Genome Engineering with ZFNs, TALENs and CRISPR/Cas9

Designer Endonucleases

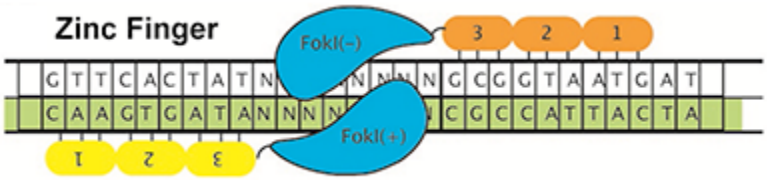
ZFNs (zinc finger nucleases), TALENs (transcription activator-like effector nucleases) and CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/CRISPR-associated system 9) are all engineered endonucleases. These designer nucleases can introduce double-strand breaks (DSBs) at desired locations in the genome. These targeted DSBs can be used to knock-out genes via DNA repair or to introduce modifications via induced homologous recombination.

ZFN technology was the first to become available. Because ZFNs are difficult to construct in ordinary research labs their design is left to commercial sources and are very expensive. TALENs are a lot easier to design. TALEN construction kits are available from Addgene and its procedure is well described in a published paper (e.g., Nucleic Acids Res 39(12): e82). TALEN construction generally requires ligation of more than two dozens of DNA fragment in multiple steps.

Rumor says, that CRISPRs are so easy to construct that anyone can do it. Although many who attempt without sufficient understanding may get into trouble, CRISPR's simplicity and speed is undeniable. Protocols for designing/constructing CRISPR/Cas9 plasmids are readily available (e.g., Nature Protocols 8(11): 2281-2308). To create a CRISPR/Cas9 that recognizes a new target site, the only step required is to insert annealed short oligos into a plasmid which is available from Addgene. With these new tools, one can easily modify the genome of cultured cells or one cell stage embryos. Thus, the field of genome engineering has now become easily accessible to all researchers.

ZFNs

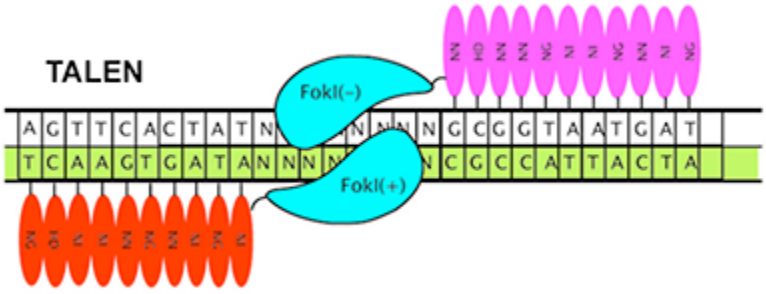
Each engineered zinc finger (shown in the figure below by numbers) recognizes unique nucleotide triplets and a combination of three or more zinc fingers provides high specificity in sequence recognition. Fusion of a FokI monomer and pairing of two such fusion proteins will induce dimerization and activation of FokI which introduces a DSB.



(Credit: Addgene)

TALENs

TALEs (transcription activator-like effectors) were initially discovered in a plant pathogen. The central regions of TALEs are occupied by repeats of nearly identical peptide sequence. The repeat units possess affinity to DNA bases and minor difference within each repeat unit determines to which DNA base it has highest affinity. A combination of 15-20 of such repeat unit provides high specificity in sequence recognition. Like ZFNs, FokI monomers are fused to TALEs. Pairing of two such fusion proteins (TALE Nucleases = TALENs) will induce dimerization and activation of FokI, which introduces a DSB.

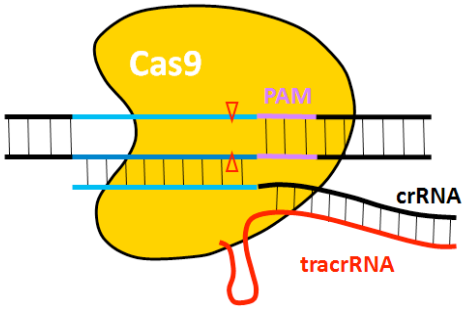


(Credit: Addgene)

CRISPR/Cas9

CRISPR technology has its origin in prokaryotic adaptive immune system. The currently widely used system is a modified version of *S. pyogenes* type II CRISPR/Cas system. In the original form, this system uses three components: two short RNAs (crRNA and tracrRNA) and Cas9 RNA-dependent nuclease. crRNA contains 20-nt

target sequence, or guide sequence, and hybridizes to complementary sequence on genome. A complex structure formed between crRNA and tracrRNA is recognized by Cas9 protein, which is brought to the target site and exerts its nuclease activity.



(Zhang lab drawing; slightly modified)

For convenience, crRNA and tracrRNAs can be fused to make sgRNA (single-guide RNA) which maintains the original activity of crRNA and tracrRNA. For further convenience, both of the two components (sgRNA and Cas9) can be encoded on same plasmid. On those plasmids, sgRNA and Cas9 are transcribed via pol III and pol II promoters, resepectively. To create a designer endonuclease that recognizes a new target site, the only step required is to insert annealed 24-25 nt oligos into a plasmid that already contains the rest of sgRNA sequence and Cas9 cDNA. That makes construction of CRISPR a lot easier than TALENs. Introduction of the finished plasmid into cells is all what is needed to edit (i.e., modify) a genomic site.

TALENs vs CRISPR/Cas9

CRISPR excels in its simplicity, but TALEN has its own advantage when it comes to finding target sites. Although TALEN monomer binding sites have to be preceded by a 5'-T, the flexibility both in TALEN monomer length and in distance between two monomers enables one to find TALEN targets practically anywhere. On the other hand, CRISPR guide sequences need to satisfy a strict rule that they should be immediately followed by a PAM sequence (NGG). Additionally, CRISPR guide sequences need to satisfy another requirement that the chance of off-target editing should be minimized. Because CRISPR/Cas9 system is known to exhibit higher probability of off-target editing, only very unique sequences should to be chosen as CRISPR guide sequences. Thus, the number of usable CRISPR guide sequences may be severely limited in some genomic regions.

Off-Target Editing: Immortalized Cell Lines vs Stem Cells

As soon as the vast potential of CRISPR technology was reported, a big concern arose over off-target editing and many researchers were dissuaded from jumping on the CRISPR bandwagon. In 2014, the September issue of Nature Biotechnology hosted 4 papers that reported almost unbearable frequency of off-target editing by CRISPR. However, as time passed, the concern over off-target site editing started to wane gradually and it was replaced by enthusiasm again. This was mainly due to people's realization that the problem of off-target editing is controllable.

Those four studies that reported significant levels of off-target editing used immortalized cell lines, such as HEK293 or K562. Those reports were stark contrast to others that reported low frequency of off-target editing in human iPSCs (e.g., a couple of papers in July issue of Cell Stem Cell) and single-cell stage mouse embryos. Since immortalized cell lines possess aberrantly high DNA repair ability, it is expected that immortalized cell lines exhibit elevated levels of off-target editing.

Remedy for Off-target Editing

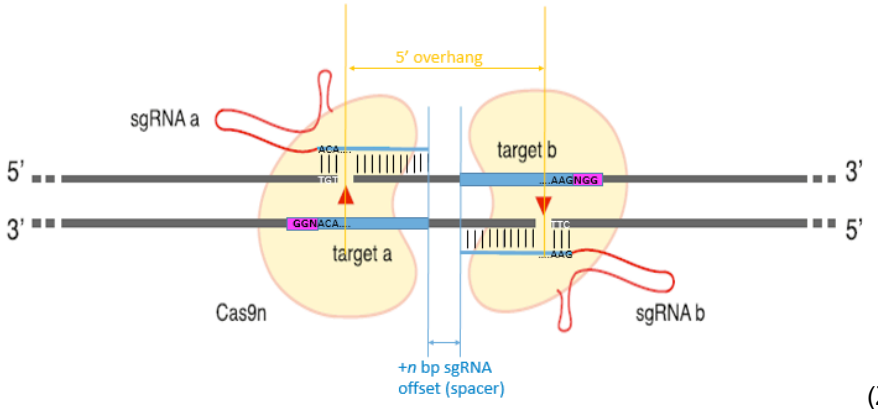
How should we deal with off-target effects, especially when immortalized cell lines are treated with CRISPR/Cas9 plasmid for the purpose of knocking-out a gene? First, one should use CRISPR guide sequences that are given "high score". Multiple online tools are available for finding guide sequences (e.g., http://crisp.mit.edu/) and those tools provide "score" to indicate probability of off-target editing. Usually, the higher the score is, the lower the chance of off-target editing will be. A problem here is that each online tool uses different scoring definitions. For example, a score 100 by one online tool might be score 70 by another tool. Online tools also predict which off-target sites are more likely edited than the other so that researchers can examine those predicted off-target sites for absence/presence of off-target editing.

As noted above, off-target editing in embryonic stem cells/iPS cells, or in single-cell stage mouse embryos, occurs only at manageable low frequency. Use of high score guides yields very specific editing with very few off-target editing. In case of mouse embryos, if off-target editing occurs, founders may exhibit confound phenotype. But even in such case, the unwanted editing can be bred away through repeated crossing.

One remedy for the unpredictable off-target editing is to generate different knock-out cell lines using two totally different guide sequences and then compare phenotype between the two knock-out lines. If identical phenotype is observed, it is likely due to the gene knockout. Another approach, which we recommend, is to use the dual nickase system. A nickase is a mutated version of Cas9 that generates a break on only one of the two DNA strands. You can expect very low off-target editing from nickases, which will be further explained below. A similarly low off-target editing was reported when "dimeric CRISPR RNA guided FokI nuclease" was used.

How Dual Nickases Work

Cas9 nuclease contains two domains each of which independently generates a single-strand break on opposing DNA strands. By mutating only one of the two domains, Cas9 protein becomes "nickase". By "guiding" two nickases close to each other in such a way that opposing strands will get nicked in close proximity, a DSB can be generated. Any off-target site that gets a single nick should be repaired immediately without unwanted mutation. Chances of two nickases come together at off-target sites in appropriate arrangement is extremely low. This enables even low score guide sequence to become a partner of dual nickase without excessively increasing chance of off-target editing. There are online tools that provide list of dual nickase from which we can choose desired pairs. One important factor we must take into consideration when choosing nickase pairs is the offset length (i.e., distance between two guide sequences) since it affects efficiency of DSB formation (e.g., Cell 154: 1380-1389).



(Zhang lab)

Low/Zero Off-target Editing by Dual Nickase

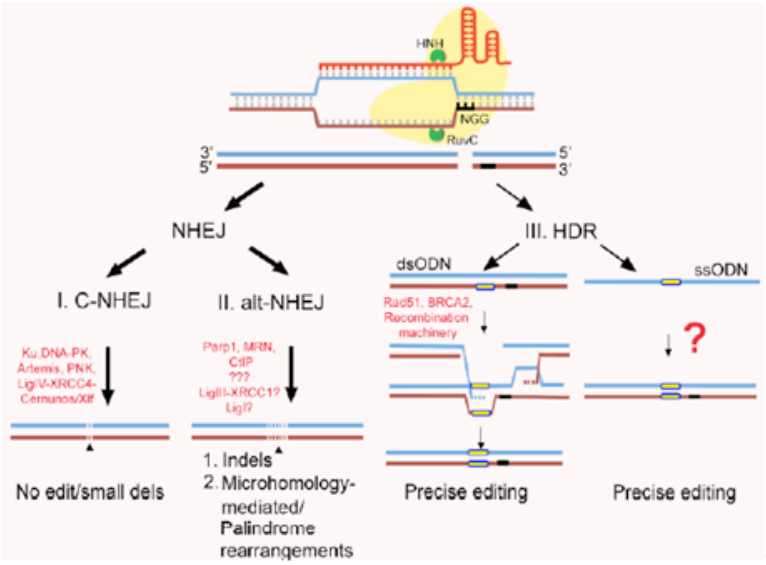
Theory that predicts zero or low off-target editing by dual nickase sounds so good. But is it true in real world? A couple of studies appeared in 2015 February issue of Nature Biotechnology confirmed this (Nat Biotechnol 33: 175-178, 179-186). Previously, locations of potential off-target sites were "predicted" by homology to guide sequence. The studies in February issue of Nature Biotechnology used a different approach. Their approach did not rely on prediction by sequence homology. Instead, they *in vivo* sequence-tagged ds-break at unspecified loci, which were subsequently identified by sequencing. Their finding indicated that dual nickases are indeed practically off-target free. A similarly encouraging result was obtained for TALENs.

Dimeric CRISPR RNA-guided FokI Nuclease

A fusion of FokI nuclease to dCas (i.e., mutated Cas9 with no nuclease activity) is a different approach to the problem of off-target editing. It was claimed that this system exhibits even lower off-target editing than dual nickase. This system has a problem, however. Because there is much stricter restriction on the offset length (around 16 +/- a few bp), finding an appropriate pair can be difficult.

Non-Homologous End-Joining (NHEJ) and Homology Directed Repair (HDR)

Designer endonucleases (ZFNs, TALENs, CRISPR/Cas9) generate DSBs on genome. DSBs are repaired either by non-homologous end-joining (NHEJ) pathways or by homology-directed repair (HDR) pathway. There are two NHEJ pathways: canonical non-homologous end-joining (C-NHEJ) and alternative non-homologous end-joining (alt-NHEJ). C-NHEJ pathway provides the quickest repair and results in either wild-type sequence (i.e., no change) or short random deletion. Alt-NHEJ pathway and HDR pathway become more prominent if C-NHEJ is blocked. Alt-NHEJ pathway yields variety of changes: from single base insertion/ replacement to large deletion/insertion or inversion etc. HDR pathway uses homologous DNA template to repair the lesion. Therefore, HDR can be utilized to introduce precise modification into genome.



(Genetics, 199(1): 1-15)

Gene Knock-out by NHEJ

NHEJ provides a very easy quick way to knock-out gene of your interest. When a DSB within a coding-exon is repaired by NHEJ, most frequent outcome is a short deletion and less frequently we may see a large deletion or an insertion of random sequence. Such deletion or insertion, often called "indels", can result in frame-shift which generates a pre-mature stop codon. If such pre-mature stop codon is created more than 50-55 bp upstream of last exon-exon boundary, phenomenon called nonsense mediated decay of mRNA (NMD) is induced and the entire message will be degraded rapidly. The result is a production of no peptide, not even a partial peptide. This phenomenon makes the basis of gene knock-out by CRISPR, as well as by ZFNs and TALENs. Because frame-shift may not always lead to NMD (e.g., indel may cause splice pattern change), complete gene knock-out must be confirmed by Western analysis.

Precise Modification of Genome by HDR

While gene knockout can be achieved by unpredictable nature of NHEJ, designed modification of genome can be achieved by sometimes finicky HDR pathway. Repair of DSB by HDR is much less frequent occurrence compared to repairs made by NHEJ. So, in order to increase chance of HDR, we block C-NHEJ pathway transiently with an inhibitor. Along with designer nucleases we co-introduce "donor DNA", which is a DNA template for designed modification. A donor DNA consists of a new sequence to insert or replace, which is sandwiched by homology arms.

An important consideration to be made when designing donor sequence is that re-cutting of target site after successful genome modification would not occur. Either placement of an insert or incorporation of a silent mutation within target sequence is a commonly used strategy to prevent re-cutting. If a very short sequence ranging from a few bases to 50 bases are to be inserted (for example, subtle mutation or insertion of epitope tag), single-stranded oligonucleotide (ssODN) provides high HDR efficiency. Frequently, 60 base long homology arms are added to each side.

When a larger insertion is planned (for example, insertion of a reporter gene), double-stranded DNA (ds DNA) is used as donor. Plasmid DNA is often used for such purpose. Homology length used in ds donor DNA is between 500 bp and 1 kb. For some reason, many people seem to prefer 800 bp for each homology arms. HDR efficiency with ds donor DNA is significantly lower than ssODNs. Maximum size of cargo which designer nucleases can insert into mouse genome by co-injection of donor DNA is reported to be around 3 kb. This number may increase as technologies gain improvements.

Other than the size of cargo to be inserted and the activity of individual designer nuclease, our own experience indicates that local genomic sequence can affect HDR efficiency as well. We have observed that suppression of C-NHEJ with an inhibitor at particular genomic site diverted DSB repair activity toward alt-NHEJ without improvement in HDR. Sequence within homology arms may also affect HDR efficiency. Another important factor to consider is the distance between intended modification site and DSB site. The larger the distance is, the lower the HDR frequency will be. Therefore, the efficiency of NHEJ at a particular site, which is often assayed when new guide sequences are designed, does not necessarily correlate with HDR efficiency. If local genomic sequence is not favorable for HDR, it may make HDR by micro-injection at particular site extremely difficult.

On the other hand, in cell-based work where a selection marker can be included, isolation of successfully recombined clones is less laborous and the limit for cargo size increases significantly. From our own experience

with ZFN technology, the maximum cargo size for cell-based work with selection marker is expected to be far beyond 10 kb.

Micro-injection of Designer Nuclease into Mouse Embryos for HDR

If the factors mentioned above do not impede HDR efficiency, microinjection of designer nuclease with donor DNA into mouse embryos provides the quickest way to generate transgenic mice with precise modification (e.g., a point mutation, insertion of a tag). Perhaps the most commonly used procedure injects RNA (sgRNA and Cas9 mRNA, or TALEN/ZFN mRNAs) along with donor DNA.

It is now well recognized that injected Cas9 mRNA has a longer life span than desired. Genome editing by surviving Cas9 may take place at 2- or 4-cell stage or beyond, resulting more than two mutations in single mouse (mosaicism). Mosaicism occurs rather frequently and it can cause some complication when analyzing founder mice. Injection of Cas9 protein may reduce such problem.

Injection of CRISPR plasmid DNA into mouse embryos has been reported to work without "run away expression" of the nuclease. However, at least in our hands, efficiency of genome editing by plasmid DNA appears to be very low. Lower genome editing efficiency by plasmid DNA injection, compared to RNA injection, has been reported for ZFNs and TALENs as well.

CRISPR Mediated Genome Engineering of Immortalized Cell Lines

As researchers have learned how easy CRISPR technology can be, many are trying to capitalize it in their own labs. Perhaps the most popular goal is to knock-out their own gene of interest in their favorate cell lines.

A number of commercial companies are also into this business. One company is offering haploid cell lines in which numerous genes have been already knocked-out. There is a big problem when using haploid cells, however. Because genetic background can dramatically alter phenotype, it is much preferred to apply CRISPR technology to the cell line that has been used in one's own research.

Many companies offer their service for popular immortalized cell lines. Those immortalized cell lines generally have extra copies of chromosomes and knocking out a gene requires knocking out all copies. This could cause a trouble since some cell line may have up to 8 to 10 copies of same chromosome. When proceeding to make a specific knockout cell line, it is important to know gene copy number. You can find chromosome numbers for popular cell lines on the web (e.g., http://www.ncbi.nlm.nih.gov/sky), but such information is not available for every cell line. Most chromosomes in popular cell lines, such as HeLa, K562 or HEK293, have around 3 or 4 copies, and are therefore within an attainable range. Because of high copy numbers, the price a company may offer is not particularly low, (between $5,000 and $15,000 for a knock-out cell line).

As mentioned earlier, immortalized cell lines exhibit off-target editing at higher than usual frequency. Therefore, it is important to consider potential off-target effects when designing a project on immortalized cell lines. If wild-type Cas9 is used, isolate at least two knock-out cell lines using different CRISPR guide sequences. At the Clara Christie Centre for Mouse Genomics, we use the dual nickase system.

Test Your Designer Nuclease before Using

The activity of individual designer nucleases can vary widely and some of them may not exhibit any detectable activity at all. For this reason, it is recommended to design at least a couple of them and test them before use. Nuclease activity can be assayed by multiple methods that detect NHEJ activity in a cell population.

Perhaps most popular method uses the SURVEYOR Mutation Detection Kit, also known as the CEL-I assay. First, we isolate a sequence surrounding the DSB by PCR. The PCR products are heat-denatured and then re-annealed. Since the edited genome and the wild-type genome share extensive homology, re-annealing yields hybrids with small number of mismatched bases in the middle of the PCR fragment. CEL-I is a single-strand specific endonuclease that recognizes and cleaves such mismatched sites. When CEL-I digest is separated on agarose or polyacrylamide gel, appearance of shorter fragments at expected positions tells the presence of NHEJ activity. An essentially identical principle applies to T7EI assay in which an enzyme similar to CEL-I is used. We have determined that the re-annealed PCR products can be directly analyzed on native PAGE. Heteroduplexes exhibit slower mobility than that of homoduplexes. Thus, the presence of NHEJ activity can be detected even without using CEL-I nuclease.

A third method, TIDE assay, requires sequence chromatograms. When sequencing reaction is carried out on a mixed population, the result is overlapping peaks of multiple bases. By comparing chromatogram of test sample to that of wild-type genome, an online tool (TIDE: http://tide.nki.nl/) decomposes the test chromatogram and provides rough estimate of genome editing efficiency.

Results from these assays will decide which nuclease (or nuclease pair) is best for further work, such as isolation of cell clones whose gene of interest has been totally knocked out. The method to identify such knock-out cell clones is western analysis because NMD is expected to disrupt protein expression.

Other Applications of Designer Nuclease

One advantage CRISPR technology offers is its ability for multiplex genome editing. Simultaneous knocking-out of multiple genes, large deletion of genomic DNA by targeting two sites on same DNA molecule, or creation of conditional allele by inserting loxP at two sites on the same DNA molecule are such examples and there have been reports of spectacular success. Such success cannot be achieved by every attempt, however. Instead, many attempts by number of people resulted in very disappointing and frustrating results. At the moment, we do not know exactly what factors are determining the outcomes. Since HDR efficiency is determined not only by activity of individual designer nucleases but also by multiple other unknown factors, prediction for outcome is extremely difficult to make.

A very different kind of application for CRISPR technology is suppression or activation of genes. Such application uses a mutated Cas9 protein that does not possess nuclease activity (dCas) and acts as RNA-guided DNA-binding protein. Placement of dCas over promoter elements could downregulate a gene activity. On the other hand, tethering of activator to dCas could activate a gene. Tethering of a chromatin modifying enzyme is another idea multiple labs are currently pursuing. We are looking forward to seeing further development in those fields.