS) via, Phenomics Australian. A.N. is supported from an Australian Government Research Training Program Scholarship.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

CRISPR/Cas9, double-stranded breaks, guide RNA design, high-throughput sequencing, machine learning, off-target effects, on-target effects

Received: March 5, 2020
Revised: May 12, 2020
Published online:

[1] R. J. Roberts, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 5905.
[2] M. Adli, *Nat. Commun.* **2018**, *9*, 1911.
[3] C. Anders, O. Niewoehner, A. Duerst, M. Jinek, *Nature* **2014**, *513*, 569.
[4] S. Shmakov, A. Smargon, D. Scott, D. Cox, N. Pyzocha, W. Yan, O. O. Abudayyeh, J. S. Gootenberg, K. S. Makarova, Y. I. Wolf, *Nat. Rev. Microbiol.* **2017**, *15*, 169.
[5] A. H. Dastjerdi, A. Newman, G. Burgio, *BioDrugs* **2019**, *33*, 503.
[6] L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, *Science* **2013**, *339*, 819.
[7] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, E. Charpentier, *Science* **2012**, *337*, 816.
[8] M. Jinek, A. East, A. Cheng, S. Lin, E. Ma, J. Doudna, *eLife* **2013**, *2*, e00471.
[9] P. Mali, L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville, G. M. Church, *Science* **2013**, *339*, 823.
[10] M. R. O'Connell, B. L. Oakes, S. H. Sternberg, A. East-Seletsky, M. Kaplan, J. A. Doudna, *Nature* **2014**, *516*, 263.
[11] B. Chen, L. A. Gilbert, B. A. Cimini, J. Schnitzbauer, W. Zhang, G.-W. Li, J. Park, E. H. Blackburn, J. S. Weissman, L. S. Qi, *Cell* **2013**, *155*, 1479.
[12] L. A. Gilbert, M. H. Larson, L. Morsut, Z. Liu, G. A. Brar, S. E. Torres, N. Stern-Ginossar, O. Brandman, E. H. Whitehead, J. A. Doudna, *Cell* **2013**, *154*, 442.
[13] A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, *Nature* **2016**, *533*, 420.
[14] R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, P. Horvath, *Science* **2007**, *315*, 1709.
[15] C. S. Nødvig, J. B. Nielsen, M. E. Kogle, U. H. Mortensen, *PLoS One* **2015**, *10*, e0133085.
[16] Q. Shan, Y. Wang, J. Li, Y. Zhang, K. Chen, Z. Liang, K. Zhang, J. Liu, J. J. Xi, J.-L. Qiu, *Nat. Biotechnol.* **2013**, *31*, 686.
[17] X. Ma, Q. Zhang, Q. Zhu, W. Liu, Y. Chen, R. Qiu, B. Wang, Z. Yang, H. Li, Y. Lin, *Mol. Plant* **2015**, *8*, 1274.
[18] R. J. Bevacqua, R. Fernandez-Martín, V. Savy, N. G. Canel, M. I. Gismondi, W. A. Kues, D. F. Carlson, S. Fahrenkrug, H. Niemann, O. A. Taboga, *Theriogenology* **2016**, *86*, 1886.
[19] J. Eyquem, J. Mansilla-Soto, T. Giavridis, S. J. van der Stegen, M. Hamieh, K. M. Cunanan, A. Odak, M. Gönen, M. Sadelain, *Nature* **2017**, *543*, 113.
[20] R. Kaminski, Y. Chen, T. Fischer, E. Tedaldi, A. Napoli, Y. Zhang, J. Karn, W. Hu, K. Khalili, *Sci. Rep.* **2016**, *6*, 28213.
[21] E. A. Stadtmauer, J. A. Fraietta, M. M. Davis, A. D. Cohen, K. L. Weber, E. Lancaster, P. A. Mangan, I. Kulikovskaya, M. Gupta, F. Chen, *Science* **2020**, *367*, eaba7365.
[22] L. Xu, J. Wang, Y. Liu, L. Xie, B. Su, D. Mou, L. Wang, T. Liu, X. Wang, B. Zhang, *N. Engl. J. Med.* **2019**, *381*, 1240.
[23] E. A. Stadtmauer, A. D. Cohen, K. Weber, S. F. Lacey, V. E. Gonzalez, J. J. Melenhorst, J. A. Fraietta, G. Plesa, J. Shea, T. Matlawski, A. Cervini, A. L. Gaymon, S. Desjardins, A. Lamontagne, J. Salas-Mckee, A. Fesnak, D. L. Siegel, B. L. Levine, J. K. Jadlowsky, R. M. Young, A. Chew, W. T. Hwang, E. O. Hexner, B. M. Carreno, C. L. Nobles, F. D. Bushman, K. R. Parker, Y. Qi, A. T. Satpathy, H. Y. Chang, *Blood* **2019**, suppl-1 *134*, 49.
[24] Y. Lu, J. Xue, T. Deng, X. Zhou, K. Yu, M. Huang, M. Liang, L. Deng, J. Huang, Z. Ding, *J Clin. Oncol.* **2018**, *36*, 3050.
[25] W. Jiang, D. Bikard, D. Cox, F. Zhang, L. A. Marraffini, *Nat. Biotechnol.* **2013**, *31*, 233.
[26] Y. Fu, J. A. Foden, C. Khayter, M. L. Maeder, D. Reyon, J. K. Joung, J. D. Sander, *Nat. Biotechnol.* **2013**, *31*, 822.
[27] T. J. Cradick, E. J. Fine, C. J. Antico, G. Bao, *Nucleic Acids Res.* **2013**, *41*, 9584.
[28] A. Hendel, E. J. Kildebeck, E. J. Fine, J. T. Clark, N. Punjya, V. Sebastiano, G. Bao, M. H. Porteus, *Cell Rep.* **2014**, *7*, 293.
[29] K. Boroviak, B. Fu, F. Yang, B. Doe, A. Bradley, *Sci. Rep.* **2017**, *7*, 12867.
[30] M. Kosicki, K. Tomberg, A. Bradley, *Nat. Biotechnol.* **2018**, *36*, 765.
[31] X. Shi, J. Shou, M. M. Mehryar, J. Li, L. Wang, M. Zhang, H. Huang, X. Sun, Q. Wu, *Cell Discov.* **2019**, *5*, 53.
[32] E. Haapaniemi, S. Botla, J. Persson, B. Schmierer, J. Taipale, *Nat. Med.* **2018**, *24*, 927.
[33] R. J. Ihry, K. A. Worringer, M. R. Salick, E. Frias, D. Ho, K. Theriault, S. Kommineni, J. Chen, M. Sondey, C. Ye, *Nat. Med.* **2018**, *24*, 939.
[34] Z. Zuo, J. Liu, *Sci. Rep.* **2016**, *6*, 37584.
[35] W. J. Cannan, D. S. Pederson, *J. Cell. Physiol.* **2016**, *231*, 3.
[36] C. D. Yeh, C. D. Richardson, J. E. Corn, *Nat. Cell Biol.* **2019**, *21*, 1468.
[37] M. Thomas, G. Burgio, D. J. Adams, V. Iyer, *PLoS Genet.* **2019**, *15*, e1007994.
[38] G. Cullot, J. Boutin, J. Toutain, F. Prat, P. Pennamen, C. Rooryck, M. Teichmann, E. Rousseau, I. Lamrissi-Garcia, V. Guyonnet-Duperat, *Nat. Commun.* **2019**, *10*, 1136.
[39] In this paper, editing efficiency refers to the proportion of edited vs non-edited target sites i.e., a Cas9/gRNA complex that edits a target site in 9/10 experiments is highly efficient. If a Cas9/gRNA complex cleaves only the target site, it is highly specific, as opposed to one

that cleaves the on-target site and many off-target sites. The ultimate goal of gene editing is a technique that is highly efficient and highly specific.

[40] K. Kraft, S. Geuer, A. J. Will, W. L. Chan, C. Paliou, M. Borschiwer, I. Harabula, L. Wittler, M. Franke, D. M. Ibrahim, *Cell Rep.* **2015**, *10*, 833.

[41] F. Adikusuma, S. Piltz, M. A. Corbett, M. Turvey, S. R. McColl, K. J. Helbig, M. R. Beard, J. Hughes, R. T. Pomerantz, P. Q. Thomas, *Nature* **2018**, *560*, E8.

[42] F. Adikusuma, N. Williams, F. Grutzner, J. Hughes, P. Thomas, *Mol. Ther.* **2017**, *25*, 1736.

[43] E. Zuo, X. Huo, X. Yao, X. Hu, Y. Sun, J. Yin, B. He, X. Wang, L. Shi, J. Ping, *Genome Biol.* **2017**, *18*, 224.

[44] K. Wang, D. de la Torre, W. E. Robertson, J. W. Chin, *Science* **2019**, *365*, 922.

[45] S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, J. A. Doudna, *Nature* **2014**, *507*, 62.

[46] S. Q. Tsai, Z. Zheng, N. T. Nguyen, M. Liebers, V. V. Topkar, V. Thapar, N. Wyvekens, C. Khayter, A. J. Iafrate, L. P. Le, *Nat. Biotechnol.* **2015**, *33*, 187.

[47] M. Jinek, F. Jiang, D. W. Taylor, S. H. Sternberg, E. Kaya, E. Ma, C. Anders, M. Hauer, K. Zhou, S. Lin, *Science* **2014**, *343*, 1247997.

[48] T. Zheng, Y. Hou, P. Zhang, Z. Zhang, Y. Xu, L. Zhang, L. Niu, Y. Yang, D. Liang, F. Yi, *Sci. Rep.* **2017**, *7*, 40638.

[49] J. G. Doench, N. Fusi, M. Sullender, M. Hegde, E. W. Vaimberg, K. F. Donovan, I. Smith, Z. Tothova, C. Wilen, R. Orchard, *Nat. Biotechnol.* **2016**, *34*, 184.

[50] Y. Lin, T. J. Cradick, M. T. Brown, H. Deshmukh, P. Ranjan, N. Sarode, B. M. Wile, P. M. Vertino, F. J. Stewart, G. Bao, *Nucleic Acids Res.* **2014**, *42*, 7473.

[51] S. Abadi, W. X. Yan, D. Amar, I. Mayrose, *PLoS Comput. Biol.* **2017**, *13*, e1005807.

[52] G. Palermo, J. S. Chen, C. G. Ricci, I. Rivalta, M. Jinek, V. S. Batista, J. A. Doudna, J. A. McCammon, *Q. Rev. Biophys.* **2018**, *51*, e9.

[53] M. A. Horlbeck, L. B. Witkowsky, B. Guglielmi, J. M. Replogle, L. A. Gilbert, J. E. Villalta, S. E. Torigoe, R. Tjian, J. S. Weissman, *eLife* **2016**, *5*, e12677.

[54] R. S. Isaac, F. Jiang, J. A. Doudna, W. A. Lim, G. J. Narlikar, R. Almeida, *eLife* **2016**, *5*, e13450.

[55] R. Clarke, R. Heler, M. S. MacDougall, N. C. Yeo, A. Chavez, M. Regan, L. Hanakahi, G. M. Church, L. A. Marraffini, B. J. Merrill, *Mol. Cell* **2018**, *71*, 42.

[56] A. S. Wang, L. Chen, R. A. Wu, C. D. Richardson, B. G. Gowen, K. R. Kazane, J. T. Vu, S. K. Wyman, J. Shin, J. C. Walter, *bioRxiv, Cell Biol.* **2019**, 705657.

[57] M. D. Newton, B. J. Taylor, R. P. Driessen, L. Roos, N. Cvetesic, S. Allyjaun, B. Lenhard, M. E. Cuomo, D. S. Rueda, *Nat. Struct. Mol. Biol.* **2019**, *26*, 185.

[58] I. E. Ivanov, A. V. Wright, J. C. Cofsky, K. D. P. Aris, J. A. Doudna, Z. Bryant, *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 5853.

[59] A. Xiao, Z. Wang, Y. Hu, Y. Wu, Z. Luo, Z. Yang, Y. Zu, W. Li, P. Huang, X. Tong, *Nucleic Acids Res.* **2013**, *41*, e141.

[60] H. Y. Shin, C. Wang, H. K. Lee, K. H. Yoo, X. Zeng, T. Kuhns, C. M. Yang, T. Mohr, C. Liu, L. Hennighausen, *Nat. Commun.* **2017**, *8*, 15464.

[61] J. Hu, R. M. Meyers, J. Dong, R. A. Panchakshari, F. W. Alt, R. L. Frock, *Nat. Protoc.* **2016**, *11*, 853.

[62] D. D. Owens, A. Caulder, V. Frontera, J. R. Harman, A. J. Allan, A. Bucakci, L. Greder, G. F. Codner, P. Hublitz, P. J. McHugh, *Nucleic Acids Res.* **2019**, *47*, 7402.

[63] M. E. Hurles, E. T. Dermitzakis, C. Tyler-Smith, *Trends Genet.* **2008**, *24*, 238.

[64] J. Zischewski, R. Fischer, L. Bortesi, *Biotechnol. Adv.* **2017**, *35*, 95.

[65] R. T. Leenay, A. Aghazadeh, J. Hiatt, D. Tse, T. L. Roth, R. Apathy, E. Shifrut, J. F. Hultquist, N. Krogan, Z. Wu, *Nat. Biotechnol.* **2019**, *37*, 1034.

[66] H. K. Kim, Y. Kim, S. Lee, S. Min, J. Y. Bae, J. W. Choi, J. Park, D. Jung, S. Yoon, H. H. Kim, *Sci. Adv.* **2019**, *5*, eaax9249.

[67] E. K. Brinkman, T. Chen, M. Amendola, B. van Steensel, *Nucleic Acids Res.* **2014**, *42*, e168.

[68] J. P. Connelly, S. M. Pruett-Miller, *Sci. Rep.* **2019**, *9*, 4194.

[69] K. Clement, H. Rees, M. C. Canver, J. M. Gehrke, R. Farouni, J. Y. Hsu, M. A. Cole, D. R. Liu, J. K. Joung, D. E. Bauer, *Nat. Biotechnol.* **2019**, *37*, 224.

[70] P. D. Hebert, T. W. Braukmann, S. W. Prosser, S. Ratnasingham, J. R. DeWaard, N. V. Ivanova, D. H. Janzen, W. Hallwachs, S. Naik, J. E. Sones, *BMC Genomics* **2018**, *19*, 219.

[71] I. Höijer, J. Johansson, S. Gudmundsson, C.-S. Chin, I. Bunikis, S. Häggqvist, A. Emmanouilidou, M. Wilbe, M. den Hoed, M.-L. Bondeson, *bioRxiv, Genomics* **2020**, 940486.

[72] E. Rayner, M.-A. Durin, R. Thomas, D. Moralli, S. M. O'Cathail, I. Tomlinson, C. M. Green, A. Lewis, *CRISPR J.* **2019**, *2*, 406.

[73] M. Paulis, A. Castelli, M. Lizier, L. Susani, F. Lucchini, A. Villa, P. Vezzoni, *Sci. Rep.* **2015**, *5*, 12327.

[74] R. Gabriel, A. Lombardo, A. Arens, J. C. Miller, P. Genovese, C. Kaeppel, A. Nowrouzi, C. C. Bartholomae, J. Wang, G. Friedman, *Nat. Biotechnol.* **2011**, *29*, 816.

[75] H. O'Geen, I. M. Henry, M. S. Bhakta, J. F. Meckler, D. J. Segal, *Nucleic Acids Res.* **2015**, *43*, 3389.

[76] P. Cameron, C. K. Fuller, P. D. Donohoue, B. N. Jones, M. S. Thompson, M. M. Carter, S. Gradia, B. Vidal, E. Garner, E. M. Slorach, *Nat. Methods* **2017**, *14*, 600.

[77] P. Akcakaya, M. L. Bobbin, J. A. Guo, J. Malagon-Lopez, K. Clement, S. P. Garcia, M. D. Fellows, M. J. Porritt, M. A. Firth, A. Carreras, *Nature* **2018**, *561*, 416.

[78] S. Q. Tsai, N. T. Nguyen, J. Malagon-Lopez, V. V. Topkar, M. J. Aryee, J. K. Joung, *Nat. Methods* **2017**, *14*, 607.

[79] D. Kim, J.-S. Kim, *Genome Res.* **2018**, *28*, 1894.

[80] Y. Zhu, A. Biernacka, B. Pardo, N. Dojer, R. Forey, M. Skrzypczak, B. Fongang, J. Nde, R. Yousefi, P. Pasero, *Nat. Commun.* **2019**, *10*, 2313.

[81] W. X. Yan, R. Mirzazadeh, S. Garnerone, D. Scott, M. W. Schneider, T. Kallas, J. Custodio, E. Wernersson, Y. Li, L. Gao, *Nat. Commun.* **2017**, *8*, 15058.

[82] N. Crosetto, A. Mitra, M. J. Silva, M. Bienko, N. Dojer, Q. Wang, E. Karaca, R. Chiarle, M. Skrzypczak, K. Ginalski, *Nat. Methods* **2013**, *10*, 361.

[83] B. Wienert, S. K. Wyman, C. D. Richardson, C. D. Yeh, P. Akcakaya, M. J. Porritt, M. Morlock, J. T. Vu, K. R. Kazane, H. L. Watry, *Science* **2019**, *364*, 286.

[84] V. Iyer, K. Boroviak, M. Thomas, B. Doe, L. Riva, E. Ryder, D. J. Adams, *PLoS Genet.* **2018**, *14*, e1007503.

[85] S. A. Verkuijl, M. G. Rots, *Curr. Opin. Biotechnol.* **2019**, *55*, 68.

[86] R. T. Walton, K. A. Christie, M. N. Whittaker, B. P. Kleinstiver, *Science* **2020**, *368*, 290.

[87] D. Kim, S. Bae, J. Park, E. Kim, S. Kim, H. R. Yu, J. Hwang, J.-I. Kim, J.-S. Kim, *Nat. Methods* **2015**, *12*, 237.

[88] Y. Li, L. M. Almassalha, J. E. Chandler, X. Zhou, Y. E. Stypula-Cyrus, K. A. Hujsak, E. W. Roth, R. Bleher, H. Subramanian, I. Szleifer, *Exp. Cell Res.* **2017**, *358*, 253.

[89] M. Haeussler, K. Schönig, H. Eckert, A. Eschstruth, J. Mianné, J.-B. Renaud, S. Schneider-Maunoury, A. Shkumatava, L. Teboul, J. Kent, *Genome Biol.* **2016**, *17*, 148.

[90] H. Mou, J. L. Smith, L. Peng, H. Yin, J. Moore, X.-O. Zhang, C.-Q. Song, A. Sheel, Q. Wu, D. M. Ozata, *Genome Biol.* **2017**, *18*, 108.

[91] A. H. Smits, F. Ziebell, G. Joberty, N. Zinn, W. F. Mueller, S. Clauder-Münster, D. Eberhard, M. F. Savitski, P. Grandi, P. Jakob, *Nat. Methods* **2019**, *16*, 1087.
[92] D. Carroll, *Genome Biol.* **2019**, *20*, 114.
[93] S. Kim, D. Kim, S. W. Cho, J. Kim, J.-S. Kim, *Genome Res.* **2014**, *24*, 1012.
[94] J. B. Miller, S. Zhang, P. Kos, H. Xiong, K. Zhou, S. S. Perelman, H. Zhu, D. J. Siegwart, *Angew. Chem., Int. Ed.* **2017**, *56*, 1059.
[95] Z. Tu, W. Yang, S. Yan, A. Yin, J. Gao, X. Liu, Y. Zheng, J. Zheng, Z. Li, S. Yang, *Sci. Rep.* **2017**, *7*, 42081.
[96] W. Mu, Y. Zhang, X. Xue, L. Liu, X. Wei, H. Wang, *Protein Cell* **2019**, *10*, 223.
[97] S. Nahar, P. Sehgal, M. Azhar, M. Rai, A. Singh, S. Sivasubbu, D. Chakraborty, S. Maiti, *Chem. Commun.* **2018**, *54*, 2377.
[98] T. Gutschner, M. Haemmerle, G. Genovese, G. F. Draetta, L. Chin, *Cell Rep.* **2016**, *14*, 1555.
[99] M. D. Hoffmann, S. Aschenbrenner, S. Grosse, K. Rapti, C. Domenger, J. Fakhiri, M. Mastel, K. Börner, R. Eils, D. Grimm, *Nucleic Acids Res.* **2019**, *47*, e75.
[100] H. Meltzer, E. Marom, I. Alyagor, O. Mayseless, V. Berkun, N. Segal-Gilboa, T. Unger, D. Luginbuhl, O. Schuldiner, *Nat. Commun.* **2019**, *10*, 2113.
[101] J. Carlson-Stevermer, A. A. Abdeen, L. Kohlenberg, M. Goedland, K. Molugu, M. Lou, K. Saha, *Nat. Commun.* **2017**, *8*, 1711.
[102] K. Kundert, J. E. Lucas, K. E. Watters, C. Fellmann, A. H. Ng, B. M. Heineike, C. M. Fitzsimmons, B. L. Oakes, J. Qu, N. Prasad, *Nat. Commun.* **2019**, *10*, 2127.
[103] K. I. Liu, M. N. B. Ramli, C. W. A. Woo, Y. Wang, T. Zhao, X. Zhang, G. R. D. Yim, B. Y. Chong, A. Gowher, M. Z. H. Chua, *Nat. Chem. Biol.* **2016**, *12*, 980.
[104] B. Maji, C. L. Moore, B. Zetsche, S. E. Volz, F. Zhang, M. D. Shoulders, A. Choudhary, *Nat. Chem. Biol.* **2017**, *13*, 9.
[105] Y. Nihongaki, F. Kawano, T. Nakajima, M. Sato, *Nat. Biotechnol.* **2015**, *33*, 755.
[106] K. Chylinski, M. Hubmann, R. E. Hanna, C. Yanchus, G. Michlits, E. C. Uijttewaal, J. Doench, D. Schramek, U. Elling, *Nat. Commun.* **2019**, *10*, 5454.
[107] J. Bondy-Denomy, A. Pawluk, K. L. Maxwell, A. R. Davidson, *Nature* **2013**, *493*, 429.
[108] B. J. Rauch, M. R. Silvis, J. F. Hultquist, C. S. Waters, M. J. McGregor, N. J. Krogan, J. Bondy-Denomy, *Cell* **2017**, *168*, 150.
[109] D. Dong, M. Guo, S. Wang, Y. Zhu, S. Wang, Z. Xiong, J. Yang, Z. Xu, Z. Huang, *Nature* **2017**, *546*, 436.
[110] M. Nakamura, P. Srinivasan, M. Chavez, M. A. Carter, A. A. Dominguez, M. La Russa, M. B. Lau, T. R. Abbott, X. Xu, D. Zhao, *Nat. Commun.* **2019**, *10*, 194.
[111] Y. Fu, J. D. Sander, D. Reyon, V. M. Cascio, J. K. Joung, *Nat. Biotechnol.* **2014**, *32*, 279.
[112] D. D. Kocak, E. A. Josephs, V. Bhandarkar, S. S. Adkar, J. B. Kwon, C. A. Gersbach, *Nat. Biotechnol.* **2019**, *37*, 657.
[113] D. E. Ryan, D. Taussig, I. Steinfeld, S. M. Phadnis, B. D. Lunstad, M. Singh, X. Vuong, K. D. Okochi, R. McCaffrey, M. Olesiak, *Nucleic Acids Res.* **2018**, *46*, 792.
[114] H. Yin, C.-Q. Song, S. Suresh, S.-Y. Kwan, Q. Wu, S. Walsh, J. Ding, R. L. Bogorad, L. J. Zhu, S. A. Wolfe, *Nat. Chem. Biol.* **2018**, *14*, 311.
[115] B. P. Kleinstiver, V. Pattanayak, M. S. Prew, S. Q. Tsai, N. T. Nguyen, Z. Zheng, J. K. Joung, *Nature* **2016**, *529*, 490.
[116] I. M. Slaymaker, L. Gao, B. Zetsche, D. A. Scott, W. X. Yan, F. Zhang, *Science* **2016**, *351*, 84.
[117] J. S. Chen, Y. S. Dagdas, B. P. Kleinstiver, M. M. Welch, A. A. Sousa, L. B. Harrington, S. H. Sternberg, J. K. Joung, A. Yildiz, J. A. Doudna, *Nature* **2017**, *550*, 407.
[118] C. A. Vakulskas, D. P. Dever, G. R. Rettig, R. Turk, A. M. Jacobi, M. A. Collingwood, N. M. Bode, M. S. McNeill, S. Yan, J. Camarena, *Nat. Med.* **2018**, *24*, 1216.
[119] P. I. Kulcsár, A. Tálas, K. Huszár, Z. Ligeti, E. Tóth, N. Weinhardt, E. Fodor, E. Welker, *Genome Biol.* **2017**, *18*, 190.
[120] F. A. Ran, P. D. Hsu, C.-Y. Lin, J. S. Gootenberg, S. Konermann, A. E. Trevino, D. A. Scott, A. Inoue, S. Matoba, Y. Zhang, *Cell* **2013**, *154*, 1380.
[121] S. W. Cho, S. Kim, Y. Kim, J. Kweon, H. S. Kim, S. Bae, J.-S. Kim, *Genome Res.* **2014**, *24*, 132.
[122] R. Gopalappa, B. Suresh, S. Ramakrishna, H. Kim, *Nucleic Acids Res.* **2018**, *46*, e71.
[123] S. Bae, J. Kweon, H. S. Kim, J.-S. Kim, *Nat. Methods* **2014**, *11*, 705.
[124] H. Zhou, M. Zhou, D. Li, J. Manthey, E. Lioutikova, H. Wang, X. Zeng, *BMC Genomics* **2017**, *18*, 826.
[125] A. O'Brien, T. L. Bailey, *Bioinformatics* **2014**, *30*, 2673.
[126] M. Stemmer, T. Thumberger, M. del Sol Keyer, J. Wittbrodt, J. L. Mateo, *PLoS One* **2015**, *10*, e0124633.
[127] J. Park, S. Bae, J.-S. Kim, *Bioinformatics* **2015**, *31*, 1093.
[128] S. Bae, J. Park, J.-S. Kim, *Bioinformatics* **2014**, *30*, 1473.
[129] A. R. Perez, Y. Pritykin, J. A. Vidigal, S. Chhangawala, L. Zamparo, C. S. Leslie, A. Ventura, *Nat. Biotechnol.* **2017**, *35*, 347.
[130] M. C. Canver, M. Haeussler, D. E. Bauer, S. H. Orkin, N. E. Sanjana, O. Shalem, G.-C. Yuan, F. Zhang, J.-P. Concordet, L. Pinello, *Nat. Protoc.* **2018**, *13*, 946.
[131] P. D. Hsu, D. A. Scott, J. A. Weinstein, F. A. Ran, S. Konermann, V. Agarwala, Y. Li, E. J. Fine, X. Wu, O. Shalem, *Nat. Biotechnol.* **2013**, *31*, 827.
[132] R. Singh, C. Kuscu, A. Quinlan, Y. Qi, M. Adli, *Nucleic Acids Res.* **2015**, *43*, e118.
[133] M. A. Moreno-Mateos, C. E. Vejnar, J.-D. Beaudoin, J. P. Fernandez, E. K. Mis, M. K. Khokha, A. J. Giraldez, *Nat. Methods* **2015**, *12*, 982.
[134] N. Wong, W. Liu, X. Wang, *Genome Biol.* **2015**, *16*, 218.
[135] K. Labun, T. G. Montague, M. Krause, Y. N. Torres Cleuren, H. Tjeldnes, E. Valen, *Nucleic Acids Res.* **2019**, *47*, W171.
[136] M. W. Shen, M. Arbab, J. Y. Hsu, D. Worstell, S. J. Culbertson, O. Krabbe, C. A. Cassa, D. R. Liu, D. K. Gifford, R. I. Sherwood, *Nature* **2018**, *563*, 646.
[137] H. Xu, T. Xiao, C.-H. Chen, W. Li, C. A. Meyer, Q. Wu, D. Wu, L. Cong, F. Zhang, J. S. Liu, *Genome Res.* **2015**, *25*, 1147.
[138] B. Farboud, B. J. Meyer, *Genetics* **2015**, *199*, 959.
[139] J. Listgarten, M. Weinstein, B. P. Kleinstiver, A. A. Sousa, J. K. Joung, J. Crawford, K. Gao, L. Hoang, M. Elibol, J. G. Doench, *Nat. Biomed. Eng.* **2018**, *2*, 38.
[140] A. R. O'Brien, G. Burgio, D. C. Bauer, *Briefings Bioinf.* **2020**, bbz145.
[141] L. O. Wilson, D. Reti, A. R. O'Brien, R. A. Dunne, D. C. Bauer, *CRISPR J.* **2018**, *1*, 182.
[142] Y. Gao, G. Chuai, W. Yu, S. Qu, Q. Liu, *Briefings Bioinf.* **2019**, bbz069.
[143] G. Chuai, H. Ma, J. Yan, M. Chen, N. Hong, D. Xue, C. Zhou, C. Zhu, K. Chen, B. Duan, *Genome Biol.* **2018**, *19*, 80.
[144] J. Bradford, D. Perrin, *BMC Genomics* **2019**, *20*, 931.
[145] L. O. Wilson, A. R. O'Brien, D. C. Bauer, *Front. Pharmacol.* **2018**, *9*, 749.