NUCLEASE-MEDIATED TARGETED GENE INSERTION AT THE ADENOSINE DEAMINASE LOCUS IN PRIMARY CELLS

APPROVED BY SUPERVISORY COMMITTEE

Matthew Porteus, MD, PhD (Advisor)

Joseph Alabanesi, PhD (Chair)

Sandeep Burma, PhD

John Abrams, PhD

Paul Sternweis, PhD

DEDICATION

I would like to thank all the members of the Porteus lab who have helped me in many different ways throughout this long journey. I would like to especially thank the crew of three that survived the move to Stanford with me: Eric, Shaina, and Richard. Not only coworkers, you've been great friends and have helped me get through the lows of graduate school, of which there have been many. I also have to mention the honorary Porteus lab member: Sam Ward. You've been a true friend. My girlfriend Michelle has been a great help these last few months, always encouraging and supporting me in my efforts to finish everything. I would like to most especially thank my family. My parents and siblings have always encouraged me, supported me, prayed for me, and most importantly, believed in me, even when I sometimes doubted. Most of all, my family loved me. Knowing that helped me continue working and finish graduate school.

NUCLEASE-MEDIATED TARGETED GENE INSERTION AT THE ADENOSINE DEAMINASE LOCUS IN PRIMARY CELLS

by

JOSHUA ALLEN CHECKETTS

DISSERTATION / THESIS

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

August, 2013

NUCLEASE-MEDIATED TARGETED GENE INSERTION AT THE ADENOSINE DEAMINASE LOCUS IN PRIMARY CELLS

Publication No. _____

Joshua Allen Checketts, PhD

The University of Texas Southwestern Medical Center at Dallas, 2013

Supervising Professor: Matthew Porteus, MD, PhD

Gene therapy is the ability to correct diseases at the DNA level and has long been a goal of science and medicine. The earliest gene therapy clinical trial was for a patient with severe combined immunodeficiency (SCID) due to adenosine deaminase (ADA) deficiency. Initial trials looked promising and the technique was extended to other forms of primary immunodeficiency. Unfortunately, some of the patients enrolled in these trials using retroviral vectors to carry replacement genes resulted in insertional oncogenesis. To avoid the insertional oncogenesis caused by random integration into the genome, we postulated that targeted insertion of the gene of interest through homologous recombination would prove to

be a safer alternative to random viral insertion of a gene. To this end, we developed several pairs of TAL effector nucleases (TALENs) designed to target exon 1 of ADA. These TALENs function as dimers, and each pair creates a different targeted double strand break near the start site of the ADA gene. The most effective pair induces a DNA double strand break immediately preceding the ADA start codon. Targeted activity of these TALENs was measured through determining the percent of alleles that undergo mutagenic non-homologous end joining upon exposure to the TALENs, with up to 14% of alleles undergoing such mutations. In order to stimulate gene targeting at the ADA locus in human cells, these TALENs were nucleofected into the cells as plasmid DNA, along with a donor plasmid that contains the DNA to be inserted flanked by 800bp arms of homology to the cut site. These TALENs were able to stimulate site-specific integration of the desired fragment at rates of up to 10% in human cell lines. Successful targeted gene insertion was verified through maintained fluorescence, western blots, and sequencing of the targeted alleles through PCR amplification. We demonstrated the ability to enrich for targeted cells through the expression of a selectable marker within the DNA cassette integrated at the ADA locus. In addition to the editing of cell lines, we showed successful stimulation of gene targeting in patient-derived fibroblasts in 1.5% of cells. We demonstrated the feasibility of using the ADA locus as a safe harbor through the targeted insertion of three therapeutically interesting genes. Finally, we demonstrated the successful targeted gene insertion in human CD34+ in up to 0.5% of cells treated. The successful targeting of human CD34+ is especially relevant, as these cells will need to undergo gene targeting in order to be therapeutically relevant as a curative therapy for SCID due to ADA deficiency.

TABLE OF CONTENTS

PRIOR PUBLICATIONS

Kildebeck E, **Checketts J**, Porteus M. Gene therapy in primary immunodeficiencies. Curr Opin Pediatr, 2012. (Review)

Link N, Chen P, Lu WJ, Pogue K, Chuong A, Mata M, **Checketts J**, Abrams JM. A collective form of cell death requires homeodomain interacting protein kinase. J Cell Biol., 2007.

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF DEFINITIONS

Gene targeting – Specifically manipulating the DNA content at a desired locus within the genome, one method used to achieve gene therapy.

Adenosine deaminase (ADA) - Enzyme that catalyzes the removal of an amine group from deoxyadenosine, converting it to inosine. Adenosine deaminase deficiency results in the loss of all immune cells.

Severe-combined immunodeficiency (SCID) – A patient lacking functional B-cells, T-cells, and natural killer cells. Unable to fight off infections by foreign agents (ie, bacteria, viruses).

Gene therapy – manipulating the human genome in order to alleviate the effects of a deleterious mutation

Primary immunodeficiency (PID) – A mutation in the genome that directly limits the patient's immune system

Homologous recombination (HR) – When a double-strand break is created in the DNA, homologous recombination is one of the ways that a cell repairs the break. This repair mechanism uses DNA that is homologous to the break site as a template in order to facilitate correct repair of the break.

Nonhomologous end-joining (NHEJ) – When a double-strand break is created in the DNA, nonhomologous end-joining is one of the ways that a cell repairs the break. The cell sticks the two ends back together, sometimes resulting in small DNA insertions and deletions.

Zinc-finger nuclease (ZFN) – Chimeric proteins consisting of two domains: zinc-finger DNA binding domain and FokI nuclease domain. Each zinc-finger recognizes a DNA triplet.

TAL-effector nucleases (TALENs) – Chimeric proteins consisting of two domains: TAL-effector DNA binding domain and FokI nuclease domain. The TAL-effector domain consists mostly of a series of repeats that can be combined in order to recognize a desired DNA sequence.

CHAPTER ONE
Introduction and Background

GENE THERAPY FOR ADENOSINE DEAMINASE DEFICIENCY

Monogenic Diseases and Gene Therapy

Monogenic diseases are those in which the disease phenotype results from a mutation in a single gene in the genome. Monogenic diseases are divided into three different categories: autosomal dominant, autosomal recessive, and X-linked (sometimes referred to as sex-linked). Autosomal dominant monogenic diseases require only one autosomal allele to be mutated in order for the disease phenotype to manifest itself, despite the presence of the remaining unmutated allele. Autosomal recessive monogenic diseases require both autosomal alleles to be mutated, as one unmutated allele is sufficient to maintain normal physiological function. X-linked monogenic diseases are located on the X chromosome and are usually recessive in nature. However, as males only possess one copy of the X chromosome, if that allele is mutated, they will present with the disease, as they have no unmutated allele to overcome the effects of the mutation. As such, males are much more likely to have X-linked monogenic diseases. Females can also present with X-linked monogenic diseases, though much less frequently, as they would need both of their X chromosomes to contain mutated alleles. Monogenic diseases vary in severity, but can be very debilitating and even lethal. In 2006, the March of Dimes (a non-profit organization focused on improving the health of both mothers and babies) released a report that estimates there are over 7,000 monogenic diseases in the world, an impressive number considering that

the estimated total number of genes in the human genome is between 20,000-25,000.[1] Furthermore, March of Dimes reports that the worldwide average for babies born with single gene defects is 15 out of 1000 live births, with slightly lower rates in the United States at 10 out of every 1000 live births. The prevalence and severity of monogenic diseases necessitate a far-reaching approach to alleviate the affects of these disorders. While both preventative measures and clinical treatment of these diseases through pharmacology are very beneficial, this paper focuses on the work being done to cure these patients of their diseases through gene therapy.

Gene therapy is a broad term that refers to the manipulation of a patient's genome in order to relieve a disease phenotype. While the methodologies for performing gene therapy are quite diverse, the overall strategy usually falls into one of three categories: 1) gene insertion, 2) gene correction, or 3) gene disruption. Gene insertion relies on the integration of an expressed gene into the human genome that ameliorates the effect of the mutated gene. These insertions can be in either a random or a target-specific location. Gene insertion is the form of gene therapy that has had the most widespread clinical success while the other two strategies have only recently become viable clinical options. Gene correction is the most straightforward idea, causing a direct conversion of the disease-causing mutation to that of the DNA sequences of unmutated alleles. Gene disruption requires the mutation of a targeted gene in order to relieve the disease phenotype, usually the mutated dominant allele or another gene in the same pathway that effectively prevents the manifestation of the disease phenotype. More complex strategies have evolved that combine these strategies in order to enhance the efficiency of gene therapy. My project focuses on developing a new method for

gene therapy for adenosine deaminase (ADA) deficiency that relies on targeted gene insertion in order to restore ADA activity to patient cells.

Adenosine Deaminase Deficiency

Adenosine deaminase is a ubiquitously expressed metabolic enzyme whose function is to, as the name states, catalyze the removal of an amine group from deoxyadenosine (dAdo) and convert it to deoxyinosine.[2] ADA-deficiency leads to the systemic toxic accumulation of dAdo in both the cells and the blood. The most deleterious phenotype of ADA-deficiency is immune deficiency, which leads to death within the first year of life if left untreated. ADA-deficient patients are classified as having severe combined immunodeficiency (SCID), as they present with a complete absence of functional T cells, B cells, and natural killer cells. While ADA-SCID is a rare disease (with an incidence of about 1 in 200,000 births), it is the second-leading cause of SCID.

ADA-deficiency leads to immune deficiency through a combination of several pathways. The primary cause for toxicity in ADA-deficient patients is the increase in levels of dAdo, which leads to a subsequent increase in the levels of dATP. High levels of dATP inhibit ribonucleotide reductase, an enzyme necessary for DNA replication and repair, leading to cell cycle arrest. This cell cycle arrest increases the rate of apoptosis in immature lymphocytes. High levels of dAdo also inhibit activity of proteins necessary for V(D)J recombination and lymphocyte activation, two processes necessary to maintain a functional adaptive immune system. The loss of immune cells is further aggravated because

lymphocytes have a higher-than-normal expression level of deoxycytidine kinase, the enzyme responsible for converting dAdo to dATP. Due to the ubiquitous expression of ADA, ADA-deficient patients also show toxicity in several other organs, leading to additional complications in the brain, liver, and lungs. ADA-SCID remains the most urgent phenotype, and restoration of ADA activity in immune cells helps lower systemic levels of dAdo, thus alleviating some of the toxic effects of ADA-deficiency in all other organs.

ADA-deficiency is currently treated clinically using three different procedures: hematopoietic stem cell transplantation (HSCT), enzyme replacement therapy (ERT), and gene therapy using viral vectors. Gene therapy for ADA-deficiency will be discussed extensively later in this introduction, so the focus here will be on the two former strategies: HSCT and ERT. HSCT (commonly referred to as a bone marrow transplant) is the preferred treatment for ADA-deficiency, provided that a suitable donor can be found. As with any organ transplant, potential donors are screened to ensure that they are human leukocyte antigen (HLA) matched with the recipient in order to lessen the chance of organ rejection by the recipient. Patients who receive an HSCT from an HLA-matched related donor (usually a sibling) have an 80-90% rate of successful engraftment and immune reconstitution. This rate falls slightly if the transplant comes from an HLA-matched unrelated donor, to around 70%. Upon successful engraftment of the HSCT, immune reconstitution continues throughout the life of the patient. Unfortunately, an HLA-matched donor can only be found for about 20% of these patients, and so other alternatives must be examined for the remaining 80%.[3] When an HLA-matched donor cannot be found a mismatched donor is sometimes used, with the rate of successful engraftment falling drastically to 30-40%. For this reason, a recent review

published by some of the world's experts on ADA-deficiency recommend the use of either ERT or gene therapy to treat ADA-SCID patients before using a mismatched donor, if those alternatives are available.[2]

Enzyme replacement therapy for ADA-deficiency has been shown to decrease levels of circulating dAdo in the blood, and to restore at least partial immune function to patients. ERT involves the weekly or twice-weekly intramuscular injection of bovine ADA that has been pegylated in order to increase its stability in the blood (referred to as PEG-ADA). ERT usually restores immune function within 2-4 months of beginning treatment, while also lessening the toxic impact of high dAdo levels. Approximately 50% of all patients who have started ERT are still following this treatment plan, while 20% have died and 30% have stopped to pursue another therapy (either HSCT or gene therapy). 40% of those who die after beginning ERT do so within the first month of treatment, likely due to pre-existing conditions incident to the disease. It is estimated that those patients who make it through the first six months of ERT treatment have a greater than 90% chance of living 12 years. Despite the benefit that ERT has provided to many patients, there are still drawbacks to the treatment. The first is cost and availability. It is estimated that the injections cost $300,000 a year, and the cost increases with age as the dose is based on the weight of the patient.[4] PEG-ADA is also not available in many parts of the world. Some patients develop antibodies against the PEG-ADA, as it is the bovine version of ADA and is recognized as foreign, thus further increasing the cost as the doses must be increased accordingly in order to overcome the immune response. It is ironic that the very success of PEG-ADA, that of restoring immune response, can also lead to eventual rejection of the protein that allows for the immune system

to persist. Patients on ERT only display partial immune reconstitution and immune efficiency often declines after years of treatment.[2] The lack of HLA-matched donors for HSCT and the less-than-optimal immune reconstitution stimulated by ERT has led to a concerted effort to develop alternative therapeutic options, namely gene therapy. Patients expressing as little as 5% of normal ADA maintain a functional immune system.[4] This fact significantly increases the feasibility of developing a successful gene therapy treatment.

Gene Therapy for Primary Immunodeficiencies

ADA-SCID is classified as a primary immunodeficiency, a large and varied group of genetic disorders resulting in an underdeveloped or dysfunctional immune system in patients. More than 150 types of primary immunodeficiencies have been molecularly characterized, with more being added over time.[5] The phenotypes range from being almost asymptomatic (i.e., selective immunoglobulin A deficiency) to potentially fatal (i.e., SCID). Similar to patients with ADA-SCID, patients with primary immunodeficiencies with severe phenotypes undergo HSCT and receive significant therapeutic benefit, with success rates around 90% when an HLA-matched related donor is available. Unfortunately, due to the low percentage of patients for which a matched donor is available, mismatched donors are often used, decreasing the rate of successful engraftment while simultaneously increasing the risk of morbidity and mortality. In patients with T-B- phenotypes (no functional T cells or B cells), such as in those with SCID, a lower probability of graft rejection improves the response to

transplantation. Other forms of primary immunodeficiencies, however, can require myeloablative conditioning regimens that cause additional complications and lower the percent of patients that experience successful outcomes.

For those patients who lack an HLA-matched donor, the best solution may lie in gene therapy. Gene therapy is the process of transplanting autologous cells with restored gene expression, usually through insertion of the cDNA of the mutated gene, back into a patient to correct the disease phenotype. The earliest gene therapy trials in humans were for the treatment of patients with ADA-SCID.[6] Patients were transplanted with autologous T-lymphocytes transduced with a retrovirus encoding ADA cDNA flanked by viral long terminal repeats (LTRs). Following transplantation with these cells, patients showed immune system reconstitution without apparent adverse side effects. This success was followed by gene therapy trials for patients with mutations in the *IL2RG* gene causing SCID-X1. In the first SCID-X1 trial, 9 patients were treated with autologous CD34+ hematopoietic stem/progenitor cells (HSPCs) transduced with a gammaretroviral vector, with 8 patients demonstrating immune reconstitution.[7] Unfortunately, 4 of these patients subsequently developed T-cell leukemia resulting from retroviral insertion. Three of these patients entered remission following standard leukemia therapies; however, one passed away because of complications due to his leukemia. Despite these complications, gammaretroviral-based gene therapy trials continued for ADA-SCID[8, 9] and SCID-X1[10], as well as for several other primary immunodeficiencies including Wiskott-Aldrich Syndrome (WAS)[11] and chronic granulomatous disease (CGD).[12] All but the CGD patients showed significant long-term immune reconstitution, with the CGD patients showing only transient benefit. As in the

SCID-X1 trials, adverse affects were observed in WAS and CGD patients, but for reasons not completely understood have yet to be observed in trials for ADA-SCID. The viral integration profile for the ADA-SCID patients in similar to those for other diseases, with integration found near LMO2, BCL2, and other oncogenes.[13] The lack of insertional oncogenesis may be due to intrinsic metabolic differences in these patients.

Improvements in Viral Vectors in Preclinical Trials

While the restoration of immune function in numerous patients illustrated the possibilities for gene therapy, the occurrence of serious adverse events in early trials highlighted the need to improve the safety of viral vectors used for gene therapy. Multiple avenues for improving the safety and efficacy profiles of viral vectors have been explored, including the use of self-inactivating (SIN) LTRs and insulating elements, using viral vectors that do not integrate near promoters, and driving transgene expression from endogenous promoters.[14]

Perhaps the most significant change in recent preclinical trials is the shift from using gammaretroviral vectors to lentiviral vectors. Lentiviral vectors show a decreased probability of integrating near the regulatory elements of actively transcribed genes[15] and do not require HSPCs to be actively dividing for transduction. Removing the need to activate HSPCs with cytokines decreases the duration that stem cells need to be cultured, reducing the probability of cell differentiation.[16] Differentiated cells do not have the potential for self-renewal and may not lead to long-term contributions to the immune system. To prevent activation of

transcription from integrated LTRs, nearly all lentiviral preclinical trials incorporate SIN LTRs and drive transgene expression with a ubiquitously expressed promoter such as SFFV, PGK, or EF1α. Using this approach, restoration of immune function in mice has been demonstrated for Rag 1 & 2 deficiencies, WAS, CGD, SCID-X1, and other diseases.[17-23] One obstacle encountered with ubiquitous promoters, particularly SFFV, is transgene silencing due to DNA methylation.[20] The addition of the ubiquitin chromatin opening element (UCOE) has been shown to decrease DNA methylation and provide consistent transgene expression from ubiquitous promoters across different cell lines.[17, 20, 24] On the other end of the spectrum from transgene silencing, over-expression of a transgene can also be detrimental.[25] The use of endogenous promoters for a specific transgene mitigates this problem and has been used to treat WAS, SCID-X1, and canine leukocyte adhesion deficiency in animal models.[19-21, 26] Interestingly, the use of codon-optimized cDNAs was critical to the success of several preclinical trials and was shown to increase transgene expression and lower the required vector copy number.[17, 20, 21, 25, 27] In addition to advances in the viral vector and control of transgene expression, pre-transplant conditioning offers a further avenue for optimization that was shown to improve transplant engraftment and immune system reconstitution in mice.[21]

Recent Clinical Trials For Primary Immunodeficiencies

Clinical trials using lentiviral vectors to transduce autologous CD34+ HSPCs have recently commenced, including an ADA-SCID trial using a SIN-lentiviral vector with an

internal EF1α promoter, SCID-X1 trials using similar vectors, and trials for WAS and CGD using SIN-lentiviral vectors that proved to be effective and safe in preclinical studies.[13, 18, 19, 28] Ongoing gammaretroviral-based ADA-SCID trials continue to show excellent results with over 70% of patients able to cease ADA enzyme replacement therapy and no reports of oncogenesis.[9] A gammaretroviral-based trial for CGD showed only transient clinical benefit with significantly decreased transgene expression over time, perhaps resulting from DNA methylation.[12] While it is our belief that SIN-lentiviral vectors will prove to have a better safety profile than gammaretroviral vectors, it is worth noting that in a recent trial to treat β-thalassemia there was an outgrowth of a clonal population, although no leukemic events have been reported as of yet.[29]

Gene Targeting: A Promising Alternative To Viral Insertion For Gene Therapy

Despite the clinical success achieved with viral vectors, challenges including insertional oncogenesis, transgene silencing, and lack of endogenous gene regulation have driven efforts to develop alternative approaches for gene therapy. Gene targeting is a process where homologous recombination between a genomic sequence and an exogenous DNA template harboring desired sequence alterations creates precise genome modifications. This method is a promising alternative because it can be used to directly correct a disease-causing mutation *in situ* or insert a therapeutic transgene at a specific location without otherwise altering the genome. Engineered nucleases such as homing endonucleases (HEs), zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and more

recently clustered regularly interspaced palindromic repeats (CRISPRs) can be designed to stimulate gene targeting at specific sites in the genome, and proof-of-principle correction of disease phenotypes in cell lines has been demonstrated for multiple monogenic diseases, including the primary immunodeficiencies SCID-X1, RAG1-SCID, and CGD.[30-32] I will now discuss in detail the method of repair of DNA double-strand breaks and the process of creating and validating nucleases that can specifically target these breaks in the genome, as these are essential processes for the success of my project.

DNA Double-Strand Break Repair

Double-strand breaks (DSBs) in the DNA are a potentially debilitating form of damage that the cell must process before proceeding on in the cell cycle. While dangerous, it is estimated that each cell experiences approximately 50 DSBs per round of cell cycle, so the cell is well equipped to handle these events.[33] Additionally, when cells undergo meiosis, DSBs are stimulated in order to induce crossing over between the homologous chromosomes to increase the genetic variability of the population. DSBs are recognized by the DNA damage response pathway, which includes the MRN complex, consisting of Mre11, Rad50 and NDS1.[34] These proteins activate the proteins necessary to prime the broken ends for repair, and also activate cell-cycle checkpoints, halting further cell cycling until the DSB is repaired, as unrepaired ends can lead to genomic instability that can be fatal to the cell. DSBs are most often repaired by one of two pathways, either non-homologous end-joining (NHEJ) or homologous recombination (HR). Pathway choice is influenced by several factors, the most notable being the phase of cell cycle the cell occupies at the time of repair.

NHEJ is the favored form of repair in G0 and G1 phases of the cell cycle, while HR is favored in S and G2.[35]

When the NHEJ repair pathway is activated, the cell repairs the break by directly ligating the two ends back together. If the DNA ends at the break have not been altered in any way, this results in a simple re-joining of the two ends. However, if the two ends are altered in any way, when they are ligated back together there can be insertions and/or deletions that occur at the site of the break. This is referred to as mutagenic NHEJ, and it is for this reason that NHEJ is sometimes referred to as an error-prone repair pathway. Mutagenic NHEJ is critical in the V(D)J recombination pathway.[36] Antibody diversity is achieved not only by different combinations of V, D and J (or just V and J in the case of the light chain) segments, but this combination process is purposefully mutagenic, thus slightly altering the antibody recognition profile and further expanding the repertoire of antibodies available to recognize foreign antigens. Mutations in the genome disrupting NHEJ function result in a SCID phenotype in the patients. Gene disruption strategies usually rely on mutagenic NHEJ, as these mutations often create frame-shift mutations in the gene, creating truncated proteins that can no longer perform their functions and can be quickly degraded by the proteosome.

The HR pathway repairs DSBs through the use of a homologous template in order to ensure high fidelity of repair. For naturally occurring DSBs, HR utilizes the sister chromatid as a template to restore the original DNA sequence.[35] There is no sister chromatid available in the G0 and G1 phases of the cell cycle, and so HR is inhibited in the cells until after DNA replication is completed. HR in G1 would of necessity use the homologous chromosome for

its repair template, resulting in a loss of heterozygousity, which is evolutionarily disfavored. Mutations in the HR pathway result in higher genomic instability and can lead to cancer, (i.e., mutations in the BRCA1 or BRCA2 genes predisposing patients to breast cancer). For gene targeting, an exogenous DNA donor with homology to the regions flanking the DSB is added to a cell and the HR machinery can use this exogenous DNA as a template to incorporate novel sequences into the chromosome (Figure 1.1).

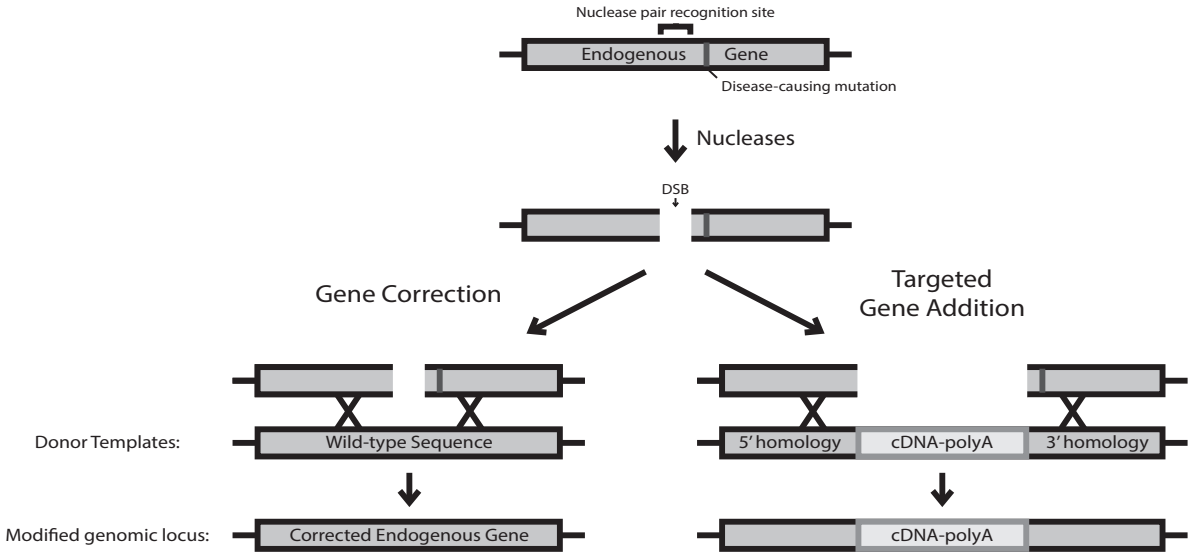

Figure 1.1: Utilizing Homologous Recombination for Gene Therapy. Outline of two strategies for correction of disease-causing mutations using nuclease-stimulated homologous recombination. Gene correction involves the use of a wild-type sequence as a donor template that directly corrects the mutation in the DNA. Targeted gene addition uses nucleases to stimulate the insertion of a full expression cassette, which can then functionally correct the disease phenotype.

The spontaneous rate of gene targeting through HR with an exogenous homologous template is low in most cells, around 10^{-6}, which is not high enough to be therapeutically useful in most systems.[37] Smithies and Capecchi were the first to demonstrate the usefulness of homologous recombination in its potential for gene targeting,[38, 39] and were awarded the 2007 Nobel Prize in Medicine for their work in this field. The discovery in the mid-1990s that induction of DSBs by the I-SceI homing endonuclease could increase the

frequency of gene targeting by five orders of magnitude from 10^{-6} up to nearly 10^{-1} provided the first evidence that therapeutic levels of gene targeting could be achieved.[40] The application of this discovery for gene therapy of various diseases, however, is complicated by the requirement to create site-specific DSBs at each genomic locus of interest. Engineered nucleases have provided a solution to this challenge by allowing targeting of DSBs through the fusion of novel DNA binding domains with previously characterized nuclease domains.

Homing Endonucleases

Homing endonucleases (HEs) are microbial proteins that create DSBs in the DNA in order to generate gene conversion events that are controlled in a site-specific manner.[41] I-SceI was the first of those discovered and is the most well known. I-SceI belongs to a class of HEs called the LAGLIDADG, due to the amino acid sequence directing cleavage of the DNA. HO, the HE responsible for mating-type switching in yeast, is another protein of this same class. As HEs possess intrinsic gene conversion attributes, it was theorized that they could be used to stimulate gene targeting. The long recognition sites of HEs posed a problem at first. LAGLIDADG HEs recognize sequences of 18-24bps in length, meaning that most mammalian cells do not contain an HE recognition site, and if they do, there is a low probability of that site being near a gene of interest. In order to bypass this hurdle, an I-SceI site was integrated into a murine cell line and subsequently demonstrated the ability to create DSBs at its target site and induce mutagenic NHEJ. A human cell line was created in which the I-SceI site was placed inside of a disrupted GFP gene and integrated into the human genome. I-SceI was able to stimulate HR-mediated correction of the GFP gene in a then

astounding rate of 10^{-2}, four orders of magnitude higher than the rate seen when cells receive the donor template alone. While more and more HEs were discovered that increased the number of potential target sties, the problem of targeting sites in the human genome still remained. The I-CreI HE was engineered such that it now recognized and cleaved a site in the human genome, and could stimulate HR at that site.[42] Several publications followed outlining methods of altering and screening engineered HEs targeting human loci. I-CreI has been engineered to successfully recognize and cleave endogenous human loci such as RAG1, XPC and dystrophin.[42-44] The engineered I-CreI was able to stimulate targeted HR in murine cells at of frequency of up to 6% of cells. Despite these advances, early engineering of HEs was difficult and so was not widely adapted for gene targeting. At the same time, engineered zinc finger nucleases arose that offered a much more user-friendly design strategy and soon became the nuclease of choice for most labs desiring to stimulate gene targeting in their system of interest.

Zinc-Finger Nucleases

ZFNs are chimeric proteins made by fusing C_2H_2 zinc finger DNA-binding domains to the non-specific nuclease domain of the *FokI* endonuclease (Figure 1.2). Each zinc finger recognizes ~3 base pairs of DNA, and tandem arrays of 3-6 individual zinc fingers can be designed to specify a wide variety of DNA target sequences 9-18 base pairs in length. Since the *FokI* nuclease domain must dimerize in order to cleave DNA, DSBs are only created when a pair of ZFNs bind DNA in reverse orientation such that their nuclease domains can dimerize in the spacer sequence between the two target half-sites. This creates a high degree

of specificity with full ZFN pair target sites ranging from 18-36 base pairs, depending on the number of individual zinc fingers arrayed in each ZFN.

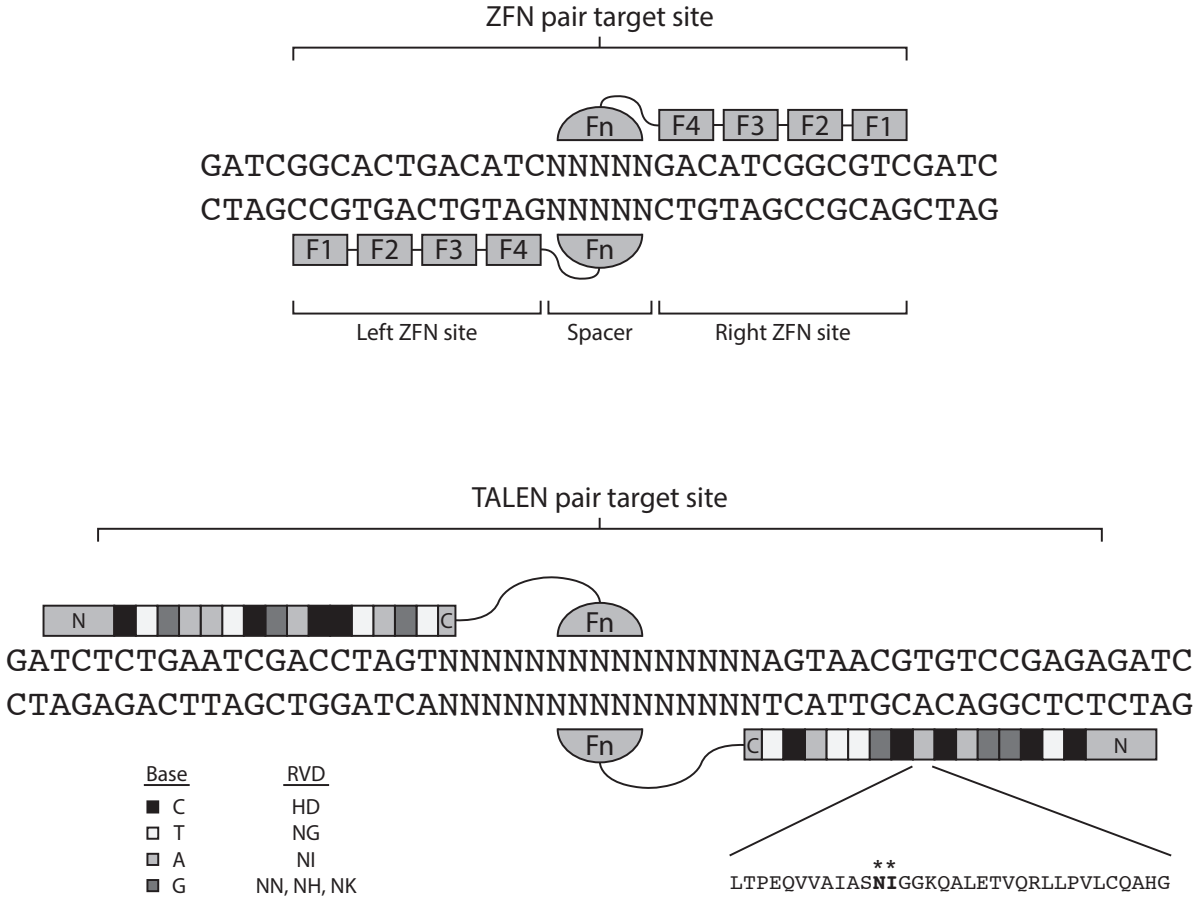

Figure 1.2: ZFN and TALEN Architecture. ZFNs are made up of two domains: DNA-binding domain consisting of a series of zinc fingers and the DNA-cutting domain consisting of the FokI nuclease. Each zinc finger binds a DNA triplet and so the ZFNs illustrated above have a DNA recognition site of 12bps each, for a total of 24bps. The spacer region varies from 5-7bps. TALENs are similarly made up of two domains, replacing zinc fingers with TAL repeats. Each repeat recognizes a single DNA base, as determined by the repeat-variable diresidue (RVD) as indicated above. The TALENs as illustrated above each have recognition sites of 15bps, for a total recognition sequence of 30bps. The spacer region usually varies from 14-20 bps.

The path from *in vitro* demonstration of gene targeting to clinical application has been pioneered by ZFNs, which provided the first demonstration of gene targeting at a novel chromosomal locus in 2003 and at an endogenous locus in 2005.[45, 46] Clinical application of gene targeting has 3 requirements: (i) development of a therapeutic targeting system

including active nucleases targeting a disease-relevant site and a therapeutic donor template, (ii) the ability to stimulate gene targeting in the clinically relevant cell type at a frequency useful for therapy, and (iii) prevention of toxicities that preclude clinical translation (Figure 1.3). Development of a highly active ZFN pair targeting exon 5 of the *IL2RG* gene, which is a hotspot for mutations causing SCID-X1, showed that development of a gene targeting system for a particular disease is indeed a solvable problem.[46] These *IL2RG*-specific ZFNs stimulated targeted gene modification in both human cell lines and primary T cells at frequencies of 18% and 5% respectively. Furthermore, by targeting a partial cDNA encoding *IL2RG* exons 5-8, phenotypic correction of SCID-X1 for any disease-causing mutation downstream of exon 5 was demonstrated.[30]

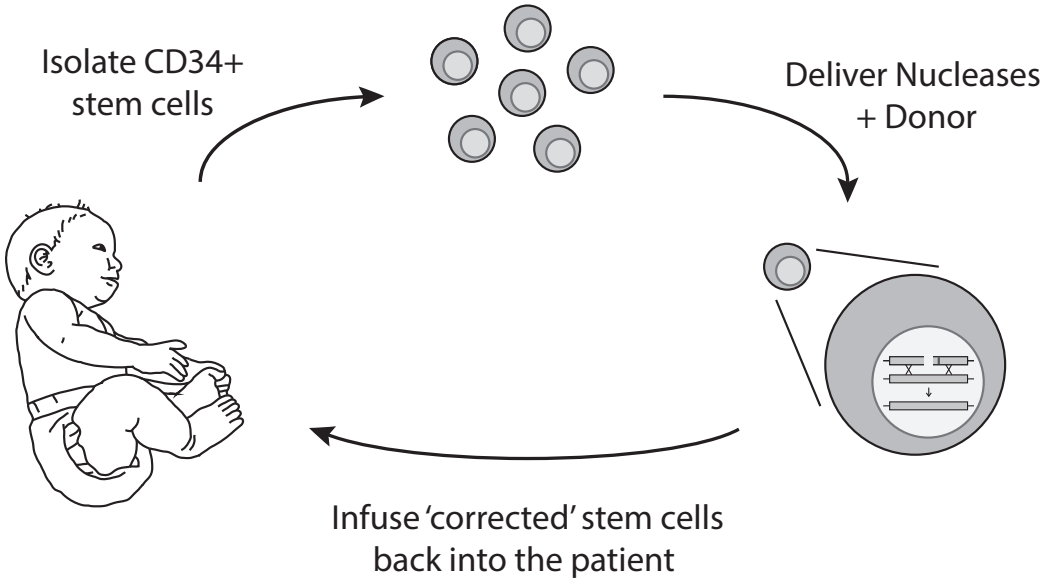

Figure 1.3: Nuclease-mediated Gene Therapy Strategy. Patient CD34+ are removed and isolated. After treatment with nucleases and correction of disease-causing mutation, cells are infused back into the patient. Successful engraftment of corrected cells will result in complete immune reconstitution in the patient.

Translation of gene targeting for SCID and many other diseases will require modification of stem cells, and efforts to develop protocols for efficient gene targeting in

stem cells have met with some success. Human induced pluripotent stem cells (hiPSCs) and human embryonic stem cells (hESCs) targeted with ZFNs specific for the *PIG-A* gene responsible for paroxysmal nocturnal hemoglobinuria (PNH) were shown to retain pluripotency and have normal karyotypes, and similar results have been shown with ZFNs targeting *CCR5* and *AAVS1*.[30, 32, 47, 48] Targeting of neural stem cells and CD34+ HSPCs has been reported using IDLV for delivery of ZFNs, but in the case of CD34+ cells only at a very low frequency of 0.11%.[30, 49] For clinical paradigms that do not require modification of stem cells, gene targeting has been demonstrated in a wide variety of primary cell types including fibroblasts, astrocytes, lymphocytes, keratinocytes, and hepatocytes.[46, 49-51] In one remarkable study, Li *et al*. reported correction of a mouse model of hemophilia B *in vivo* by co-injection of hepatotropic AAV8 encoding human factor IX-specific ZFNs driven by a liver-specific promoter and a donor.[51] ZFN-mediated targeting of a splice acceptor and partial cDNA to intron 1 of the *F9* gene was able to restore endogenously-regulated expression of factor IX from a gene rendered non-functional by a mutation in exon 6. *In vivo* targeting of 1-3% of hepatocytes restored circulating levels of human factor IX to 2-3% of normal, resulting in clinically significant correction of the coagulation defect. While ZFN-mediated gene targeting is yet to be used in clinical trials, phase I trials using ZFNs to disrupt the CCR5 co-receptor required for HIV-1 infection are currently underway. In pre-clinical studies, primary CD4+ cells and CD34+ HSPCs were treated with *CCR5*-specific ZFNs and transplanted into mice. Following infection with HIV-1, ZFN-mutated cells were greatly enriched due to their resistance to infection, and treated mice were protected from CD4+ cell depletion and had significantly lower levels of viremia.[52, 53]

ZFNs have revealed the potential of gene targeting for gene therapy applications, but after nearly a decade of development only a handful of ZFN pairs have been shown to achieve therapeutically relevant levels of gene modification. The obstacles to more rapid translation of ZFNs have included low success rates for making highly active ZFN pairs, limitations on the potential target sequences ZFNs can be designed for,[54] and significant toxicity due to the creation of 'off-target' DSBs.

TAL-Effector Nucleases

Recently, an exciting new class of engineered nucleases has emerged known as transcription activator-like effector nucleases, or TALENs.[55] TAL effectors were discovered as virulence factors produced by the plant pathogen *Xanthomonas*, which bind to specific DNA sequences and alter transcription of target genes. TALEs were discovered to bind DNA through an array of repeats where each repeat binds to a single nucleotide of DNA, which is unique among known DNA binding proteins.[56, 57] A single TALE repeat is composed of ~34 amino acids, which are highly conserved except for two amino acids at positions 12 and 13, known as the repeat-variable diresidue (RVD), that determine the nucleotide binding specificity (Figure 1.2). This simple 1:1 code for protein-DNA binding allows for easy prediction of the binding specificity of naturally occurring and engineered TALEs.

Efforts to create novel TALE-nuclease fusions for gene targeting have met with early success, with multiple groups demonstrating high frequencies of targeted gene modification in human cells.[58-60] One study by Miller *et al.* showed that truncation variants of TALEs with novel binding specificities fused at the C-terminus to the *FokI* nuclease domain could

modify endogenous genes at frequencies up to 25% in human cell lines.[59] It is exciting that TALENs can achieve these levels of activity, which are in the same range as the most active ZFNs reported in the literature. Where TALENs differ from ZFNs, however, is in the ease of engineering active TALEN pairs and the wide range of DNA sequences they can target. A typical TALEN pair target site is comprised of two 13-17 base pair TALEN binding sites separated by a 14-21 base pair spacer. Initial estimates using design criteria derived from the features of naturally-occurring TALEs suggested that on average a suitable TALEN target site is found every 35 base pairs in genomic DNA.[61] In addition, multiple groups have taken advantage of the modular and repetitive structure of TALENs to develop synthesis protocols that are simple, fast, and inexpensive. Methods utilizing Golden Gate cloning, FLASH assembly, and iterative capped assembly have all been reported for TALEN synthesis, allowing new TALE arrays to be generated in ~1 week, 1 day, and 3 hours respectively.[61-63] Using the FLASH assembly method, Reyon *et al.* were able to assemble 96 different TALE arrays in 1 day and generate sequence-verified TALEN expression constructs for < $100 each including the cost of labor.[63] Using this high-throughput approach to generate TALENs targeting 96 endogenous genes, an astonishing 84 pairs were shown to efficiently modify their target site with an average modification frequency of 22.2%. Surprisingly, these highly active TALENs broke the majority of the design criteria suggested by earlier studies, and using revised guidelines it was estimated that on average more than 3 TALEN pairs can be targeted per base pair of genomic DNA.

Since the code for TALE binding of DNA was cracked in 2009 and efficient architectures for generating TALENs established in 2010-11, use of TALENs for gene

targeting applications has progressed rapidly. TALENs have already been used for gene modification in multiple species and have achieved similar targeting frequencies in hiPSCs and hESCs as has been achieved with ZFNs.[64, 65] TALENs designed to target disease-causing mutations, such as the sickle cell mutation in β-globin,[66] have also been reported, but the application of TALENs for gene therapy is really just beginning.

CRISPRs

In 2013, two labs published findings showing they could target novel DSBs in the DNA through the utilization of the CRISPR system in human cells.[67, 68] Bacteria and archaea are constantly incorporating new DNA into their genome through horizontal gene transfer in order to increase their chances of survival. Working in contrast to this desire to uptake potentially helpful DNA fragments is the desire to prevent the invasion of "selfish" DNA, such as that from infection by phage. To combat this, many bacteria and almost all archaea have developed the CRISPR system, which is akin to our adaptive immune system. These cells take up harmful DNA and incorporate a piece of the DNA into a series of repeats in the CRISPR region.[69] These pieces of foreign DNA are then transcribed into RNA and act as a guide for a nuclease, which will then cleave and destroy any foreign DNA containing this sequence, analogous to how the siRNA system functions. There are three types of CRISPR systems that have been discovered, with the type II system being adapted by both of the aforementioned labs for engineered gene targeting in human cells. Type II CRISPRs utilize the Cas9 nuclease in order to cleave the targeted DNA, so the gene encoding for this protein was codon-optimized for human expression and a nuclear localization signal added in

order to ensure that the protein was targeted to the nucleus. Custom guide RNAs were expressed in the cells along with the modified Cas9 and were found to be able to stimulate mutagenesis at the targeted locus of up to 38% in human cell lines, similar to that achieved by very active ZFNs and TALENs. No overt toxicity was observed in these cells. These CRISPRs were also able to stimulate the repair of an integrated mutated GFP in 3-8% of cells, again comparing favorably with active ZFNs and TALENs.

The CRISPR system promises great ease of use and design, as one needs but to order custom oligos that fit the design criteria, clone them into the RNA expression plasmid, and transfect them in along with the Cas9 plasmid in order to generate DSBs at the locus of interest. As with the other nuclease platforms, off-target mutagenesis is again a concern, as the guide RNA targeting area may be restricted to only the first 12-15bps, below the 17 that would make it statistically unique in the human genome. Also of concern is the initial inability to achieve gene targeting in human primary cells, such as HSPCs, that may be due to the cells' innate immune response. Despite these concerns, the gene targeting community is abuzz with the potential and ease of design of the CRISPR system.

Looking at 'Off-Target' Effects of Gene Targeting

The price of avoiding insertional oncogenesis with gene targeting is the introduction of nucleases capable of generating DSBs. Even highly optimized engineered nucleases generate 'off-target' DSBs at sites similar to the sequence they are designed to target, which can result in undesired mutagenesis in the genome. Multiple studies have also demonstrated

the ability of DSBs created by engineered nucleases to induce translocations between the DSB and sites spread throughout the genome.[70, 71] Potential off-target sites for specific ZFNs have been determined using unbiased assays for DSB creation.[72, 73] Pattanayak *et al*. created a method to test the ability of ZFNs to cleave 10^{11} DNA sequences *in vitro* to determine the true specificity of ZFN pairs.[73] For the highly optimized *CCR5* ZFNs, this technique identified hundreds of thousands of DNA sequences that can be cut, 37 of which occur in the human genome. Excluding the known binding sites at *CCR5* and *CCR2*, mutations were detected in a human cell line at 8 off-target sites at frequencies from 1:300 to 1:5,300, including one site in the promoter of the malignancy-associated *BTBD10* gene. While the specificity of TALENs has yet to be rigorously studied, multiple reports have suggested that TALENs are less toxic, and therefore potentially more specific, than ZFNs.[60, 66] TALENs expressed in human cell lines have been shown to cause fewer γH2AX foci, a marker of DSBs, and less overt cell death than commonly used ZFNs. In one study, TALENs designed to target *CCR5* were shown to have very little activity at the highly homologous *CCR2* locus, as compared to CCR5-specific ZFNs that had similar activity at the two sites.[60] The crystal structure of TALEs binding DNA also suggests that TALEN specificity will be more complex than a simple number of mismatches tolerated for 'off-target' binding, with repeats for different nucleotides interacting with DNA in different ways.[74-76] A recent paper confirmed that the CRISPR system also exhibits measurable off-target cutting.[77]

For clinical use engineered nucleases must have both high activity *and* high specificity, and multiple strategies are being utilized to limit the toxic effects of nucleases. Modification of the *FokI* nuclease domain to create obligate heterodimer variants has been

shown to significantly reduce off-target cutting by homodimers of a single nuclease, in effect decreasing potential off-target sites by 50%.[72, 78] Nucleases can also be delivered as mRNA to prevent random integration of an expression plasmid and decrease the duration of nuclease expression, thus decreasing the probability of undesired mutagenesis. Furthermore, instead of characterizing the toxicity profiles of new nuclease pairs, the simpler alternative of utilizing a 'safe harbor' site with a known nuclease pair may be useful for many diseases. Lombardo *et al*. recently showed that the *AAVS1* site located within intron 1 of the *PPP1R12C* gene, which is a common integration site of adeno-associated virus, is a promising candidate for such a safe-harbor.[49] ZFN-mediated transgene insertion at the *AAVS1* locus provided stable transgene expression without perturbing the expression of the targeted *PPP1R12C* gene or 26 other genes in the 400kb region flanking the integration site, even when transgene expression was driven by viral, tissue-specific, or strong ubiquitously expressed promoters. The utility of this approach for gene therapy was demonstrated by Zou *et al*. with the targeting of a $gp91^{phox}$ minigene to *AAVS1* to correct the phenotype of CGD.[32] iPSCs were derived from a patient with CGD and differentiated neutrophils displayed the reactive oxygen species-negative phenotype that is pathognomonic for CGD. Following ZFN-mediated insertion of the $gp91^{phox}$ minigene, neutrophils differentiated from 'corrected' iPSC lines showed full restoration of oxidase activity.

Overview of My Project

At the beginning of my graduate career, it became clear that ADA was an ideal candidate for targeted gene therapy. Disruption of the ADA gene leads to ADA-deficiency, a disease that has a severe phenotype (ADA-SCID) and is a well-defined monogenic disease. Furthermore, the gold standard for treatment of ADA-SCID (that of HSCT from an HLA-matched donor) is only available to a small subset of patients, and while current alternative treatments are beneficial, they have a less-than-optimal efficiency of immune reconstitution and come with certain risks, such as the possibility of insertional oncogenesis through the use of lentiviral vectors. ZFNs have been engineered to recognize novel loci in the human genome and create DSBs at the targeted locus with great efficiency. These targeted DSBs also demonstrated the ability to greatly stimulate targeted HR at the site of the break in 5-10% of cells. This laid the foundation for my project: engineer nucleases that create targeted DSBs in the ADA gene and stimulate targeted gene insertion into the endogenous ADA locus to functionally correct ADA mutations and restore ADA activity to these cells.

CHAPTER TWO
Methodologies

Materials and Methods

Human Cells and Cell Culture

HEK 293T cells and K562 cells were obtained from ATCC. HEK 293T cells were maintained in DMEM (Hyclone) supplemented with 10% bovine growth serum (Hyclone), 100 units/mL penicillin, 100 μg/mL streptomycin, and 2 mM L-glutamine. K562 cells were maintained in RPMI 1640 (Hyclone) supplemented with 10% bovine growth serum, 100 units/mL penicillin, 100 μg/mL streptomycin, and 2 mM L-glutamine. ADA deficient patient derived fibroblasts were obtained from Coriell Cell Repositories, catalog #GM02605. The primary fibroblasts were maintained in DMEM supplemented with 20% bovine growth serum, 100 units/mL penicillin, 100 μg/mL streptomycin, and 2 mM L-glutamine and cultured at 5% O_2. Human cord blood CD34+ cells were obtained from StemCell Technologies, or isolated from fresh cord blood. CD34+ cells were maintained in X-VIVO 15 (Lonza) supplemented with 0.1 μg/mL SCF, 0.1 μg/mL TPO, 0.1 μg/mL Flt3, 0.1 μg/mL !L-6 and 0.75 μM SR1 and cultured at 5% O_2. All cytokines obtained from Peprotech.

OPEN Method of Selecting ZFNs

ZFPs recognizing the target sites in ADA exon 5 and 6 were selected and cloned into a nuclease backbone using the OPEN protocol outlined previously.[79] The recognition sites

are as follows: ADA5A Left- 5'-GGTGAGGTC-3', ADA5A Right- 5'-GAGGTGGTG-3', ADA5B Left- 5'-GAGGTCCCC-3', ADA5B Right- 5'-GACGAGGTG-3', ADA6 Left- 5'-GCTGCTGGT-3', ADA6 Right- 5'-GTGGTAGCC-3'. The sequences of the selected ZFPs were determined through standard DNA sequencing methods.

Transfections

K562 cells and primary fibroblasts were nucleofected using the Amaxa IIS machine. One million cells were resuspended in 100 μL nucleofection buffer containing 100 mM KH_2PO_4, 15 mM $NaHCO_3$, 12 mM $MgCl_2 \cdot 6H_20$, 8 mM ATP, and 2 mM glucose, at pH 7.4. K562 cells were nucleofected using program T-016, while program A-033 was used for primary fibroblasts. Nucleofection cuvettes are 0.1 cm width (VWR). HEK 293T cells were transfected using calcium phosphate. Human CD34+ cells were nucleofected using the Amaxa 4D machine, in the P3 buffer. 200,000 cells were resuspended in 20 μL buffer and nucleofected by program EH 100 in strip cuvettes.

Nuclease Toxicity Assay

We followed the established protocol as outlined previously by our lab.[80] Briefly, HEK 293T were transfected using calcium phosphate with a GFP-expressing plasmid as well as plasmids expressing the nuclease pairs. The decrease in GFP+ cells over time in nuclease-treated cells was compared to that of cells that were transfected with I-SceI, a non-toxic control. An increased rate in the loss of GFP+ cells transfected with nucleases as compared to the non-toxic control was converted into a measurement of toxicity for the nuclease.

Single-Strand Annealing Assay

We performed the single-strand annealing (SSA) assay as outlined previously.[81] The murine ADA SSA target region was as follows: 5'-GACGTCACCCCTGATGACGTTGTGACCATCTTCTAATAAGACGACGGC-3', with the zebrafish ADA SSA target region changed only to reflect the differences at the genomic locus: 5'-GACATCACTCCTGATGATGTGGTGACCATCTTCTAATAAGACGACGGC-3'. Both the murine and zebrafish ADA SSA targets also include recognition sites for the GFP ZFNs previously characterized,[80] used as an internal control.

CelI Enzymatic Mutation Detection Assay

The CelI enzymatic mutation assay was performed as previously outlined.[82] Briefly, for both K562 cells and primary fibroblasts, one million cells are nucleofected with 2.5 μg of each TALEN (along with 1 μg pMaxGFP plasmid from Lonza for measurement of transfection efficiency). On day 3 after nucleofection, the cells were collected and genomic DNA isolated using the Qiagen DNeasy Kit. The locus of interest was PCR amplified with the following primers: Fwd Primer- 5'-GTCCAGAATTCCAGGAAATGCGCGATCCAG-3' and Rev Primer- 5'-GAAGGAAGAACTCGCCTGCAGGAGC-3', with a T_m=67.8°C. Accuprime Pfx (Invitrogen) was the polymerase used for all PCR amplifications. 200 ng of PCR product were treated with the surveyor nuclease (Transgenomic) following the manufacturer's protocol.

TALEN Assembly

Our TALENs were assembled utilizing the Golden Gate method of assembly outlined by the Voytas lab,[83] with the plasmids purchased from Addgene. The TAL repeats were dropped into a final vector containing the Δ152 N-terminal domain and +63 C-terminal domain as previously described.[84] TALEN recognition sites are outlined in Table 2.1.

TALEN Pair	Left TALEN Target Site	Right TALEN Target Site
-77 TALENs	GGCCGGCCGCGGCCACCGCT	CTCTGCCGGCTCGGT
-1 TALENs	AGCGCCGGGGCGCACG	CGAAGGCGGGCGTCT
+28 TALENs	GGCCCAGACGCCCGCCT	CCCCGGAGCCCCCGCGCGCGCT

Table 2.1: ADA TALEN Target Sites. Each of the three main targets for the engineered ADA TALENs are outlined above, with both the left and right TALEN recognition site for each pair of TALENs used.

Western Blots

Western blots were performed following standard western blot protocols. Primary antibodies used: M2 mouse anti-FLAG (Sigma) 1:2000, rabbit anti-actin (Santa Cruz) 1:2000. HRP conjugated secondary antibodies used: goat anti-mouse (Santa Cruz) 1:10000, goat anti-rabbit (Santa Cruz) 1:10000.

P140K MGMT Drug Selection

Targeted cells can be enriched for those expressing the P140K MGMT construct through drug selection. The efficacy of this strategy has been outlined previously.[85] O6-benzylguanine (Sigma-Aldrich) was added to samples to a final concentration of 50 μM.

Cells are returned to 37°C cell incubator for one hour. Carmustine (Sigma-Aldrich) is added to a final concentration of 40 μM and again returned to the 37°C cell incubator for one hour. Cells are washed with PBS, resuspended in normal growth medium, and returned to 37°C cell incubator and grown as normal.

Genomic PCR Analysis of Early Donors

Early donors contained an I-SceI recognition site between the 5' arm of homology and the gene to be inserted into the genome. This allowed for specific amplification of targeted alleles. The primers used for this amplification are as follows: Fwd Primer- 5'-GGTGTGAATATTCCGGTTTA-3', Rev Primer- 5'-CATTAGGGATAACAGGGTAAT-3', T_m=53°C.

ADA-GFP Donor Template Assembly

ADA arms of homology were amplified from genomic DNA isolated from stock K562 cells with the following primers: 5' Arm Fwd Primer- 5'-CCCACTCGCCCGTGGAGGGG-3', 5' Arm Rev Primer- 5'-GGTGCCCTCGTGCGCCCCG-3', 3' Arm Fwd Primer- 5'-GCCCAGACGCCCGCCTTCGAC-3', 3' Arm Rev Primer- 5'-GAGGCTGACTGCAGCTGACTCCTTCCC-3', T_m=68°C. ADA cDNA construct was obtained through OriGene. A 3x FLAG peptide was inserted immediately preceding the stop codon through PCR. Alterations to the 6^{th}, 9^{th}, and 12^{th} bases of exon 1 were also altered through site-directed PCR mutagenesis. Modified ADA cDNA was inserted in-frame with

the 5' arm of homology through overlap PCR. The bGH polyA and Ubc-eGFP fragments were added to the donor through standard cloning techniques due to the presence of a multiple cloning site in the plasmid. See Appendix A for the full donor sequence.

PCR Analysis of Targeted Integration of ADA-GFP Donor

For analysis of the integration of the ADA-GFP donor cassette, a PCR strategy was designed to amplify both the wild-type and targeted alleles. Sequencing of the PCR products and the subsequent analysis would allow for the quantitative measurement of the frequency of gene targeting. The primers used for this PCR amplification are as follows: Fwd Primer- 5'-GCTGAAGTTCATTCCGTTTCACTGCTGTC-3', Rev Primer- 5'-TTTGGGCTTGTCGAAGGCGGG-3', T_m=66.7°C. PCR products from the single-cell clones were analyzed directly by Sanger sequencing. PCR products from populations were first cloned using the Zero Blunt TOPO PCR Cloning Kit (Invitrogen) following the manufacturer's protocol. Bacteria colonies were grown overnight in 2mL LB with appropriate antibiotic, and DNA was isolated from these cultures using standard mini-prep procedures. Isolated DNA was then analyzed by Sanger sequencing.

CHAPTER THREE
Results

NUCLEASE-MEDIATED TARGETED GENE INSERTION FOR FUNCTIONAL CORRECTION OF ADA DEFICIENCY

Engineering ZFNs That Target ADA Gene

As previously outlined, the goal of my project is to develop nucleases that efficiently target DSBs to the ADA locus and stimulate targeted functional correction of mutations in the ADA gene. This targeted correction is done in a human cell line as well as primary cells. ZFNs are have been shown to be effective tools in stimulating efficient nuclease-mediated gene targeting across several different organisms and within several types of human cells[37] and were selected as the initial nuclease platform for my project.

Design and Creation of ZFNs Targeting ADA Locus

Upon close examination of the endogenous ADA locus, three sites in the ADA gene were chosen for which we engineered ZFN pairs that we hoped would stimulate gene targeting: two adjacent sites in exon 5 and a separate site in exon 6. These sites adhere to the constraints imposed by the OPEN system of ZFN construction, namely: two GNNGNNGNN potential binding sites in reverse orientation, separated by a spacer of 5-7bps. One triplet in the exon 5 recognition site is CCC, so for this zinc-finger we used a previously published zinc-finger domain that had shown efficient binding for this target.[86]

All other zinc-finger domains for the targets were selected through the OPEN system. The two exon 5 sites contain a 6bp spacer and are shifted by only one triplet, and the exon 6 site contains a 5bp spacer between the ZFN recognition sites. As exon 5 of the ADA gene is one of the hot spots of mutation for ADA deficiency, this target was especially interesting (Figure 3.1).

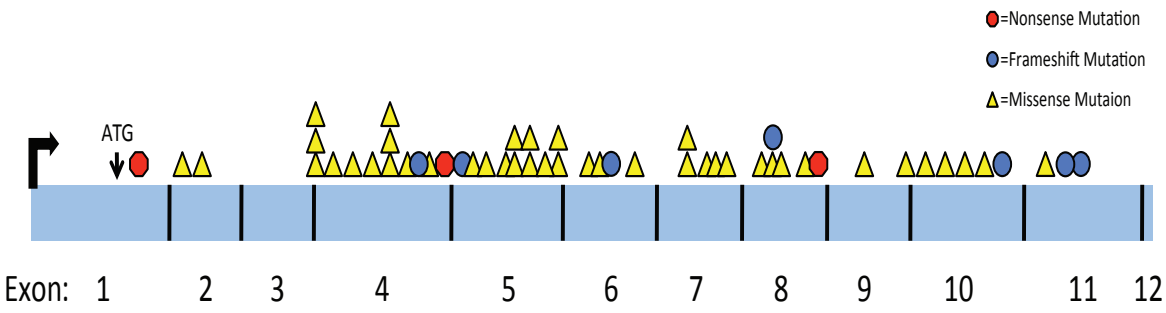

Figure 3.1: Summary of Known Mutations That Cause ADA Deficiency. All known mutations that cause ADA deficiency are mapped to the ADA gene (with introns removed). Map plots relative locations and is not to scale. Red octogons represent nonsense mutations, blue circles frameshift muations, and yellow triangles missense mutations. List of ADA mutations obtained from ADAbase webiste: http://bioinf.uta.fi/base_root/.

Using the OPEN system of selecting for zinc-finger binding proteins that efficiency bind their target sites, I engineered ZFNs that would bind my ADA target sites. In short, the OPEN system utilizes overlap PCR to combine libraries of previously characterized single zinc-fingers (each finger recognizing one triplet) into a three-fingered ZFP (zinc-finger protein) library. Each of these ZFPs contain a three-fingered DNA-binding domain that recognizes 9bps of DNA, three bases for each zinc-finger. These ZFP libraries are fused to a transcription-activating domain and utilized in a bacterial 2-hybrid system in which only those bacteria containing a ZFP able to bind the target site and activate transcription will survive the selection process. After selection, the ZFPs are isolated from the selected bacteria for further analysis and further manipulations for downstream processes (such as

gene targeting). I obtained a pool of ZFPs for each of my six target half-sites, and further characterized the activity of at least six of these ZFPs for each half-site (Appendix B). After fusing these zinc-finger binding proteins with the *FokI* nuclease domain, I now had ZFN pairs to each of my three target sites ready to be tested for their ability to create targeted DSBs.

Verification of ADA ZFN Activity

In order to verify targeted activity of my newly created ZFN pairs, I created a single-strand annealing (SSA) assay reporter plasmid that contained all three target sites. ZFNs that actively cut their target site will cleave the reporter plasmid and stimulate SSA-directed repair of the plasmid, creating a functional GFP expression plasmid that can then be detected through flow cytometry. ZFNs require dimerization in order to cleave their targets, and so must work in pairs to efficiently create DSBs at the desired locus. I tested the ZFN pairs obtained from the OPEN selection in every different combination to determine the best functional pair of ZFNs for each target site. I transfected the many ZFN combinations with their corresponding SSA reporter plasmid into HEK 293T cells and measured the percent of GFP+ cells by flow cytometry on day 2. Those cells transfected with the SSA reporter plasmid alone showed very low levels of GFP+ cells. The percent of GFP+ cells increased greatly with the co-transfection of the newly created ZFNs, some showing as much as an 80-fold enrichment of GFP+ cells over background (Figure 3.2).

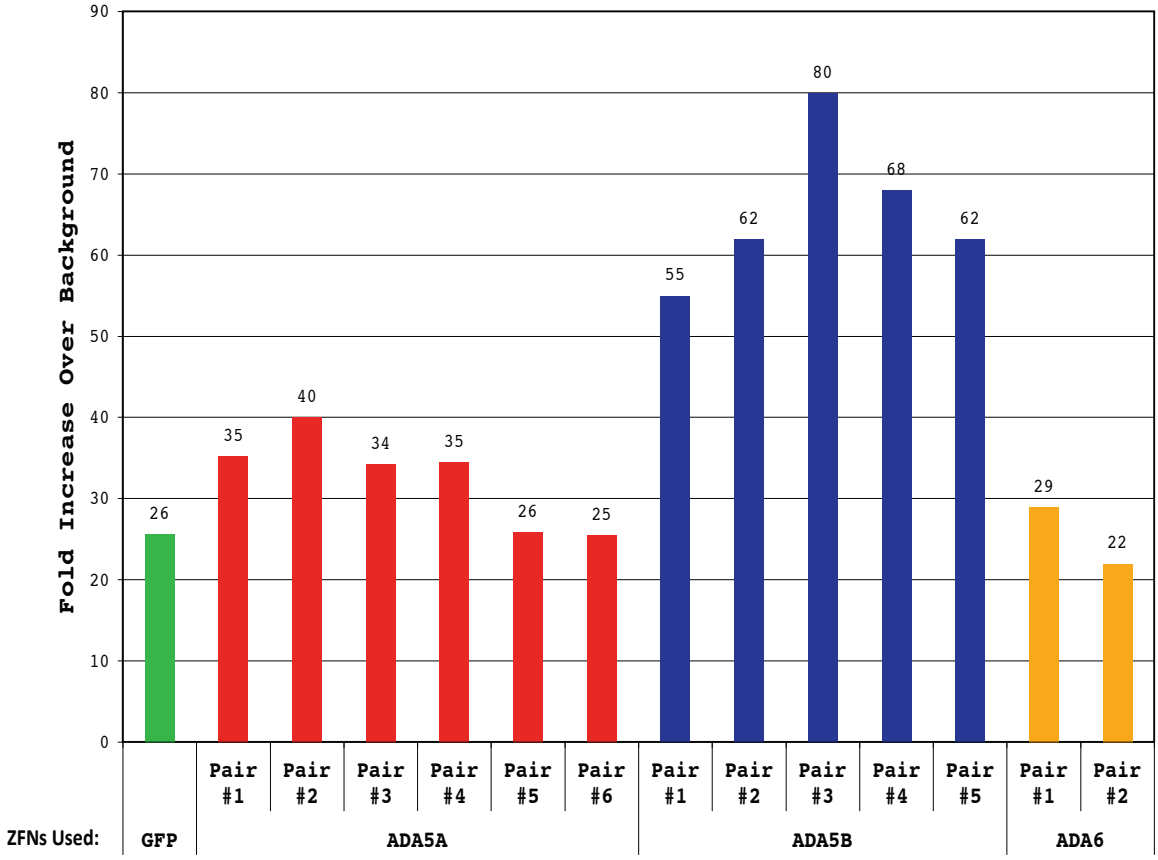

Figure 3.2: ADA ZFNs Show High Activity in the SSA Assay. HEK 293T cells were nucleofected with various ZFNs and the SSA reporter plasmid. GFP ZFNs were used as an internal control for ZFNs with high activities. Only those ADA ZFN pairs that stimulated SSA recombination at a rate similar to that of the GFP ZFNs were included in this graph. Those combinations who failed to reach this threshold were excluded from further experimentation.

Determining Specificity of ADA ZFNs

Developing ZFNs for gene therapy require showing that the nucleases have both high activity and high specificity. While both the OPEN system and the SSA reporter assay provided evidence that the engineered ADA ZFNs were capable of binding and cutting their designed target sequence, they offered no evidence as to the specificity of the ZFNs. I addressed this issue through indirectly looking at specificity in two separate assays: the cell toxicity assay and through manipulation of the SSA reporter assay.

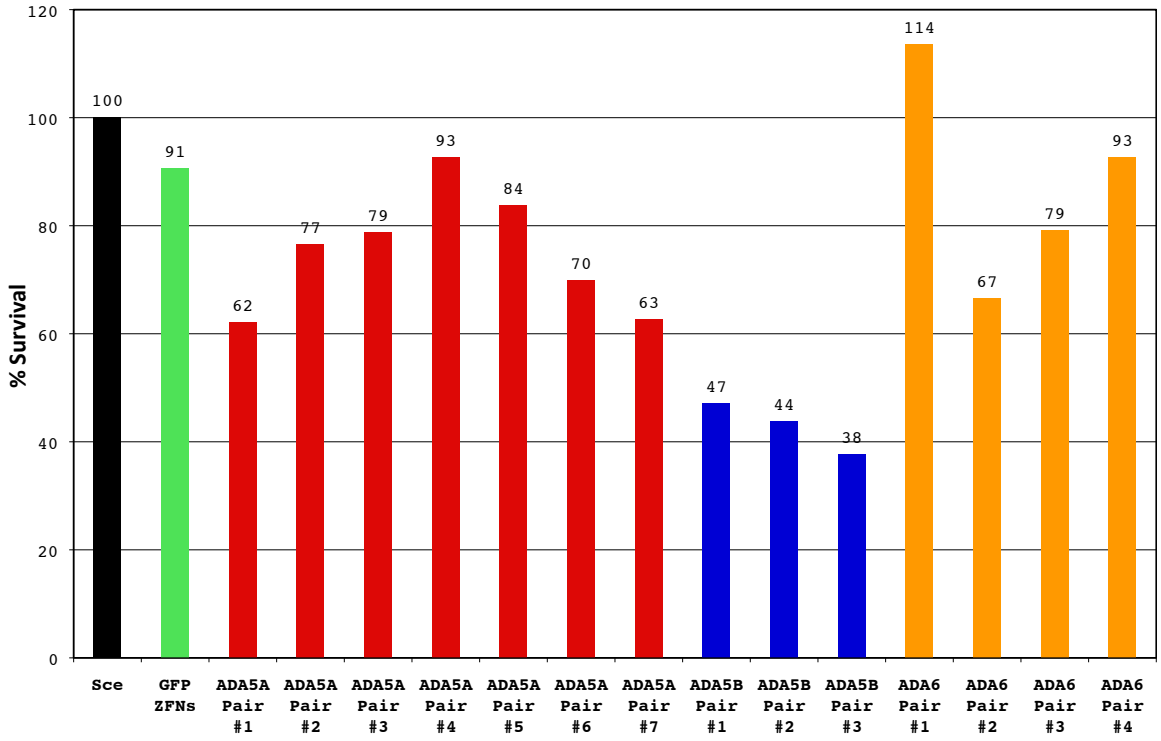

Figure 3.3: ADA ZFNs Demonstrate Mild Cytotoxicity. ADA ZFNs were transfected into HEK 293T and tested in our toxicity assay, as compared to Sce (I-SceI, a non-toxic control) and our GFP ZFNs, which have been shown to have low toxicity and high activity for their target sites. ADA ZFNs showed a variety of toxicities, with most demonstrating only mild cytotoxicity. Only the ZFNs targeting the ADA5B locus demonstrated <50% survival. The ADA5B ZFNs were also the pairs that showed the highest rate of stimulation in the SSA assay, as shown in Figure 3.2.

The cell toxicity assay measures the relative cytotoxicity of different nucleases in human cells and so indirectly assesses the relative specificity of nucleases. High cytotoxicity in this assay is likely due to a lower specificity of the nucleases. Lower specificity leads to an increased number of DSBs in the cells, activating cellular checkpoints and often leading to the initiation of apoptosis. HEK 293T cells are transfected with a GFP expression plasmid as well as the plasmids expressing each nuclease pair. The percent of GFP+ cells is measured at day 2 and then again at day 6 through flow cytometry. The percentage of GFP+ cells naturally decreases as the cells divide and the GFP expression plasmids are diluted out and

degraded. In cells with cytotoxic nucleases, the decrease in GFP+ cells is greatly accelerated as the nucleases increase the rate of cell death in transfected cells over that for non-transfected cells in the sample. The relative cytotoxicity is measured by comparing the decline of %GFP+ cells from day 2 to day 6 in samples with non-cytotoxic nucleases as compared to samples containing experimental nucleases. The active ADA ZFN pairs showed a range of cytotoxicity, with some displaying no significant cytotoxicity and the majority maintaining more than 50% survival (Figure 3.3). I chose those ADA ZFNs that demonstrated the highest activity and lowest cytotoxicity for further experimentation, eliminating the use of the less active or highly toxic nucleases.

The SSA reporter assay was also used to further investigate the potential specificity of the ADA ZFNs. ADA is a relatively well-conserved protein across multiple species and so I looked at the conservation across species of my specific target sites. I observed that both murine and zebrafish ADA genes contained sites for one of the ADA exon 5 targets where 15 out of 18 possible residues were conserved. I created SSA reporter plasmids that contained either the murine or zebrafish ADA locus and used them in the SSA assay with the corresponding ADA ZFNs. The ADA ZFNs stimulated a 2-5 fold increase in GFP+ cells over background when co-transfected with either the murine or zebrafish ADA SSA reporter plasmids (as compared to the 30-50 fold increase when the wild-type target is used). This low degree of stimulation shows that the ADA ZFNs had a high degree of specificity and the mutation in as little as three base pairs in the target site would abrogate nearly all activity. Combined with the cell toxicity assay these results were encouraging results, demonstrating that the ADA ZFNs were both active and specific. The next step was to demonstrate activity

at the endogenous ADA locus, as the SSA assay measures ZFN activity on ectopically transfected plasmids.

Measuring Gene Targeting Efficiency of ADA ZFNs

Gene targeting efficiency of the ADA ZFNs was measured in two ways, stimulation of mutagenic NHEJ at their target sites and stimulation of targeted integration of a Ubc-eGFP construct. K562 cells, a human chronic myelogenous leukemia cell line, were used for these assays. This cell line allows for the generation of high gene targeting frequencies.

Measuring mutagenic NHEJ at the target site is a good initial measurement of ZFN activity at the target site. Highly active nucleases generate correspondingly high rates of mutagenic NHEJ at their target sites. I nucleofected K562s cells with ADA ZFNs using the Amaxa-IIS nucleofection machine. Transfection efficiency in these cells was measured through co-transfection of a GFP expressing plasmid and subsequent analysis by flow cytometry. The cells nucleofected with the ADA ZFNs were collected on day 3 and the genomic DNA isolated. The regions surrounding the target sites were amplified through PCR and then analyzed in the CelI enzymatic mutation detection assay. PCR products are melted and then slowly annealed in order to allow mixing of the two strands from different PCR products. If there are any differences in the PCR strands (due to the amplification of alleles that have undergone mutagenic NHEJ), small bubbles form in the double-stranded DNA in the re-annealing process. These re-annealed PCR products are treated with the CelI enzyme, which recognizes any bubbles due to DNA mismatches and cleaves the products that contain a mismatch. Subsequent analysis through polyacrylamide gel electrophoresis

(PAGE) can reveal multiple bands of DNA in the gel. The WT, uncleaved band, and then the two bands that are formed when a mismatch is recognized and the WT band is cleaved. Unfortunately, there was no measurable mutagenic NHEJ stimulated by the ADA ZFNs observed in this assay.

Concurrent to the CelI assay, I tested the ability of the ADA-ZFNs to target integration of a Ubc-eGFP cassette to the ADA locus. I first created a donor template that contained a Ubc-eGFP cassette flanked by arms of homology to the target sites (Figure 3.4). The arms of homology were created by using PCR to amplify the 800bp region immediately 5' of the targeted cut site for the 5' arm, and PCR amplifying the 800bp region immediately 3' of the target for the 3'arm of homology. The Ubc-eGFP cassette was cloned between the two arms in such a way as to disrupt potential cleavage of the plasmid by the ADA ZFNs. Two such donor plasmids were created, one with arms of homology to the ADA exon 5 target site, and the other to the exon 6 target site. The donor plasmid was nucleofected into K562 cells both with and without co-transfection of ADA ZFNs, and the percentage of GFP+ cells measured over time. By day 21, the percentage of GFP+ cells for which only the donor plasmid was included in the nucleofection leveled out at around 1%. This reflects the tendency of plasmids to randomly integrate into the host genome. When ADA ZFNs were co-transfected in with the donor plasmid, the final percentage of GFP+ cells remained around 1%. Given the absence of mutagenic NHEJ, this was the expected result.

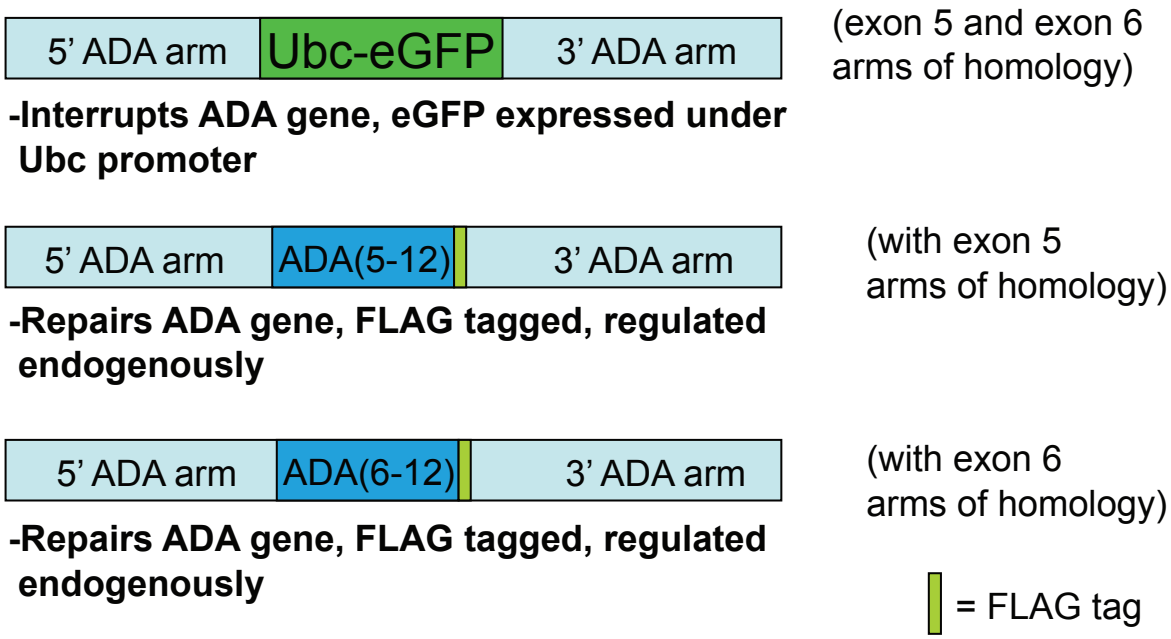

Figure 3.4: Donor Construct Designs for ADA ZFNs. Ubc-eGFP cassette will be inserted as a proof-of-principle for targeted gene insertion at the ADA locus. The other two constructs insert the remaining ADA cDNA in-frame with the targeted locus for functional correction of any downstream mutation. FLAG tag added to differentiate between expression from targeted locus from the endogenous expression of ADA.

Troubleshooting Inability of ADA ZFNs to Target Endogenous Locus

Despite the activity of the ADA ZFNs in the SSA assay, they showed no measurable activity at their endogenous loci. In an effort to stimulate this activity, I set about increasing the levels of ZFN expression in the cells. The original *FokI* nuclease is a bacterial protein and the original expression was not optimized for expression in human cells. Altering the *FokI* DNA such that each codon reflected the most-commonly used codon for that amino acid in human cells (a process known as codon-optimization) was previously shown to increase expression of ZFNs and also increased the activity of ZFNs at their target site.[87] I replaced the *FokI* DNA in my ZFN expression plasmids with the codon-optimized version,

creating new ADA ZFNs that have much better expression profiles in human cells than the previous versions (Figure 3.5).

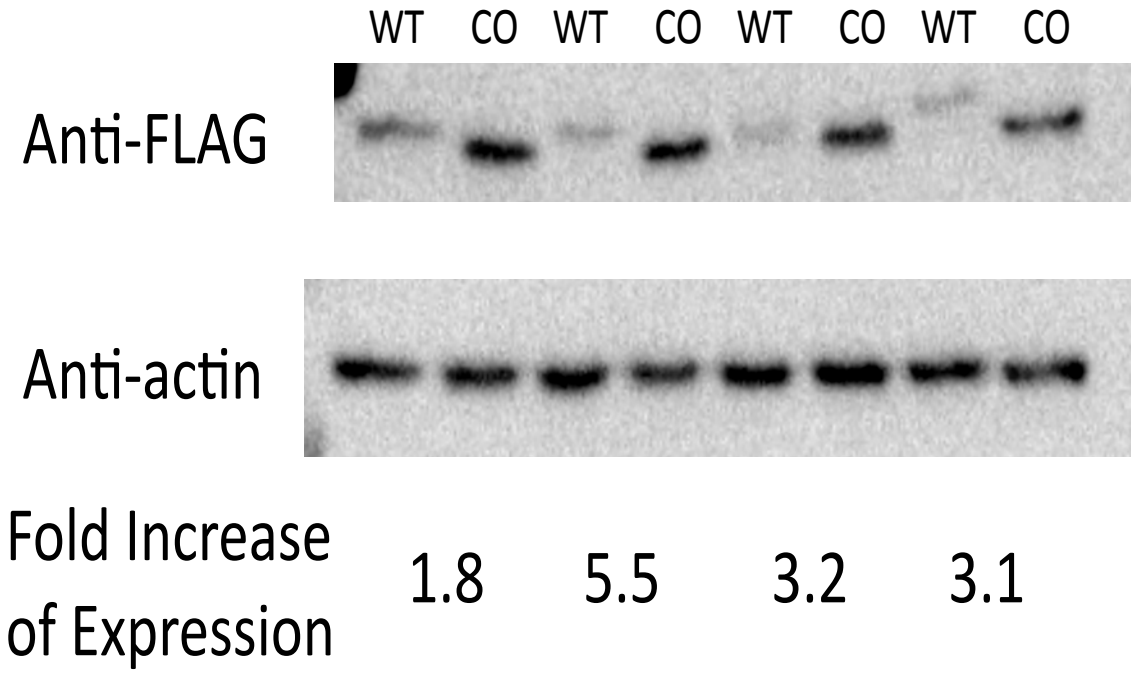

Figure 3.5: Codon-optimized ZFNs Show Increased Expression. HEK 293T cells were transfected with plasmids encoding the ZFNs with the wild-type FokI nuclease domain (WT) or with the codon-optimized FokI nuclease domain (CO). CO ZFNs showed increased levels of expression as compared to the WT ZFNs.

I nucleofected the codon-optimized ADA ZFNs into K562 cells and assayed for mutagenic NHEJ using the CelI assay. After three days, these ZFNs were able to demonstrate a frequency of mutagenic NHEJ at 2% of the alleles in the sample (Figure 3.6). This number remained well below the frequency of mutagenic NHEJ stimulated through the use of the previously described IL2RG ZFNs (nearly 40%). These ADA ZFNs also failed to stimulate increased integration of the donor template in cells co-transfected with both the

donor plasmid and the codon-optimized ADA ZFN plasmids.

- Lane 1: DNA Ladder
- Lane 2: No ZFNs, No Cel1
- Lane 3: No ZFNs, Plus Cel1
- Lane 4: ADA5A ZFNs, Plus Cel1

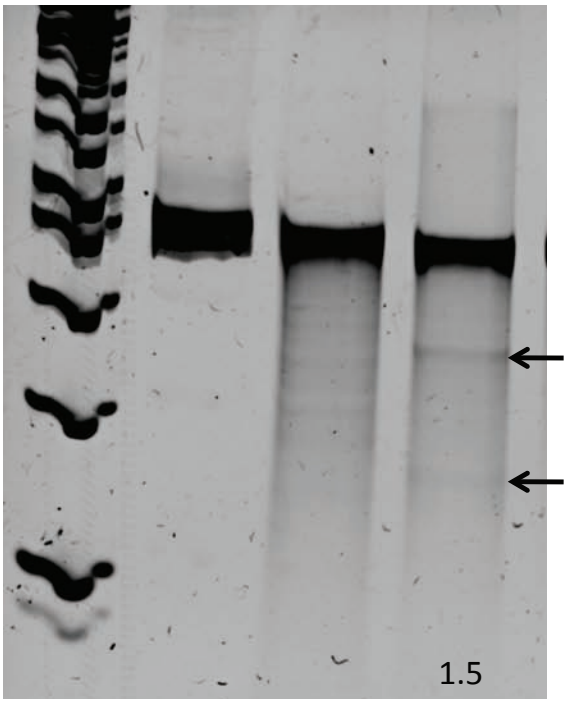

Figure 3.6: ADA5A ZFNs Stimulate Low Amount of Mutagenic NHEJ. Nucleofection of ADA5A codon-optimized ZFNs into K562 cells and subsequent analysis shows that approximately 1.5% of the alleles have been modified by mutagenic NHEJ. Samples nucleofected with no ZFNs do not reveal any bands at the correct length.

It is still not entirely clear why these ADA ZFNs showed robust efficiency of cleavage in the SSA reporter assay but showed minimal cutting at the endogenous ADA loci. The SSA reporter assay relies on episomal cleavage of the target sites contained within plasmid DNA, often having many plasmids contained within the same cell. These sites are more readily available for cleavage by the ADA ZFNs than the endogenous locus. Each diploid K562 cell only contains two potential target sites, thus reducing the chance that the ADA ZFNs would find their target. Additionally, the heterochromatin and methylation states of the ADA gene are unknown, and the presence of either heterochromatin or CpG methylation at the target site would likely inhibit the ability of the ZFNs to bind and cleave

the target sites, resulting in a much lower observed frequency of mutagenic NHEJ. It is also possible that due to any of the above factors, any sub-optimal binding affinity or specificity in the zinc-finger binding domains was magnified in the endogenous setting of the cell. Whatever the cause, it was clear that these ADA ZFNs were not sufficiently active at the ADA gene for the purposes of gene therapy.

Engineered TALENs Efficiently Target Endogenous ADA Locus

In 2011, the Voytas lab published a paper introducing a new architecture for engineered nucleases: transcription activator-like effector nucleases (TALENs).[83] Although the initial paper did not make any claims to the efficiency of their engineered TALENs, they demonstrated that their engineered TALENs did stimulate targeted mutagenesis at the endogenous target site for which they were designed to cut. Importantly, this paper also included a protocol to create newly designed TALENs and made the reagents publicly accessible. The protocol utilized a modification of the Golden Gate assembly method that promised ease of use and the creation of new TALENs in as little as a week using a relatively simple protocol. For comparison's sake, the OPEN method for engineered ZFNs takes 1-3 months and is much more labor-intensive. Due to the low efficiency of the engineered ADA ZFNs, I determined to engineer TALENs to the ADA locus using the Golden Gate assembly method as outlined by Cermak *et al.*[83]

Design and Assembly of ADA TALENs

With the advent of TALENs and their increased repertoire of potential binding sites, our strategy for targeted gene therapy for ADA deficiency also changed. Originally, we engineered ZFNs to target ADA exons 5 and 6, chosen in part due to constraints imposed by the limited number of potential ZFN binding sites. Targeted gene correction of mutations in exons 5 and 6 would only be useful to any patient suffering a mutation in either of those exons, useful to a very limited number of patients suffering from ADA deficiency. To expand the pool of patients that could potentially be helped by our strategy, we designed a strategy to fuse the remaining ADA cDNA onto the end of the targeted exon, and so functionally be able to correct any mutation that occurred downstream of the targeted exon. Following the failure of the ADA ZFNs to stimulate gene targeting and the concurrent arrival of TALENs with their expanded repertoire of potential sites, the strategy of my project shifted to developing a method of gene therapy that could functionally correct almost all known mutations that cause ADA deficiency. I engineered TALENs that could efficiently target the start site of the ADA gene and use targeted gene insertion in order to insert a full copy of ADA cDNA under control of the endogenous ADA promoter (Figure 3.7). Not only would this functionally correct any known mutation disrupting the ADA gene, it would also leave the cDNA under control of the endogenous promoter, thus maintaining some level of endogenous regulation of expression. This strategy could restore ADA enzymatic function to almost all ADA deficient patients, excepting those that have bi-allelic deletion of the ADA promoter region.

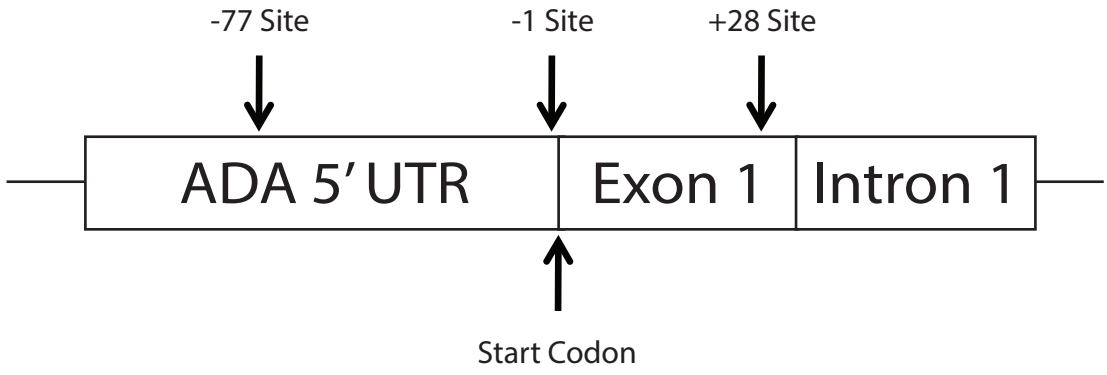

Figure 3.7: Diagram of ADA TALEN Recognition Sites. The -77 TALENs are designed to recognize a site centered 77bps upstream of the ADA translational start site. -1 TALENs recognize a site centered at the start site, and the +28 TALENs recognize a site 28bps downstream of the start site.

Using the Golden Gate method of assembling TALENs developed by the Voytas lab, I designed and assembled five pairs of TALENs that are engineered to target sites around the start codon of the ADA gene. Initial reports concerning the design of successful TALENs suggested that the target for each individual TALEN must be preceded by a 5' T residue, and that the most 3' residue must also be a T. Due to a surprising dearth of T residues surrounding the ADA start codon, I was initially unable to design a TALEN pair that directly targeted the ADA translational start site. Accordingly, I designed several TALENs pairs that would target regions surrounding the ADA start codon. After successfully assembling these pairs of TALENs, I set out to verify the activity of these new ADA TALENs to bind and cleave their target sites. All five pairs showed minimal cytotoxicity in the cell toxicity assay, maintaining at least 75% cell survival (Figure 3.8).

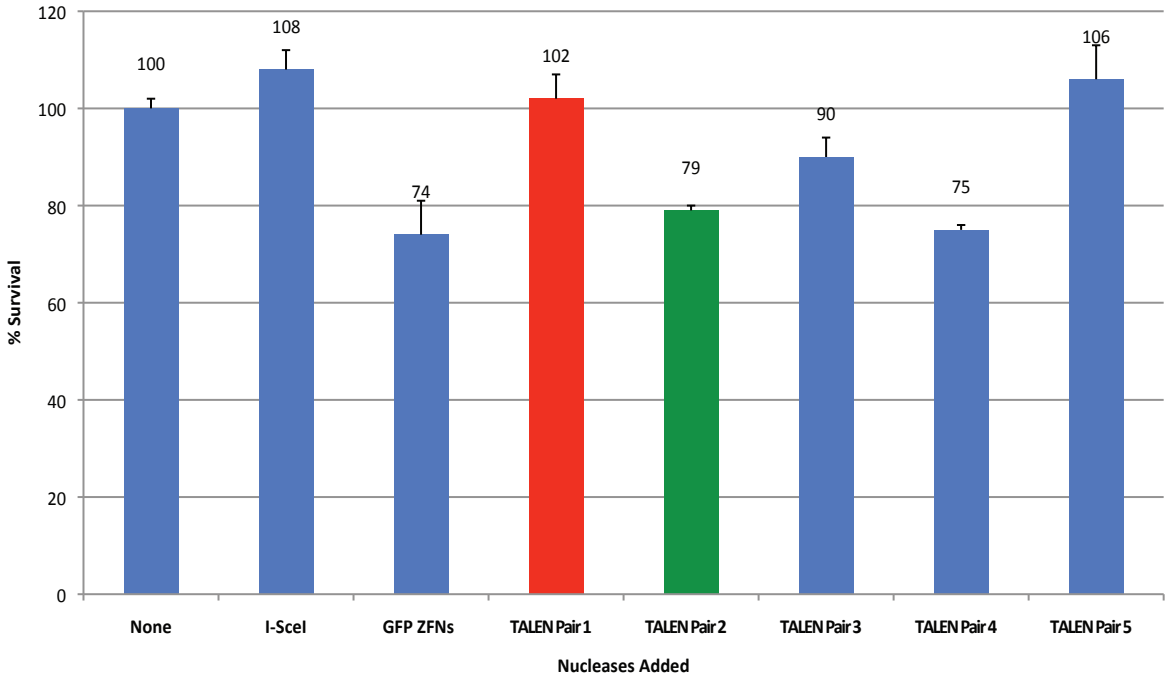

Figure 3.8: ADA TALENs Show Only Mild Cytotoxicity. HEK 293T cells were transfected with the indicated nucleases and the relative toxicity measured after 6 days. The red bar is the -77 TALENs used in subsequent assays, and the green bar represents the +28 TALENs used.

Verification of TALEN Cutting Efficiency at Target Loci

I tested the ADA TALENs in the CelI enzymatic mutation detection assay to measure TALEN activity at each of the endogenous target loci. I nucleofected the plasmids encoding each pair of TALENs into K562 cells, along with a plasmid encoding GFP expression as a measure of transfection efficiency. Three days after nucleofection, I collected the cells and isolated the genomic DNA. I amplified the targeted region through PCR and assayed the PCR products for mutations. Two of the pairs showed the ability to stimulate mutagenic NHEJ at their target loci with high efficiency, at frequencies ranging from 10-15% of alleles in the population (Figure 3.9). These TALEN pairs' target sites are centered at the residues 77bp upstream and 28bp downstream of the translational start. These TALEN pairs will be

referred to as the -77 TALENs and the +28 TALENs respectively, in reference to their relation to the translational start site of the ADA gene.

- Lane 1: DNA Ladder
- Lane 2: No TAELNs, No Cel1
- Lane 3: No TALENs, Plus Cel1
- Lane 4: -77 TALENs, No Cel1
- Lane 5: -77 TALENs, Plus Cel1
- Lane 6: No TALENs, No Cel1
- Lane 7: No TALENs, Plus Cel1
- Lane 8: +28 TALENs, No Cel1
- Lane 9: +28 TALENs, Plus Cel1

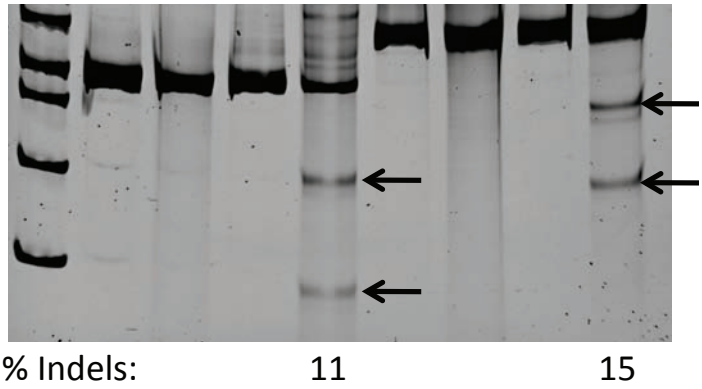

Figure 3.9: ADA Exon1 TALENs Stimulate Mutagenic NHEJ at ADA Locus. K562 cells were nucleofected with the -77 TALENs and the +28 TALENs and assayed for mutagenic NHEJ events on day 3. The signal is only observed when both the TALENs and the Cel1 enzyme are present, validating specificity of bands.

To validate the accuracy of the signals obtained in the CelI assay, I sequenced the PCR products. Sanger sequencing of individual PCR products was accomplished by cloning the products into a bacterial plasmid and sequencing each subsequent bacterial colony individually. The sequencing results confirmed site-specific mutagenic events at a frequency similar to that showed in the CelI assay. The mutagenic NHEJ events observed were mostly small to large deletions at the target site with a smaller percentage of clones showing small insertions (Figure 3.10).

Genotype	DNA Sequence
Wild Type	GGCCGGCCGCGGCCACCGCTGGCCCCAGGGAAAGCCGAGCGGCCACCGAGCCGGCAGAG
1bp del	GGCCGGCCGCGGCCACCGCTGGCCCCAGGGA-AGCCGAGCGGCCACCGAGCCGGCAGAG
1bp del	GGCCGGCCGCGGCCACCGCTGGCCCCAGGGA-AGCCGAGCGGCCACCGAGCCGGCAGAG
2bp del	GGCCGGCCGCGGCCACCGCTGGCCCCAGGG--AGCCGAGCGGCCACCGAGCCGGCAGAG
42bp del	----------CCACCGAGCCGGCAGAG
45bp del	----------ACCGAGCCGGCAGAG
48bp del	----------CGAGCCGGCAGAG
83bp del	----------
54bp del, 73bp ins	AGGCTGAGGCAGGAGAATGGTGTGAACCCTGGAGCCGAGCGGCCACCGAGCCGGCAGAG

Figure 3.10: Sample Mutations From Mutagenic NHEJ Stimulated By ADA -77 TALENs. A selection of mutants obtained when treating K562 cells with the ADA -77 TALENs. Orange bases represent the TALEN recognition sites, dashes represent deleted bases, magenta bases represent an insertion. All mutations seem to originate at the middle of the spacer region, which is the predicted cut site for the TALENs.

Establishing Successful Gene Targeting in K562 Cells Using ADA TALENs

After successfully engineering nucleases that efficiently cleave the endogenous ADA locus, I set out to demonstrate that these TALENs can also stimulate targeted gene insertion through homologous recombination. The efficiency of these ADA TALENs to stimulate gene targeting was first tested in the K562 cell line, due to the ability of many other nuclease pairs to demonstrate high levels of gene targeting in these cells.

Targeted Insertion of a Ubc-eGFP cassette by ADA TALENs

The first step was to demonstrate proof-of-principal gene insertion at the ADA locus, using a Ubc-eGFP cassette for ease of subsequent analysis of the samples through flow cytometry. I created a donor template with 850bp arms of homology centered at the ATG

start codon of the ADA gene that contained a Ubc-eGFP cassette inserted in between the arms of homology (Figure 3.11). In this manner, the stimulation of targeted integration could initially be measured through observing the increase in the percentage of GFP+ cells upon co-transfection of the donor and the ADA TALENs, as compared to those cells nucleofected with the donor template alone.

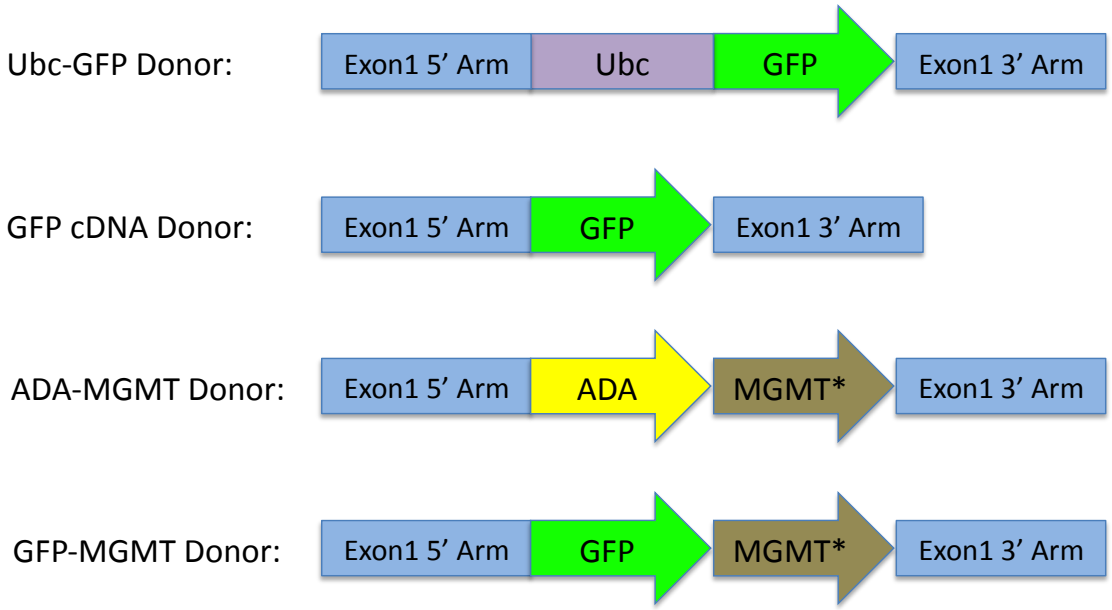

Figure 3.11: Donor Constructs for Targeted Gene Insertion at the ADA Start Locus. 5' arm of homology ends with the base preceding the ATG start codon, and the 3' arms start with the ATG start codon. GFP cDNA and ADA constructs are in-frame so as to be expressed from the endogenous ADA promoter. MGMT* codes for the P140K version of MGMT, which allows for drug selection of targeted cells, as described in the mthods section. The T2A linker allows co-transcription of the P140K MGMT cDNA with the ADA cDNA but results in translation of two distinct proteins.

Upon co-transfection with the Ubc-eGFP donor template into K562 cells, both pairs of TALENs stimulated integration of the Ubc-eGFP cassette into the ADA locus. The samples co-transfected with the ADA TALENs and the donor template maintained a stable GFP+ population of around 10% of cells, compared to those nucleofected with the donor template alone with a stable GFP+ population of around 1% of cells (Figure 3.12).

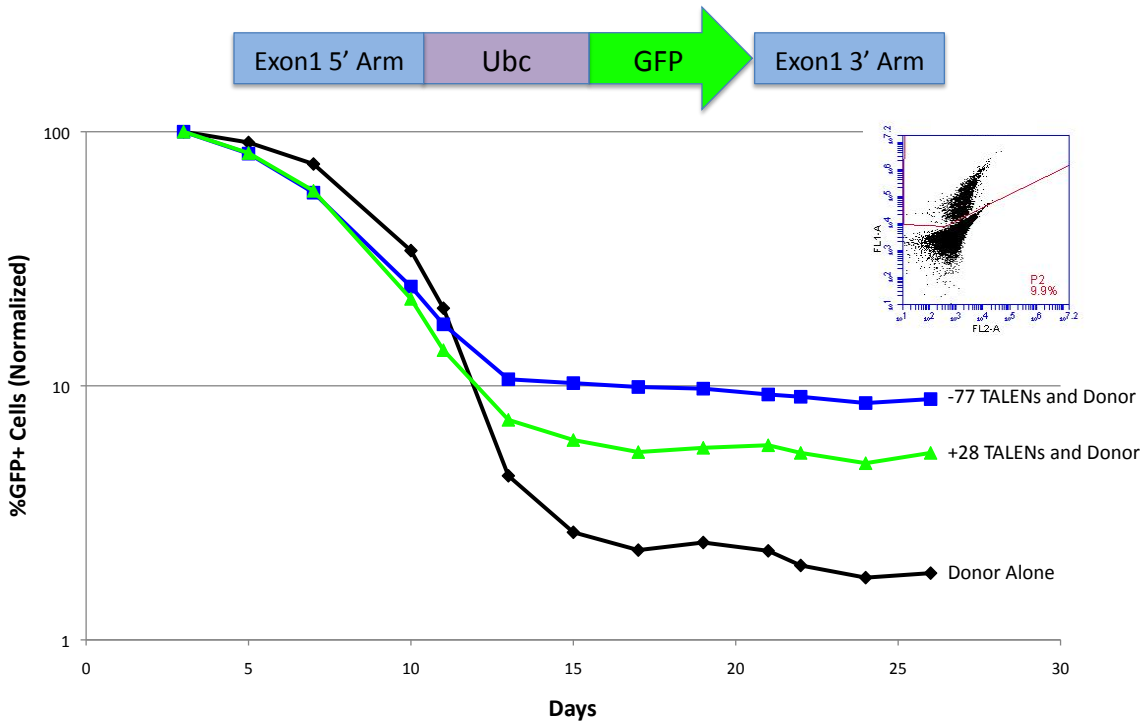

Figure 3.12: ADA TALENs Stimulate Integration of Ubc-GFP Cassette. K562s were nucleofected with either the Ubc-GFP donor alone, or with the donor plus ADA TALENs. The TALENs demonstrated the ability to stimulate targeted gene insertion of the donor cassette, as evidenced by continual GFP expression. A representative plot is shown.

Previous studies using nucleases to stimulate targeted integration utilized donor templates with arms of homology centered on the cut site, while the Ubc-eGFP donor arms of homology were centered on the ADA translational start site and so slightly offset from the target sites. I postulated that the efficiency of the ADA TALENs to stimulate targeted gene insertion was underrepresented with the current donor. To assess this, I created donor templates with arms of homology centered on the -77 site and the +28 site, again with the Ubc-eGFP cassette splitting the two arms of homology in the plasmid. I nucleofected K562 cells with the ADA TALENs and donor templates in order to determine if efficiency of insertion of the Ubc-eGFP cassette was enhanced using the new donors. Both the -77 TALENs and the +28 TALENs showed a slightly higher efficiency of integration with the

new donors, stimulating targeted integration at about 120% of the efficiency displayed by the original Ubc-eGFP donor. Despite this modest enhancement of integration, it was decided to continue using donors whose arms of homology are centered at the ADA start codon to minimize the disturbance to the endogenous locus in future experiments.

Targeting In-frame Expression of GFP cDNA From Endogenous ADA Promoter

The ultimate goal is the integration and expression of ADA cDNA from the endogenous promoter, so I next set out to demonstrate targeted expression of GFP cDNA from the ADA locus as a proof-of-principle. I took the original Ubc-eGFP donor template (with arms of homology centered at the ADA start codon) and exchanged the Ubc-eGFP cassette for a cassette containing the full GFP cDNA with a bGH polyA signal (for enhanced expression in mammalian cells).

I co-transfected K562 cells with the new GFP cDNA donor and plasmids expressing the ADA TALENs and followed GFP expression over time. It should be noted here that in the absence of a strong promoter (i.e., Ubc), GFP expression is much lower in these cells and GFP expression more difficult to distinguish from autofluorescence using flow cytometry, leading to artificially low numbers of transfected and targeted cells. Nevertheless, the ADA TALENs still demonstrated a noticeable increase in percentage of GFP+ cells in the samples when compared to those samples nucleofected with the GFP cDNA donor template alone, 2% and 0.4% respectively (Figure 3.13).

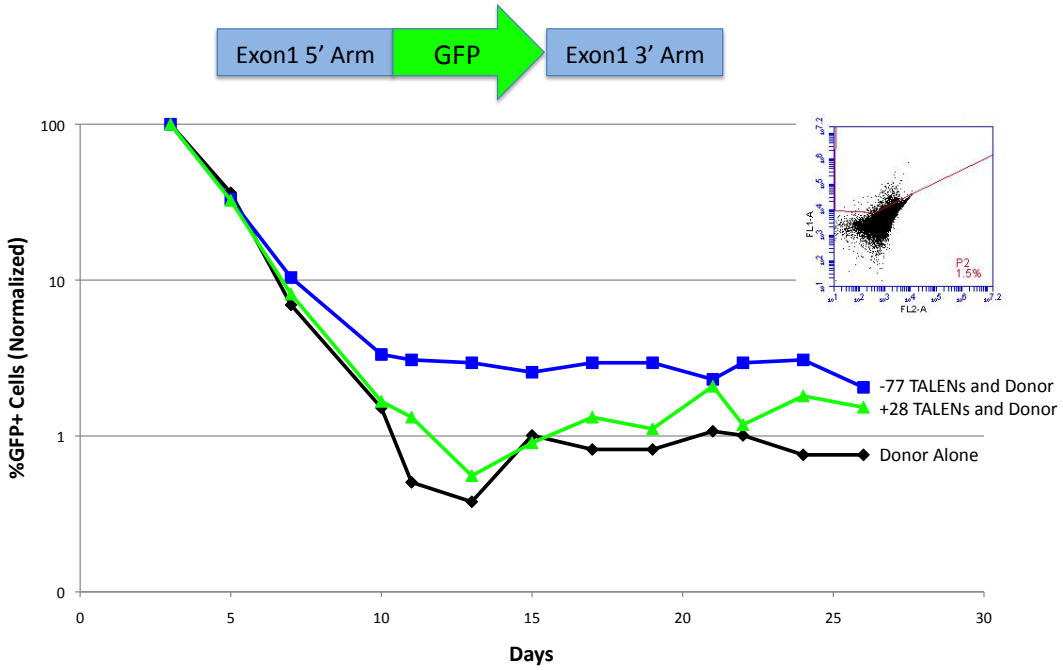

Figure 3.13: ADA TALENs Stimulate Targeted Gene Insertion of In-Frame GFP cDNA. K562 cells were nucleofected with either the GFP cDNA donor alone, or with the donor and the ADA TALENs. Both the -77 and +28 TALEN pairs were able to stimulate targeted integration of GFP cDNA.

To confirm the ADA TALENs had stimulated targeted gene integration of the Ubc-eGFP and GFP cDNA cassettes (as opposed to random integration), a PCR strategy was designed that would amplify only correctly integrated cassettes. This strategy utilizes one primer inside of the predicted insert and the other primer outside of the insert, located just upstream of where the 5' arm of homology begins. I will refer to this PCR strategy as genomic PCR, as it will only amplify cells that contain at least one targeted integration in the genome at the ADA locus. The cells were first enriched for targeted cells by sorting for GFP+ cells by fluorescence assisted cell sorting (FACS). Genomic DNA was isolated from samples nucleofected with the donor template alone and also the GFP+ enriched samples co-transfected with both donor template and ADA TALENs. Only those samples co-transfected

with both the donor and the ADA TALENs showed amplification using this PCR strategy, confirming on-target integration of both the Ubc-eGFP and GFP cDNA cassettes to the ADA locus. It should be noted that while this genomic PCR strategy confirms successful integration of the donor templates, it does not quantify the percent of targeted alleles (Figure 3.14).

Lane 1: DNA Ladder
Lane 2: Negative PCR Control
Lane 3: Ubc-eGFP Donor Only
Lane 4: Ubc Donor, -77 TALENs
Lane 5: Ubc Donor, +28 TALENs
Lane 6: GFP cDNA Donor Only
Lane 7: cDNA Donor, -77 TALENs
Lane 8: cDNA Donor, +28 TALENs

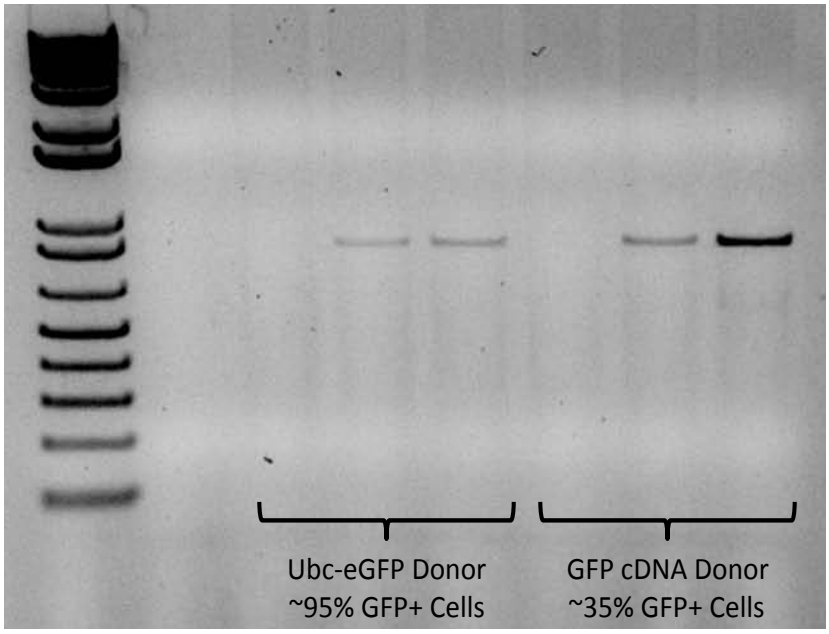

Figure 3.14: Confirmation of Targeted Gene Insertion in K562 Cells. K562 cells from Figures 3.12 and 3.13 were sorted for GFP+ cells. Genomic DNA was isolated from sorted cells and amplified using the startegy outlined above that specifically amplifies only targeted alleles.

To quantify the percentage of cells that had undergone successful site-specific integration of the GFP cDNA cassette at the ADA locus, I isolated 82 single-cell clones for further characterization from a sample co-transfected with the donor template and the -77 TALENs. After isolation of the single-cell clones, I expanded the clones in order to have sufficient DNA for subsequent analysis. Of the 82 single-cell clones, 14 contained GFP+

cells. These 14 clones were analyzed by genomic PCR and subsequent Sanger sequencing, and all 14 were shown to contain at least one allele that had successfully integrated the GFP cDNA construct at the ADA locus. These observations show that 17% of the clones had undergone targeted gene integration. This number is consistent with the approximately 10% of GFP+ cells observed when using the Ubc-eGFP cassette as the donor template.

Selection of Targeted Cells Using P140K MGMT

While using GFP to enrich for targeted cells serves as a valuable tool in the laboratory, it is not a strategy that can be used therapeutically. Davis *et al.* demonstrated that expression of the P140K mutant version of the methyl-guanine methyl transferase (MGMT) can overcome the toxic effects of serial treatment with O6-benzylguanine (O6BG) and carmustine (also referred to as BCNU).[85] O6BG and BCNU are chemotherapy drugs already in clinical use. Inclusion of the P140K MGMT expression cassette in our donor template allows for selection of targeted cells, both *in vitro* and *in vivo*.[88] The P140K MGMT expression was linked to the expression of the in-frame GFP cDNA through the use of the T2A system. This system allows for multicistronic expression (multiple genes from one translational start site). In this manner, both the GFP and the P140K MGMT can be driven from the endogenous ADA promoter.

I also created a donor template that replaces the in-frame GFP cDNA with a modified in-frame ADA cDNA. The first modification to the ADA cDNA was the addition of a 3x-FLAG tag to the C-terminus. This will allow for the differentiation of targeted expression of ADA from the wild-type expression of ADA. The C-terminus of ADA was chosen for the

addition of the FLAG residues as it is not well conserved across species, and so the FLAG peptides are less likely to interfere with ADA enzymatic activity. I also altered 3bps in exon 1 of the tagged ADA cDNA in order to more easily distinguish an allele that has undergone targeted gene insertion from one that has not through DNA sequencing. The bases altered were the 6th, 9th and 12th bases in exon 1. Each of the alterations were chosen based on the "wobble" theory, that there are multiple codons that encode for the same amino acid, and changing the last base pair of the triplet often does not change the amino acid incorporated into the subsequent protein. We can use these wobbles in order to obtain a more quantitative measurement of the efficiency of targeting.

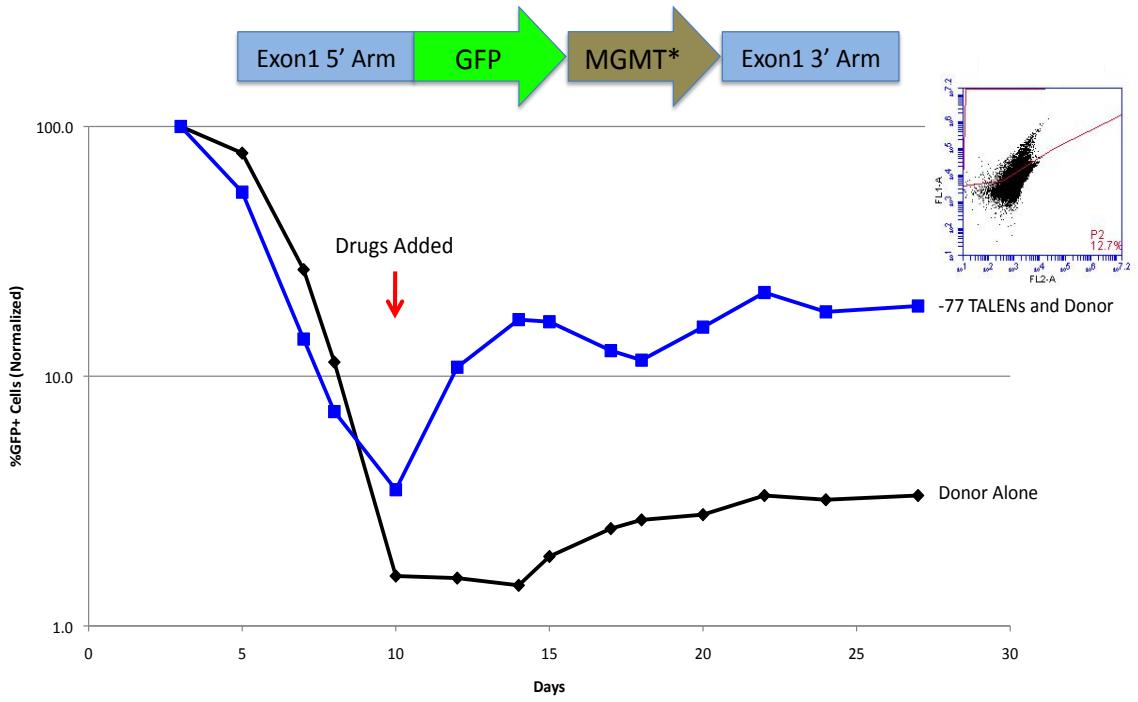

Figure 3.15: P140K MGMT Allows for Selection for Targeted Cells. ADA TALENs were used to target integration of a GFP-T2A-P140K MGMT cassette at the endogenous ADA locus. On day 10 (indicated by the red arrow), K562 cells were treated with drugs that select for cells expressing the P140K mutant of MGMT, as outlined previously. Corresponding GFP expression was increased following treatment, indicating enrichment for targeted cells.

K562 cells were nucleofected with these two P140K MGMT donors along with the two ADA TALEN pairs, with cells nucleofected with donor alone as controls. After ten days, the cells were treated with O6BG and BCNU as described in the methods section. After treatment, the samples co-transfected with TALENs and the GFP-MGMT donor showed an increase in GFP+ cells (Figure 3.15), demonstrating that the P140K MGMT selection is able to enrich for targeted cells.

1: DNA Ladder
2: GFP-MGMT Donor Only
3: GFP-MGMT, -77 TALENs
4: GFP-MGMT, +28 TALENs
5: ADA-MGMT Donor Only
6: ADA-MGMT, -77 TALENs
7: ADA-MGMT, +28 TALENs
8: Blank
9: GFP-MGMT Donor Only
10: GFP-MGMT, -77 TALENs
11: GFP-MGMT, +28 TALENs
12: ADA-MGMT Donor Only
13: ADA-MGMT, -77 TALENs
14: ADA-MGMT, +28 TALENs

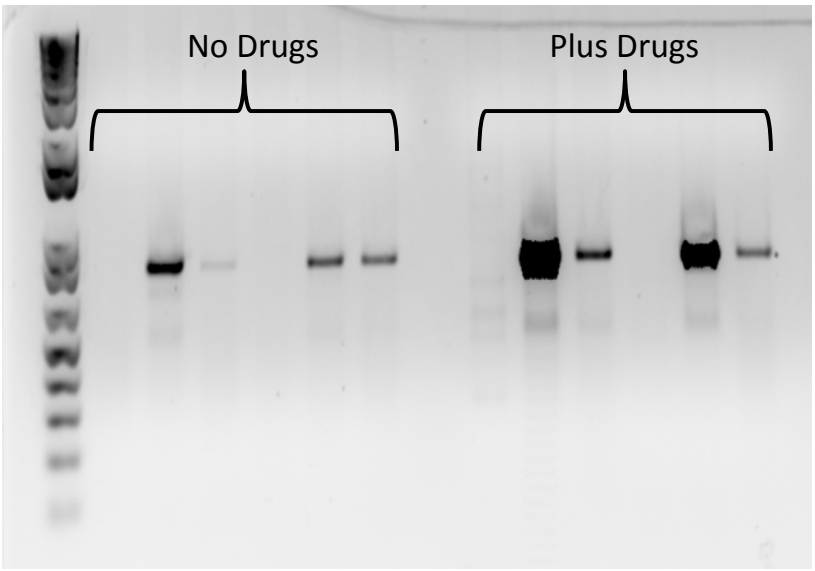

Figure 3.16: Targeted Gene Insertion of P140K MGMT Allows for Selection of Targeted Cells. K562 cells were nucleofected with donor alone or donor and TALENs. Twelve days after nucleofection, cells were treated O6BG and Carmustine as detailed previously. Targeted locus amplified by using primers that are specific for alleles having undergone targeted gene insertion.

The ADA-MGMT donor does not have a fluorescent protein, but the samples were treated in the same manner as the GFP-MGMT cells. Non-drug treated samples were also maintained as a control. Two weeks after drug selection, genomic DNA was isolated from all of the samples and the targeted region amplified through genomic PCR. Only those samples co-transfected with TALENs and donor template were amplified, and the samples treated with drugs showed increased levels of amplification (Figure 3.16). The P140K

MGMT system was thus shown to be an effective tool for enriching for targeting cells can be used therapeutically.

Gene Targeting in Human Patient-Derived Fibroblasts

Having established efficient gene targeting procedures in the human K562 cell line, we next sought to establish the feasibility of using the ADA TALENs to stimulate gene targeting in primary cell lines. Accordingly, we obtained primary fibroblasts from Coriell Cell Repositories that were derived from a patient suffering from ADA deficiency. Both ADA alleles in these cells were mutated, one in exon 4 and the other in exon 7. Clinically, this patient presented with <1% normal ADA activity and exhibited the SCID phenotype.

Verifying ADA TALEN Activity in Fibroblasts

After establishing the correct culturing and nucleofection protocols for the patient-derived fibroblasts, I tested the ability of the ADA TALENs to stimulate gene targeting in these cells. I nucleofected the ADA TALENs into the fibroblasts in order to determine their ability to create DSBs at the endogenous ADA locus in these cells, as measured by mutagenic NHEJ. 3 days after nucleofecting the fibroblasts, I collected the cells and isolated the genomic DNA. After amplifying the ADA locus through PCR, I used the PCR products in the CelI assay in order to quantify the percentage of PCR products that had undergone mutagenic NHEJ. Analysis revealed that the -77 TALENs and the +28 TALENs were able to stimulate mutagenic NHEJ at a frequency of 14% and 18% respectively, similar to the

frequencies observed in K562 cells. Site-specific mutagenesis was confirmed through Sanger sequencing individual PCR products (as described above), with deletions again being the most common form of mutation.

Targeted Integration of Ubc-eGFP Cassette in Primary Fibroblasts

Now that the ability of the ADA TALENs to stimulate mutagenic NHEJ at the ADA locus in primary fibroblasts was established, I tested the efficiency with which they could stimulate targeted gene insertion in these same cells. I co-transfected the ADA TALENs into the fibroblasts with the original Ubc-eGFP donor plasmid and followed the percentage of GFP+ cells over time. After four weeks, the ADA TALENs were shown to stimulate integration of the Ubc-eGFP cassette at a frequency approximately ten-fold above that of the samples nucleofected with the donor alone (Figure 3.17). This stimulation is very similar to that seen in K562 cells (1.1% stimulated to 9.5% in K562 cells versus 0.19% stimulated to 1.77% in fibroblasts), although the absolute percentages of GFP+ cells are significantly lower. The K562 cell line has shown to have high genomic instability, which instability is probably a leading cause of the increased frequency of integration observed in those cells, in addition to reduced eGFP expression from the Ubc promoter in fibroblasts. Genomic PCR was performed on the genomic DNA of the fibroblast samples confirming targeted integration of the Ubc-eGFP cassette to the ADA locus, with only the samples co-transfected with the ADA TALENs and donor plasmid able to be amplified by this PCR assay.

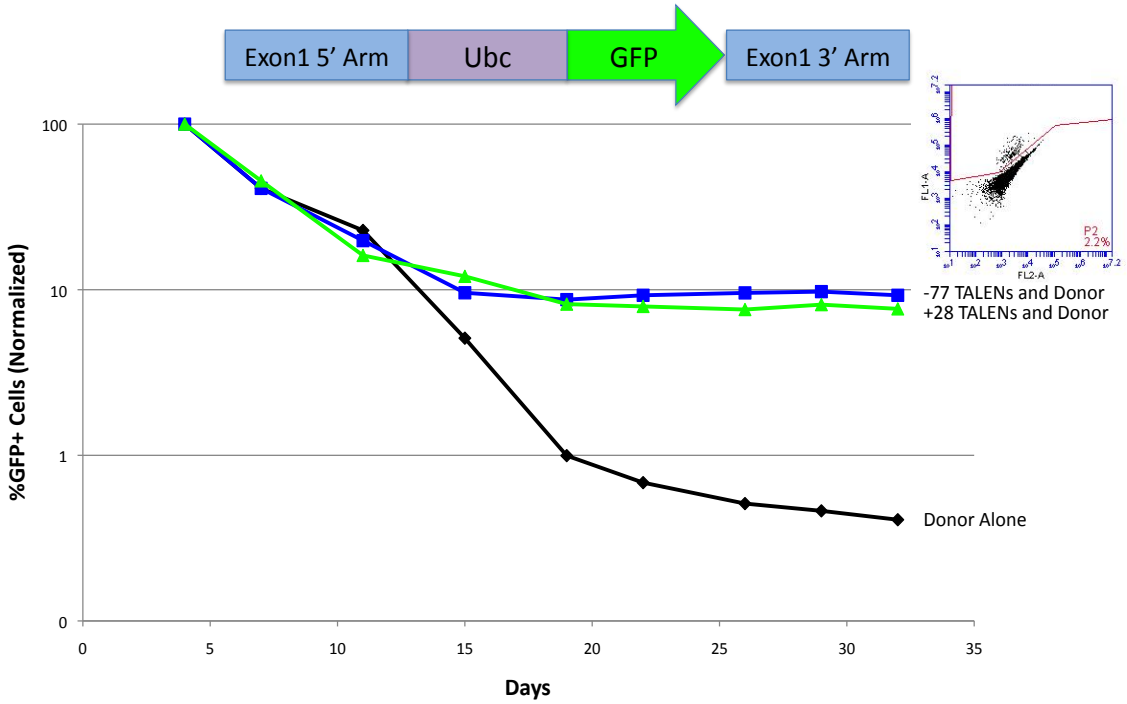

Figure 3.17: Targeted Gene Insertion in Primary Fibroblasts. Patient-derived fibroblasts were nucleofected with either the Ubc-GFP donor alone or with the donor and ADA TALENs. Cells were analyzed for %GFP+ cells using flow cytometry.

Targeted Insertion of In-Frame GFP cDNA Cassette in Primary Fibroblasts

I next nucleofected the primary fibroblasts with the GFP cDNA donor template to verify that I could achieve targeted expression of an in-frame cDNA insertion. Despite following the same protocol that demonstrated efficient gene addition using the Ubc-eGFP cassette, no increase in the percentage of GFP+ cells was observed when compared to those nucleofected with donor alone. It should be noted that the percentage of GFP+ cells was very low in these samples (<0.1%) and those that cells that were GFP+ exhibited a low level of GFP expression. Genomic PCR confirmed that there was targeted integration in the TALEN and donor co-transfected cells, demonstrating that the expression of GFP from the

endogenous ADA locus fell below the levels of detection of our flow cytometer. The low expression levels were probably due to a combination of lower targeting frequencies that occur in primary cells and of lower levels of endogenous ADA expression in fibroblasts as compared to lymphoblasts (which K562 cells originated as, before becoming transformed).

I co-transfected the ADA TALENs and the GFP-MGMT and ADA-MGMT donors into primary fibroblasts to determine if I could use this system in primary fibroblasts. No enrichment of targeted cells was shown after treatment with O6BG and BCNU. This failure of the drug selection was thought to be due to the low endogenous expression of the targeted cassette from the ADA locus. Low levels of P140K MGMT expression is insufficient to allow for enrichment of targeted cells.

Utilizing the ADA Locus as a "Safe Harbor" for Gene Insertion

We next explored the ability of the ADA locus to act as a "safe harbor" for targeted gene insertion. As explained earlier, a safe harbor site is one in which targeted gene insertion is used to specifically integrate a desired transgene into one locus in the human genome. While the process of designing and assembling efficient nucleases to a target site has been dramatically simplified in the last several years, it is still a labor-intensive process to screen through engineered nucleases and verify nucleases that are both highly active and highly specific. It would be much more efficient to have one pair of well-characterized nucleases that could specifically and consistently target a variety of transgenes to its locus. In this way, only donor template would need to be changed to insert an expression cassette for you gene

of interest, whether it be ADA, β-globin (for sickle-cell disease and thalassemias), human growth hormone (for growth hormone deficiency), or factor VIII (for hemophilia).

It may seem counter-intuitive to use the ADA locus as a safe harbor, having already discussed the many deleterious affects of ADA deficiency in patients. Despite this fact, there are several reasons why we believe that the ADA locus is a good candidate for use as a safe harbor. ADA is ubiquitously expressed in all cell types, ensuring the chromatin status of the site will be amenable for transgene expression. Most gene targeting events only occur on one of the two chromosomes, and people that are heterozygous for ADA mutations display no phenotype. ADA is a systemic disease, so knocking out ADA expression in any individual cell is not toxic, so long as there are other ADA expressing cells in to maintain the systemic levels of dAdo. Additionally, to increase the safety profile of the use of the ADA locus as a safe harbor, the full ADA cDNA will be inserted in-frame with the endogenous promoter, which will be followed by the cassette expressing the transgene of interest. I use the modified ADA cDNA here, with the 3 wobble mutations and the 3x-FLAG tag. In order to facilitate recognition of, and selection for, targeted clones, I included in the transgene cassette either tdTomato (a fluorescent marker) or P140K MGMT (allowing for drug selection) (Figure 3.18). TdTomato and P140K MGMT were linked to the expression of the gene of interest (GOI) through a T2A sequence, which allows for the multicistronic expression of genes. By linking the expression of the selectable marker to that of the GOI and placing the selectable marker as the downstream of the two genes, we ensure that any cell expressing the selectable marker is also expressing the GOI. Using the T2A system also

allows us to only use one promoter in the donor plasmid, decreasing the size of the insert and increasing the efficiency of both nucleofection and targeted gene insertion.

I used three different genes to test the feasibility of using the ADA locus as a safe harbor for targeted integration of transgenes. The three genes chosen are human Factor VIII (fVIII), human growth hormone (hGH), and a mutated form of human platelet-derived growth factor (PDGF) that displays higher activity than the wild-type version. FVIII is a clotting factor in the human blood that, when deficient, leads to hemophilia. Growth hormone deficiency results in both short stature and general pituitary gland disregulation. Treatments through regular hGH injections are available, but expensive, costing over $900 a month. PDGF is involved in wound healing and the mutated version of PDGF allows for quicker cell proliferation and subsequent acceleration of the rate of wound healing in mice. These targeted gene insertion experiments were done in the ADA deficient fibroblasts for the purpose of showing successful targeting of the ADA cDNA to the endogenous start site, as well as the fact that they are primary cells whose rates of targeting will be much more similar to patient cells than the K562 cell line. Additionally, wound-healing assays can be done on mice with human fibroblasts in order to demonstrate function activity of the transgene.

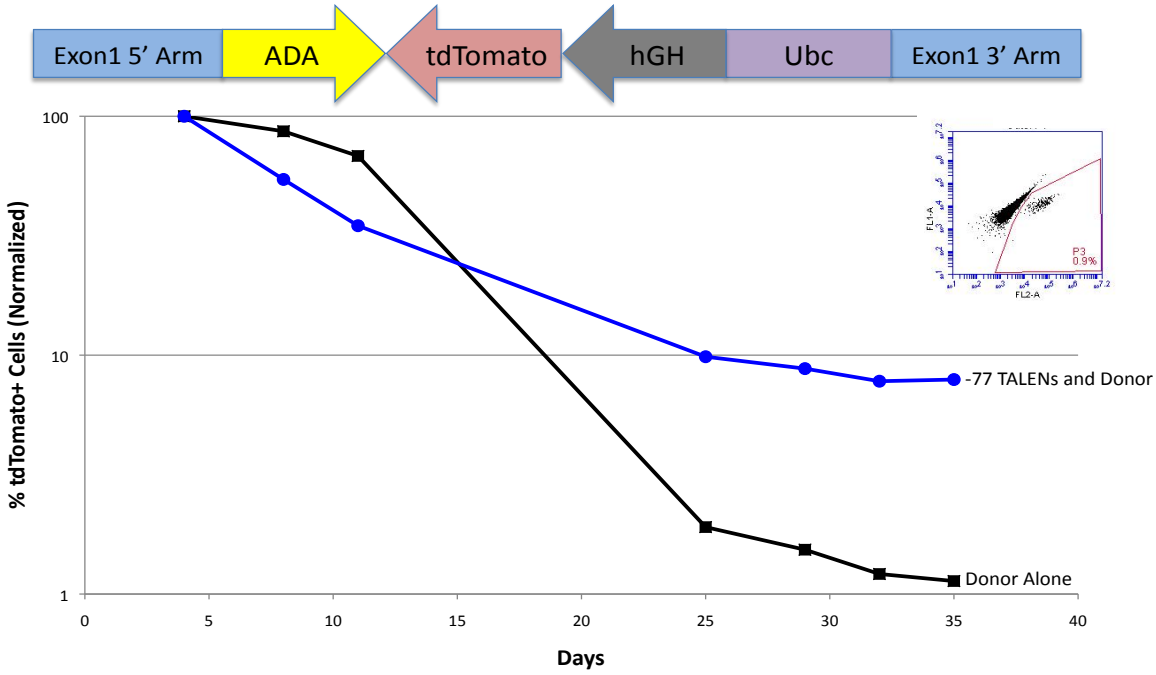

Figure 3.18: Targeted Gene Insertion of hGH construct. Primary fibroblasts were nucleofected with donor alone or with donor and ADA -77 TALENs. Samples analyzed for tdTomato+ cells using flow cytometry.

Each of these transgenes was cloned into the donor plasmid used previously, replacing the Ubc-eGFP cassette with an in-frame ADA cDNA-Ubc-GOI-T2A-tdTomato cassette. These donors were nucleofected in the fibroblasts along with the ADA TALENs. TdTomato expression was followed over the course of four weeks through flow cytometry. All samples co-transfected with the donor plasmid and ADA TALENs demonstrated a higher level of stable %tdTomato+ cells than those samples nucleofected with donor alone (Figure 3.18, Figure 3.19). The samples containing the fVIII transgene showed a lower transfection efficiency (as measured by tdTomato expression) and greater cytotoxicity than the samples nucleofected with either the hGH or PDGF donors. This increased cytotoxicity for the fVIII donors is probably due to the large size of the fVIII cDNA, being over 8kb in length, while hGH and PDGF are both under 1kb. FVIII protein may also be toxic to these cells. FVIII

expression in these cells was also very low, so use of this donor plasmid was discontinued in further experiments. The cells nucleofected with the hGH or PDGF donor templates grew rapidly and appeared much healthier than those with the fVIII donor (possibly due to the increased expression of the growth factors in the cells). Despite the increase in percentage of tdTomato+ cells in those samples nucleofected with ADA TALENs compared to those nucleofected with donor alone, further characterization of these samples was necessary in order to determine the efficiency of targeted gene insertion.

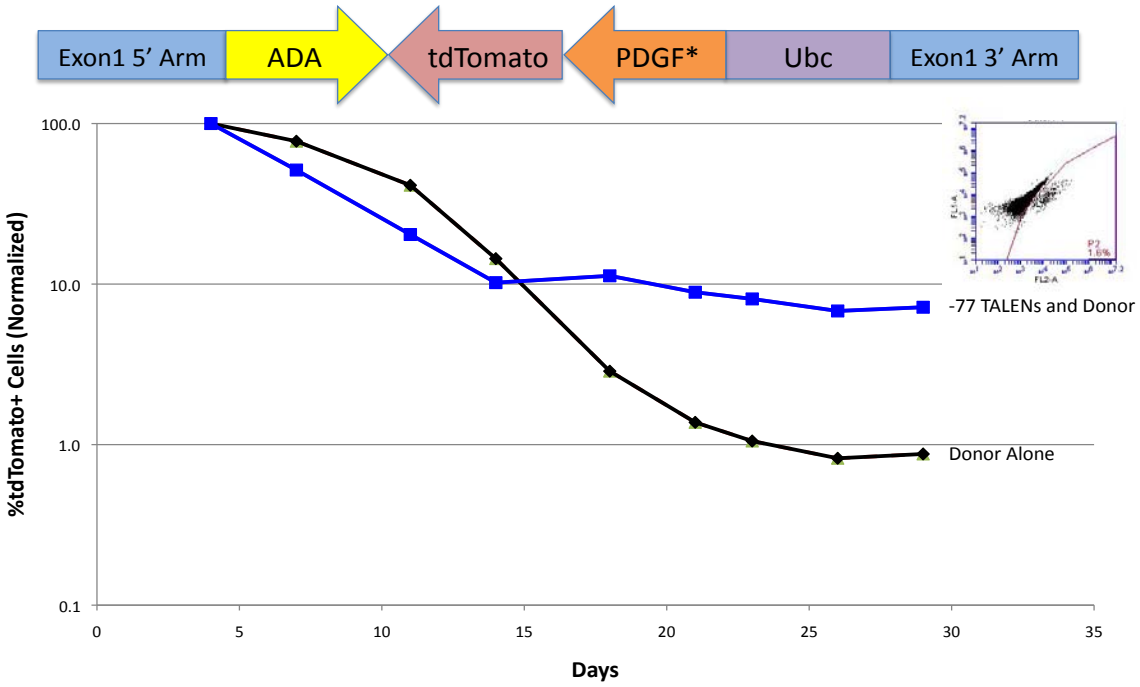

Figure 3.19: Targeted Insertion of PDGF Cassette. Primary fibroblasts were nucleofected with the donor alone or with both the donor and ADA -77 TALENs. %tdTomato cells measured by flow cytometry.

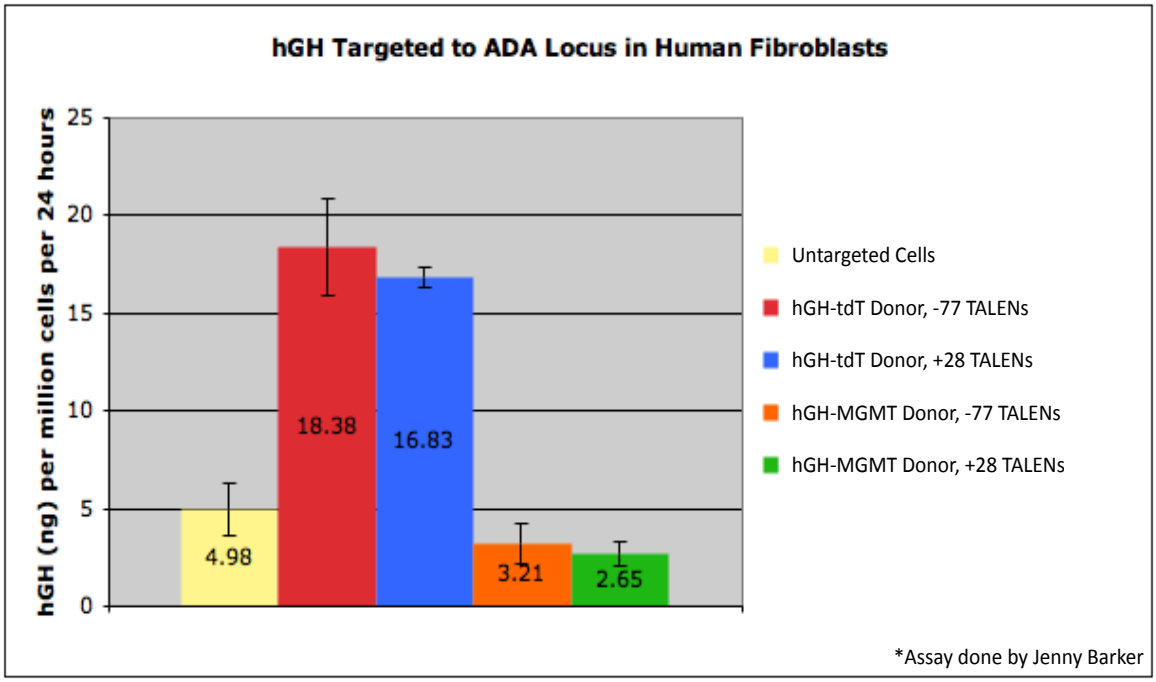

Figure 3.20: ELISA Assay Confirming Increased hGH Levels in Targeted Cells. Primary fibroblasts were nucleofected with indicated donor and corresponding ADA TALENs. After four weeks, samples nucleofected with tdTomato-containing donors were sorted for tdTomato+ cells. At the same time, samples nucleofected with P140K MGMT-containing donors were treated with O6BG and BCNU as described above. Selected cells were then grown for two weeks to increase cell number. ELISA activity done according to manufacturers protocol.

To enrich for targeted cells, I sorted these samples for tdTomato+ cells, increasing the percentage of tdTomato+ cells from 1-2% to over 80% of cells. I screened these cells for targeted gene insertion using genomic PCR and observed PCR amplification of the locus only in those samples that were nucleofected with the ADA TALENs and donor plasmid. Subsequently, I lysed these samples and performed a western blot on the cell lysates, blotting for the presence of the FLAG epitope, as expressed from the tagged ADA cDNA cassette. Only those samples nucleofected with TALENs and donor plasmid displayed the band, which ran at around 44 kDa, the predicted size of the tagged ADA protein. These two assays confirmed that the integration of the transgene cassette was inserted specifically at the ADA locus. The enriched cells targeted with the hGH donor underwent subsequent hGH ELISA

analysis and revealed that the enriched cells expressed more than 3-fold more hGH than the untargeted cells (Figure 3.20).

Primary fibroblasts targeted co-transfected with the ADA TALENs and donors containing P140K MGMT constructs were subjected to treatment with O6BG and BCNU. Subsequent analysis of the targeted fibroblasts demonstrated no enrichment of targeted cells following treatment showing that P140K MGMT selection is not efficient in primary fibroblasts at our current dosage levels.

ADA TALENs stimulate targeted gene insertion of multiple genes into the ADA locus while maintaining expression of the ADA protein. This process was successful using several different strategies, with many more possible. While the strategy of using the ADA locus as a safe harbor may provide additional dividends in the future, the remainder of my efforts focused on achieving efficient gene targeting in human HSPCs as a precursor to gene therapy for ADA deficiency.

Engineered ADA TALENs Target ADA Start Site Directly

My ideal TALENs for the ADA locus would target the ATG start site directly, but this was not possible with the original TALEN design constraints. Due to these constraints, the original ADA TALENs targeted sites 77bps upstream and 28bps downstream of the ADA start site. Subsequent publications on the feasibility of engineering TALENs, most notably from Reyon *et al.* outlining the fast ligation-based automatable solid-phase high-throughput (FLASH) method of assembling TALENs,[89] showed that the 3' T residue was not necessary

for TALEN activity, and that the residue immediately 5' of the TALEN binding site could be either a T or a C. This new information eased some of the constraints limiting TALEN design, allowing me to engineer a third pair of TALENs whose recognition sites were centered at the beginning of the ATG start codon of the ADA gene (referred to as the -1 site hereafter).

Verification of Gene Targeting Efficiency of -1 TALENs

I first needed validate that the -1 TALENs can efficiently target the endogenous ADA locus. The -1 TALENs were nucleofected into K562 cells, and the samples were collected on day 3 in order to analyze using the CelI enzymatic mutation detection assay. The -1 TALENs were showed to stimulate mutagenic NHEJ at a frequency of 15%, which is comparable to that of the -77 TALENs and +28 TALENs. Next, the -1 TALENs were co-transfected with the original Ubc-eGFP donor plasmid and their efficiency to stimulate targeted integration compared with that of the +28 TALENs previously characterized. Again, the -1 TALENs showed a similar efficiency of targeted gene insertion of the Ubc-eGFP cassette to that of +28 TALENs, stimulating integration of the cassette nearly 10-fold over that observed in cells nucleofected with the donor plasmid alone (Figure 3.21).

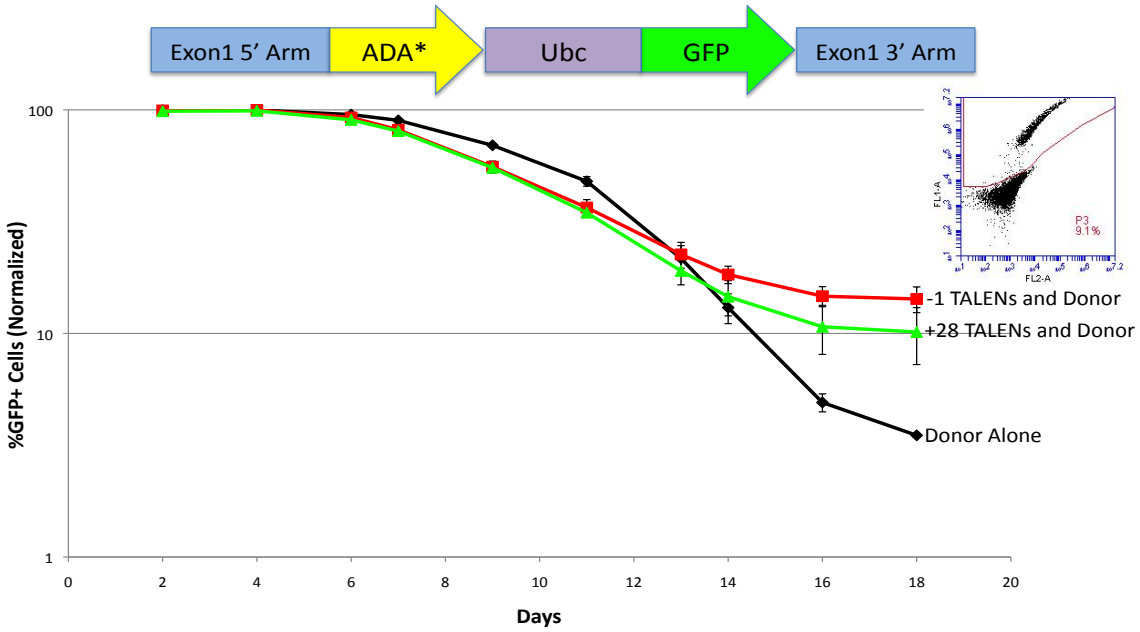

Figure 3.21: -1 TALENs Stimulate Targeted Gene Insertion at Least as Well as +28 TALENs. K562 cells were nucleofected with the Ubc-GFP donor alone, or with both the donor and indicated ADA TALENs. %GFP+ cells were measured by flow cytometry. Black line represents donor alone samples, green line represents donor and +28 TALENs samples, and red line represents donor and -1 TALENs. ADA* is ADA cDNA with wobbles at bases 6, 9, and 12 and also contains 3x-FLAG tag at C-terminus.

Determining the "Best" RVD to Recognize Guanine

Although the TALEN architecture was presented as a platform in which one RVD recognized one DNA base, the truth is a little more complex. Canonical TALEN design encodes HD RVD to recognize C, the NI RVD to recognize A, the NG RVD to recognize T, and the NN RVD to recognize G. Each RVD has differing specificity and affinity profiles for the target base, and there are many alternative RVDs that appear in nature to recognize these same bases. The NN RVD has high affinity for guanine, but its specificity is a little lower as it has also been shown to recognize adenine with a lower degree of affinity. Alternative RVDs have been proposed for guanine recognition, namely the NK and NH RVDs. Both of these RVDs show potential increased specificity for guanine. Prior

publications had shown that when the NK RVD was substituted for the NN RVD, the TALENs exhibited a lower efficiency at creating DSBs at their target loci. A recent publication showed that TALEs (no nuclease domain) containing the NH RVD showed increased specificity for guanine as compared to the NN RVD.[90]

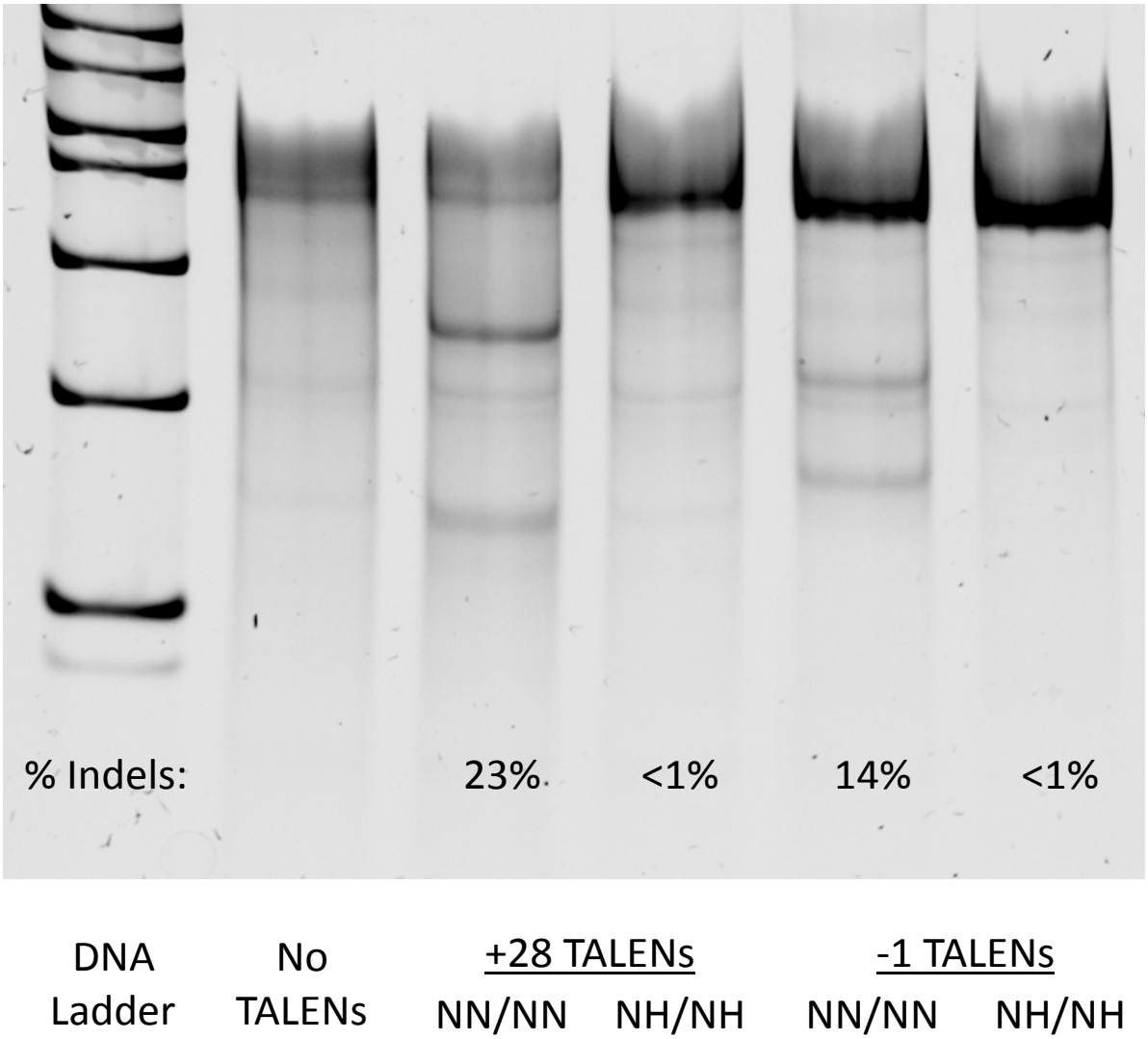

Figure 3.22: NH TALENs Have Greatly Reduced Activity As Compared to NN TALENs. K562 cells were nucleofected wtih indicated ADA TALENs and cells collected for analysis on day 3. TALENs using the NH RVD to recognize guanosine show drastic reduction in activity at their sites when compared to TALENs made using the NN RVD to recognize guanosine, all other things being kept constant.

I determined to test the ability of NH RVDs to create DSBs at the ADA locus by engineering new TALENs that recognize the same target sites as the -1 and +28 TALENs, substituting an NH RVD for every NN RVD. I then nucleofected the -1 TALENs, -1 (NH) TALENs, +28 TALENs, and +28 (NH) TALENs into K562 cells and measured the efficiency with which they stimulate mutagenic NHEJ through use of the CelI assay. The NH TALENs almost completely abolished the frequency of mutagenic NHEJ at these two targets (Figure 3.22). This trend was confirmed at two other TALEN sites as well (ADA exon 2 and β-globin), indicating that this was not a site-specific phenomenon. NH TALENs also were unable to stimulate targeted gene insertion. The NH TALENs were not used in future experiments.

Redesign of Donor Plasmid for Gene Targeting Experiments

For my final experiments, I created a new donor plasmid that would be most suited for further screening processes. This new donor plasmid still contains the original Ubc-eGFP cassette. This donor plasmid also contains the earlier ADA cDNA construct in-frame with the 5' arm of homology, containing both the wobble mutations and the 3x FLAG tag. To distinguish this donor plasmid from all previous constructs, I will refer to the new donor plasmid as the ADA-GFP donor, as it contains both a full ADA cDNA segment and the Ubc-eGFP cassette. Earlier donor constructs included an I-SceI recognition site between the 5' arm of homology and the ADA cDNA, due to previous cloning constraints. The ADA-GFP donor removes this site, aligning our donor construct more closely to the sequence found at

the endogenous ADA locus. Primers were ordered that would amplify both the wild-type and targeted loci in a non-biased manner, with the only differences being those three wobble sites in the targeted allele. PCR products could then be sequenced and classified as wild-type or targeted, and a frequency of targeted gene insertion obtained.

ADA TALENs Efficiently Target Integration of In-Frame ADA cDNA

With the ADA-GFP donor plasmid, I set out to obtain a more quantitative measure of the ability of the -1 TALENs to stimulate targeted gene insertion at the ADA locus. I nucleofected K562 cells and ADA deficient patient-derived fibroblasts with the -1 TALENs and the ADA-GFP donor plasmid. After three weeks for the K562 cells and four weeks for the fibroblasts, the -1 TALENs were able to stimulate integration of the ADA-GFP donor at frequencies similar to what was observed in earlier targeting experiments with previous donors.

I enriched for the targeted population by sorting the co-transfected samples for GFP+ cells. The samples in K562 cells showed an enrichment in the percentage of GFP+ cells from 8% to greater than 98%, while the percentage of GFP+ cells in the fibroblast samples was enriched from 1.4% to ~94%. I collected the cells from three populations for each cell type: donor alone; donor plus TALENs; and donor plus TALENs (sorted). I lysed a subset of cells from each of these samples and performed a western blot on the lysates, blotting for the FLAG peptide, which is an indicator of expression of the tagged ADA cDNA, presumably from the endogenous ADA locus. There was no FLAG signal observed in the samples nucleofected with donor alone, while the donor plus TALENs nucleofected samples

showed a weak FLAG signal, and the enriched donor plus TALENs samples showed a strong FLAG signal (Figure 3.23). This correlates with the prediction that the tagged ADA cDNA will only be expressed when the cell has undergone a successful gene targeting event, and that the entire ADA-GFP cassette is integrated upon initiation of HR incorporating the donor template into the endogenous locus, as tagged ADA expression increases as the percentage of GFP+ cells increases. Actin was used as a loading control.

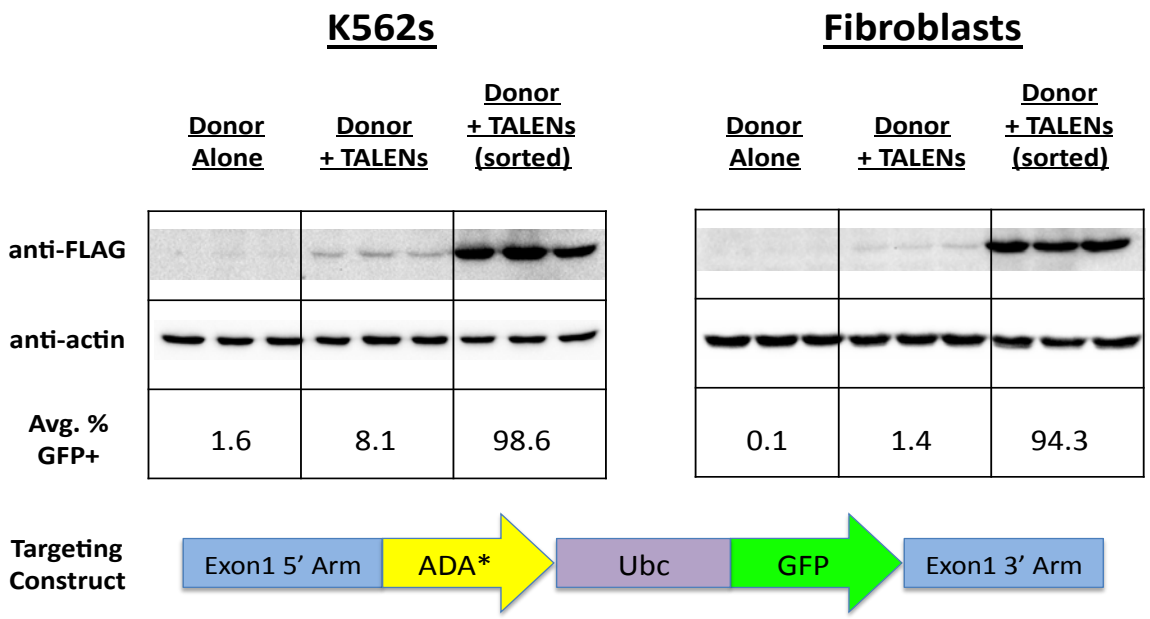

Figure 3.23: Targeted Gene Insertion of In-Frame ADA cDNA by ADA TALENs. Cells were nucleofected with either donor alone or donor and the ADA -1 TALENs. One month after nucleofection, and after stabilization of % GFP+ cells in the population, samples that were transfected with donor and TALENs were sorted to enrich for GFP+ cells. After recovery of the population, cells were lysed and a western blot done against FLAG (tied to ADA cDNA expression) and actin (as a loading control).

Genomic DNA was isolated from the remaining cells for each of these samples. The targeted region of donor plus TALENs (sorted) samples was amplified by PCR. These PCR products were individually sequenced using the method outlined earlier. The sequences were screened for successful gene targeting events. Analysis of 131 colonies from the K562 cells

enriched for GFP+ cells revealed that 49% of all PCR products analyzed demonstrated a successful targeting event. The percentage of targeted PCR products corresponds with the percentage of GFP+ cells, with each cell containing (on average) one targeted allele and one wild-type allele. Analysis of 116 PCR products from the fibroblasts that had been enriched showed a similar number of targeted alleles, although slightly reduced to 38%. The percentage of targeted alleles in fibroblasts again represented an average of nearly one targeted gene insertion event per GFP+ cell.

To ensure that there was no PCR bias in our results, I expanded single-cell clones isolated from the targeted K562 cells that had been enriched for GFP+ cells. I isolated the genomic DNA from each clone and PCR amplified the ADA locus and Sanger sequenced the PCR product from each clonal population. After characterizing the PCR products from 97 clones, Sanger sequencing revealed that 87% of all clones contained at least one targeted allele, and 50% of all alleles had been successfully targeted. 19% of clones analyzed had undergone successful targeting on both alleles. This data corresponds with the sequencing data obtained from analyzing individual PCR products from the sample population.

Gene Targeting in Human CD34+ Cells

Now that I had firmly established that the ADA TALENs efficiently stimulate targeted gene insertion in K562 cells and primary fibroblasts, I set out to establish targeting in human CD34+ cells. CD34 is a cell surface protein that can serve as a marker for hematopoietic stem/progenitor cells (HSPCs), as CD34+ cells are greatly enriched for

HSPCs. For gene therapy purposes, the patient's bone marrow will be collected (being rich in HSPCs) and will undergo targeted gene insertion before being transfused back into the patient. To achieve both long-term reconstitution of the immune system, HSPCs must be among those cells that are successfully targeted and then successfully engraft in the patient's bone marrow cavity. Otherwise, any benefit to the patient would be temporary or incomplete. The ability to successfully target gene insertion to CD34+ cells is thus a critical step towards achieving therapeutic benefit.

We obtained the human CD34+ cells by either purchasing them or isolating the CD34+ cells from fresh cord blood. The CD34+ cells purchased from a company contained a very pure population of CD34+ cells (>99%), but were obtained through pooling the cord blood from several donors, which can occasionally confound subsequent analysis. The vials of CD34+ cells have been frozen and so also suffer a little in viability. These vials are expensive, at nearly $1000 per one million cells. Alternatively, we isolated CD34+ from fresh cord blood obtained from new births in the nearby hospital. We used the Miltenyi beads and columns to purify the population for CD34+ cells to a purity usually between 75-90% CD34+ cells. This process is much more labor intensive and the yield can vary due to the variability of the quality of the cord blood. It is much less expensive (in terms of reagents) and ultimately is a process more similar to that which will take place therapeutically. We obtained similar experimental results from both sources of CD34+ cells; so all of the following experiments were done in CD34+ cells isolated from fresh cord blood.

Delivering DNA to CD34+ cells in an efficient manner has been one of the main hurdles to stimulating gene targeting in these cells. Many labs have tried to use an integrase-

deficient lentivirus (IDLV) in order to deliver both the donor DNA and the TALEN plasmids to CD34+ cells, and have recently begun to experience some success using this method. The IDLV has a greatly reduced capacity to integrate its viral DNA into the human genome and so is greatly favored over wild-type lentiviral vectors, which can integrate near proto-oncogenes and may cause leukemic events. We are still concerned about the potential for these IDLV vectors to cause insertional oncogenesis and so have chosen to pursue alternative methods of DNA delivery. Another graduate student in our lab, Eric Kildebeck, tested a variety of nucleofection protocols until he found one that gave both a high level of transfection efficiency and a low level of cell death subsequent to the nucleofection procedure. These nucleofections utilize the new Amaxa 4D machine, which consistently demonstrates a higher level of cell viability than the older versions of Amaxa machines used for nucleofection, especially in CD34+ cells.

I nucleofected CD34+ cells with the -1 TALENs in order to measure their ability to stimulate gene targeting. Three days after nucleofection, I collected the cells and isolated the genomic DNA in order to use in the CelI enzymatic mutation detection assay. Results from the CelI assay demonstrated that the -1 TALENs could stimulate mutagenic NHEJ at a frequency of about 2% of the alleles in the sample. This was a much lower percentage of mutagenic NHEJ than observed in both K562 cells and primary fibroblasts, which were both around 15%. The lower efficiency of gene targeting is probably mostly due to a reduced nucleofection efficiency. CD34+ cells are also sensitive to the amount of DNA nucleofected in to them, and so we must use lower amounts of DNA in order to maintain sufficient

viability of the nucleofected cells. There may also be some innate cell processes that restrict the mutagenic NHEJ process to a greater extent in CD34+ cells than in other cell types.

I next assessed the ability of the -1 TALENs to stimulate targeted gene insertion in CD34+ cells. I nucleofected the isolated CD34+ cells with the -1 TALENs and the ADA-GFP donor plasmid. After seven days, the samples nucleofected with the donor alone show no GFP+ cells. However, those samples nucleofected with TALENs and donor plasmid showed a percentage of GFP+ cells of up to 0.5% (Figure 3.24). Despite the lower frequency of mutagenic NHEJ in these cells, the -1 TALENs were still able to show a stimulation of integration in the CD34+ cells above the background rate (which is 0% by GFP fluorescence). To ensure that these GFP+ cells were due to targeted integration of the ADA-GFP donor template at the ADA locus, I isolated the genomic DNA from these cells for further analysis. I amplified the target region by PCR and sequenced the PCR product. Analysis of the sequencing results showed that I had achieved targeted gene insertion in these CD34+ cells.

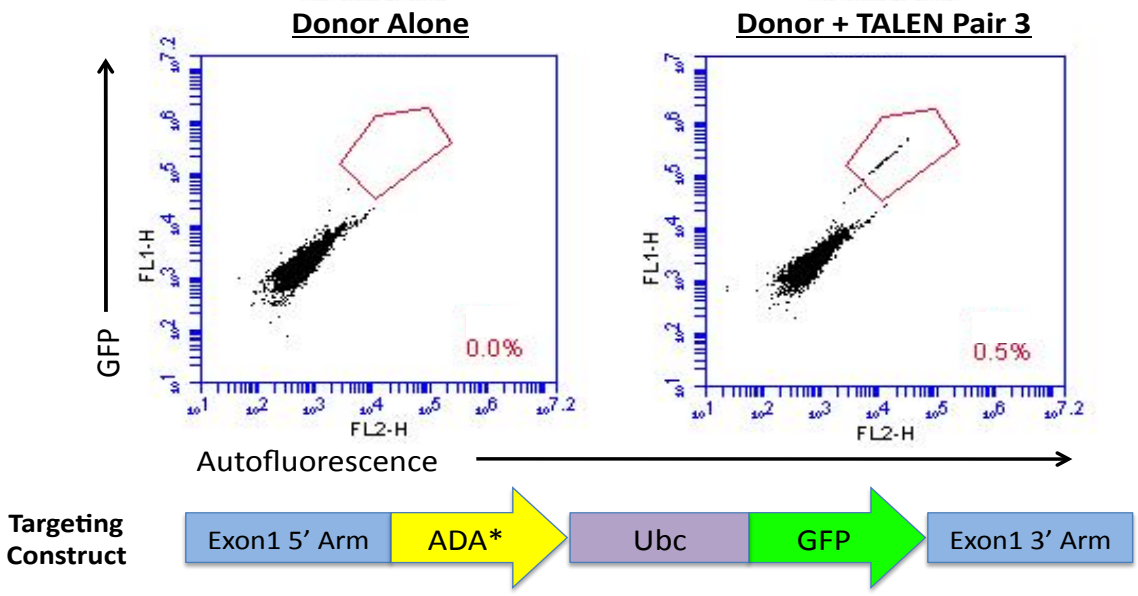

Figure 3.24: Targeted Gene Insertion in Human CD34+ Cells. Human CD34+ cells isolated from fresh cord blood were nucleofected with donor alone or with donor and the ADA -1 TALENs. %GFP+ cells were analyzed by flow cytometry and stayed constant from day 7 on. The dots inside the polygon are representative of targeted cells.

CHAPTER FOUR
Conclusions and Recommendations

TARGETED GENE INSERTION TO THE ADA LOCUS IS A REALITY

Nuclease-Mediated Targeted Gene Insertion

Science and medicine are continually striving to improve the breadth and the efficiency with which they treat the many different human disorders. While current treatments for ADA deficiency have prolonged the lives of many of the patients, most patients are not able to undergo the procedures that would allow for a cure of their disease; namely, HLA-matched HSCT or gene therapy using a viral vector to insert a functional copy of ADA into their HSPCs. Additionally, the risk of insertional oncogenesis remains for ADA-SCID patients treated through viral-mediated gene therapy. Site-specific insertion of a functional ADA cDNA construct expressed under the endogenous promoter into patient CD34+ HSPCs would allow for a functional cure to be available for all patients while eliminating the risk of viral-mediated insertional oncogenesis. This work demonstrates the design and validation of engineered TALENs that efficiently stimulate targeted gene insertion into K562 cells, ADA deficient patient-derived fibroblasts, and human CD34+ cells. This work also shows the potential of these TALENs to stimulate the integration of transgenes that could be used in the treatment of a variety of monogenic recessive disorders. Nuclease-mediated targeted gene insertion may one day become the treatment of choice for a

variety of disorders, improving and extending the lives of many, many patients who currently have little to no hope of a cure.

ADA TALENs Stimulate Targeted Gene Insertion of ADA-GFP Donor

Using the Golden Gate method of assembly, TALENs were engineered that successfully target the endogenous ADA locus at three different loci surrounding the ATG start codon of exon 1. The efficiency with which these TALENs can stimulate gene targeting at their targets was measured through the frequency of mutagenic NHEJ in cells nucleofected with TALENs, as measured by the CelI enzymatic mutation detection assay. The frequency of mutagenic NHEJ observed is proportional to the ability of the TALENs to create DSBs at their target sites- the more breaks there are, the more total breaks will be repaired in a mutagenic manner. The ADA TALENs were able to simulate mutagenic NHEJ at a rate of 15-20% in K562 cells and primary fibroblasts, and at a rate of about 2% in human CD34+ cells. The CelI assay signals were verified by Sanger sequencing the individual PCR amplicons and confirming that the mutations created in the PCR products were at the target site for each pair of TALENs. The vast majority of mutations generated were deletion events, ranging from as small as a 1bp deletion to as large as several hundred base pairs. It is possible that even larger deletions were created by the TALENs but remained undetected if the deletion events spanned one, or both, of the primer sites required for the creation of the PCR amplicon.

The ADA TALENs were able to stimulate targeted gene insertion into K562s, primary fibroblasts, and human CD34+ cells at the ADA locus. The integrated construct

included modified ADA cDNA followed by a Ubc-eGFP expression cassette. The ADA cDNA was modified in order to enable distinction of the targeted ADA locus from the endogenous ADA locus on both the DNA and protein level. Co-transfection of the ADA TALENs and the ADA-GFP donor plasmid showed increased frequency of integration in these samples as compared to those samples nucleofected with donor alone in all three cell types, as measured by the percentage of GFP+ cells in each sample. DNA analysis of samples from all three cell types revealed successful targeted gene insertion events. After sorting the K562 cells and primary fibroblasts for GFP+ cells, the frequency of targeted gene insertion events increased in conjunction with the increase in %GFP+ cells, confirming that the entire donor construct was inserted at the target site. Furthermore, targeted samples from both K562 cells and primary fibroblasts showed expression of the tagged ADA cDNA that was absent in cells nucleofected with the donor plasmid alone. Expression of the ADA cDNA construct was increased in the samples that had been sorted for GFP+ cells, providing further evidence that the entire donor construct is integrated into the ADA locus as one cassette.

My work puts forth evidence that nuclease-mediated targeted gene insertion could one day used for gene therapy in patients. ADA cDNA targeted to and expressed from the endogenous ADA locus in both K562 cells and primary fibroblasts, while targeted gene insertion was also confirmed in human CD34+ cells. For therapeutic uses, the Ubc-eGFP cassette could be replaced with a cassette that allows for selection of targeted cells, such as a construct containing the P140K MGMT gene. This would allow for an enrichment of targeted cells before being transfused back into the patient, or even post-transfusion

enrichment. While the work I did here was important in laying the foundation for gene therapy using ADA TALENs, there is still much to be done in before use in humans can be considered.

Future Directions For Nuclease-Mediated Gene Therapy for ADA Deficiency

The next experiments to be done on this project would be to first optimize the gene targeting efficiency of the ADA TALENs. While the efficiency is high in K562 cells, it is still very low in CD34+ cells, and that needs to improve in order to be therapeutically useful. I here provide a list of several steps that can be taken in order to increase that efficiency in a relatively straightforward manner. The ideal nucleases for gene therapy are both highly active and highly specific for their targets, so the "off-target" activity would need to be measured and minimized. The final step is to show that ADA TALEN-mediated targeted gene insertion can restore the immune function in immuno-compromised mice transfused with targeted human cells.

Codon Optimization of Expression Vectors

A major determining factor in the efficiency of gene targeting that can be stimulated by TALENs is the amount of TALEN protein that is present. We have shown through numerous experiments that decreasing the amount of TALEN nucleofected into cells results in a related decrease in gene targeting events observed. While the inverse is also true, that you get more gene targeting events when nucleofecting in more TALEN plasmid, it is only

true up to a limit. As you increase the amount of DNA nucleofected into the cells, you also increase the cytotoxicity in the cells and the reduced viability of the cells is easily observed. This cytotoxicity is partially independent of the expression of TALENs, as the nucleofection of high amounts of carrier DNA also reduces the viability of the cells significantly. This toxicity results from the activation of the interferon response of the cells as DNA plasmids have many bacterial components that can be recognized as foreign elements. Therefore, if we were able to increase the expression of the TALENs without increasing the amount of DNA, we should be able to observe an increased frequency of gene targeting in our samples. Codon optimization is one way to achieve this increased expression. TALEs originally come from plant pathogens, and so are not optimized for expression in human cells. Our lab compared a codon optimized version of a highly active pair of TALENs to a pair targeting the same site made using the Golden Gate method of assembly. The Golden Gate TALENs were only about 80% as efficient as the codon optimized TALENs at stimulating mutagenic NHEJ in K562 cells. Another potentially critical candidate for codon optimization is the ADA cDNA itself. Even though we are using human ADA, not all of the codons are the most efficient choices for the desired amino acid. Codon optimization should increase the level of expression from each targeted ADA cDNA and thus increase the overall therapeutic effectiveness. The success of several pre-clinical trials for SCID gene therapy have depended on the use of the codon optimized version of the transgene, as outlined in the introduction of this paper.

mRNA TALEN Expression

Due to the toxicity of DNA plasmid nucleofection into cells, some labs have chosen to eliminate the transfection of DNA entirely when delivering TALENs into cells. For ZFNs, some labs found have found success through delivery of the actual ZFN proteins to the cells, but this has not yet proven effective for TALENs, due to the difference in DNA binding domains. The alternative many labs are now taking is to nucleofect the ZFN and TALEN mRNA into the cells. Expression from mRNA TALEN constructs has exhibited robust gene targeting capabilities while enabling the transfection of orders of magnitude more nucleic acid into the cells without negatively impacting cell viability. Preliminary experiments using mRNA for ADA TALENs in human CD34+ cells has shown the ability to stimulate gene targeting, but the protocol must be further optimized as the overall efficiency was low. In addition to the alleviation of the cytotoxicity caused by DNA plasmids, using mRNAs to express the TALENs also allows for a tighter window of TALEN expression, possibly improving the specificity profile of the TALENs. Prolonged expression of the TALENs from the DNA plasmids is thought to increase the amount of "off-target" breaks occurring in the cells, while only minimally increasing the percentage of targeted cells.[91]

Alternative Methods of Delivery of Donor DNA

Another way to increase the frequency of targeted gene insertion is to increase the amount of donor template present in the cells. This increases the frequency with which the donor is used as the template for repair, and results in more cells that have undergone successful targeting. Unfortunately, increasing the amount of donor DNA added to the cells

is not advisable due to the aforementioned cytotoxicity caused by DNA plasmids. For targeted gene insertion through HR, there's no way around it: you need the donor construct in the cells, and in DNA form, in order to be used as a template for DSB repair. One strategy for increased targeting in CD34+ cells is to minimize the size of the donor DNA vector through the use of the minicircle DNA vector. DNA plasmids contain many bacterial elements that are not necessary, and indeed are likely detrimental, for targeted gene insertion in human cells. The bacterial elements are needed in order to produce the plasmids, as they code both for an origin of replication and an antibiotic resistance gene. Minicircle DNA vectors basically excise those fragments and leave a circularized piece of DNA containing almost exclusively the arms of homology and insertion cassette.[92] This reduction in size of the donor DNA allows nucleofection of a higher molar level of donor template while not increasing the overall level of DNA added. For example, the ADA-GFP donor plasmid is a little over 8kb. If transferred to a minicircle vector, after excision of the backbone, the ADA-GFP minicircle DNA vector would be 6kb. If 5 μg of DNA is nucleofected per sample, each sample will now contain 33% more copies of the donor template. Also, through eliminating most of the bacterial elements of the plasmids, it is probable that more DNA can be nucleofected into the cells without increased cytotoxicity, as it is the bacterial elements that activate the interferon response.

Another alternative to explore is that of donor delivery through the use of viral vectors. Successful nuclease-mediated gene targeting has been shown using adeno-associated virus and integrase-deficient lentivirus (IDLV) to deliver the donor template. IDLV is of particular interest as it has been shown to be able to infect human HSPCs, which

is the cell type necessary to use these methods therapeutically. It is also more similar to that of the current gene therapy trials, and may pose an easier path to clinical acceptance than nucleofection.

Measuring and Minimizing "Off-Target" Breaks

The creation of "off-target" breaks in the cell, and the resultant mutations in the genome is a cause of serious concern in the nuclease community. It is not yet known if these mutations will prove harmful in humans, but many efforts are being put forth to both identify and to minimize the amount of "off-target" breaks that TALENs make in order to improve their safety profile. A collaborator at Georgia Tech has created a computer program that predicts the most-likely "off-target" sites that will be recognized by any pair of TALENs and designs primers to PCR amplify the region surrounding each of those sites. In conjunction with work done in our lab, a high-throughput method of analyzing the frequency of mutagenesis at each of these sites is being developed. The specificity of the ADA TALENs could be indirectly measured through the use of this system. Additionally, a necessary step in order to reduce "off-target" breaks is the use of the obligate heterodimer version of the *FokI* nuclease domain. The use of the obligate heterodimer reduces the number of potential "off-target" sites to 50% of that of the wild-type *FokI* domains, while only slightly lowering the activity of the TALENs. Despite the reduction in activity, these obligate heterodimers are advantageous due to their increased target/off-target cutting ratio.

Demonstrating Immune Reconstitution in Mice

The final step before bringing any new therapy to clinical trials is to show efficacy and safety in pre-clinical trials in an animal model, and developing ADA TALENs for potential therapeutic use is no different. Human CD34+ cells that have successfully undergone targeted gene insertion using the ADA TALENs will be injected into immunodeficient NSG mice. After a period of time, these mice will be sacrificed and their immune cells analyzed. It will be determined if the CD34+ cells that had undergone targeted gene insertion were able to successfully engraft and contribute to all branches of the mouse immune system. If it is determined that the targeted cells were engrafted and contributed to all immune cell types in the mouse, ADA-SCID patient cells could then be sought out in order to repeat this experiment. As ADA-SCID patients are rare, patient CD34+ cells should not be used for experiments until all branches of the experiment have been shown to function properly with wild-type human CD34+ cells. Experiments with ADA-SCID CD34+ cells will show if ADA TALEN-mediated targeted gene insertion can functionally correct the ADA deficiency and reconstitute a fully functioning immune system in mice.

The Future is Bright for Gene Therapy

There is truly a great amount of reason to hope for a successful cure of many recessive monogenic diseases to be developed through the use of gene therapy in the relatively near future. While competition between different research groups can sometimes

get very intense, in the end, it doesn't matter who "wins" as long as someone "wins". It doesn't matter if it is lentiviral vectors or TALENs or CRISPRs or some other platform that turns out to be the best treatment for monogenic diseases. If a safe, reliable, curative treatment is found, thousands to millions of patients will benefit through the lengthening of life and improving the quality of life. The very fact that there are multiple approaches to solving the same problem increases the chances that an effective treatment will be developed in the near future. It is an exciting time to be involved in the gene therapy field, and I only hope to have contributed one little piece to the puzzle that is well on its way to being solved.

APPENDIX A
ADA-GFP Donor Template

5’-

TAACTATAACGGTCCTAAGGTAGCGATTAATTAACCCACTCGCCCGTGGAGG

GGCATTTGAGGGTGTTTCCAATGTTCCTGTTATTCGGAATAGCGCTGGTGTGA

ACATTCTGCACAGGTCTCTGGCTGCGCCTGGGCGGGTTTCTTAAAGGTGAATG

CCCAGGAGGGGACTGTCTGTGTTCTCCCTCCCTCCGAGCTCCAGCCTTCCTCG

CCTCCTTTCACTCCCAGCTCCCTGGAGTCTCTCACGTAGAATGTCCTCTCCAC

CCCCACCCACCCCTGATGAACTCCTGCAGGTTCTGCAGGCCACGGCTGGCCC

CCCTCGAAAGTTCCTTAACTATACAATTATGGTGTGTGTTTCTGCGACGAGCG

TCCGTCTATCCGGTGGAAGGCACGCCGCTCGAGGCTTGCGATGCTCCCGGGG

TCCCCGCTTCTAGCTTGGGCCTGGCGCACAGCAGCGCCCAGACTGCAGGGGG

ACGCTTGAAAGTTGCTGGAGGAGCCGGGGGGAAGGCAGCGCCCAGCGAGGC

GGCTGGAGCGCGCGCCCACAGGTGGGTCCGGTCGGGCGCCGCGGGGCCGTA

GTTTTCGGGTCGGCGGGCGAGGACGCCGGGTCCAGAATTCCAGGAAATGCGC

GATCCAGGCCGGCGGGCGGGGCGGGGGCTCCGGCGAGAGGGCGGGCCCCGG

GAACGGCGGCGGGCGGGGCGGGAAGCGGGGCCCGGCCCGTTAAGAAGAGCG

TGGCCGGCCGCGGCCACCGCTGGCCCCAGGGAAAGCCGAGCGGCCACCGAG

CCGGCAGAGACCCACCGAGCGGCGGCGGAGGGAGCAGCGCCGGGGCGCACG

AGGGCACCATGGCTCAAACCCCCGCCTTCGACAAGCCCAAAGTAGAACTGCA

TGTCCACCTAGACGGATCCATCAAGCCTGAAACCATCTTATACTATGGCAGG

AGGAGAGGGATCGCCCTCCCAGCTAACACAGCAGAGGGGCTGCTGAACGTC

ATTGGCATGGACAAGCCGCTCACCCTTCCAGACTTCCTGGCCAAGTTTGACTA

CTACATGCCTGCTATCGCGGGCTGCCGGGAGGCTATCAAAAGGATCGCCTAT

GAGTTTGTAGAGATGAAGGCCAAAGAGGGCGTGGTGTATGTGGAGGTGCGGT

ACAGTCCGCACCTGCTGGCCAACTCCAAAGTGGAGCCAATCCCCTGGAACCA

GGCTGAAGGGGACCTCACCCCAGACGAGGTGGTGGCCCTAGTGGGCCAGGG

CCTGCAGGAGGGGGAGCGAGACTTCGGGGTCAAGGCCCGGTCCATCCTGTGC

TGCATGCGCCACCAGCCCAACTGGTCCCCCAAGGTGGTGGAGCTGTGTAAGA

AGTACCAGCAGCAGACCGTGGTGGCCATTGACCTGGCTGGAGATGAGACCAT

CCCAGGAAGCAGCCTCTTGCCTGGACATGTCCAGGCCTACCAGGAGGCTGTG

AAGAGCGGCATTCACCGTACTGTCCACGCCGGGGAGGTGGGCTCGGCCGAAG

TAGTAAAAGAGGCTGTGGACATACTCAAGACAGAGCGGCTGGGACACGGCT

ACCACACCCTGGAAGACCAGGCCCTTTATAACAGGCTGCGGCAGGAAAACAT

GCACTTCGAGATCTGCCCCTGGTCCAGCTACCTCACTGGTGCCTGGAAGCCG

GACACGGAGCATGCAGTCATTCGGCTCAAAAATGACCAGGCTAACTACTCGC

TCAACACAGATGACCCGCTCATCTTCAAGTCCACCCTGGACACTGATTACCA

GATGACCAAACGGGACATGGGCTTTACTGAAGAGGAGTTTAAAAGGCTGAAC

ATCAATGCGGCCAAATCTAGTTTCCTCCCAGAAGATGAAAAGAGGGAGCTTC

TCGACCTGCTCTATAAAGCCTATGGGATGCCACCTTCAGCCTCTGCAGGGCA

GAACCTCGACTACAAAGACGATGACGACAAAGACTACAAAGACGATGACGA

CAAAGACTACAAAGACGATGACGACAAATGAGGCGCGCCCCGCTGATCAGC

CTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCC

TTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGG

AAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTG

GGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGG

GATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACGTTTAAACATAACT

TCGTATAGCATACATTATACGAAGTTATCGGCCGGGCCTCCGCGCCGGGTTTT

GGCGCCTCCCGCGGGCGCCCCCCTCCTCACGGCGAGCGCTGCCACGTCAGAC

GAAGGGCGCAGCGAGCGTCCTGATCCTTCCGCCCGGACGCTCAGGACAGCGG

CCCGCTGCTCATAAGACTCGGCCTTAGAACCCCAGTATCAGCAGAAGGACAT

TTTAGGACGGGACTTGGGTGACTCTAGGGCACTGGTTTTCTTTCCAGAGAGCG

GAACAGGCGAGGAAAAGTAGTCCCTTCTCGGCGATTCTGCGGAGGGATCTCC

GTGGGGCGGTGAACGCCGATGATTATATAAGGACGCGCCGGGTGTGGCACAG

CTAGTTCCGTCGCAGCCGGGATTTGGGTCGCGGTTCTTGTTTGTGGATCGCTG

TGATCGTCACTTGGTGAGTAGCGGGCTGCTGGGCTGGCCGGGGCTTTCGTGG

CCGCCGGGCCGCTCGGTGGGACGGAAGCGTGTGGAGAGACCGCCAAGGGCT

GTAGTCTGGGTCCGCGAGCAAGGTTGCCCTGAACTGGGGGTTGGGGGGAGCG

CAGCAAAATGGCGGCTGTTCCCGAGTCTTGAATGGAAGACGCTTGTGAGGCG

GGCTGTGAGGTCGTTGAAACAAGGTGGGGGGCATGGTGGGCGGCAAGAACC

CAAGGTCTTGAGGCCTTCGCTAATGCGGGAAAGCTCTTATTCGGGTGAGATG

GGCTGGGGCACCATCTGGGGACCCTGACGTGAAGTTTGTCACTGACTGGAGA

ACTCGGTTTGTCGTCTGTTGCGGGGGCGGCAGTTATGGCGGTGCCGTTGGGCA

GTGCACCCGTACCTTTGGGAGCGCGCGCCCTCGTCGTGTCGTGACGTCACCCG

TTCTGTTGGCTTATAATGCAGGGTGGGGCCACCTGCCGGTAGGTGTGCGGTA

GGCTTTTCTCCGTCGCAGGACGCAGGGTTCGGGCCTAGGGTAGGCTCTCCTGA

ATCGACAGGCGCCGGACCTCTGGTGAGGGGAGGGATAAGTGAGGCGTCAGTT

TCTTTGGTCGGTTTTATGTACCTATCTTCTTAAGTAGCTGAAGCTCCGGTTTTG

AACTATGCGCTCGGGGTTGGCGAGTGTGTTTTGTGAAGTTTTTTAGGCACCTT

TTGAAATGTAATCATTTGGGTCAATATGTAATTTTCAGTGTTAGACTAGTAAA

TTGTCCGCTAAATTCTGGCCGTTTTTGGCTTTTTTGTTAGACGAAGCTTGGTAC

CGAGCTCGGATCCATCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACC

GGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGT

TCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCT

GAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTG

ACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGA

AGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCG

CACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAG

TTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCA

AGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCC

ACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTT

CAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTAC

CAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACT

ACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATC

ACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGA

CGAGCTGTACAAGTAACTCGAGGCGGCCGCCCGCTGATCAGCCTCGACTGTG

CCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACC

CTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATC

GCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGAC

AGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTG

GGCTCTATGGCTTCTGAATAACTTCGTATAGCATACATTATACGAAGTTATCC

TGCAGGGCCCAGACGCCCGCCTTCGACAAGCCCAAAGTGAGCGCGCGCGGG

GGCTCCGGGGACGGGGGTCCGGCGCCTGGGCGGCCCGAGGGGCTTAGCGGG

GCCCAGCCCGGGGCGTCCAAACCCTGGGAACGAACGGGGGCTCCTGCAGGC

GAGTTCTTCCTTCGGCTTAGGCCGTGGCTTGCTTGCGGGCTAATCAGGGACAA

TGGGGCAGAGAAGGTCCAGAACCCGGAGGCCTCCAGAGTCTGCTTCTGCCCC

TGACTTGACCCCTCTGGGTCTCAGTTTCGCTGTCTGTCAAGTGGGCATCCTAG

CACCGCTGAGCGCTGTGTGGGCCTGGGCAGGGACTTGAGGTCTCTGAAGCTC

AGCTGTATGATCAGGCCCGATGTCTACGCCGGATAGGCACCTAGTGCTGTGC

CCGGCGCCTACTGAGTGCTCAGTGAATGGAAGCAGCTTTGTACGCCAGCGTT

ATGGTGGTGAGCGCCAAGGAGCTCAGGTTTGTGGATGCGCCCCGGGGAAGAA

CCGTGAGCCCTGCCAGAAAGGGGAGGGAGGGGAGCAGAGCACCCCCCTTCC

CCCGCGCGGGAAGAACAGGAGCTAGGTAGGCCCTGGGTTTGGGGCCCTAGC

AGGGTTCACTCGAGGCCAAGCCATGGCCACTGGCCCCAGGGGAGAATCCCCT

TGTTTCTCCGCCCACCAGCTGTGGCGTCTTGGGACTGTTGGGATCAGGGAGGG

TCTGGACCCCCTTGGCCTGTCTCAGAGTCCGAGAGGAGGGGCCCAGGAGTCT

GCCAAGCAGGGTGAGTCAGCCAGTAGGGTGTGAGAGTGGTTGGGGAAGGAG

TCAGCTGCAGTCAGCCTCTGCCATTTCATTACCTCTTTCTCCGCACCCGACAT

AGAT-3’

APPENDIX B
ADA ZFN Recognition Helices

ADA5A ZFNs

Target Site: GACCTCACCCCAGACGAGGTGGTG

Left:	GGT	GAG	GTC
L1	DPGALVR	RHDNLAR	HRHGLGR
L2	DPGALVR	RHDNLDR	QNHGLLR
L3	DPGALVR	RSDNLLR	HRHGLGR
L4	DPGALVR	RGDNLTR	HRHGLGR
L5	DPGALVR	RHDNLRR	VMHHLAR
L6	DPGALVR	RHDNLAR	LTQSLQR
L7	DPGALVR	RHDNLAR	VMHHLAR
L8	DPGALVR	RRDNLGR	VMHHLAR

Right:	GAG	GTG	GTG
R1	RQDNLQR	RLEVLAN	RNTVLMR
R2	RRDHLLR	RLEVLAN	RNTTLLR
R3	RQDNLQR	RREVLMN	VNRSLTR
R4	RIDNLGR	RRNILQN	RRHILHN
R5	RQDNLQR	RNAILVN	RNTTLLR
R6	QASELRL	RNAILVN	VNRSLTR
R7	RQDNLQR	RREVLMN	RRHILHN
R8	RRDNLLR	RRPVLVN	VNRSLTR
R9	QASELRL	RNAILVN	RNTTLLR

ADA5B ZFNs

Target Site: GGGGACCTCACCCCA**GACGAGGTG**

Left:	GAG	GTC	CCC
L1	RRDNLNR	DPSVLKR	SKKQLAE
L2	GAANLLE	DSAVLTR	SKKQLAE
L3	REDSLTR	DREVLRR	SNKHLAE
L4	REDNLPR	DSAVLTR	RRDELNV
L5	RRDHLRR	DREVLRR	SKNHLAE
L6	RSDGLRR	DREVLRR	SKKHLAE

Right:	GAC	GAG	GTG
R1	DPSNLRR	RKDNLPR	RRHILLN

R2	DPSNLRR	RRDNLPR	KSHLLNR
R3	DQGNLIR	RRDNLPR	RSTTLQR
R4	DPSNLRR	RKDNLPR	RLTTLHV
R5	DPSNLRR	RKDNLPR	RQTVLAR
R6	DPSNLRR	RKDNLPR	RAESLRI
R7	DPSNLRR	RRDNLPR	RSTSLHR

ADA6
ZFNs
Target Site: **ACCAGCAGC**AGACC**GTGGTAGCC**

Left:	**GCT**	**GCT**	**GGT**
L1	LGHTLNR	LSQTLNR	LRHHLEA
L2	LRHDLQR	LSQTLNR	MKHHLDT
L3	LRASLRR	LSQTLNR	LRHHLEA
L4	LKHDLRR	LSQTLKR	TKQKLDV
L5	LKHDLRR	LSQTLKR	TKQKLGI
L6	LRASLRR	LSQTLKR	TKQKLGI
L7	LKHDLRR	LSQTLKR	LRHHLEA
L8	VGGSLNR	LSQTLRR	MKHHLDA

Right:	**GTG**	**GTA**	**GCC**
R1	RGDSLTR	QGTTLRN	DSSTLAR
R2	RADSLRR	QSSSLVR	DGRGLRR
R3	RGDSLTR	QRSSLVR	AKRDLDR
R4	RSDGLAR	QSASLRR	DRRDLTV
R5	RRDSLNR	QRSSLVR	DRRTLDR
R6	RADSLTR	QSASLRR	DARGLLR
R7	RMDALRR	QSASLRR	DRRDLDR
R8	RGDSLTR	QNTVLRN	DGGTLHR
R9	RGDSLRR	QSSSLVR	DRRGLVR

BIBLIOGRAPHY

1. Christianson, A., C.P. Howson, and B. Modell, *Global Report on Birth Defects: The Hidden Toll of Dying and Disabled Children.* March of Dimes, 2006.

2. Gaspar, H.B., et al., *How I treat ADA deficiency.* Blood, 2009. **114**(17): p. 3524-32.

3. Ferrua, F., I. Brigida, and A. Aiuti, *Update on gene therapy for adenosine deaminase-deficient severe combined immunodeficiency.* Current opinion in allergy and clinical immunology, 2010. **10**(6): p. 551-6.

4. Montiel-Equihua, C.A., A.J. Thrasher, and H.B. Gaspar, *Gene therapy for severe combined immunodeficiency due to adenosine deaminase deficiency.* Current gene therapy, 2012. **12**(1): p. 57-65.

5. Booth, C., H.B. Gaspar, and A.J. Thrasher, *Gene therapy for primary immunodeficiency.* Current opinion in pediatrics, 2011. **23**(6): p. 659-66.

6. Blaese, R.M., et al., *T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years.* Science, 1995. **270**(5235): p. 475-80.

7. Hacein-Bey-Abina, S., et al., *Efficacy of gene therapy for X-linked severe combined immunodeficiency.* The New England journal of medicine, 2010. **363**(4): p. 355-64.

8. Aiuti, A., et al., *Gene therapy for immunodeficiency due to adenosine deaminase deficiency.* The New England journal of medicine, 2009. **360**(5): p. 447-58.

9. Gaspar, H.B., et al., *Hematopoietic stem cell gene therapy for adenosine deaminase-deficient severe combined immunodeficiency leads to long-term immunological recovery and metabolic correction.* Science translational medicine, 2011. **3**(97): p. 97ra80.

10. Gaspar, H.B., et al., *Long-term persistence of a polyclonal T cell repertoire after gene therapy for X-linked severe combined immunodeficiency.* Science translational medicine, 2011. **3**(97): p. 97ra79.

11. Boztug, K., et al., *Stem-cell gene therapy for the Wiskott-Aldrich syndrome.* The New England journal of medicine, 2010. **363**(20): p. 1918-27.

12. Kang, H.J., et al., *Retroviral gene therapy for X-linked chronic granulomatous disease: results from phase I/II trial.* Molecular therapy : the journal of the American Society of Gene Therapy, 2011. **19**(11): p. 2092-101.

13. Rivat, C., et al., *Gene therapy for primary immunodeficiencies (PIDs).* Human gene therapy, 2012.

14. Naldini, L., *Ex vivo gene transfer and correction for cell-based therapies.* Nature reviews. Genetics, 2011. **12**(5): p. 301-15.

15. Cattoglio, C., et al., *High-definition mapping of retroviral integration sites identifies active regulatory elements in human multipotent hematopoietic progenitors.* Blood, 2010. **116**(25): p. 5507-17.

16. Cooray, S., S.J. Howe, and A.J. Thrasher, *Retrovirus and lentivirus vector design and methods of cell conditioning.* Methods in enzymology, 2012. **507**: p. 29-57.

17. Pike-Overzet, K., et al., *Correction of murine Rag1 deficiency by self-inactivating lentiviral vector-mediated gene transfer.* Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K, 2011. **25**(9): p. 1471-83.

18. Santilli, G., et al., *Biochemical correction of X-CGD by a novel chimeric promoter regulating high levels of transgene expression in myeloid cells.* Molecular therapy : the journal of the American Society of Gene Therapy, 2011. **19**(1): p. 122-32.

19. Scaramuzza, S., et al., *Preclinical Safety and Efficacy of Human CD34(+) Cells Transduced With Lentiviral Vector for the Treatment of Wiskott-Aldrich Syndrome.* Molecular therapy : the journal of the American Society of Gene Therapy, 2012.

20. van Til, N.P., et al., *Correction of Murine Rag2 Severe Combined Immunodeficiency by Lentiviral Gene Therapy Using a Codon-optimized RAG2 Therapeutic Transgene.* Molecular therapy : the journal of the American Society of Gene Therapy, 2012.

21. Huston, M.W., et al., *Correction of murine SCID-X1 by lentiviral gene therapy using a codon-optimized IL2RG gene and minimal pretransplant conditioning.* Molecular therapy : the journal of the American Society of Gene Therapy, 2011. **19**(10): p. 1867-77.

22. Astrakhan, A., et al., *Ubiquitous high-level gene expression in hematopoietic lineages provides effective lentiviral gene therapy of murine Wiskott-Aldrich syndrome.* Blood, 2012. **119**(19): p. 4395-407.

23. Avedillo Diez, I., et al., *Development of novel efficient SIN vectors with improved safety features for Wiskott-Aldrich syndrome stem cell based gene therapy.* Molecular pharmaceutics, 2011. **8**(5): p. 1525-37.

24. Zhang, F., et al., *A ubiquitous chromatin opening element (UCOE) confers resistance to DNA methylation-mediated silencing of lentiviral vectors.* Molecular therapy : the journal of the American Society of Gene Therapy, 2010. **18**(9): p. 1640-9.

25. Ng, Y.Y., et al., *Correction of B-cell development in Btk-deficient mice using lentiviral vectors with codon-optimized human BTK.* Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K, 2010. **24**(9): p. 1617-30.

26. Hunter, M.J., et al., *Gene therapy of canine leukocyte adhesion deficiency using lentiviral vectors with human CD11b and CD18 promoters driving canine CD18 expression.* Molecular therapy : the journal of the American Society of Gene Therapy, 2011. **19**(1): p. 113-21.

27. Moreno-Carranza, B., et al., *Transgene optimization significantly improves SIN vector titers, gp91phox expression and reconstitution of superoxide production in X-CGD cells.* Gene therapy, 2009. **16**(1): p. 111-8.

28. *Abstracts of the American Society of Gene & Cell Therapy 15th Annual Meeting. May 16-19, 2012. Philadelphia, Pennsylvania, USA.* Molecular therapy : the journal of the American Society of Gene Therapy, 2012. **20 Suppl 1**: p. S1-322.

29. Cavazzana-Calvo, M., et al., *Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia.* Nature, 2010. **467**(7313): p. 318-22.

30. Lombardo, A., et al., *Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery.* Nat Biotechnol, 2007. **25**(11): p. 1298-306.

31. Munoz, I.G., et al., *Molecular basis of engineered meganuclease targeting of the endogenous human RAG1 locus.* Nucleic Acids Res, 2011. **39**(2): p. 729-43.

32. Zou, J., et al., *Oxidase-deficient neutrophils from X-linked chronic granulomatous disease iPS cells: functional correction by zinc finger nuclease-mediated safe harbor targeting.* Blood, 2011. **117**(21): p. 5561-72.

33. Porteus, M., *Homologous recombination-based gene therapy for the primary immunodeficiencies.* Annals of the New York Academy of Sciences, 2011. **1246**: p. 131-40.

34. Langerak, P. and P. Russell, *Regulatory networks integrating cell cycle control with DNA damage checkpoints and double-strand break repair.* Philosophical

transactions of the Royal Society of London. Series B, Biological sciences, 2011. **366**(1584): p. 3562-71.

35. Amunugama, R. and R. Fishel, *Homologous recombination in eukaryotes.* Progress in molecular biology and translational science, 2012. **110**: p. 155-206.

36. Boboila, C., F.W. Alt, and B. Schwer, *Classical and alternative end-joining pathways for repair of lymphocyte-specific and general DNA double-strand breaks.* Advances in immunology, 2012. **116**: p. 1-49.

37. Carroll, D., *Genome engineering with zinc-finger nucleases.* Genetics, 2011. **188**(4): p. 773-82.

38. Capecchi, M.R., *Altering the genome by homologous recombination.* Science, 1989. **244**(4910): p. 1288-92.

39. Smithies, O., et al., *Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination.* Nature, 1985. **317**(6034): p. 230-4.

40. Rouet, P., F. Smih, and M. Jasin, *Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells.* Proc Natl Acad Sci U S A, 1994. **91**(13): p. 6064-8.

41. Stoddard, B.L., *Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification.* Structure, 2011. **19**(1): p. 7-15.

42. Grizot, S., et al., *Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease.* Nucleic acids research, 2009. **37**(16): p. 5405-19.

43. Arnould, S., et al., *Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells.* Journal of molecular biology, 2007. **371**(1): p. 49-65.

44. Chapdelaine, P., et al., *Meganucleases can restore the reading frame of a mutated dystrophin.* Gene therapy, 2010. **17**(7): p. 846-58.

45. Porteus, M.H. and D. Baltimore, *Chimeric nucleases stimulate gene targeting in human cells.* Science, 2003. **300**(5620): p. 763.

46. Urnov, F.D., et al., *Highly efficient endogenous human gene correction using designed zinc-finger nucleases.* Nature, 2005. **435**(7042): p. 646-51.

47. van Rensburg, R., et al., *Chromatin structure of two genomic sites for targeted transgene integration in induced pluripotent stem cells and hematopoietic stem cells.* Gene Ther, 2012.

48. Zou, J., et al., *Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells.* Cell stem cell, 2009. **5**(1): p. 97-110.

49. Lombardo, A., et al., *Site-specific integration and tailoring of cassette design for sustainable gene transfer.* Nat Methods, 2011. **8**(10): p. 861-9.

50. Connelly, J.P., et al., *Gene correction by homologous recombination with zinc finger nucleases in primary cells from a mouse model of a generic recessive genetic disease.* Mol Ther, 2010. **18**(6): p. 1103-10.

51. Li, H., et al., *In vivo genome editing restores haemostasis in a mouse model of haemophilia.* Nature, 2011. **475**(7355): p. 217-21.

52. Perez, E.E., et al., *Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases.* Nat Biotechnol, 2008. **26**(7): p. 808-16.

53. Holt, N., et al., *Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo.* Nat Biotechnol, 2010. **28**(8): p. 839-47.

54. Sander, J.D., et al., *Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA).* Nat Methods, 2011. **8**(1): p. 67-9.

55. Mussolino, C. and T. Cathomen, *TALE nucleases: tailored genome engineering made easy.* Curr Opin Biotechnol, 2012.

56. Boch, J., et al., *Breaking the code of DNA binding specificity of TAL-type III effectors.* Science, 2009. **326**(5959): p. 1509-12.

57. Moscou, M.J. and A.J. Bogdanove, *A simple cipher governs DNA recognition by TAL effectors.* Science, 2009. **326**(5959): p. 1501.

58. Christian, M., et al., *Targeting DNA double-strand breaks with TAL effector nucleases.* Genetics, 2010. **186**(2): p. 757-61.

59. Miller, J.C., et al., *A TALE nuclease architecture for efficient genome editing.* Nat Biotechnol, 2011. **29**(2): p. 143-8.

60. Mussolino, C., et al., *A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity.* Nucleic Acids Res, 2011. **39**(21): p. 9283-93.

61. Cermak, T., et al., *Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting.* Nucleic Acids Res, 2011. **39**(12): p. e82.

62. Briggs, A.W., et al., *Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers.* Nucleic Acids Res, 2012.

63. Reyon, D., et al., *FLASH assembly of TALENs for high-throughput genome editing.* Nat Biotechnol, 2012. **30**(5): p. 460-5.

64. Hockemeyer, D., et al., *Genetic engineering of human pluripotent cells using TALE nucleases.* Nat Biotechnol, 2011. **29**(8): p. 731-4.

65. Wood, A.J., et al., *Targeted genome editing across species using ZFNs and TALENs.* Science, 2011. **333**(6040): p. 307.

66. Sun, N., et al., *Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease.* Mol Biosyst, 2012. **8**(4): p. 1255-63.

67. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-23.

68. Mali, P., et al., *RNA-guided human genome engineering via Cas9.* Science, 2013. **339**(6121): p. 823-6.

69. Westra, E.R., et al., *CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3.* Molecular cell, 2012. **46**(5): p. 595-605.

70. Brunet, E., et al., *Chromosomal translocations induced at specified loci in human stem cells.* Proc Natl Acad Sci U S A, 2009. **106**(26): p. 10620-5.

71. Chiarle, R., et al., *Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells.* Cell, 2011. **147**(1): p. 107-19.

72. Gabriel, R., et al., *An unbiased genome-wide analysis of zinc-finger nuclease specificity.* Nat Biotechnol, 2011. **29**(9): p. 816-23.

73. Pattanayak, V., et al., *Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection.* Nat Methods, 2011. **8**(9): p. 765-70.

74. Deng, D., et al., *Structural basis for sequence-specific recognition of DNA by TAL effectors.* Science, 2012. **335**(6069): p. 720-3.

75. Mak, A.N., et al., *The crystal structure of TAL effector PthXo1 bound to its DNA target.* Science, 2012. **335**(6069): p. 716-9.

76. Streubel, J., et al., *TAL effector RVD specificities and efficiencies.* Nat Biotechnol, 2012. **30**(7): p. 593-5.

77. Fu, Y., et al., *High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells.* Nature biotechnology, 2013.

78. Doyon, Y., et al., *Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures.* Nat Methods, 2011. **8**(1): p. 74-9.

79. Maeder, M.L., et al., *Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification.* Molecular cell, 2008. **31**(2): p. 294-301.

80. Pruett-Miller, S.M., et al., *Comparison of zinc finger nucleases for use in gene targeting in mammalian cells.* Molecular therapy : the journal of the American Society of Gene Therapy, 2008. **16**(4): p. 707-17.

81. Szczepek, M., et al., *Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases.* Nature biotechnology, 2007. **25**(7): p. 786-93.

82. Guschin, D.Y., et al., *A rapid and general assay for monitoring endogenous gene modification.* Methods in molecular biology, 2010. **649**: p. 247-56.

83. Cermak, T., et al., *Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting.* Nucleic acids research, 2011. **39**(12): p. e82.

84. Miller, J.C., et al., *A TALE nuclease architecture for efficient genome editing.* Nature biotechnology, 2011. **29**(2): p. 143-8.

85. Davis, B.M., et al., *Characterization of the P140K, PVP(138-140)MLK, and G156A O6-methylguanine-DNA methyltransferase mutants: implications for drug resistance gene therapy.* Human gene therapy, 1999. **10**(17): p. 2769-78.

86. Dreier, B., et al., *Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors.* The Journal of biological chemistry, 2005. **280**(42): p. 35588-97.

87. Wilson, K.A., et al., *Expanding the Repertoire of Target Sites for Zinc Finger Nuclease-mediated Genome Modification.* Molecular therapy. Nucleic acids, 2013. **2**: p. e88.

88. Beard, B.C., et al., *Efficient and stable MGMT-mediated selection of long-term repopulating stem cells in nonhuman primates.* The Journal of clinical investigation, 2010. **120**(7): p. 2345-54.

89. Reyon, D., et al., *FLASH assembly of TALENs for high-throughput genome editing.* Nature biotechnology, 2012. **30**(5): p. 460-5.

90. Streubel, J., et al., *TAL effector RVD specificities and efficiencies.* Nature biotechnology, 2012. **30**(7): p. 593-5.

91. Pruett-Miller, S.M., et al., *Attenuation of zinc finger nuclease toxicity by small-molecule regulation of protein levels.* PLoS genetics, 2009. **5**(2): p. e1000376.

92. Chen, Z.Y., et al., *Minicircle DNA vectors devoid of bacterial DNA result in persistent and high-level transgene expression in vivo.* Molecular therapy : the journal of the American Society of Gene Therapy, 2003. **8**(3): p. 495-500.