Trends in Biotechnology

CellPress

Focus on Genome Editing

Review

Genome Editing of Structural Variations: Modeling and Gene Correction

Chul-Yong Park,[1] Jin Jea Sung,[1] and Dong-Wook Kim[1,*]

The analysis of chromosomal structural variations (SVs), such as inversions and translocations, was made possible by the completion of the human genome project and the development of genome-wide sequencing technologies. SVs contribute to genetic diversity and evolution, although some SVs can cause diseases such as hemophilia A in humans. Genome engineering technology using programmable nucleases (e.g., ZFNs, TALENs, and CRISPR/Cas9) has been rapidly developed, enabling precise and efficient genome editing for SV research. Here, we review advances in modeling and gene correction of SVs, focusing on inversion, translocation, and nucleotide repeat expansion.

Engineered Nucleases and Structural Variations

Genome engineering methods have evolved dramatically over the past few decades. The oldest genome modification methods are site-specific recombinase systems using the cyclization recombinase (Cre) in conjunction with the *loxP* sequence, or flippase (Flp)–*FRT*. These systems first insert the *loxP* and *FRT* flanked sequences to specific target loci in the genome, and then induce site-specific recombination by introducing Cre or Flp enzymes, which enable the rearrangement of genomic DNA [1–3]. Following these early methods, genome modification using endonucleases was introduced. Some of the earliest introduced endonucleases were meganucleases. Originally, they were mobile elements that created site-specific **double-strand breaks** (DSBs; see Glossary) by recognizing relatively long sequences (12–45 bp), and they thus have been recognized as an efficient method of gene manipulation [4,5]. Next, as a customized engineered nuclease, the **zinc-finger nuclease** (ZFN), was produced by the fusion of tandem repeats of sequence-specific DNA-binding zinc fingers and the *Fok*I endonuclease domain [6–8]. The next generation of engineered nuclease was **transcription activator-like effector nuclease** (TALEN), which is similar to ZFN but instead recognize targets via the TALE DNA-binding domains [9,10]. In both systems, the local dimerization of two monomers of *Fok*I, assembled in tail-to-tail orientation, cleaves its target.

The most recently developed engineering method uses **clustered regularly interspaced short palindromic repeats** (CRISPR) and the **CRISPR-associated protein 9** (Cas9) system, also known as the RNA-guided engineered nuclease (RGEN) system. CRISPR/Cas9 is based on the adaptive immune system of prokaryotes, and it consists of endonuclease Cas9 and guide RNA (gRNA) [11–15]. The CRISPR/Cas9 system differs from its predecessors in that its target recognition ability depends on the spacer sequence in gRNA. However, all engineered nucleases share a common feature – the separation of the endonuclease and the DNA recognition activities. Therefore, it is easy to construct custom endonucleases by redesigning DNA-binding sites for different purposes, such as target-specific gene disruption or insertion, regulation of

Trends

SVs are found in healthy individuals and have a positive effect on genetic diversity. However, through development of genome-wide sequencing technologies, it has been revealed that SVs can be involved in some diseases.

The development of engineered nucleases such as ZFNs, TALENs, and CRISPR/Cas9 has led to breakthroughs in the studies of diseases with SVs in terms of understanding mechanisms, disease modeling, and gene correction.

We discuss current advances and future strategies in modeling and gene correction of SVs using engineered nucleases.

[1]Department of Physiology and Brain Korea 21 Plus Project for Medical Science, Yonsei University College of Medicine, Seoul 03722, Korea

*Correspondence: dwkim2@yuhs.ac (D.-W. Kim).

 http://dx.doi.org/10.1016/j.tibtech.2016.02.011

Table 1. Comparison of Different Nuclease Systems

	Meganuclease	ZFN	TALEN	Cas9
Length of Target Site	12–45 bp	18–36 bp	30–40 bp	22 bp
Off-Target Effects	Low	Variable[a]	Low	Variable[a]
Nuclease Size	1 kb	1 kb	3 kb	4.2 kb + 0.1 kb[b]
Accessibility for Engineering	Difficult	Difficult	Moderate	Easy
Multiple Targeting	Difficult	Difficult	Difficult	Easy
***In vivo* Delivery**	Easy	Easy	Difficult	Moderate
DNA Targeting System	Protein	Protein	Protein	Protein/RNA

[a]In ZFN and Cas9 systems, off-target effects are variable depending on their target sites.
[b]In the Cas9 system, the gRNA sequence is about 100 nts.

gene expression, and correction of genetic mutations [16–19]. The advantages and limitations of these nucleases are summarized in Table 1.

Because the specificity of nucleases is not absolute, sequences that are similar to a target sequence may be cut, which may result in unintended mutations caused by **off-target** cleavage. Recently, genome-wide unbiased identification of DSBs enabled by deep sequencing after tagging phosphorothioate-protected double-stranded oligonucleotides to intracellularly generated DSBs (GUIDE-seq) and digested genome sequencing (Digenome-seq) that can map DSBs developed directly by Cas9 at the genome-wide level were developed for unbiased off-target site searching and the assessment of nuclease accuracy [20,21]. When these methods were used to analyze off-target effects, they were better able to detect off-target effects that were not predicted by the conventional *in silico* method. In addition, the use of ggX_{20} single-guided RNAs (sgRNAs) was shown to increase specificity by a fold of 660 relative to conventional sgRNAs [21], and the use of truncated gRNA reduced the number of off-target cleavages [20], enabling more accurate nuclease design. These advances are important in terms of clinical application of the nucleases.

Structural variations (SVs) are large-scale variations such as translocations and inversions that occur when the broken ends of the DSBs rejoin during the repair process [22]. The introduction of SVs results from non-allelic **homologous recombination** (HR), **non-homologous end-joining** (NHEJ), fork stalling and template switching (FoSTeS), and retrotransposition [23]. SVs appear in DNA segments longer than 1 kb [24], but recently their definition has been expanded to encompass smaller events (i.e., those greater than 50 bp in length) [25]. Although SVs are found in healthy individuals, and have a positive effect on genetic diversity, they can be the cause of genetic diseases in some cases [22]. The diversity and importance of SVs have been in the spotlight as genome-wide sequencing technologies advance. Furthermore, targeted genome engineering technology using site-directed nucleases has contributed greatly to the field of SVs. We describe here how genome-editing technology with engineered nucleases is bringing new progress in studying inversion, translocation, and **short nucleotide repeat** expansions among the various SVs.

Genome Editing of Inversions

Chromosomal inversion is a phenomenon where the homologous sequences between two loci on a chromosome are reversed 180 degrees and rejoined when breakages occur. Chromosomal inversions in the human genome vary in size from 1 kb to several Mb [26]. Chromosomal inversions sometimes cause genetic diseases, such as hemophilia A. Before the development of ZFNs, chromosomal inversions were induced using systems such as Cre–*loxP* recombination, meganuclease I-*Sce*I, and Flp–*FRT* [27–29], which enabled the generation and development of SV models and the study of the function of genes involved in cancer and other

Glossary

Chromosomal inversion: a chromosomal rearrangement caused by the breakage and reinsertion of a specific chromosomal segment in a reversed orientation.
Chromosomal translocation: a chromosomal rearrangement caused by the exchange of specific segments in non-homologous chromosomes.
Clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated (Cas) system: an adaptive immune system found in bacteria and archaea which enables efficient genome editing in various types of cells and organisms. This system is also known as an RNA-guided engineered nuclease (RGEN) system consisting of Cas9 endonucleases and guide RNA. The guide RNA determines DNA specificity.
Double-strand break (DSB): a break in a double-stranded segment of DNA. DSBs are induced by stimuli such as ultraviolet light and chemicals, but can also be caused by nucleases in a specific manner.
Homologous recombination (HR): a DNA break repair pathway resulting in error-free repair through the involvement of homologous sequences. This occurs during the G2 and S phases of the cell cycle.
Indels: the insertion or deletion of a small portion of the chromosome, which occurs mainly during error-prone DNA repair processes.
Induced pluripotent stem cells (iPSCs): cells reverted to an undifferentiated state through the artificial expression of reprogramming factors in a somatic cell. iPSCs have characteristics similar to those of embryonic stem cells.
Non-homologous end-joining (NHEJ): a DSB repair mechanism by which two non-homologous ends are directly ligated. Indels are occasionally induced by this mechanism.
Off-target effects: unwanted modifications that occur when engineered nucleases recognize and cut an unwanted sequence.
Short nucleotide repeats: the repetition of 3–6 nts as a unit. Depending on their expansion size, short nucleotide repeats may cause diseases.
Transcription activator-like effector nucleases (TALENs):