CRISPR-Cas9 Genome Engineering

Precise genome engineering/editing can help us understand how the genome functions in mammals. The development of genome engineering technology can benefit a broad range of industries, such as materials, fuel, food, surgery, and drug development [1]. In this mini-review, we focus on the emerging technology CRISPR (clustered regularly interspaced short palindromic repeats) - Cas9 (CRISPR-associate-9). CRISPR is a segment of DNA that has short repetitions of base sequences. Cas9 is a nuclease that cleaves DNA. We review its working principle in the immune system of type II bacteria and mammals, its advantages over other techniques, as well as the major challenges in its development.

1 Working principle

1.1 Overall process

The whole process of genome engineering includes two major parts: specific cutting and repair of the DNA sequence.

Specific cutting has a few pathways: double-stranded break (DSB) (Section 1.2Specific locus cleavage by Cas9), nicking, double-nicking (Section 3.1.1) [1].

The repair process has two pathways, nonhomologous end-joining- (NHEJ)- and homology-directed repair- (HDR)-mediated genome editing. HDR is based on homologous recombination (HR). Homologous recombination (HR) refers to genetic exchange between a pair of homologous DNA sequences [1].

1.2 Specific locus cleavage by Cas9

In the Type II bacteria, the immune system with Cas9 works as follows. CRISPR gene is composed of repeating sequences and spacers. The spacers are copies of short sequences of the phage that invaded the bacteria in the past, as “embedded memories” to trigger the defense action (Figure 1). The target DNA sequences in the invading phage that will be cleaved are named protospacers.

CRISPR Array is transcribed into a long CRISPR RNA (crRNA) array. The spacers in the crRNA array are converted into crRNAs that act as small guide RNAs. The guide RNA is a 20-nucleotide guide sequence that can direct the Cas9 to find specific locus (specific site on the target DNA) to cleave (Figure 1). Another strand of RNA is formed by complementary base-pairing with crRNA, named trans-acting RNA (tracrRNA). The Cas9 protein is linked to the dual RNAs to form a complex. Protospacer adjacent motif (PAM) is the motif (short sequence, 2-6 base pairs) in the DNAs adjacent to the protospacers (Figure 1).

the Cas9 nuclease will cleave a target DNA sequence, if

- the target DNA sequence (protospacer) is equal to the guide RNA sequence (spacer), and
- the PAM has a sequence of NGG, where N is any nucleotide.

The PAM on the virus is NGG in Streptococcus pyogenes, whereas the PAM on its CRISPR Array is GTT. Such difference makes the Cas9 nuclease distinguish between the DNA sequence of the invading virus and the sequence in CRISPR array in the DNA of the bacteria. PAM is short and faster to recognize than just by matching the long target DNA sequence that requires to unwind the DNA [1].

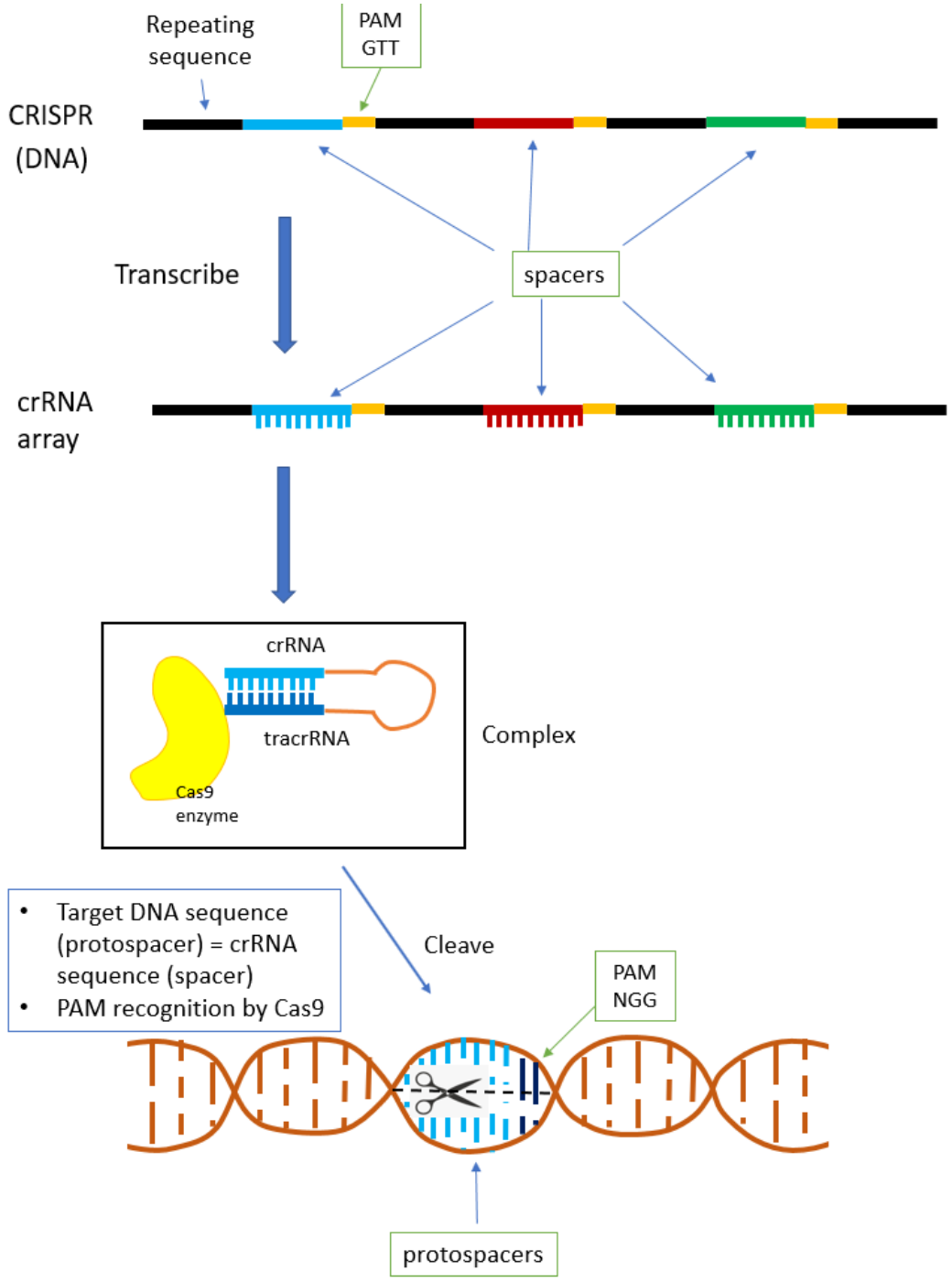

Figure 1 Working principle of Cas9 genomic editing.

1.3 Genome engineering in mammalian cells by Cas9

For genome engineering in mammals, scientists generate the crRNA array and tracrRNA by heterologous expression in mammalian cells. In the case of Cas9, they utilize human codon to obtain an optimized Cas9. The crRNA and tracrRNA can fuse into a single-guide RNA (sgRNA) that has a 20-nucleotide guide

sequence (Figure 2). It is also feasible to inject sgRNA and mRNA that encodes Cas9 into embryos [2]. These experiments hold big promise for genome editing in mammalian cells [1].

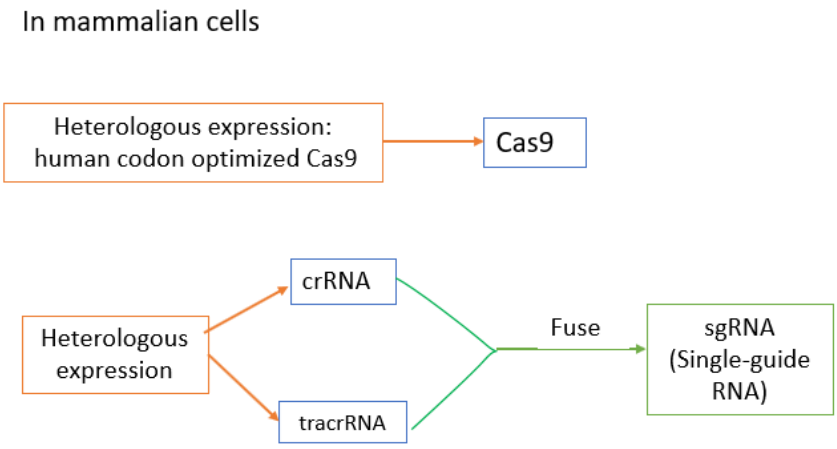

Figure 2 Working principle of the synthesis of Cas9 system in mammalian cells.

1.4 HR after Cas9 cleavage

To pursue the pathway of HR-mediated repair, we need a donor DNA that gives its sequence to recombine with the genome in the target DNA (Figure 3). In the work of DiCarlo et al. [3], the authors synthesized a donor DNA with its PAM removed. They obtained nearly 100% HR efficiency in haploid yeast cells *Saccharomyces cerevisiae*. Removing the PAM sequence from the donor DNA can protect recombinant cells from further activity and toxicity from CRISPR system [3].

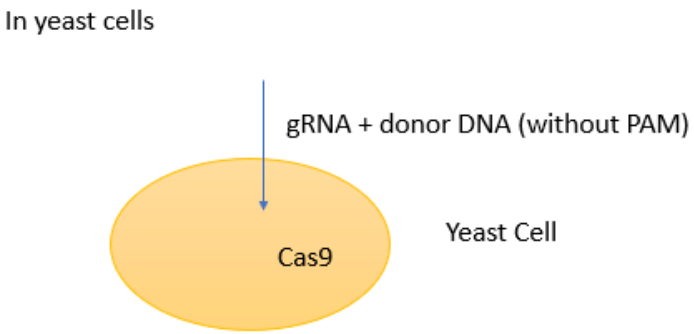

Figure 3 Designed protocol for high HR frequency in the yeast cell.

1.5 Limitation of NHEJ-mediated repair

NHEJ induces indels (insertion or deletion of bases) of random lengths. It leads to undesired genetic mosaicism in mice [1].

2 Advantages of Cas9 over other genome editing technologies

2.1 Easy to synthesize

The guide molecule for Cas9 is an RNA molecule that is easier to synthesize than a DNA-binding protein, such as zinc finger (ZF) protein and transcription activator-like effector (TALE) [3]. Such convenience

further enables Cas9 to easily retarget a new DNA sequence, as well as target multiple genomic loci simultaneously.

3 Challenges

3.1 Off-target activities

Cas9 has off-target activities because it tolerates mismatches. Off-target activities include off-target binding and off-target cleavage [1].

Cas9 has many off-target binding sites but cleaves only a small fraction of them. Its concentration is an important factor to fine-tune its ratio of on-/ off-target binding sites and efficiency of on-target cleavage [4]. At high concentrations, Cas9 tolerates more mismatches and has more off-target binding sites with high efficiency of on-target cleavage. And vice versa [5].

Some researchers used computational tools as experimental methods to comprehensively predict off-target sites [6]. The main goals are to understand deeply the working principle of Cas9 and to reduce the off-target cleavage.

3.1.1 Nuclease mutation

Cas9 has two nuclease domains, HNH and RuvC, for strand-specific cleavage. The two domains can be mutated for other functions. RuvC domain can have an aspartate-to alanine (D10A) mutation that generates the Cas9 nickase mutant (Cas9n) to nick rather than to cleave the DNA sequence. The double-nicking strategy using Cas9n can reduce the amount of off-target cleavage [7].

3.2 Homologous recombination efficiency

HR efficiency with Cas9 (1 in 10^6 -10^9 cells) needs to be improved. The protocol of Dicarlo et al. in yeast cells can be a good reference for protocols in mammals (Section 1.4) [3].

3.3 Synthesis of guide RNA

The guide RNA needs have a specific 20-nucleotide sequence. To avoid undesired modifications during the synthesis of guide RNA, some researchers preferentially used RNA polymerase III regulatory elements instead of RNA polymerase II system [7].

4 Reference

[1] P. D. Hsu, E. S. Lander, and F. Zhang, “Development and Applications of CRISPR-Cas9 for Genome Engineering,” *Cell*, vol. 157, no. 6, pp. 1262–1278, 2014.

[2] L. Cong *et al.*, “Multiplex Genome Engineering Using CRISPR/Cas Systems,” *Science (80-.).*, vol. 339, pp. 819–824, 2013.

[3] J. E. Dicarlo, J. E. Norville, X. Rios, J. Aach, and G. M. Church, “Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems,” *Nucleic Acids Res.*, vol. 41, no. 7, pp. 4336–4343, 2013.

[4] X. Wu *et al.*, “Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells,” *Nat. Biotechnol.*, vol. 32, no. 7, pp. 670–676, 2014.

[5] Y. Fu *et al.*, “High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells,” *Nat. Biotechnol.*, vol. 31, no. 9, pp. 822–826, 2013.

[6] R. Gabriel *et al.*, “An unbiased genome-wide analysis of zinc-finger nuclease specificity,” *Nat. Biotechnol.*, vol. 29, no. 9, pp. 816–823, 2011.

[7] F. A. Ran, P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott, and F. Zhang, “Genome engineering using the CRISPR-Cas9 system,” *Nat. Protoc.*, vol. 8, no. 11, pp. 2281–2308, 2013.