Fine-tuning next-generation genome editing tools

Fine-tuning next-generation genome editing tools

Chidananda Nagamangala Kanchiswamy[1*], Massimo Maffei[2], Mickael Malnoy[1], Riccardo Velasco[1], Jin-Soo Kim[3, 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, TALLENs, 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

ZFN

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'-GNNGNNGNN-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. pyogenesis* 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 phophothiorate 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.

198 Cell-based methods suffer from DSBs that occur spontaneously in the cells even in the absence of
199 an engineered nuclease. GUIDE-seq, HTGTS and BLESS fails to distinguish these naturally-
200 occurring events resulting from nuclease-induced cleavages, resulting in false positives.
201 Digenome-seq is not limited by naturally-occurring DSBs, which cannot produce uniform
202 cleavage patterns, signatures of nuclease-induced events in vitro. Furthermore, DSB ends are
203 trimmed or resected by endogenous repair enzymes in cells but not in vitro. To identify off-target
204 sites using GUIDE-seq and HTGTS, bioinformatics filters are applied to search for sequences
205 around the capture sites that are homologous to the on-target site. Up to 95% of captured sites are
206 discarded during this filtering step. In contrast, Digenome-seq can pinpoint off-target sites because
207 DSB ends are not processed in vitro. In addition, homology-based sequence search for off-target
208 sites is unnecessary with Digenome-seq.

209 Each programmable nuclease produces its own DSB pattern. SpCas9 yields blunt ends because
210 it cuts both strands in a DNA molecule at the same position. ZFNs produce 5' 4 or 5-nt overhangs
211 because they cut the DNA molecule asymmetrically by leaving several single stranded bases. Cpf1
212 and c2c1 is a recently identified RNA-guided nuclease derived from the class II CRISPR system
213 that produces 5' 5-nt overhangs [36, 37]. Blunt-ended oligonucleotides used in GUIDE-seq or
214 BLESS may not be efficiently ligated with DSB ends. HTGTS and Digenome-seq are not limited
215 by cohesive ends because no oligonucleotide tags are used. Both cohesive ends and blunt ends are
216 resected in cells. As a result, the overhang patterns produced by a novel nuclease cannot be inferred
217 by cell-based methods. Fortunately, these patterns are preserved in vitro and can be revealed by
218 Digenome-seq. Finally, Digenome-seq is multiplexible without increasing sequencing depth [35]:
219 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].

3. 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 Fokl nuclease domain to make dimeric nucleases [41-43], similar to ZFNs or TALENs. The Fokl 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.

4. 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.

5. Cas9 variants

Cas9 can be engineered to reduce off-target effects. Slaymaker et al. [48] replaced positively charged amino acid residues in Cas9 to weaken its interaction with a non-target DNA strand. Likewise, Kleinstiver et al. [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 [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 section of the Cas9 protein complex like “two blades of scissors” to precisely align the active sites that cut double stranded DNA [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 [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 *Fok1* 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 *Fok1* 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)

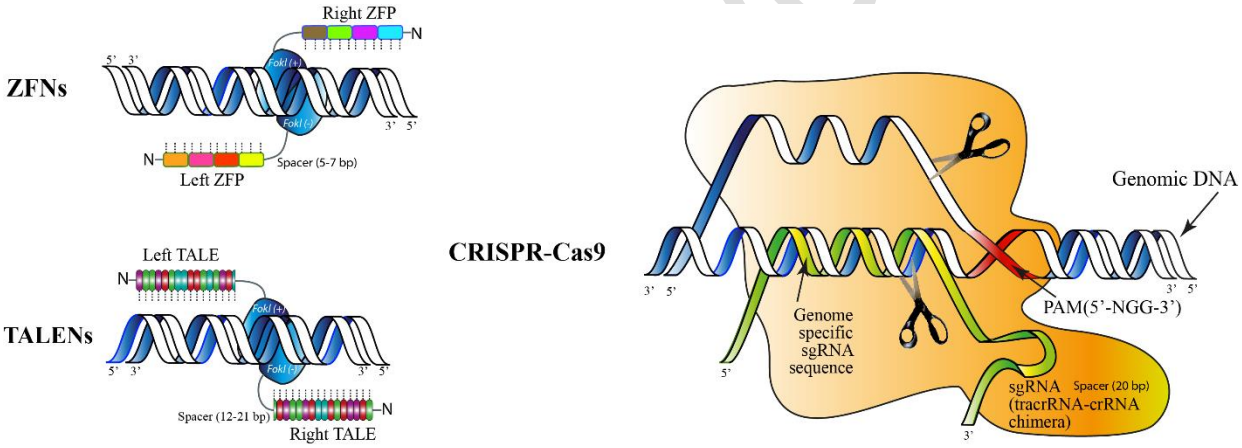


Figure 2 – Outline of four different methods of unbiased genome-wide profiling of nuclease cleavage sites; Integrase-deficient lentivirus (IDLV) capture or genome-wide, unbiased identification of DSBs enabled by sequencing (GUIDE-seq); High-throughput genomic translocation sequencing (HTGTS); Breaks 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.

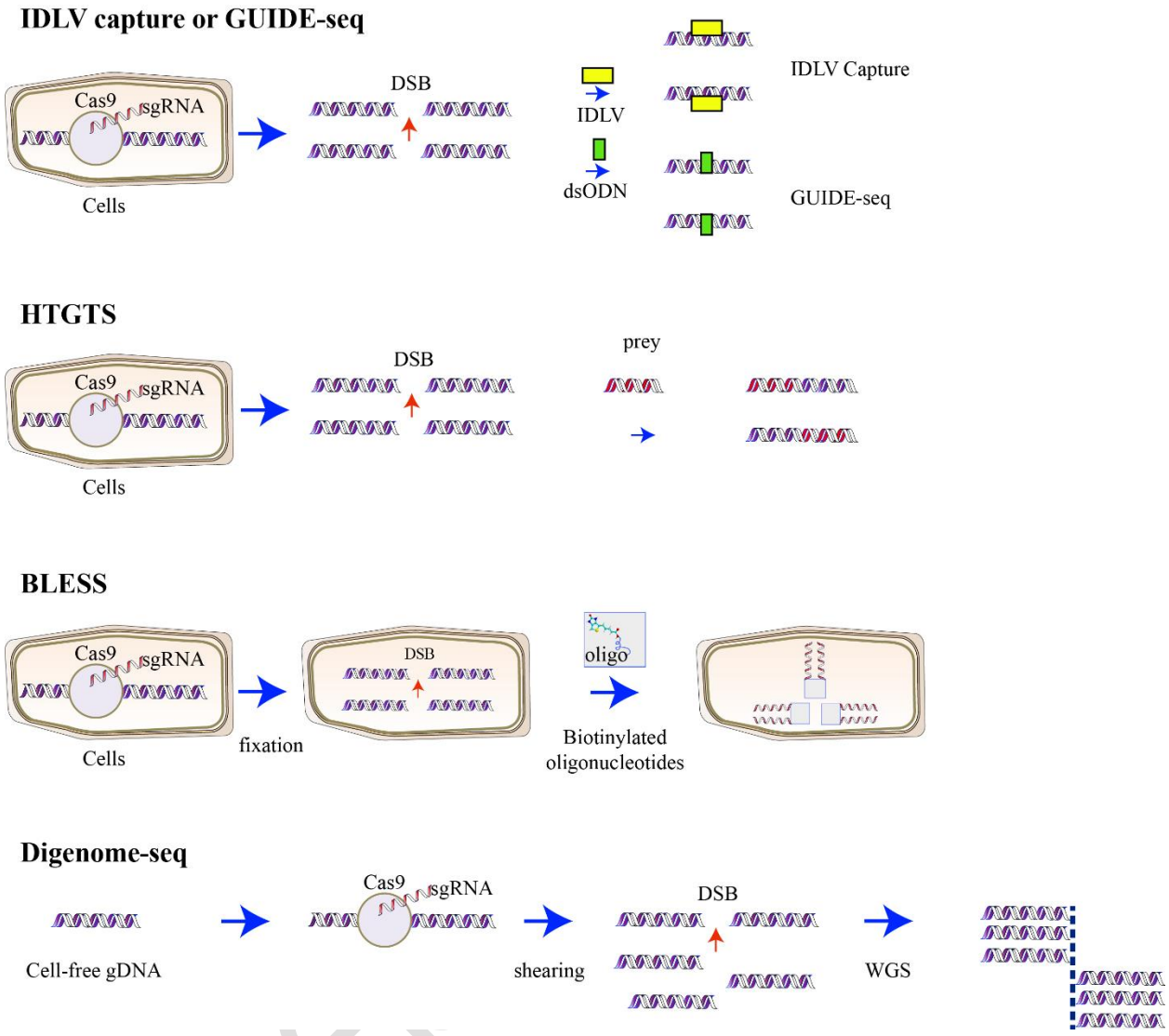
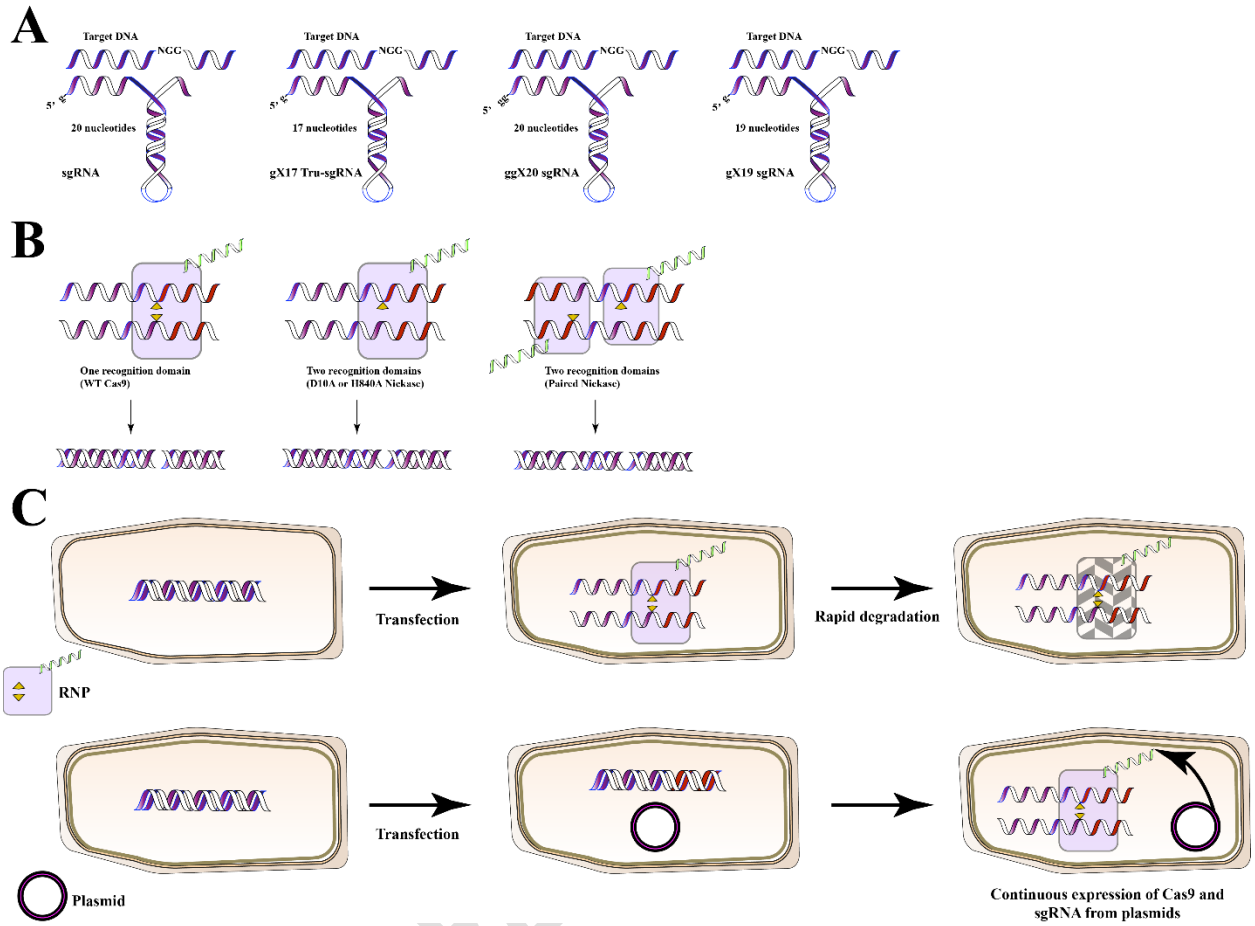


Figure 3 – Graphical representation of strategies to minimize off-target mutations in CRISPR/Cas9; A) SgRNA Variants; sgRNAs with two extra guanines (ggX_{20}) or truncated sgRNAs (gX_{17}) enhance the on-target specificity, compared to conventional sgRNAs (gX_{19} or gX_{20}). B) Cas9 variants; Use 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].



328

329

330

331

Table 1: Comparison of the limitations of GE tools

	METHODS		
	ZFNs	TALENs	CRISPR/Cas9
OFF-TARGET ACTIVITY	Low to moderate	Low	Low to moderate
EASE OF APPLICATION TO GENETARTE TARGETTED GENOME EDITING	Laborious, difficult and substantial cloning protein engineering required	Laborious, moderately difficult and substantial cloning required	Easy, simple cloning steps required
EASE OF MULTIPLEXING	Low	Low	High
EASE OF GENERATING LARGE SCALE LIBRARIES	Low; laborious and complex protein engineering required	Moderate; laborious and substantial cloning required	Easy; Simple oligo synthesis and cloning required

Table 2. Comparisons of various methods for profiling genome-wide nuclease off-target sites.

Methods	GUIDE-seq	HTGTS	BLESS	Digenome-seq
DNA cleaved in vivo or in vitro	In vivo	In vivo	In vivo	In vitro
DSB captured in vivo or in vitro	In vivo	In vivo	In vitro	In vitro
Quantitative?	Yes	No	?	?
Sensitivity	High	Low	Low	High
Pre-sequencing PCR required?	Yes	Yes	Yes	No
Homology-based search required?	Yes	Yes	No	No
Compatible with cohesive ends?	Less sensitive with cohesive ends	Yes	Yes	Yes
Overhang patterns inferred?	No	No	?	Yes
Multiplexible	No	No	No	Yes

Table 3 – Web-based tools for guide RNA synthesis.

Tools	Web address	Throughput	Input	Scoring	Support for Cas9 nickase	Application	Species supported	Ref
SgRNA designer	http://www.broadinstitute.org/rnai/public/analysis-tools/sgrna-design	Medium to high	Sequence	Yes	No	Picks and ranks sgRNA sequences	Human, mouse	[52]
Cas-OFFinder/ Cas-Designer	http://www.rgenome.net/cas-offinder/ http://rgenome.net/cas-designer/	Medium to high	Sequence	Yes	No	Fast and versatile algorithms based search for potential off-targets	20	[28, 53]
SSFinder	https://code.google.com/archive/p/ssfinder/	High	Sequence	No	No	High throughput prediction of CRISPR/Cas9 binding site from huge nucleotide dataset	N/A	[54]
Cas9 design	http://cas9.cbi.pku.edu.cn/	Low	Sequence	No	No	Find target sequences and OTs for single sequence	10	[55]

CRISPR Multitargeter	http://www.multicrispr.net/	Low	Gene symbol or sequences	Yes	No	Algorithm based unique target sequence prediction from multiple genes or transcripts	12	[56]
ZiFit	http://zifit.partners.org/ZiFiT/	Low	Sequence	No	Yes	Find target sequences and OTs for single sequence	9	[57]
E-CRISP	http://www.e-crisp.org/E-CRISP/	Low	Gene symbol or sequences	Yes	Yes	Added options for Cas9 nickase design	More than 30	[58]
CRISPR Direct	http://crispr.dbcls.jp/	Low	Sequence, transcript or genome location	Yes	No	Find target sequences with limited information on off-target sites	20	[59]
CCTop	http://crispr.cos.uni-heidelberg.de/	Low	Sequence	Yes		Fast and easy to generate sgRNAs with comprehensive information on- and off-target sites	15	[60]

CROP-IT	http://www.adlilab.org/CROP-IT/homepage.html	Low	sgRNA	Yes	No	Comprehensive off-target details	Mouse and human	None
CHOP CHOP	https://chopchop.rc.fas.harvard.edu/	Medium	Sequence, transcript or gene i.d.		Yes	Easier and fast synthesizing sgRNAs with complete info OTs for a single target sequence	20	[61]
Crispr.mit	http://crispr.mit.edu/	Low to medium	Sequence or FASTA files	Yes	Yes	Easier and faster with comprehensive information on- and off-target. Provide option to synthesize paired sgRNAs for nickases	15	[15]
GT-Scan	http://gt-scan.braembl.org.au/gt-scan/	Low	Sequence or gene i.d.	No		Find target sequence and OTs for single sequence	20	[62]
Cas OT	http://eendb.zfgenetics.org/casot/	Low to medium	FASTA file	Yes		Finds target sequences and OTs	User specified	[63]

WU-CRISPR	http://crispr.wustl.edu/	Low	Sequence or gene i.d.	Yes		Finds efficient target site based on OTs	Mouse and human	[64]
sgRNACas9	http://www.biootools.com/col.jsp?id=103	High	Software package	Yes		Finds target sequences with limited information on- and off-target sites	User specified	[65]
sgRNA Scorer 1.0	https://crispr.med.harvard.edu/sgRNAScorer/	Low	Sequence or FASTA file	Yes		Finds target sequence OTs and also provides information on on-target scoring	12	[66]
Protospacer	http://www.protospacer.com/	Medium to high	Sequence, gene i.d. and many other inputs	Yes		Finds target sequence along with sgRNA ranking	User specified	[67]
CRISPRseek	http://www.bioconductor.org/packages/release/bioc/html/CRISPRseek.html	High	Software package	Yes	Yes	Performs OTs and target sequence for multiple sequences	Several common genomes	[68]

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

References

1. **Method of the Year 2011**. *Nat Meth* 2012, **9**(1):1-1.
2. Travis J: **Making the cut**. *Science* 2015, **350**(6267):1456-1457.
3. Kim YG, Cha J, Chandrasegaran S: **Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain**. *Proceedings of the National Academy of Sciences of the United States of America* 1996, **93**(3):1156-1160.
4. Bitinaite J, Wah DA, Aggarwal AK, Schildkraut I: **FokI dimerization is required for DNA cleavage**. *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**(18):10570-10575.
5. Koo T, Lee J, Kim JS: **Measuring and Reducing Off-Target Activities of Programmable Nucleases Including CRISPR-Cas9**. *Molecules and cells* 2015, **38**(6):475-481.
6. Kim HJ, Lee HJ, Kim H, Cho SW, Kim JS: **Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly**. *Genome research* 2009, **19**(7):1279-1288.
7. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U: **Breaking the code of DNA binding specificity of TAL-type III effectors**. *Science* 2009, **326**(5959):1509-1512.
8. Moscou MJ, Bogdanove AJ: **A simple cipher governs DNA recognition by TAL effectors**. *Science* 2009, **326**(5959):1501.
9. Mussolino C, Morbitzer R, Lutge F, Dannemann N, Lahaye T, Cathomen T: **A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity**. *Nucleic acids research* 2011, **39**(21):9283-9293.
10. Kim Y, Kweon J, Kim A, Chon JK, Yoo JY, Kim HJ, Kim S, Lee C, Jeong E, Chung E *et al*: **A library of TAL effector nucleases spanning the human genome**. *Nature biotechnology* 2013, **31**(3):251-258.
11. van der Oost J, Westra ER, Jackson RN, Wiedenheft B: **Unravelling the structural and mechanistic basis of CRISPR-Cas systems**. *Nature reviews Microbiology* 2014, **12**(7):479-492.
12. Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S, Kim JS: **Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases**. *Genome research* 2014, **24**(1):132-141.
13. Cradick TJ, Fine EJ, Antico CJ, Bao G: **CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity**. *Nucleic acids research* 2013, **41**(20):9584-9592.
14. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD: **High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells**. *Nature biotechnology* 2013, **31**(9):822-826.
15. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O *et al*: **DNA targeting specificity of RNA-guided Cas9 nucleases**. *Nature biotechnology* 2013, **31**(9):827-832.
16. Hendel A, Fine EJ, Bao G, Porteus MH: **Quantifying on- and off-target genome editing**. *Trends in biotechnology* 2015, **33**(2):132-140.
17. Zhang XH, Tee LY, Wang XG, Huang QS, Yang SH: **Off-target Effects in CRISPR/Cas9-mediated Genome Engineering**. *Molecular therapy Nucleic acids* 2015, **4**:e264.
18. Lee HJ, Kim E, Kim JS: **Targeted chromosomal deletions in human cells using zinc finger nucleases**. *Genome research* 2010, **20**(1):81-89.
19. Lee HJ, Kweon J, Kim E, Kim S, Kim JS: **Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases**. *Genome research* 2012, **22**(3):539-548.
20. Park CY, Kim J, Kweon J, Son JS, Lee JS, Yoo JE, Cho SR, Kim JH, Kim JS, Kim DW: **Targeted inversion and reversion of the blood coagulation factor 8 gene in human iPS cells using TALENs**. *Proc Natl Acad Sci U S A* 2014, **111**(25):9253-9258.

21. Tebas P, Stein D, Tang WW, Frank I, Wang SQ, Lee G, Spratt SK, Surosky RT, Giedlin MA, Nichol G *et al*: **Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV**. *The New England journal of medicine* 2014, **370**(10):901-910.
22. Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ *et al*: **A TALE nuclease architecture for efficient genome editing**. *Nature biotechnology* 2011, **29**(2):143-148.
23. Wang X, Wang Y, Wu X, Wang J, Wang Y, Qiu Z, Chang T, Huang H, Lin RJ, Yee JK: **Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors**. *Nature biotechnology* 2015, **33**(2):175-178.
24. Kuscu C, Arslan S, Singh R, Thorpe J, Adli M: **Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease**. *Nature biotechnology* 2014, **32**(7):677-683.
25. Wu X, Scott DA, Kriz AJ, Chiu AC, Hsu PD, Dadon DB, Cheng AW, Trevino AE, Konermann S, Chen S *et al*: **Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells**. *Nature biotechnology* 2014, **32**(7):670-676.
26. Gabriel R, Lombardo A, Arens A, Miller JC, Genovese P, Kaeppel C, Nowrouzi A, Bartholomae CC, Wang J, Friedman G *et al*: **An unbiased genome-wide analysis of zinc-finger nuclease specificity**. *Nature biotechnology* 2011, **29**(9):816-823.
27. Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR: **High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity**. *Nature biotechnology* 2013, **31**(9):839-843.
28. Bae S, Park J, Kim JS: **Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases**. *Bioinformatics* 2014, **30**(10):1473-1475.
29. Fine EJ, Cradick TJ, Zhao CL, Lin Y, Bao G: **An online bioinformatics tool predicts zinc finger and TALE nuclease off-target cleavage**. *Nucleic acids research* 2014, **42**(6):e42.
30. Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar VV, Thapar V, Wyvekens N, Khayter C, Iafrate AJ, Le LP *et al*: **GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases**. *Nature biotechnology* 2015, **33**(2):187-197.
31. Frock RL, Hu J, Meyers RM, Ho YJ, Kii E, Alt FW: **Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases**. *Nature biotechnology* 2015, **33**(2):179-186.
32. Crosetto N, Mitra A, Silva MJ, Bienko M, Dojer N, Wang Q, Karaca E, Chiarle R, Skrzypczak M, Ginalski K *et al*: **Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing**. *Nature methods* 2013, **10**(4):361-365.
33. Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y *et al*: **Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity**. *Cell* 2013, **154**(6):1380-1389.
34. Kim D, Bae S, Park J, Kim E, Kim S, Yu HR, Hwang J, Kim JI, Kim JS: **Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells**. *Nature methods* 2015, **12**(3):237-243, 231 p following 243.
35. Kim D, Kim S, Kim S, Park J, Kim JS: **Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq**. *Genome research* 2016, **26**(3):406-415.
36. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A *et al*: **Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System**. *Cell* 2015, **163**(3):759-771.
37. Shmakov S, Abudayyeh OO, Makarova KS, Wolf YI, Gootenberg JS, Semenova E, Minakhin L, Joung J, Konermann S, Severinov K *et al*: **Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems**. *Molecular cell* 2015, **60**(3):385-397.
38. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK: **Improving CRISPR-Cas nuclease specificity using truncated guide RNAs**. *Nature biotechnology* 2014, **32**(3):279-284.

39. Kim E, Kim S, Kim DH, Choi BS, Choi IY, Kim JS: **Precision genome engineering with programmable DNA-nicking enzymes**. *Genome research* 2012, **22**(7):1327-1333.
40. Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, Yang L, Church GM: **CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering**. *Nature biotechnology* 2013, **31**(9):833-838.
41. Dahlman JE, Abudayyeh OO, Joung J, Gootenberg JS, Zhang F, Konermann S: **Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease**. *Nature biotechnology* 2015, **33**(11):1159-1161.
42. Kleinstiver BP, Prew MS, Tsai SQ, Nguyen NT, Topkar VV, Zheng Z, Joung JK: **Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition**. *Nature biotechnology* 2015, **33**(12):1293-1298.
43. Long L, Guo H, Yao D, Xiong K, Li Y, Liu P, Zhu Z, Liu D: **Regulation of transcriptionally active genes via the catalytically inactive Cas9 in C. elegans and D. rerio**. *Cell research* 2015, **25**(5):638-641.
44. Kim S, Kim D, Cho SW, Kim J, Kim JS: **Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins**. *Genome research* 2014, **24**(6):1012-1019.
45. Woo JW, Kim J, Kwon SI, Corvalan C, Cho SW, Kim H, Kim SG, Kim ST, Choe S, Kim JS: **DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins**. *Nature biotechnology* 2015, **33**(11):1162-1164.
46. Chidananda Nagamangala Kanchiswamy MM, Riccardo Velasco , Jin-Soo Kim, Roberto Viola: **Non-GMO genetically edited crop plants**. *Trends in biotechnology* 2015.
47. Liu J, Gaj T, Yang Y, Wang N, Shui S, Kim S, Kanchiswamy CN, Kim JS, Barbas CF, 3rd: **Efficient delivery of nuclease proteins for genome editing in human stem cells and primary cells**. *Nature protocols* 2015, **10**(11):1842-1859.
48. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F: **Rationally engineered Cas9 nucleases with improved specificity**. *Science* 2015.
49. Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, Joung JK: **High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects**. *Nature* 2016, **529**(7587):490-495.
50. Knight SC, Xie L, Deng W, Guglielmi B, Witkowsky LB, Bosanac L, Zhang ET, El Beheiry M, Masson JB, Dahan M *et al*: **Dynamics of CRISPR-Cas9 genome interrogation in living cells**. *Science* 2015, **350**(6262):823-826.
51. Sternberg SH, LaFrance B, Kaplan M, Doudna JA: **Conformational control of DNA target cleavage by CRISPR-Cas9**. *Nature* 2015, **527**(7576):110-113.
52. Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE: **Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation**. *Nature biotechnology* 2014, **32**(12):1262-1267.
53. Park J, Bae S, Kim JS: **Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites**. *Bioinformatics* 2015.
54. Upadhyay SK, Sharma S: **SSFinder: high throughput CRISPR-Cas target sites prediction tool**. *BioMed research international* 2014, **2014**:742482.
55. Ma M, Ye AY, Zheng W, Kong L: **A guide RNA sequence design platform for the CRISPR/Cas9 system for model organism genomes**. *BioMed research international* 2013, **2013**:270805.
56. Prykhozhij SV, Rajan V, Gaston D, Berman JN: **CRISPR multitargeter: a web tool to find common and unique CRISPR single guide RNA targets in a set of similar sequences**. *PloS one* 2015, **10**(3):e0119372.
57. Sander JD, Zaback P, Joung JK, Voytas DF, Dobbs D: **Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool**. *Nucleic acids research* 2007, **35**(Web Server issue):W599-605.

58. Heigwer F, Kerr G, Boutros M: **E-CRISP: fast CRISPR target site identification**. *Nature methods* 2014, **11**(2):122-123.
59. Naito Y, Hino K, Bono H, Ui-Tei K: **CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites**. *Bioinformatics* 2015, **31**(7):1120-1123.
60. Stemmer M, Thumberger T, Del Sol Keyer M, Wittbrodt J, Mateo JL: **CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool**. *PloS one* 2015, **10**(4):e0124633.
61. Montague TG, Cruz JM, Gagnon JA, Church GM, Valen E: **CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing**. *Nucleic acids research* 2014, **42**(Web Server issue):W401-407.
62. O'Brien A, Bailey TL: **GT-Scan: identifying unique genomic targets**. *Bioinformatics* 2014, **30**(18):2673-2675.
63. Xiao A, Cheng Z, Kong L, Zhu Z, Lin S, Gao G, Zhang B: **CasOT: a genome-wide Cas9/gRNA off-target searching tool**. *Bioinformatics* 2014.
64. Wong N, Liu W, Wang X: **WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system**. *Genome biology* 2015, **16**(1):218.
65. Xie S, Shen B, Zhang C, Huang X, Zhang Y: **sgRNAcas9: a software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites**. *PloS one* 2014, **9**(6):e100448.
66. Chari R, Mali P, Moosburner M, Church GM: **Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach**. *Nature methods* 2015, **12**(9):823-826.
67. MacPherson CR, Scherf A: **Flexible guide-RNA design for CRISPR applications using Protospacer Workbench**. *Nature biotechnology* 2015, **33**(8):805-806.
68. Zhu LJ, Holmes BR, Aronin N, Brodsky MH: **CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems**. *PloS one* 2014, **9**(9):e108424.