Emma Chory

BIOC258 Genomics, Bioinformatics, and Medicine

Final Project, Spring 2013

Friday, June 7, 2013

Recent advances in engineered nucleases for human genome editing and their therapeutic potential

Introduction

The ability to selectively edit specific loci in a genome has been instrumental in biotechnology, medicine, basic science, and drug discovery for the last half of a century. However, until recently, generating knock-in, knock-out, point-mutants, and disease models has been laborious, time consuming, and limited to a handful of model organisms: *Drosophila melanogaster, Arabidopsis thaliana*, zebrafish, and mice. This has limited advancement of both the basic and translational sciences and slowed the transition of genome editing into useful therapies. In the last decade, "Recombineering" via bacterial artificial chromosomes has greatly improved the efficiency of bench-top transformations, however these genome modifications require a trained expert in recombination for success, are generally limited to well-known model organisms, and unsuited for human therapies.

In the last half-decade, a variety of strategies have been developed to improve not only the efficiency of genome engineering, but also the ability to translate laboratory

findings into clinical solutions. Genome engineering enables precise modifications to the genome through the introduction of a double-stranded break (DSB) at a specific target sequence. This break is then subsequently repaired by either non-homology end joining, or homologous recombination (**Figure 1**). This mechanism, when translated to eukaryotic species such as Arabidopsis, mice, and eventually humans has the ability to transform the world of genome editing in the same way that the discovery of restriction endonucleases transformed the world of molecular biology. Clearly these novel and exciting techniques, that will be discussed in depth below, will have massive implications in the world of human genome engineering and therapies for Mendelian inherited diseases, stem cell therapies, and cancer treatments.

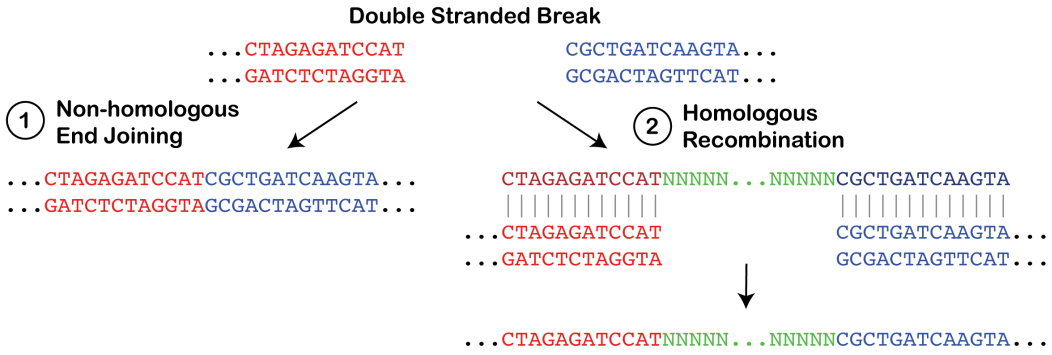

Figure 1: Repairing Double Stranded Breaks

Recent Advances in Genome Engineering Technologies

Genome engineering originated through the discovery of "zinc finger nucleases" (ZNF), which are a class of DNA binding proteins fused to a nuclease with DSB activity. Zinc finger proteins are a class of eukaryotic transcription factors that specifically bind codons unique codons. By fusing the DNA binding domain of numerous zinc finger proteins, a small protein array may be generated that binds to a specific DNA sequence of

the user's choosing. The ZF array is then fused to the Fok1 nuclease catalytic domain to create a DNA sequence specific cutter—ZFN (Figure 2) (Urnov et al. 2010; Kim, Lee, and Carroll 2010). This enzyme then acts in the same mechanism as site-specific blunt-end cutters do in bacteria and can be used to generate a double stranded break that can be repaired in a variety of mechanisms to selectively edit—add, delete or mutate—a genome of interest. It is important to note that ZFN are limited by their ability to recognize only a few codons which lowers recognition specificity, are difficult to engineer, and may have many off target effects due to their length limitations. However, the laboratory successes of ZFNs paved the way for other genome engineering technologies in recent years, and have been indispensable to the field of targeted genome editing.

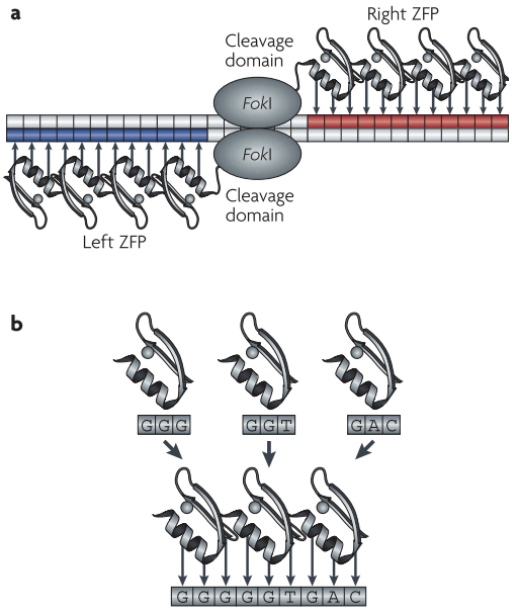

Figure 2: Zinc Finger Nucleases (Urnov et al. 2010)

A breakthrough in nuclease specific genome editing came from the discovery of designer Transcription Activator-like Effector Nucleases (dTALENs) which are sequence specific nucleases that can be tailored to target 12-30 bp regions of the users' choosing. First, a TALE is generated by assembly of C,T,G,A-specific TALE repeats (RVDs) in a sequential fashion. There are 4 universal RVDs (HD, NG, NN and NI) that target C,T,G, and A, respectively. Each RVD consists of 34 amino acids whose nucleotide specificity is governed by two-αα in the monomer peptide (HD, NG, NN, NI) (Figure 3**A**) (Cermak et al. 2011; Wood et al. 2011). By using this TALE "code", designer TALE proteins

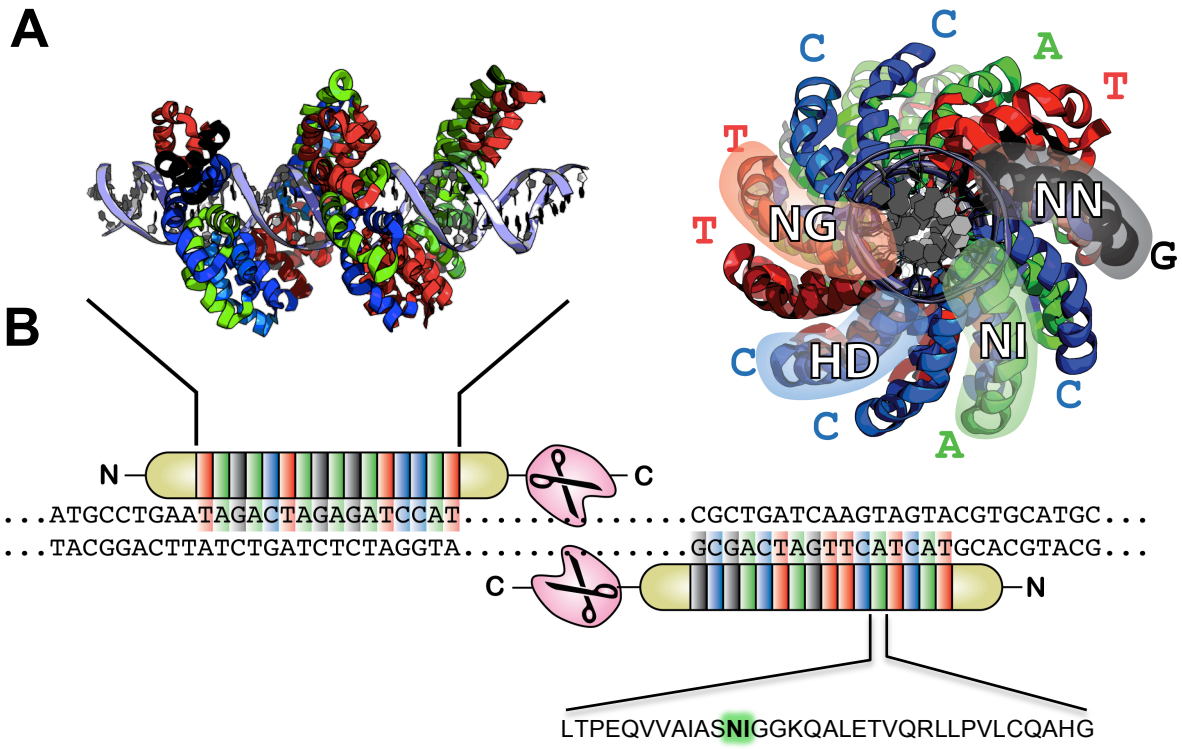

Figure 3: Genome Engineering by Designer TALE Nucleases (dTALENs)
A) Trasnscription activatior-like effectors (TALEs) consist of DNA binding array of tandem monomeric repeats (34aa in length/2 α-helices each), with specificity for a single nucleotide. Engineered TALEs fused with a nuclease catalytic domain target genomic regions *ad libitum.*
B) TALEs form a helical complex with DNA with high sequence specificity.

consisting of 12-30 monomers can be assembled in tandem to target specific DNA sequences in the genome, *ad libitum* **(Figure 3B**). This assembly can be achieved by various modular-cloning techniques (often golden gate assembly). Further, by fusing the catalytic domain of the FokI nuclease to the C-terminus of the dTALE, the sequence specific protein then gains nuclease function and can generate highly specific double stranded breaks in DNA (dsBreaks) in the genome.

Natural TALEs were discovered as transcriptions factors in plant pathogens with highly repetitive and unique DNA binding regions. The TALE code was broken by investigating the AvrBs3 transcription factor in *Xanthomonus*. Much work has since been studied in TALEs since, including the discovery of additional TALE repeats that exhibit

varying specificity v. binding strength for guanine, which has been troublesome (NK, NK repeats), and the helical structure has recently been elucidated by complexing with DNA. Further, by making use of type IIs restriction enzymes, the number of cloning steps required to generate a TALE (>1 per monomer) has been greatly reduced as up to 10 monomers can now be assembled in a single one-pot digestion/ligation reaction (Cermak et al. 2011).

While dTALENs are incredibly attractive means of genome editing, it should be noted that much optimization is required prior to use. While new cloning kits are being rapidly made available for researchers interested in therapy development, the repetitive nature of the sequence and inefficiency of restriction/ligation reactions do make these assemblies non-trivial. Also numerous TALEs for a single DNA region must be screened in order to find a TALE with desired sequence specificity and minimal off-target effects which will slow the progress of these nucleases to the clinic, and introduce potential patient variability which is undesirable for therapies.

Of the various targeted nuclease technologies that exist, by far the most promising and exciting for both human gene therapy/editing and basic biology is the Cas9/CRISPR system, published just this year (Mali et al. 2013). Unlike TALENs and ZFNs which require engineering of whole proteins in order to target a

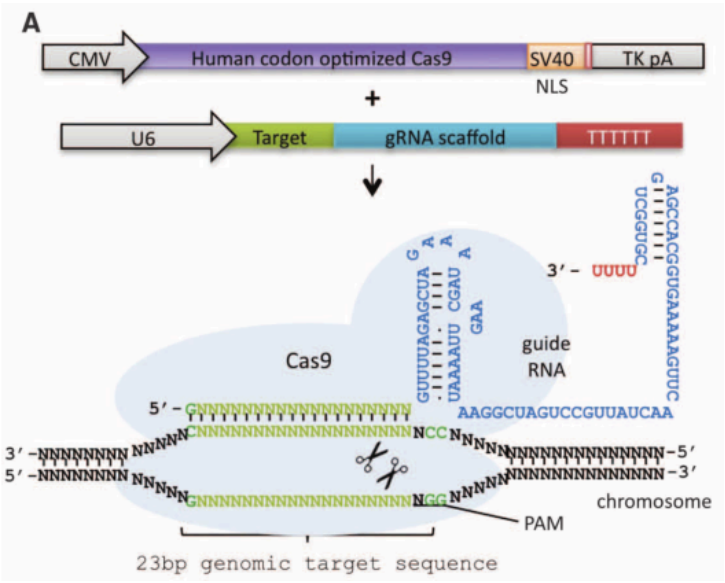

Figure 4: CRISPR/Cas9 Nuclease cleavage and recognition (Mali et al. 2013)

specific gene sequence, CRISPR/Cas9 targets a specific DNA sequence in an RNA-guided manner. Bacteria and archaea have evolved an adaptive immune system that allows for the targeting and cleavage of foreign nucleic acids via clustered regularly interspaced short palindromic repeats (CRISPR) that associate with CRISPR-assosicated protein (Cas). The CRISPR/Cas protein complex consists of a large protein with helicase and nuclease activity that is directed to its target sequence by a short guide RNA. Cas9 consists of a nuclear localization singnal (NLS), which allows it to penetrate the nucleus, and is directed to a site-specific region a DNA by the guide RNA which complexes with the protein in a hairpin structure. Upon binding to DNA via RNA-guiding, Cas9 unwinds, binds to, and then cleaves the DNA to create a DSB. Recently, a type II bacterial CRISPR system was engineered to function with custom guide RNA of the user's choosing, and human optimized (Mali et al. 2013). Upon transfection with both the Cas9 system and an engineered guide RNA, it is now possible to directedly cleave DNA and produce DSBs by simply modifying the 23 base pair genomic target sequence. The development of this system for human site cleavage may eliminate historically laborious genome engineering techniques and generate a sequence specific technology that is as simple as synthesizing a unique strand of DNA. It is clear that this technique alone, due to the specificity that results from RNA/DNA hybridization and versatility will pave the wave for translating these biological techniques into gene therapies. Ideally through the emergence of iPSC technology it is conceivable that in the near future, patient samples may be easily isolated, converted to stem cells, and quickly engineered with desirable therapeutic properties by CRISPR/Cas9 gene editing.

Therapeutic potential of engineered nucleases

Upon the introduction of a double stranded break, a variety of genome editing outcomes, many of which are desirable for human genome therapies, become possible. Non-homologous end joining (NHEJ), one pathway that results from an DSB, is a conserved pathway in all eukaryotic cells in which two broken ends are rapidly and quite efficiently ligated back together. This occurs by NHEJ enzymes recognizing microhomologies, or single stranded over-hangs, that often result from the formation of double stranded breaks and guide repair. NHEJ is a critical mechanism conserved across kingdoms and malfunctions in NHEJ proteins result in a variety of human diseases. NHEJ gene knock-out in mice is also embryonically lethal, showing that this is a wide used and critical mechanism for survival (Urnov et al. 2010).

Taking advantage of NHEJ in genome editing can yield three main outcomes that are therapeutically desirable. Due to the fact that there are occasionally small gains or losses of genetic material that result from a homology end join, a small deletion or insertion at a desirable therapeutic site may result in gene disruption. An example in which biallelic gene disruption is therapeutically relevant is conferring HIV-resistance in humans. It has been shown that a naturally occurring mutation in the CCR5 gene confers resistance to the HIV virus without detectable side-effects or pathophysiology, and that

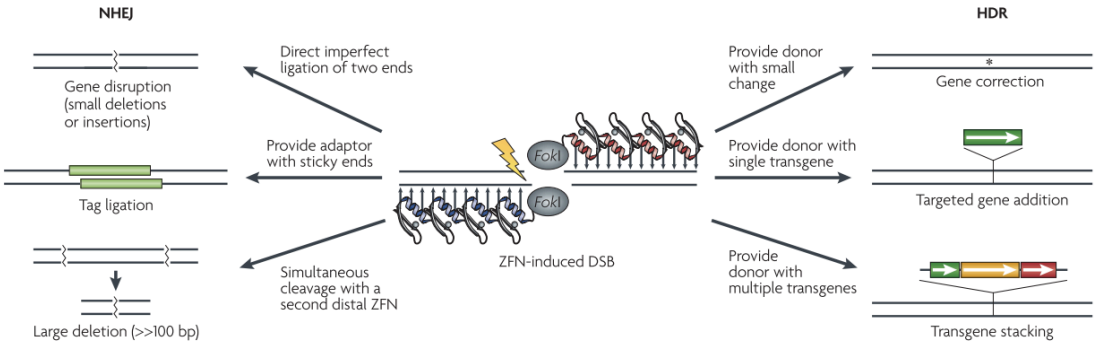

Figure 5: Methods of repairing a DSB and implications in Genome Editing

introduction of a small deletion by NHEJ can disrupt the CCR5 gene and incite resistance in greater than 50% of cells (Urnov et al. 2010). Tag ligation, another instance that can result from NHEJ, may be desirable in instances where the introduction of a stop codon is beneficial, such as in aggressive or incurable cancers in which gene fusions of are the driving oncogenic force. In tag ligation, a short double stranded oligonucleotide acts as an adapter to the overhang scars left from double stranded cleavage by a nuclease. In this case, the small oligo may contain a stop codon and may be inserted at a specific site that results in termination of transcription of the oncogenic fusion. Finally, one may also achieve a large deletion by creating two simultaneous DSBs to remove a large sequence of interest. Sequence deletion by means of NHEJ may be of use to treat diseases such as Huntington's Disease that is caused by the insertion of extra trinucleotide repeats in the middle of the gene. In the case of Huntington's Disease, in which the age of onset and severity of the disease are directly correlated with increased insertions, selective nucleotide deletion would be revert HD to a healthy gene, and preserve functionality in the patient (Urnov et al. 2010).

Homologous recombination is a second DSB repair mechanism that may be exploited in gene editing therapies. Homologous recombination, or the recombining of two dsDNA strands with sequence homology, is a rare event that occurs primarily during cell division and has historically been used to generate most knock-in and knock out mouse and human cell lines. Fortunately, the efficiency homologous recombination is greatly improved by the introduction of a DSB and sequences of the user's choosing can be introduced into the genome by placing said sequence between two homology arms. The main therapeutic advantage of homologous recombination comes from the ability to

introduce a corrected gene at the site of mutation. By providing the host with a corrected gene sequence, effectively any disease that results from a single point mutation may in theory, be converted to a normal gene (Figure 5) (Urnov et al. 2010). A perfect example of how homologous recombination may translate to therapeutics is sickle cell disease, in which the mutated protein, human beta-globin (Hbb) is necessary for livelihood, but results from a single point mutation. Gene addition, naturally then, is a viable therapeutic strategy for diseases that result from deletions such as Duchenne Muschular Dystrophy, and for that matter, any disease resulting from the absence or deletion of an intron, exon, or entire gene.

References

Cermak, Tomas, Erin L Doyle, Michelle Christian, Li Wang, Yong Zhang, Clarice Schmidt, Joshua a Baller, Nikunj V Somia, Adam J Bogdanove, and Daniel F Voytas. 2011. "Efficient Design and Assembly of Custom TALEN and Other TAL Effector-based Constructs for DNA Targeting." *Nucleic Acids Research* 39 (12) (July): e82. doi:10.1093/nar/gkr218. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3130291&tool=pmcentrez&rendertype=abstract.

Kim, Jin-Soo, Hyung Joo Lee, and Dana Carroll. 2010. "Genome Editing with Modularly Assembled Zinc-finger Nucleases." *Nature Methods* 7 (2) (February): 91; author reply 91–2. doi:10.1038/nmeth0210-91a. http://dx.doi.org/10.1038/nmeth0210-91a.

Mali, Prashant, Luhan Yang, Kevin M Esvelt, John Aach, Marc Guell, James E DiCarlo, Julie E Norville, and George M Church. 2013. "RNA-guided Human Genome Engineering via Cas9." *Science (New York, N.Y.)* 339 (6121) (February 15): 823–6. doi:10.1126/science.1232033. http://www.ncbi.nlm.nih.gov/pubmed/23287722.

Urnov, Fyodor D, Edward J Rebar, Michael C Holmes, H Steve Zhang, and Philip D Gregory. 2010. "Genome Editing with Engineered Zinc Finger Nucleases." *Nature Reviews. Genetics* 11 (9) (September): 636–46. doi:10.1038/nrg2842. http://dx.doi.org/10.1038/nrg2842.

Wood, Andrew J, Te-Wen Lo, Bryan Zeitler, Catherine S Pickle, Edward J Ralston, Andrew H Lee, Rainier Amora, et al. 2011. "Targeted Genome Editing Across Species Using ZFNs and TALENs." *Science (New York, N.Y.)* 333 (6040) (July 15): 307. doi:10.1126/science.1207773. http://www.sciencemag.org/content/333/6040/307.abstract.