Advances in Genome Engineering Approaches

Arvind Chhabra*

Department of Medicine, University of Connecticut Health Center, Farmington, Connecticut, USA

Abstract

Targeted genome editing is essential for functional characterization of a gene of interest. Targeted gene inactivation via homologous recombination made it feasible to create gene knockout animal models to ascertain the physiological role of the target genes; however, lower efficiency of site specific insertion of the genetically modified construct through homologous recombination has limited a wider applicability of this approach. Development of targeted gene knockdown through RNA interference (RNAi) offered a cost effective, high-throughput alternative to homologous recombination, however, RNAi-mediated gene knockdown is incomplete, produces experiment to experiment variation, and could provide only a temporary inhibition of the gene function. Development of genome engineering methodologies utilizing nucleases linked to the guide sequences targeting a gene of interest, such as Zinc Finger Nucleases (ZFN), Transcription Activator like Effector Nucleases (TALEN) and Clustered Palindromic Repeats (CRISPR), are quite encouraging. A brief overview of recent advances in genome engineering approaches is provided with their respective advantages and limitations.

Keywords: Genome engineering; MicroRNAs; ZFN; TALEN; CRISPR

Abbreviations: RNAi: RNA interference; ZFN: Zinc Finger Nucleases; TALEN: Transcription Activator like Effector Nucleases; CRISPR: Clustered Palindromic Repeats

Small RNA-Mediated Genome Engineering

RNA has long been considered a messenger of the genetic code imprinted in the DNA of a cell. However, the discovery of first small RNA based gene silencing of lin-4 gene in C. elegans [1,2], offered novel insights towards a regulatory role of RNA in the cellular physiology, and set the foundation for the identification of small RNAs modulating gene expression. It is now well known that eukaryotes produce many different kinds of small RNAs (19-30 nucleotides in length) that play essential roles in regulation of developmental cascades as well as cellular stress conditions, for example viral infections, through multiple mechanisms, such as direct mRNA degradation or translational repression. Among the most common small RNAs include small interfering RNAs (siRNA) and microRNAs. After initial processing in the nucleus by Drosophila, precursors RNAs (pre-RNAs) are transported to the cytoplasm, where Dicer cleavage generates mature microRNAs and siRNAs [3-5]. While miRNAs are generated from the double stranded RNA (dsRNA) of the hairpin-shaped precursors, siRNAs are generated from long ds RNAs. These double-stranded products then assemble with Argonaute proteins in a protein complex known as the RNA-induced silencing complex (RISC) [3-5], such that their one strand is preferentially selected to guide the sequence-specific modulation of complementary target mRNA through mRNA cleavage, translational repression, or cleavage independent mRNA decay [3-5]. Systematic studies characterizing the sequence requirements for developing an efficient siRNA resulted in development of tools for creation of custom siRNAs for targeting gene of interest [6,7], and this coupled with the development of vector as well as non-vector based small RNA delivery methods, made siRNAs a preferred method to ascertain the functional relevance of any given gene of interest, in-vitro as well as in-vivo. However, RNAi-mediated approaches provide an incomplete ablation of target gene and potential off-target effects make it difficult to draw accurate conclusions.

Although several natural siRNAs have been reported in different eukaryotic species, in mammals, microRNAs represent the major natural small RNAs that have been identified. Among the approaches used for the identification of microRNAs include northern hybridization, microarray based approaches and small RNA cloning etc. Several microRNAs have been found to be expressed in embryonic stem cells as well as in distinct hematopoietic cell populations [8,9]. Many microRNAs have been associated with modulation of different developmental pathways [10,11]. MicroRNAs have also been suggested to modulate functional characteristics of host immune system, for example microRNAs have been identified in different effector T cell population with distinct functional profiles, such as Th1/Th2/ regulatory CD4 T cells as well as CD8 effector T cells [12-14]. While on one hand these findings demonstrate an active role for microRNAs in modulation of cellular physiology, it also offers a unique opportunity for targeted genome engineering. Identification of naturally occurring microRNAs that can regulate specific physiologic pathways would be quite significant in developing microRNA based genome engineering methodologies. For example, identification of naturally occurring microRNAs that can regulate specific developmental pathways from human hematopoietic cells (HSC) can be utilized for generating cells of a specific lineage from engineered HSCs. Similarly, identification of microRNAs regulating the generation of specific effector function in T cells can help generate immune effectors with pre-defined functional profile. However, identification of specific microRNAs that can meet these functional end points without any off-target effects is essential to achieve these goals.

Nuclease-Mediated Genome Engineering Approaches

Targeted genome editing in embryonic stem cells through homologous recombination revolutionized the field by allowing the creation of different knock-in/knockout mice strains to ascertain the functional relevance of a gene of interest [15,16], however, as...
monomer repeats, each recognizing a single base utilize DNA binding domain comprised of 33-35 amino acid tandem each TALEN repeat (33-35 amino acids) recognizes only 1 nucleotide. The key difference between ZFNs and TALENs is that each and resulting in dimerization of Fok1 and cleavage of the target DNA fused TALENs binding in the opposite directions of the target gene, allowing on target specificity. Binding on the zinc finger modules to the target sequence results in Fok1 dimerization, activation of its nuclease activity and double stranded break (DSB) in the target sequence spacer region. Once DSB is introduced, it leads to activation of cellular DNA repair mechanisms, such as the error prone non homologous end joining (NHEJ) or the DNA homology mediated repair by homologous recombination (HR), culminating in gene disruption by homologous template introduced or NHEJ mediated nucleotide insertion, deletion or frame-shift mutations. Among the key constraint in designing an effective ZFN is their natural preference for G-rich sequences. Furthermore, making a single designer DNA-binding domain is easier than joining two domains in an appropriate orientation, with correct spacing, to yield a functional nuclease. These technological challenges make in house development of an effective ZFN quite difficult for most laboratories.

**ZFNs-mediated genome engineering**

Cys2-His2 zinc-finger domain in among the most common types of DNA-binding motifs found in eukaryotes and represents the second most frequently encoded protein domain in the human genome. ZFNs utilize the DNA binding ability and specificity of zinc finger proteins (ZFPs), along with the nuclease function of Fok1 nuclease for targeted genome engineering. The zinc finger domain is comprised of 3-6 zinc finger modules linked together, with each module comprised of 30 amino acids and recognizing 9-18 base pairs. Since DNA cleavage activity of Fok1 nuclease functions only in dimer form, two zinc finger modules are created recognizing 9-18 base pairs on opposite sides of a target gene, allowing on target specificity. Binding on the zinc finger modules to the target sequence results in Fok1 dimerization, activation of its nuclease activity and double stranded break (DSB) in the target sequence spacer region. Once DSB is introduced, it leads to activation of cellular DNA repair mechanisms, such as the error prone non homologous end joining (NHEJ) or the DNA homology mediated repair by homologous recombination (HR), culminating in gene disruption by homologous template introduced or NHEJ mediated nucleotide insertion, deletion or frame-shift mutations. Among the key constraint in designing an effective ZFN is their natural preference for G-rich sequences. Furthermore, making a single designer DNA-binding domain is easier than joining two domains in an appropriate orientation, with correct spacing, to yield a functional nuclease. These technological challenges make in house development of an effective ZFN quite difficult for most laboratories.

**TALEN-mediated genome engineering**

Transcription Activator Like Effectors (TALE) are natural DNA binding proteins of plant pathogenic bacterium, Xanthomonas, that utilize DNA binding domain comprised of 33-35 amino acid tandem monomer repeats, each recognizing a single base pair [18]. TALEN utilizes the DNA binding ability of TALEs with the cleavage function of Fok1 nuclease for targeted genome editing, with two custom Fok1 fused TALENs binding in the opposite directions of the target gene and resulting in dimerization of Fok1 and cleavage of the target DNA [17,19,20]. The key difference between ZFNs and TALENs is that each ZFN finger module (30 amino acids) recognizes three nucleotides while each TALEN repeat (33-35 amino acids) recognizes only 1 nucleotide. The DNA recognition specificity is determined by two hyper variable amino acids, known as the repeat variable diresidues (RVDs), at position 12 and 13. For example HD (His, Asp) targets Cytosine (C), NI (Asn, Ile) targets adenine (A), NG (Asn, Gly) targets thymine (T), and NN (Asn, Asn) targets guanine (G) and adenine (A). Following RVDs binding to the complementary sequences, Fok1 monomers on opposite TALENs dimerize to cause DSB at the spacer region. Repair of DSBs through externally introduced homologous elements can lead to site-specific insertions, while repair of DSBs by NHEJ pathway can result in nucleotide deletions, insertions, causing frame shift mutations and gene disruption. While single base recognition requirement for TALEN-DNA binding repeat provides it greater flexibility than the triplet confined zinc-finger proteins, high sequence similarity between each repeat (difference of only two RVD residues between TALEN repeat modules) present a challenge to assemble a functional TALEN. However, methods such as “Golden Gate” molecular cloning have made rapid assembly of TALENs possible [19] and several TALENs have been shown to produce site specific genome editing in different cell types.

**CRISPR/Cas9 based system for genome engineering**

CRISPRs are short direct repeats of 21-47 nucleotides length interspersed with short intervening spacers in the bacterial genome. The CRISPR repeats are identical but spacers sequences vary. The CRISPR locus is surrounded by a cohort of CRISPR-associated (Casi) genes. The transcribed CRISPR RNA (crRNA) is processed and the mature crRNA binds to the Cas protein/nuclease and guides the complex to DNA target, complementary to the spacer region, to cause Cas-mediated cleavage, followed by DNA repair through NHEJ or HR, as in ZFNs and TALENs. Several types of CRISPR/Cas systems, comprised of small, non-coding CRISPR RNAs and a set of Cas proteins, have been identified in different bacterial species [17,21,22]. Of these, Type II CRISPR/Cas system uses less components than the others, such that it requires a small CRISPR RNA (consisting of palindrome repeats flanking sequence recognizing spacers) and a partially complementary trans-acting (trac RNA) that associate with a single Cas9 protein to mediated target degradation. Availability of CRISPR based dual recombinant vector system that expresses the target gene specific customized short crRNA/tracRNA under polymerase III promoter in one vector, and the Cas9 protein in the second vector, makes it feasible to design custom CRISPRs for targeting a gene of interest. This gives CRISPR an edge over technological challenges under ZFNs and TALENs, as it requires construction of only one variable vector encoding target specific crRNA, with Cas9 being the common moiety between different CRISPRs. However, additional studies are needed to ascertain the efficacy of CRISPR-mediated gene targeting approaches as well as their off-target effects.

**Concluding Remarks**

MicroRNAs, ZFNs, TALENs and CRISPRs are efficient targeted genome engineering tools. Characterization of microRNAs that can facilitate the development of a specific cell lineage or imprint predefined functional characteristics in a cell of choice will have enormous translational implications. While limited studies with TALENs and CRISPRs improved genome engineering efficiency, additional studies are needed for establishing their toxicity and off-target effect profiles. However, despite these challenges, it is suffice to say that the field of targeted genome engineering has made significant progress in the last few years, and with systematic cross-disciplinary studies their potential could be translated into therapeutic approaches for the benefit of the society.

**References**


