Fine-tuning next-generation genome editing tools

Chidananda Nagamangala Kanchiswamy1*, Massimo Maffei2, Mickael Malnoy1, Riccardo Velasco1, Jin-Soo Kim3,4

1 Research and Innovation Centre, Genomics and Biology of Fruit Crop Department, Fondazione Edmund Mach (FEM), Via Mach 1, 38010 San Michele all’Adige (TN), Italy;
2 Department of Life Sciences and Systems Biology, Innovation Centre, University of Turin, Via Quarello 15/A, 10135 Turin, Italy;
3 Center for Genome Engineering, Institute for Basic Science, Gwanak-ro 1, Gwanak-gu, Seoul 151-747, South Korea.
4 Department of Chemistry, Seoul National University, Gwanak-ro 1, Gwanak-gu, Seoul 151-747, South Korea.

The availability of genome sequences of numerous organisms and the revolution brought about by genome editing (GE) tools (e.g., ZFNs, TALENs, and CRISPR/Cas9 or RGENs) has provided a breakthrough in introducing targeted genetic changes both to explore emergent phenotypes and to introduce new functionalities. However, the wider application of these tools in biology, agriculture, medicine and biotechnology is limited by off-target mutation effects. In this review, we compare available methods for detecting, measuring and analyzing off-target mutations. Furthermore, we particularly focus on CRISPR/Cas9 regarding various methods, tweaks and software tools available to nullify off-target effects.

Key words: ZFNs; TALENs; CRISPR/Cas9; genome editing; on-target; off-target; target specificity.
Sequence specific genome editing

A recent revolution in sequence-specific programmable nucleases has led to the development of zinc-finger nucleases (ZFNs) (see Glossary), transcription activator-like effector nuclease (TALENs) and RNA-guided engineered nuclease (RGENs) derived from type II clustered, regularly interspaced, short palindromic repeats (CRISPR/Cas9). These nucleases have emerged as exciting tools to edit genes of interest with unprecedented control and accuracy in eukaryotic cells, paving the way for next-generation biotechnology. These GE tools cleave targeted chromosomal DNA by producing site-specific DNA double strand breaks (DSBs). The host endogenous DNA repair mechanism repairs DSBs via homologous recombination (HR) and non-homologous end joining (NHEJ). The discovery of the most recent GE tools, particularly the CRISPR system, revolutionized genome engineering applications due to ease with which they can be adopted to target specific gene sequences. The basic details of different GE tools are listed in Glossary and Figure 1.

GE tools have been highly appreciated for their numerous applications in biology, medicine, biotechnology and agriculture. These tools have attracted considerable attention from a broad range of research topics for their wider application, contributing to their selection as method of the year by Nature Methods in 2011 [1] and as a breakthrough of the year by Science in 2015 [2]. However, GE tools are limited by off-target mutations and each GE tools has its own pros and cons. In this review, we compare current GE tools for their off-target effects. Furthermore, we emphasize the best suitable GE technology, methods, tweaks and available software aimed to nullify off-target mutations.

Specificity of GE tools
ZFNs consist of a nuclease domain derived from Fok1, a type of IIS restriction enzyme, and a DNA-binding domain [3]. These binding domains can be engineered to target specific DNA sequences. The Fok1 nuclease domain must dimerize to cleave DNA [4]; these Fok1s function as pairs contributing their high specificities. ZFNs recognize 18- to 36-bp DNA sequences; statistically, they form unique sites in many eukaryotic genome sizes.

Compared to TALEN and CRISPR/Cas9, ZFNs are more expensive and laborious to design, and they use preferentially guanine-rich repeat (GNN) sequences, such as 5’-GNGNGNGNN-3’ which occurs rarely in most of the target sequences, thus limiting targetable sites [5, 6]. The use of ZFNs originated from publically available sources often causes cytotoxicity due to off-target effects [6] (Table-1).

TALEN

Similar to ZFNs, the second generation programmable nucleases, TALENs, consist of a nuclease domain derived from Fok1, but they have a distinct DNA-binding domain and employ transcription activator-like (TAL) effectors derived from the plant pathogen Xanthomonas sp. to cleave targeted DNA sequences[7, 8].

TALENs recognize 30- to 40-bp DNA sequences, and they can be designed to target almost any DNA sequence, which represents a significant advantage over ZFNs and CRISPR/Cas9. TAL effector modules recognize single bases, whereas zinc fingers recognize 3-bp sub-sites, thus minimizing context-dependent DNA recognition and constituting a key advantage over ZFNs. Four different modules, each specific to one of the four bases, are used to construct TALENs. However, a TAL effector array often consists of 20 modules: it is time-consuming and laborious
to construct plasmids that encode TALENs. Although TALENs are not considered to be cytotoxic, they can induce off-target mutations like other GE tools [9]. However, off-target effects can be mitigated by designing unique target sequences that differ by at least 7 nucleotides from any other site in the human genome [10]. Researchers can also utilize a web-based resource (www.talenlibrary.net) to identify such unique sequences in the human genome.

CRISPR/Cas9

The third generation programmable nuclease, Cas9, is an RNA-guided DNA endonuclease that targets foreign DNA for destruction as part of a bacterial adaptive immune system mediated by CRISPR [11]. The specificity of CRISPR/Cas9, derived from S. pyogenes bacteria, depends on gRNA, which hybridizes with 20-bp target DNA sequences, and Cas9, which recognizes 5’-NGG-3’ sequences known as protospacer adjacent motifs (PAMs). CRISPR/Cas9 from other species recognize different PAMs, and their gRNAs are variable in size. Unlike ZFNs and TALENs, CRISPR/Cas9 are scalable and affordable, and the past two years of research on CRISPR/Cas9 has been revolutionary in genome engineering. However, CRISPR/Cas9 can induce off-target mutations [11-14] and off-target chromosomal rearrangements [11], raising concerns for their wider application in medicine, agriculture and other biological sciences [12-15] (Table-1).

Importance of target specificity

GE tools lack target specificity; that is, they are able to target and bind the sequences in the genome that are similar but not identical, thus inducing undesirable genome modifications. Target specificity is a critical point for all GE tools for their broader application in biology, medicine and agriculture. GE tools can cut on-target sites efficiently, inducing site-specific DSBs in the genome, but they can also induce off-target mutations at sites homologous to on-target sites. Zinc finger
and TAL effector arrays can bind to highly homologous sites, resulting in on-target and off-target mutations, whereas both Cas9 and gRNAs can contribute to CRISPR/Cas9 off-target effects [5, 16, 17].

Off-target mutations may lead to cytotoxicity, apoptosis, and gross chromosomal rearrangements such as inversions, deletions, and translocations [16, 18-20]. Major concerns of off-target mutations have been observed in medical and clinical studies (Box - 1).

One such example is a ZFN pair targeted to the C-C chemokine receptor 5 (CCR5) gene that encodes a co-receptor of human immunodeficiency virus (HIV) [21]. This ZFN pair also cleaves a highly homologous site in CCR2 gene, leading to ~15-kbp chromosomal deletions, duplications and inversions of the intervening DNA segment in human cells [18, 19]. Nevertheless, off-target mutations in the CCR2 gene do not cause adverse side effects in patients with HIV infection. Another critical point to consider is that chromosomal re-arrangements are one of the hallmarks of cancer, which may activate oncogenes. Hence, off-target mutations should be monitored carefully to avoid such incidences.

**Assessing nuclease target specificities**

Several approaches have been developed to identify off-target sites of GE tools.

**SELEX**

The systematic evolution of ligands by exponential enrichment (SELEX) has been used to predict the sequences that GE tools prefer to bind [22]. Target DNA sequences in a pool of randomized oligonucleotide duplexes are identified after alternating cycles of ligand selection and amplification.
SELEX provides unbiased results of all of the potential off-target sites for a given GE tool, but most of the results obtained are based on experimental conditions. This in vitro technique does not consider several important factors such as chromatin structures and locus accessibility because semi randomized library oligodeoxynucleotide libraries are exposed to nucleases to identify the sequences that can be cleaved in vitro.

*In vivo methods*

Integrase-deficient lentiviruses (IDLVs) or adeno-associated viruses (AAVs) integrate at the sites of DSBs, which can be mapped to quantify off-target sites [23]. On the other hand, chromatin immunoprecipitation coupled with deep sequencing (Chip-Seq) can be used to track CRISPR/Cas9 and map the binding sequences [24, 25]. In-vivo methods account for the chromatin structures and locus accessibility. However, IDLV capture is not sensitive enough to capture low-frequency off-target sites. Chip-Seq using catalytically inactive or dead Cas9 (dCas) is limited by the fact that the DNA binding and cleavage events are uncoupled. Thus, Chip-seq fails to capture *bona fide* off-target sites, while producing many false positive sites [26, 27].

*In silico methods*

In silico methods are based on sequence homology rather than experimental data. Most of these programs list potential off-target sites with 3 or fewer mismatches [28]. These methods can be employed to synthesize nucleases on-target specific sequences while nullifying any off-target sites. Example of such *in silico* algorithm-based platforms are PROGNOS [29], which can be used for off-target prediction for ZFNs and TALENs; CRISPR design tools (http://crispr.mit.edu) for the CRISPR/Cas9 system; and CHOPCHOP (http://chopchop.re.fas.harvard.edu), an algorithm suitable for both CRISPR/Cas9 and TALEN off-target prediction. Although we have several in-
vitro, in-vivo and in silico methods to analyze on- and off- target specificity, we still lack a comprehensive, unbiased, genome-wide method to identify on- and off- target sites created by GE tools. In the following section, we review recent progress made on unbiased genome-wide profiling of nuclease cleavage sites with special focus on the CRISPR/Cas9 system.

**Unbiased genome-wide profiling of nuclease cleavage sites including CRISPR/Cas9**

We and several other groups have used whole genome/exome sequencing (WGS/WES) to analyze the on- and off-target mutations in single cell-derived clones [Cho et al. Genome Res. 2014; Kim et al. Nature Methods 2015; Smith et al. Cell Stem Cell 15, 12 (2014); Veres et al. Cell Stem Cell 15, 27 (2014)] or animals [Lyer et al. Nature Methods 12, 479 (2015)] and reported that off-target mutations are rarely induced by Cas9 or other GE tools. However, WGS are not sensitive enough to detect indels in a bulk population of cells. This sensitivity matters especially when GE tools are used for gene therapy or clinical studies, where millions of cells are treated with a nuclease. If one single cell has an oncogenic off-target mutation, it may lead to cancer. In order to address these issues, we and others have developed various methods for identifying genome-wide off-target sites in a bulk population of cells. Four different methods have been recently reported for unbiased genome wide comprehensive profiling of on- and off- target sites of CRISPR/Cas9 platform in a bulk population of cells (Figure-2).

**Genome-wide, unbiased identification of DSBs enabled by sequencing (GUIDE-seq)**

This method represents an improvement over IDLV capture. Blunt-ended, double stranded phosphorothiate oligodeoxynucleotides (dsODNs) are captured at on- and off- target DNA cleavage sites in cells. These dsODN integration sites are mapped in the genome via PCR amplification followed by deep sequencing [30].
High-throughput genomic translocation sequencing (HTGTS)

HTGTS exploits translocations that are induced in cells by erroneous ligations of on- and off-target sites in the genome. It uses on-target DSBs as a ‘bait’ to catch ‘prey’ sequences that are translocated to the on-target site. HTGTS is used to determine the prey sequences that corresponds to off-target sites [31]. Unlike other methods, HTGTS requires two concurrent DSBs, rather than one DSB, in a single cell.

Breaks labelling, enrichments on streptavidin and next-generation sequencing (BLESS)

BLESS is based on the principle of labelling DSBs present in the fixed cells using biotinylated oligonucleotides, which are then enriched and subjected to deep sequencing [32, 33]. BLESS provides a snapshot of DSBs at the time of cell fixation, resulting in poor sensitivity.

Digested genome sequencing (Digenome-seq)

Digenome-seq identifies off-target sites using nuclease digested genomic DNA (digenome) which is subjected to whole genome sequencing [34, 35]. In-vitro digestion of genomic DNA with Cas9 or other nucleases yields sequence reads with the same 5’ ends at cleavage sites, which can be computationally identified using WGS data.

Comparisons of genome-wide off-target profiling methods

These methods can be classified based on whether DNA is cleaved in cells or in vitro and whether DSBs are captured in cells or in vitro (Table-2). Both GUIDE-seq and HTGTS are cell-based methods: DNA is cleaved in cells and DSBs are captured in cells. In contrast, Digenome-seq is an in vitro method using cell-free genomic DNA: DSB sites are identified computationally using
WGS data. BLESS is a method in between: DNA is cleaved in vivo but DSBs are captured in vitro after cell fixation.

Cell-based methods are advantageous over Digenome-seq in that off-target DNA cleavage sites are identified in cells of interest under given experimental conditions. Off-target sites identified in one cell type may be different from those in other cell type, owing to the discrepancy in chromatin state or in nuclease expression levels. Because cell-free, chromatin-free genomic DNA is used, Digenome-seq cannot identify cell-specific off-target sites. In fact, Digenome-seq is more comprehensive than other methods, identifying many additional off-target sites [35].

GUIDE-seq and Digenome-seq are highly sensitive, often capturing off-target sites with indel frequencies below 0.1%. BLESS is not sensitive because it provides a snapshot of DSBs at the time of cell fixation. HTGTS is limited by rare events of two concurrent DSBs, rather than one DSB, in a cell. GUIDE-seq is most quantitative: there is a good correlation between the numbers of captured sequence reads and mutations frequencies [30]. HTGTS is unlikely to be quantitative because translocation efficiencies are highly variable upon DSB sites. For example, intra-chromosomal translocations tend to occur much more frequently than are inter-chromosomal translocations.

Among the four methods, Digenome-seq is the only method without any pre-sequencing PCR steps. The other methods require oligonucleotide tag-specific amplifications (GUIDE-seq) or linear amplification-mediated (LAM)-PCR (HTGTS and BLESS), prior to high-throughput sequencing. To carry out BLESS, biotinylated oligonucleotide adaptors must be ligated to DSB ends in vitro. These steps are technically challenging and can also produce PCR primer-dependent artifacts or false positives.
Cell-based methods suffer from DSBs that occur spontaneously in the cells even in the absence of an engineered nuclease. GUIDE-seq, HTGTS and BLESS fails to distinguish these naturally-occurring events resulting from nuclease-induced cleavages, resulting in false positives. Digenome-seq is not limited by naturally-occurring DSBs, which cannot produce uniform cleavage patterns, signatures of nuclease-induced events in vitro. Furthermore, DSB ends are trimmed or resected by endogenous repair enzymes in cells but not in vitro. To identify off-target sites using GUIDE-seq and HTGTS, bioinformatics filters are applied to search for sequences around the capture sites that are homologous to the on-target site. Up to 95% of captured sites are discarded during this filtering step. In contrast, Digenome-seq can pinpoint off-target sites because DSB ends are not processed in vitro. In addition, homology-based sequence search for off-target sites is unnecessary with Digenome-seq.

Each programmable nuclease produces its own DSB pattern. SpCas9 yields blunt ends because it cuts both strands in a DNA molecule at the same position. ZFNs produce 5’ 4 or 5-nt overhangs because they cut the DNA molecule asymmetrically by leaving several single stranded bases. Cpf1 and c2c1 is a recently identified RNA-guided nuclease derived from the class II CRISPR system that produces 5’ 5-nt overhangs [36, 37]. Blunt-ended oligonucleotides used in GUIDE-seq or BLESS may not be efficiently ligated with DSB ends. HTGTS and Digenome-seq are not limited by cohesive ends because no oligonucleotide tags are used. Both cohesive ends and blunt ends are resected in cells. As a result, the overhang patterns produced by a novel nuclease cannot be inferred by cell-based methods. Fortunately, these patterns are preserved in vitro and can be revealed by Digenome-seq. Finally, Digenome-seq is multiplexible without increasing sequencing depth [35]: Up to hundreds of guide RNAs can be mixed to digest cell-free genomic DNA in vitro.
It is of note that none of these methods are comprehensive. For example, one sgRNA specific to VEGF-A site has been tested by GUIDE-seq, HTGTS, and Digenome-seq [30, 31, 34], which revealed potential off-target sites that differed by up to 6 nucleotides from the on-target site. Off-target sites with 5 or more mismatches cannot be chosen by in silico methods because there are more than thousands of such sites in the human genome. Importantly, most sites were identified commonly by all of the three methods. However, each method revealed potential off-target sites missed by other methods. Some of these sites could be false positives that arise from PCR primer-dependent artifacts and naturally-occurring DSBs. To validate off-target sites, it is important to perform targeted deep sequencing and to detect nuclease-induced indels at candidate sites.

**Improving on-target specificity of CRISPR/Cas9**

Considering the recent progress made in last 2 years to improve the on-target specificity of CRISPR/Cas9 compared to ZFNs and TALENs, we focus on recent research updates for improving CRISPR/Cas9 on-target specificity (Figure -3).

1. **Target sequences**

   This method designs unique target sequences that differ from any other site in the genome by at least 2 or 3 nucleotides in 20-nt sequences [12]. CRISPR/Cas9 discriminates efficiently against potential off-target sites with mismatched PAM sequences and seed regions upstream of the PAM sequence. Alternatively, a web-based computer algorithm (www.rgnome.net/casoffinder) can be used to search potential off-target sites and unique target sequences in more than 20 organism genomes, including the human genome. There are also several web-based tools (Table-3) to synthesize sgRNA with improved on-target specificity.

2. **Different versions of sgRNAs**
Different versions of sgRNAs were synthesized to reduce off-target activity by an order of magnitude without sacrificing on-target specificity [12]. sgRNAs with two extra, target independent guanine nucleotides at the 5' terminus can be less active at on-target sites but they are significantly more specific compared to conventional sgRNAs [12]. Truncated sgRNAs (tru-sgRNAs) with 17 nts rather than 20 nts increases the specificity [38].

**Paired nickases and nickases**

Cas9 can be converted to a nickase that generate single-strand breaks rather than DSBs by mutating one of the nuclease active sites. Paired nickases generate two single-strand breaks or nicks on different DNA strands, resulting in a composite DSB and doubling the specificity of genome editing [12, 33, 39, 40]. Catalytically inactive or dead Cas9 (dCas9) created by inactivating two nuclease active sites is fused to the FokI nuclease domain to make dimeric nucleases [41-43], similar to ZFNs or TALENs. The FokI domain must dimerize to cleave DNA. Although these approaches are quite efficient to enhance the on-target specificity, they require two active sgRNAs to make functional pairs. In addition, target sequences must contain two PAM sequences in an inverted repeat configuration, limiting the choice of targetable sites.

**Cas9 protein and direct delivery of nucleases**

Using the Cas9 recombinant protein (commercially available from www.toolgen.com) rather than Cas9 encoding plasmid further reduces off-target mutations [44-47]. The direct delivery of Cas9-sgRNA ribonucleoprotein (RNP) complexes induces mutations at target sites immediately after the delivery and decomposes rapidly by endogenous proteases, reducing off-target mutations without compromising on-target efficiency.

**Cas9 variants**
Cas9 can be engineered to reduce off-target effects. Slaymaker et al. \cite{48} replaced positively charged amino acid residues in Cas9 to weaken its interaction with a non-target DNA strand. Likewise, Kleinstiver et al. \cite{49} mutated amino acid residues that form hydrogen bonds with the phosphate backbone. The resulting variants, termed enhanced SpCas9 (eSpCas9) and SpCas9 high fidelity (SpCas9-HF), respectively, showed genome-wide reduction of off-target effects.

Consequently, two recent papers provided new insight into CRISPR/Cas9 targeting and specificity \cite{50,51}. These studies provided much needed answers to questions about the CRISPR/Cas9 target and specificity; according to these reports, CRISPR/Cas9 performs three checks before cutting the target sequences. Cas9 exerts specific control: once it binds to a region of DNA, it performs another check before bringing two sections of the Cas9 protein complex like “two blades of scissors” to precisely align the active sites that cut double stranded DNA \cite{50,51}. Alternatively, these active sites are consciously mispositioned at off-target sites, so that DNA cannot be cut. Furthermore, two different active region of Cas9 on either strand communicate via structural changes to ensure Cas9 to cut accurate and precise regions of target sequences. To support this hypothesis, recent studies on Cas9 variants with alanine substitution at various location showed reduced off-target activities \cite{48}. These reports shed light on conformational control of Cas9 on-target specificity and help researchers to synthesize more specific Cas9 variants.

**Conclusion and future perspectives**

GE tools including CRISPR/Cas9 have revolutionized genome engineering, and a major goal is to develop therapeutic applications of GE tools, particularly CRISPR/Cas9, to treat and cure genetic human and animal diseases or to use them to modulate novel traits in agriculture. Before this
technology can be adapted in humans or other organisms, researchers all around the world are endeavoring to ensure its precision and accuracy in order to avoid any unintended consequences arising from off-target mutations. Hence, in the last two years, research on CRISPR/Cas9 has made remarkable progress in gene editing, with particular a focus in reducing off-target mutations without sacrificing on-target specificity (see Outstanding questions).

In the future, researchers should emphasize the different versions of sgRNA synthesis; certain sgRNAs are remarkably specific, resulting in no measurable off-target effects, as revealed by Digenome-seq and GUIDE-seq. To better understand the specificity and accuracy of different versions of sgRNAs, it is important to profile the off-target effects of as many sgRNAs as possible at the genome-wide level. Currently available genome-wide off-target profiling methods can detect indel frequencies up to 0.01% to 1% (0.1% on average); a more sensitive, cost effective method is needed to detect indel frequencies below 0.01% in the entire genome to determine the efficiencies of the various sgRNAs.
Legend for figures

**Figure 1** – Overview of the nuclease-mediated genome engineering using ZFNs, TALENs and CRISPR/Cas9; ZFN is composed of zink-finger protein (ZFP) at the amino terminus and FokI nuclease at the carboxyl terminus, target sequence of ZFN is typically 10-36 bp in length excluding spacers; TALEN is composed of transcription activator like effectors (TALEs) at the amino terminus and FokI nuclease at the carboxyl terminus, target sequence of ZFN is typically 30-40 bp in length excluding spacers; CRISPR/Cas9 composed of Cas9 and a sgRNA, guide sequence in sgRNA is complementary to 20 bp of target DNA sequence (protospacers), next to the 5’-NGG-3’ (N represents any nucleotide) referred as protospacer adjacent motif (PAM)

**Figure 2** – Outline of four different methods of unbiased genome-wide profiling of nuclease cleavage sites; Integrase-deficient lentivirus (IDLV) capture or genome-wide, unbiased identification of DSBs enabled by sequencing (GUIDE-seq); High-throughput genomic translocation sequencing (HTGTS); Breaks labelling, enrichments on streptavidin and next-generation sequencing (BLESS); In vitro nuclease-digested genome sequencing (Digenome-seq); DSB, double-strand break; ODN, oligodeoxynucleotide; sgRNA, small-guide RNA; WGS, whole-genome sequencing; gDNA, genomic DNA.
Figure 3 – Graphical representation of strategies to minimize off-target mutations in CRISPR/Cas9: A) SgRNA Variants; sgRNAs with two extra guanines (ggX$_{20}$) or truncated sgRNAs (gX$_{17}$) enhance the on-target specificity, compared to conventional sgRNAs (gX$_{19}$ or gX$_{20}$). B) Cas9 variants; Use of paired nickases to generate two single-strand breaks or nicks on different DNA strands. C) Method of delivery; Use of Cas9-sgRNA ribonucleoprotein (RNP) complexes, rather than the Cas9 and sgRNA-encoding plasmids enhances target specificity while significantly minimizing off-target activities with continuous expression of Cas9 and sgRNA from plasmids (Modified from Koo et al.) [5].
Table 1: Comparison of the limitations of GE tools

<table>
<thead>
<tr>
<th></th>
<th>METHODS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ZFNs</td>
</tr>
<tr>
<td>OFF-TARGET ACTIVITY</td>
<td>Low to moderate</td>
</tr>
<tr>
<td>EASE OF APPLICATION TO</td>
<td>Laborious, difficult and substantial cloning protein engineering required</td>
</tr>
<tr>
<td>GENETARTE TARGETTED GENOME</td>
<td></td>
</tr>
<tr>
<td>EDITING</td>
<td></td>
</tr>
<tr>
<td>EASE OF MULTIPLEXING</td>
<td>Low</td>
</tr>
<tr>
<td>EASE OF GENERATING LARGE</td>
<td>Low; laborious and complex protein engineering required</td>
</tr>
<tr>
<td>SCALE LIBRARIES</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Comparisons of various methods for profiling genome-wide nuclease off-target sites.

<table>
<thead>
<tr>
<th>Methods</th>
<th>GUIDE-seq</th>
<th>HTGTS</th>
<th>BLESS</th>
<th>Digenome-seq</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA cleaved in vivo or in vitro</td>
<td>In vivo</td>
<td>In vivo</td>
<td>In vivo</td>
<td>In vitro</td>
</tr>
<tr>
<td>DSB captured in vivo or in vitro</td>
<td>In vivo</td>
<td>In vivo</td>
<td>In vitro</td>
<td>In vitro</td>
</tr>
<tr>
<td>Quantitative?</td>
<td>Yes</td>
<td>No</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Pre-sequencing PCR required?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Homology-based search required?</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Compatible with cohesive ends?</td>
<td>Less sensitive with cohesive ends</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Overhang patterns inferred?</td>
<td>No</td>
<td>No</td>
<td>?</td>
<td>Yes</td>
</tr>
<tr>
<td>Multiplexible</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Table 3 – Web-based tools for guide RNA synthesis.

<table>
<thead>
<tr>
<th>Tools</th>
<th>Web address</th>
<th>Throughput</th>
<th>Input</th>
<th>Scoring</th>
<th>Support for Cas9 nickase</th>
<th>Application</th>
<th>Species supported</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>SgRNA designer</td>
<td><a href="http://www.broadinstitute.org/rna/public/analysis-tools/sgrna-design">http://www.broadinstitute.org/rna/public/analysis-tools/sgrna-design</a></td>
<td>Medium to high</td>
<td>Sequence</td>
<td>Yes</td>
<td>No</td>
<td>Picks and ranks sgRNA sequences</td>
<td>Human, mouse</td>
<td>[52]</td>
</tr>
<tr>
<td>Cas-OFFinder/Cas-Designer</td>
<td><a href="http://www.rgenome.net/cas-offinder/">http://www.rgenome.net/cas-offinder/</a> <a href="http://rgenome.net/cas-designer/">http://rgenome.net/cas-designer/</a></td>
<td>Medium to high</td>
<td>Sequence</td>
<td>Yes</td>
<td>No</td>
<td>Fast and versatile algorithms based search for potential off-targets</td>
<td>20</td>
<td>[28, 53]</td>
</tr>
<tr>
<td>SSFinder</td>
<td><a href="https://code.google.com/archive/p/ssfinder/">https://code.google.com/archive/p/ssfinder/</a></td>
<td>High</td>
<td>Sequence</td>
<td>No</td>
<td>No</td>
<td>High throughput prediction of CRISPR/Cas9 binding site from huge nucleotide dataset</td>
<td>N/A</td>
<td>[54]</td>
</tr>
<tr>
<td>Cas9 design</td>
<td><a href="http://cas9.cbi.pku.edu.cn/">http://cas9.cbi.pku.edu.cn/</a></td>
<td>Low</td>
<td>Sequence</td>
<td>No</td>
<td>No</td>
<td>Find target sequences and OTs for single sequence</td>
<td>10</td>
<td>[55]</td>
</tr>
<tr>
<td>CRISPR Multitargeter</td>
<td><a href="http://www.multicrispr.net/">http://www.multicrispr.net/</a></td>
<td>Low</td>
<td>Gene symbol or sequences</td>
<td>Yes</td>
<td>No</td>
<td>Algorithm based unique target sequence prediction from multiple genes or transcripts</td>
<td>12</td>
<td>[56]</td>
</tr>
<tr>
<td>ZiFit</td>
<td><a href="http://zifit.partners.org/ZiFiT/">http://zifit.partners.org/ZiFiT/</a></td>
<td>Low</td>
<td>Sequence</td>
<td>No</td>
<td>Yes</td>
<td>Find target sequences and OTs for single sequence</td>
<td>9</td>
<td>[57]</td>
</tr>
<tr>
<td>E-CRISP</td>
<td><a href="http://www.e-crisp.org/E-CRISP/">http://www.e-crisp.org/E-CRISP/</a></td>
<td>Low</td>
<td>Gene symbol or sequences</td>
<td>Yes</td>
<td>Yes</td>
<td>Added options for Cas9 nickase design</td>
<td>More than 30</td>
<td>[58]</td>
</tr>
<tr>
<td>CRISPR Direct</td>
<td><a href="http://crispr.dbcls.jp/">http://crispr.dbcls.jp/</a></td>
<td>Low</td>
<td>Sequence, transcript or genome location</td>
<td>Yes</td>
<td>No</td>
<td>Find target sequences with limited information on off-target sites</td>
<td>20</td>
<td>[59]</td>
</tr>
<tr>
<td>CCTop</td>
<td><a href="http://crispr.cos.uni-heidelberg.de/">http://crispr.cos.uni-heidelberg.de/</a></td>
<td>Low</td>
<td>Sequence</td>
<td>Yes</td>
<td></td>
<td>Fast and easy to generate sgRNAs with comprehensive information on- and off-target sites</td>
<td>15</td>
<td>[60]</td>
</tr>
<tr>
<td>Tool</td>
<td>Website</td>
<td>Cost</td>
<td>Input Type</td>
<td>Directly Generates</td>
<td>Comprehensive Off-target Details</td>
<td>Species Support</td>
<td>Notes</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>----------------------------------------------</td>
<td>------</td>
<td>---------------------</td>
<td>--------------------</td>
<td>-----------------------------------</td>
<td>-----------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>CROP-IT</td>
<td><a href="http://www.adlilab.org/CROP-IT/homepage.html">http://www.adlilab.org/CROP-IT/homepage.html</a></td>
<td>Low</td>
<td>sgRNA</td>
<td>Yes, No</td>
<td>Comprehensive off-target details</td>
<td>Mouse and human</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>CHOP</td>
<td><a href="https://chopchop.rc.fas.harvard.edu/">https://chopchop.rc.fas.harvard.edu/</a></td>
<td>Medium</td>
<td>Sequence, transcript or gene id.</td>
<td>Yes</td>
<td>Easier and fast synthesizing sgRNAs with complete info OTs for a single target sequence</td>
<td>20</td>
<td>[61]</td>
<td></td>
</tr>
<tr>
<td>Crispr.mit</td>
<td><a href="http://crispr.mit.edu/">http://crispr.mit.edu/</a></td>
<td>Low to medium</td>
<td>Sequence or FASTA file</td>
<td>Yes, Yes</td>
<td>Easier and faster with comprehensive information on- and off-target. Provide option to synthesize paired sgRNAs for nickases</td>
<td>15</td>
<td>[15]</td>
<td></td>
</tr>
<tr>
<td>GT-Scan</td>
<td><a href="http://gt-scan.braembl.org.au/gt-scan/">http://gt-scan.braembl.org.au/gt-scan/</a></td>
<td>Low</td>
<td>Sequence or gene id.</td>
<td>No</td>
<td>Find target sequence and OTs for single sequence</td>
<td>20</td>
<td>[62]</td>
<td></td>
</tr>
<tr>
<td>Cas OT</td>
<td><a href="http://eendb.zfgenetics.org/casot/">http://eendb.zfgenetics.org/casot/</a></td>
<td>Low to medium</td>
<td>FASTA file</td>
<td>Yes</td>
<td>Finds target sequences and OTs</td>
<td>User specified</td>
<td>[63]</td>
<td></td>
</tr>
<tr>
<td>Tool</td>
<td>Website</td>
<td>Input Format</td>
<td>Run Support</td>
<td>Specific Information</td>
<td>Platforms</td>
<td>Reference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------------------------------</td>
<td>--------------</td>
<td>-------------</td>
<td>-----------------------------------------------------------</td>
<td>--------------------</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WU-CRISPR</td>
<td><a href="http://crispr.wustl.edu/">http://crispr.wustl.edu/</a></td>
<td>Low</td>
<td>Yes</td>
<td>Finds efficient target site based on OTs</td>
<td>Mouse and human</td>
<td>[64]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sgRNA Cas9</td>
<td><a href="http://www.biooools.com/col.jsp?id=103">http://www.biooools.com/col.jsp?id=103</a></td>
<td>High</td>
<td>Yes</td>
<td>Finds target sequences with limited information on- and off-target sites</td>
<td>User specified</td>
<td>[65]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sgRNA Scorer 1.0</td>
<td><a href="https://crispr.med.harvard.edu/sgRNAScorer/">https://crispr.med.harvard.edu/sgRNAScorer/</a></td>
<td>Low</td>
<td>Yes</td>
<td>Finds target sequence with limited information on- and off-target sites</td>
<td>Mouse and human</td>
<td>[66]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protoscaler</td>
<td><a href="http://www.protoscaler.com/">http://www.protoscaler.com/</a></td>
<td>Medium to high</td>
<td>Yes</td>
<td>Finds target sequence along with sgRNA ranking</td>
<td>User specified</td>
<td>[67]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRISPRseek</td>
<td><a href="http://www.bioconductor.org/packages/release/bioc/html/CRISPRseek.html">http://www.bioconductor.org/packages/release/bioc/html/CRISPRseek.html</a></td>
<td>High</td>
<td>Yes</td>
<td>Performs OTs and target sequence for multiple sequences</td>
<td>Several common genomes</td>
<td>[68]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Low: input format and run supports one gene at a time queries. Medium: Supports small batches of gene or tens to hundreds of sgRNAs queries. High: supports genome-scale queries. OTs; off-targets, MMs; mismatches.


