

**DEVELOPMENT OF A NOVEL GENE THERAPY STRATEGY
FOR SCID-X1 AND A METHOD FOR MEASURING GENE
TARGETING OUTCOMES AT ENDOGENOUS LOCI**

APPROVED BY SUPERVISORY COMMITTEE

Matthew H. Porteus, M.D., Ph.D. (Mentor)

Kathlynn Brown, Ph.D. (Chair)

Woodring Wright, M.D., Ph.D.

Andrew Zinn, M.D., Ph.D.

DEDICATION

To my sister Katie and her husband Chris, for their extraordinary support and friendship. To my brother Scott and his wife Brittany, my parents, and my close friends for their amazing love and support throughout my life.

**DEVELOPMENT OF A NOVEL GENE THERAPY STRATEGY
FOR SCID-X1 AND A METHOD FOR MEASURING GENE
TARGETING OUTCOMES AT ENDOGENOUS LOCI**

by

ERIC JAMES KILDEBECK

DISSERTATION

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

September 2013

Copyright

by

ERIC JAMES KILDEBECK, 2013

All Rights Reserved

**DEVELOPMENT OF A NOVEL GENE THERAPY STRATEGY
FOR SCID-X1 AND A METHOD FOR MEASURING GENE
TARGETING OUTCOMES AT ENDOGENOUS LOCI**

ERIC JAMES KILDEBECK, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2013

SUPERVISING PROFESSOR: MATTHEW H. PORTEUS, M.D., Ph.D.

Two decades of gene therapy trials for primary immunodeficiencies have seen tremendous clinical success with a significant majority of patients developing functional immune systems. The development of leukemia in some patients has led to the development of precise gene targeting tools to correct genetic deficits without inducing genomic instability. In this thesis I report the development of a novel gene therapy strategy for SCID-X1 and the development

of a useful method for measuring gene editing outcomes at endogenous loci in any cell type. TALENs designed to target *IL2RG* exon 1 are shown to be highly active and stimulate precise integration of *IL2RG* cDNA under the control of the endogenous *IL2RG* promoter. Activity levels of IL2R γ in cells targeted with a codon-optimized cDNA and an artificial intron are also shown to be as high or higher than WT levels, demonstrating the potential for this approach to correct the functional deficit seen in SCID-X1. Furthermore, these TALENs successfully stimulate gene targeting in CD34⁺ hematopoietic stem and progenitor cells at frequencies 10-fold higher than the highest levels previously reported, while displaying less toxicity than ZFNs already in use in clinical trials. The high activity and low toxicity of these TALENs in combination with the potential for gene targeting at exon 1 to correct more than 98% of SCID-X1-causing mutations make this a promising strategy for gene therapy, which could one day form the basis for a safe and effective cure for SCID-X1.

TABLE OF CONTENTS

<i>DEDICATION</i>	<i>ii</i>
<i>ABSTRACT</i>	<i>v</i>
<i>TABLE OF CONTENTS</i>	<i>vii</i>
<i>PRIOR PUBLICATIONS</i>	<i>x</i>
<i>LIST OF FIGURES</i>	<i>xi</i>
<i>LIST OF TABLES</i>	<i>xiv</i>
<i>LIST OF ABBREVIATIONS</i>	<i>xv</i>
<i>CHAPTER I: INTRODUCTION AND REVIEW OF THE LITERATURE</i>	<i>1</i>
GENE THERAPY FOR PRIMARY IMMUNODEFICIENCIES	1
IMPROVEMENTS IN VIRAL VECTORS IN PRECLINICAL TRIALS	3
CLINICAL TRIALS FOR PRIMARY IMMUNODEFICIENCIES	5
GENE TARGETING: A PROMISING ALTERNATIVE FOR GENE	
THERAPY	6
ZINC-FINGER NUCLEASES	9
TALE EFFECTOR NUCLEASES	14
OFF-TARGET TOXICITIES WITH GENE TARGETING	17
SUMMARY	19
PATHOPHYSIOLOGY AND CLINICAL MANAGEMENT OF SCID-X1	21
PATHOPHYSIOLOGY OF SCID-X1	21
CLINICAL PRESENTATION AND DIAGNOSIS OF SCID-X1	22
EPIDEMIOLOGY OF SCID-X1	23
TREATMENT OF SCID-X1	23

CHAPTER II: DEVELOPMENT OF A METHOD FOR MEASURING GENE EDITING OUTCOMES AT ENDOGENOUS LOCI	25
ABSTRACT	25
INTRODUCTION	25
RESULTS	29
DEVELOPMENT OF A HIGH-THROUGHPUT STRATEGY FOR MEASURING TARGETING AT ENDOGENOUS LOCI	29
MEASUREMENT OF GENE EDITING OUTCOMES AT <i>IL2RG</i>	31
MEASUREMENT OF GENE EDITING OUTCOMES AT A SILENT LOCUS IN HUMAN STEM CELLS	37
OPTIMIZATION OF GENE TARGETING PARAMETERS AT AN ENDOGENOUS LOCUS	40
DISCUSSION	48
MATERIALS AND METHODS	49
 CHAPTER III: DEVELOPMENT OF A NOVEL GENE THERAPY STRATEGY FOR SCID-X1	 56
ABSTRACT	56
INTRODUCTION	57
RESULTS	61
DEVELOPMENT OF OPEN ZFNs TARGETING <i>IL2RG</i> EXON 5	61
DEVELOPMENT OF HIGHLY-ACTIVE TALENS FOR GENE TARGETING AT <i>IL2RG</i> EXON 1	71
DEVELOPMENT OF A NOVEL GENE THERAPY STRATEGY FOR SCID-X1	81
DISCUSSION	95
MATERIALS AND METHODS	98

CHAPTER IV: CONCLUSIONS AND FUTURE DIRECTIONS	107
MEDICAL APPLICATIONS OF GENE TARGETING:	
A PARADIGM SHIFT FOR GENETIC THERAPIES	108
APPLICATION OF GENE TARGETING TECHNOLOGIES TO THE TREATMENT OF HUMAN DISEASE.....	108
DEVELOPMENT OF A SAFE AND EFFECTIVE CURE FOR SCID-X1	111
OVERCOMING TOXICITY IN HEMATOPOIETIC STEM CELLS:	
THE PATH TO PRECISE GENE THERAPY FOR SCID-X1	113
GENE TARGETING: A NEW PARADIGM FOR GENE THERAPY.	117
OVERALL CONCLUSIONS	119
 BIBLIOGRAPHY	 120

PRIOR PUBLICATIONS

Kildebeck EJ, Hendel A, Fine EJ, Clark J, Sebastiano V, Bao G, Porteus MH.

(2013) Measuring genome editing outcomes at endogenous loci using SMRT sequencing. (Submitted)

Kildebeck E, Checketts J, Porteus M. (2012) Gene therapy for primary immunodeficiencies. *Curr Opin Pediatr*. 24(6): 731-8.

Wilson KA, McEwen AE, Pruett-Miller SM, Zhang J, **Kildebeck EJ**, Porteus MH. (2013) Expanding the repertoire of target sites for zinc finger nuclease-mediated genome modification. *Mol Ther Nucleic Acids*. 2(4): published online ahead of print.

Bu D, Tomlinson G, Lewis CM, Zhang C, **Kildebeck E**, Euhus DM. (2006) An intronic polymorphism associated with increased XRCC1 expression, reduced apoptosis and familial breast cancer. *Breast Cancer Res Treat*. 99(3): 257-65.

LIST OF FIGURES

<i>FIGURE 1.1: Therapeutic Gene Targeting by Homologous Recombination</i>	8
<i>FIGURE 1.2: Engineered Nuclease Platforms For Site-Specific Gene Targeting</i>	10
<i>FIGURE 1.3: Clinical Paradigm for Gene Targeting</i>	12
<i>FIGURE 1.4: Distribution of mutations in the IL2RG gene known to cause SCID-X1</i>	21
<i>FIGURE 2.1: Gene Targeting strategy for IL2RG</i>	30
<i>FIGURE 2.2: Measuring gene editing at an endogenous locus with SMRT sequencing</i>	33
<i>FIGURE 2.3: Comparison of SMRT sequencing to RFLP and single cell clone analysis</i>	35
<i>FIGURE 2.4: Reproducibility of SMRT sequencing for gene editing analysis</i> ...	30
<i>FIGURE 2.5: Mutation spectra caused by TALENs at IL2RG in the K562 cell line and primary CD34+ hematopoietic stem and progenitor cells</i>	39
<i>FIGURE 2.6: Interrogation of gene editing parameters at an endogenous locus with SMRT sequencing</i>	41
<i>FIGURE 2.7: Ratio of HDR to NHEJ with varying amounts of donor DNA for samples in Figure 2.6b</i>	42

<i>FIGURE 3.1: Context-dependant selection of ZFNs for IL2RG exon 5 using the OPEN platform</i>	66
<i>FIGURE 3.2: Measuring the activity of OPEN ZFNs with fluorescent reporter assays</i>	67
<i>FIGURE 3.3: Measurement of standard OPEN ZFN expression levels and activity at the endogenous target site compared to codon-optimized OPEN ZFNs and MA ZFNs</i>	69
<i>FIGURE 3.4: TALENs designed to target IL2RG exon 1</i>	75
<i>FIGURE 3.5: Specific mutations created by IL2RG exon 1 TALENs</i>	76
<i>FIGURE 3.6: Gene targeting of restriction sites to IL2RG exon 1</i>	77
<i>FIGURE 3.7: Comparison of toxicity between IL2RG exon 1 TALENs and commonly-used ZFNs</i>	79
<i>FIGURE 3.8: Novel gene therapy strategy for SCID-X1</i>	87
<i>FIGURE 3.9: High throughput sequencing approach using SMRT sequencing to confirm targeted gene addition</i>	88
<i>FIGURE 3.10: Measurement of IL2Rγ activity in gene targeted RAMOS cells</i>	90
<i>FIGURE 3.11: Comparison of activity levels with 'high-expression' and standard TALEN expression constructs</i>	92
<i>FIGURE 3.12: Measurement of gene targeting efficiencies with DNA and protein expression levels relevant to CD34$^{+}$ hematopoietic stem and progenitor cells</i>	93

<i>FIGURE 3.13: Gene targeting in CD34+ hematopoietic stem and progenitor cells</i>	94
<i>FIGURE 3.14: List of primers used in Chapter III</i>	106
<i>FIGURE 4.1: Gene targeting at IL2RG exon 1 with CRISPR-Cas9 nucleases</i> .	110
<i>FIGURE 4.2: Gene targeting in CD34+ HSPCs with mRNA delivery of TALENs</i>	114
<i>FIGURE 4.3: Gene targeting using miniVector DNA as a donor template</i>	116

LIST OF TABLES

TABLE 2.1: <i>Summary of SMRT sequencing data for Figure 2.2c –</i>	
<i>Targeting IL2RG in K562 cells</i>	43
TABLE 2.2: <i>Summary of SMRT sequencing data for Figure 2.2d –</i>	
<i>Targeting IL2RG in CD34+ HSPCs</i>	43
TABLE 2.3: <i>Summary of SMRT sequencing data for gene targeting in hESCs...</i>	44
TABLE 2.4: <i>Summary of SMRT sequencing data for Figure 2.6a –</i>	
<i>Titration of TALEN amounts</i>	45
TABLE 2.5: <i>Summary of SMRT sequencing data for Figure 2.6b –</i>	
<i>Titration of Donor template amount</i>	46
TABLE 2.6: <i>Summary of SMRT sequencing data for Figure 2.6c –</i>	
<i>Homology arm length comparison</i>	47
TABLE 2.7: <i>List of primers used in Chapter II</i>	54

LIST OF ABBREVIATIONS

AAV – adeno-associated virus

AAVS1 – AAV integration site 1

ADA – adenosine deaminase

CGD – chronic granulomatous disease

cDNA – complementary deoxyribonucleic acid

CLAD – canine leukocyte adhesion deficiency

CMV – cytomegalovirus promoter

CRISPRs – clustered regularly interspaced short palindromic repeats

DNA – deoxyribonucleic acid

DSB – double strand break

EF1 α – elongation factor 1-alpha promoter

eGFP – enhanced green fluorescent protein

FACS – fluorescence-activated cell sorting

Fn – FokI nuclease domain

GFP – green fluorescent protein

GVHD – graft-versus-host disease

HDR – homology-directed repair

HE – homing endonuclease

hESCs – human embryonic stem cells

hiPSCs – human induced pluripotent stem cells

HIV – human immunodeficiency virus

HLA – human leukocyte antigen

HR – homologous recombination

HSC – hematopoietic stem cell

HSCT – hematopoietic stem cell transplantation

HSPCs – hematopoietic stem and progenitor cells

IDLV – integration defective lentivirus

IL – interleukin

IL2R γ – interleukin 2 receptor gamma chain

indel – insertion or deletion mutation

iPSC – induced pluripotent stem cell

IRES – internal ribosomal entry site

JAK – Janus kinase

KO - knockout

LTR – long terminal repeat

MA – modular assembly

mRNA – messenger ribonucleic acid

NHEJ – non-homologous end joining

NK – natural killer

OPEN – Oligomerized pool engineering

PAGE – polyacrylamide gel electrophoresis

PCR – polymerase chain reaction

PGK – phosphoglycerate kinase I promoter

PID – primary immunodeficiency

PNH – paroxysmal nocturnal hemoglobinuria

polyA – polyadenylation site

QV – quality value

RAG – recombination activating gene

RFLP – restriction fragment length polymorphism

RNA – ribonucleic acid

RVD – repeat variable diresidue

SCF – stem cell factor

SCID – severe combined immunodeficiency

SCID-X1 – X-linked severe combined immunodeficiency

SFFV – spleen focus-forming virus

SIN – self-inactivating

SMRT – single molecule real time

SNP – single nucleotide polymorphism

SSA – single-strand annealing

ssODN – single-stranded oligodeoxynucleotide

STAT – signal transducer and activator of transcription

TALE – transcription activator-like effector

TALEN – transcription activator-like effector nuclease

TPO – thrombopoietin

TREC – T-cell receptor excision circle

Ubc – ubiquitin C promoter

UCOE – ubiquitin chromatin opening element

UTR – untranslated region

WAS – Wiskott-Aldrich Syndrome

WT – wild type

X-CGD – X-linked chronic granulomatous disease

γ H2AX – phosphorylated histone 2AX

ZFN – zinc-finger nuclease

ZFP – zinc-finger protein

CHAPTER I:

INTRODUCTION AND REVIEW OF THE LITERATURE

Gene Therapy For Primary Immunodeficiencies

Primary immunodeficiencies (PIDs) are an often devastating class of genetic disorders that can be effectively treated by hematopoietic stem cell transplantation (HSCT), but the lack of a suitable donor precludes this option for many patients. Gene therapy overcomes this obstacle by restoring gene expression in autologous hematopoietic stem cells (HSCs) and has proven effective in clinical trials, but widespread use of this approach has been impeded by the occurrence of serious complications. Here I will discuss recent advances in gene therapy with an emphasis on strategies to improve safety including the emergence of gene targeting technologies for the treatment of PIDs.

PIDs are a large and varied group of genetic disorders resulting in an underdeveloped or dysfunctional immune system in patients. Over 150 types of PID have been molecularly characterized, with more being added over time (Booth et al., 2011). The phenotypes range from being almost asymptomatic (ie, selective immunoglobulin A deficiency) to potentially fatal (ie, severe combined immunodeficiency (SCID)). For PIDs with severe phenotypes, HSCT has resulted in significant therapeutic benefit, with success rates over 90% when a human

leukocyte antigen (HLA) matched donor is available. Unfortunately, an HLA-matched donor is available for only about one in three patients, and the use of a mismatched donor decreases the frequency of successful engraftment to about 60% while simultaneously increasing the risk of morbidity and mortality. In patients with T-B- phenotypes, such as in SCID, the lack of functional T cells and B cells results in a lower probability of graft rejection and improves the response to transplantation. Other forms of PID, however, can require myeloablative conditioning regimens that cause additional complications and lower the percentage of successful engraftment.

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 SCID due to adenosine deaminase deficiency (ADA-SCID) (Blaese et al., 1995). Immune cells are especially susceptible to the toxic levels of deoxyadenosine that build up in ADA deficiency, resulting in a lack of T, B, and NK cells. Patients were transplanted with autologous T-lymphocytes transduced with a retrovirus encoding ADA cDNA flanked by viral long terminal repeats (LTRs). As hoped, 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, nine patients were treated with autologous CD34+ hematopoietic stem and progenitor cells (HSPCs) transduced with a gammaretroviral vector, with eight patients demonstrating immune reconstitution (Hacein-Bey-Abina et al., 2010). Unfortunately, four of these patients subsequently developed T-cell leukemia resulting from retroviral insertion. Three of these patients entered remission following standard therapies, while one passed away due to complications incident to his leukemia. Despite these complications, gammaretroviral-based gene therapy trials continued for ADA-SCID (Aiuti et al., 2009; Gaspar et al., 2011) and SCID-X1 (Gaspar et al., 2011), as well as for several other PIDs including Wiskott-Aldrich Syndrome (WAS) (Boztug et al., 2010) and chronic granulomatous disease (CGD) (Kang et al., 2011). 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 effects were observed in WAS and CGD patients, but for reasons not completely understood have yet to be observed in trials for ADA-SCID.

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 (Naldini, 2011).

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 (Cattoglio et al., 2010), 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 (Cooray et al., 2012). 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 EF1 α . Using this approach, restoration of immune function in mice has been demonstrated for Rag 1 & 2 deficiencies, WAS, CGD, SCID-X1, and other diseases (Astrakhan et al., 2012; Avedillo Diez et al., 2011; Huston et al., 2011; Pike-Overzet et al., 2011; Santilli et al., 2011; Scaramuzza et al., 2013; van Til et al., 2012). One obstacle encountered with ubiquitous promoters, particularly SFFV, is transgene silencing due to DNA methylation (van Til et al., 2012). 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 (Pike-Overzet et al., 2011; van Til et al., 2012; Zhang et al., 2010). On the other end of the spectrum from transgene silencing, overexpression of a transgene can also be detrimental (Ng et al., 2010). The use of endogenous promoters for a specific transgene mitigates this problem and has been used to treat WAS, SCID-X1, and CLAD in animal models (Hunter et al., 2011; Huston et al., 2011; Scaramuzza et al., 2013; van Til et al., 2012). 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 (Huston et al., 2011; Moreno-Carranza et al., 2009; Ng et al., 2010; Pike-Overzet et al., 2011; van Til et al., 2012). 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 (Huston et al., 2011).

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 EF1 α promoter (clinicaltrials.gov, NCT01852071), 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 (Rivat et al., 2012; Santilli et al., 2011; Scaramuzza et al., 2013). 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 (Gaspar et al., 2011). A gammaretroviral-based trial for CGD, on the other hand, showed only transient clinical benefit with significantly decreased transgene expression over time, perhaps resulting from DNA methylation (Kang et al., 2011). While I believe 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 (Cavazzana-Calvo et al., 2010).

Gene Targeting: A Promising Alternative 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, as used in this review, 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 zinc-finger nucleases (ZFNs), homing endonucleases (HEs), and transcription activator-like effector nucleases (TALENs), and most recently clustered regularly interspaced short palindromic repeats (CRISPRs) can be designed to stimulate gene targeting at specific sites in the genome. Proof-of-principle correction of disease phenotypes using engineered nucleases has been demonstrated in cell lines for multiple monogenic diseases, including the PIDs SCID-X1, RAG1-SCID, and CGD (Lombardo et al., 2007; Munoz et al., 2011; Zou et al., 2011).

Double-strand breaks (DSBs) in DNA are typically repaired by either non-homologous end-joining (NHEJ), which joins DNA ends in an imprecise manner that can disrupt genes, or homologous recombination (HR), which utilizes a homologous donor template for precise repair. For naturally-occurring DSBs, HR utilizes the sister chromatid as a template to restore the original DNA sequence. 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**).

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 1 in 10^6 up to 3-5% provided the first evidence that

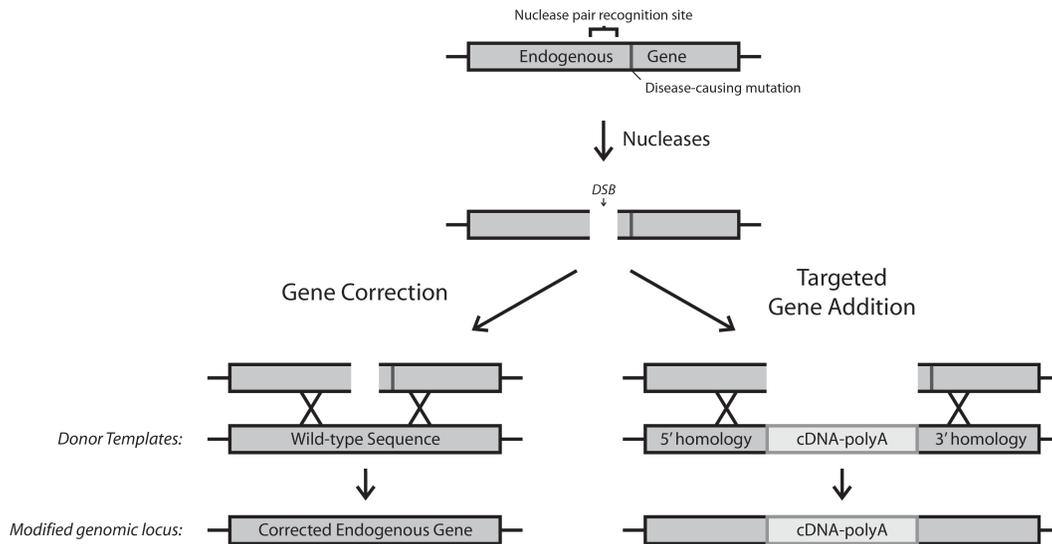


Figure 1.1: Therapeutic Gene Targeting by Homologous Recombination.

Gene targeting is a process where the sequence of an endogenous gene is precisely altered to encode a therapeutically useful gene product. The desired sequence for the gene is encoded in an exogenous donor template. When a DSB is created by targeted nucleases, the natural DSB repair machinery of the cell can recognize the donor template by regions of homology flanking the site of the DSB. If the donor template is utilized to repair the DSB, the genomic locus is permanently modified to the new sequence, which can directly correct a disease-causing mutation (gene correction) or insert a therapeutic transgene (targeted gene addition). In targeted gene addition, the cDNA can be inserted such that it is driven by the endogenous promoter, in which case no exogenous regulatory elements are needed, or it can be inserted with an exogenous regulatory element at a ‘safe harbor’ site.

therapeutic levels of gene targeting could be achieved (Rouet et al., 1994). 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 nuclease domains.

Zinc-Finger Nucleases

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 (Porteus et al., 2003). ZFNs are chimeric proteins made by fusing C₂H₂ 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 zinc fingers arrayed in each ZFN.

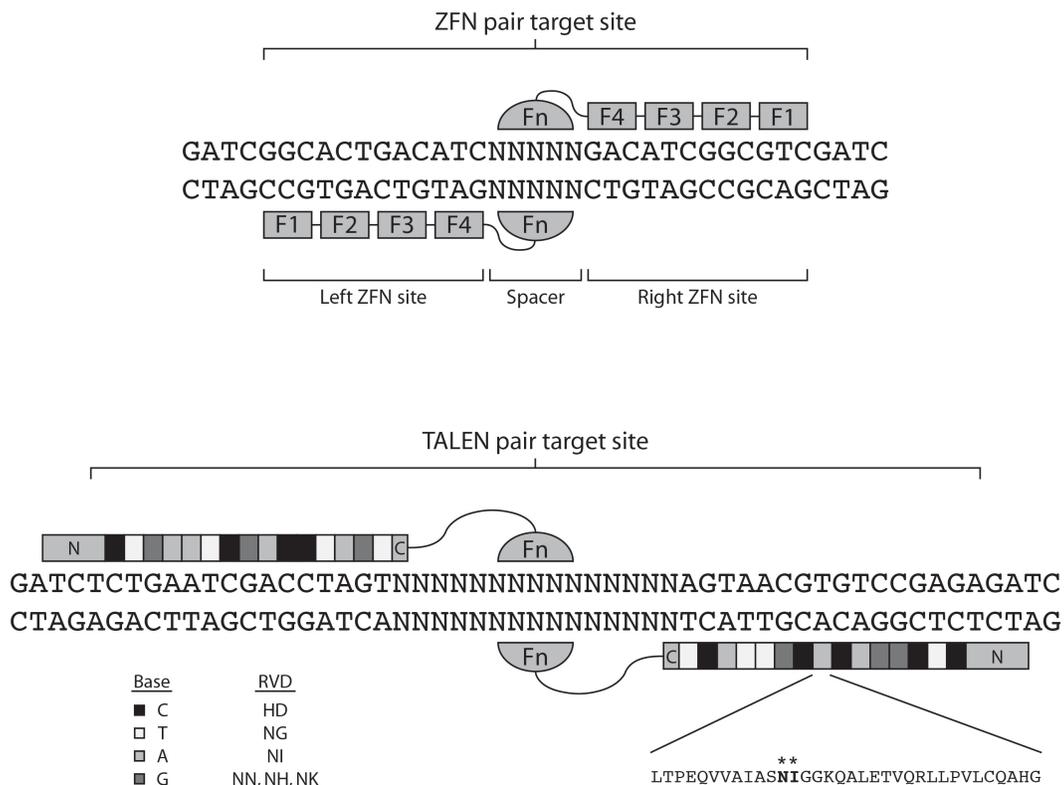


Figure 1.2: Engineered Nuclease Platforms For Site-Specific Gene Targeting.

Engineered nucleases are chimeric proteins made by fusing specific DNA binding domains to nuclease domains, which can then create DSBs at desired sites in the genome. ZFNs bind DNA through an array of 3-6 zinc finger motifs that each recognize ~3 base pairs of DNA. The ZFN pair depicted here is comprised of ZFNs with 4 zinc fingers each, labeled F1-F4. TALENs bind DNA through a series of repeats where each repeat binds to a single base pair of DNA. The amino acid sequence of an individual TALE repeat is shown with the amino acids responsible for nucleotide binding specificity, known as the repeat-variable diresidue (RVD), indicated with asterisks. RVDs for each nucleotide are shown (Boch et al., 2009; Miller et al., 2011; Moscou et al., 2009; Streubel et al., 2012). Fn: *FokI* nuclease domain, N: N-terminus, C: C-terminus.

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 (Urnov et al., 2005). 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 (Lombardo et al., 2007).

Translation of gene targeting for SCID and many other diseases will require modification of stem cells, and efforts to develop gene targeting protocols for 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 (Zou et al., 2009). Similar results have been shown with ZFNs targeting *CCR5* and *AAVS1* (Lombardo et al., 2007; van Rensburg et al., 2012; Zou et al., 2011). Targeting of

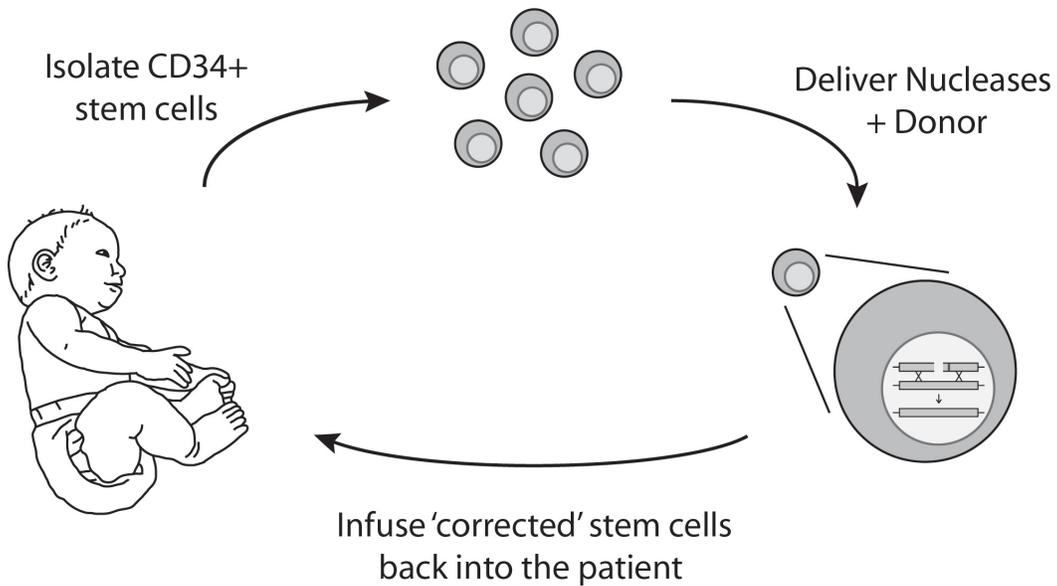


Figure 1.3: Clinical Paradigm for Gene Targeting. Clinical application of gene targeting will follow a similar approach as is used in current viral gene therapy trials. First, CD34+ HSPCs are isolated from bone marrow or mobilized peripheral blood. The cells are then cultured *ex vivo* and gene targeting reagents are delivered into the cells by transduction or transfection. The population of treated cells, some of which have been ‘corrected’, is then infused back into the patient. For PIDs where ‘corrected’ cells have a significant growth advantage, such as SCID-X1, even very small numbers of modified cells can be curative for patients.

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% (Lombardo et al., 2011; Lombardo et al., 2007).

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, and hepatocytes (Connelly et al., 2010; Li et al., 2011; Lombardo et al., 2011; Urnov et al., 2005). In one remarkable study, Li *et al.* (Li et al., 2011) 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. 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 (Holt et al., 2010; Perez et al., 2008).

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 (Sander et al., 2011), 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 (Mussolino et al., 2012). TAL effectors (TALEs) 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 (Boch et al., 2009; Moscou et al., 2009). 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 (Christian et al., 2010; Miller et al., 2011; Mussolino et al., 2011). One study by Miller *et al.* (Miller et al., 2011) 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. 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 (Cermak et al., 2011). 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 (Briggs et al., 2012; Cermak et al., 2011; Reyon et al., 2012). Using the FLASH assembly method, Reyon *et al.* (Reyon et al., 2012) 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. 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 (Hockemeyer et al., 2011; Wood et al., 2011). TALENs designed to target disease-causing mutations, such as the sickle cell mutation in β -globin (Sun et al., 2012), have also been reported, but the application of TALENs for gene therapy is really just beginning.

Off-Target Toxicities With 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 mutations. 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 (Brunet et al., 2009; Chiarle et al., 2011). Potential off-target sites for specific ZFNs have been determined using unbiased assays for DSB creation (Gabriel et al., 2011; Pattanayak et al., 2011). Pattanayak *et al.* (Pattanayak et al., 2011) 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. 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 (Mussolino et al., 2011; Sun et al., 2012). TALENs expressed

in human cell lines have been shown to cause fewer γ H2AX 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 (Mussolino et al., 2011). The crystal structure of TALEs binding DNA also shows that the RVDs for different nucleotides interact with DNA in different ways, possibly with ‘HD’ (cytosine) and ‘NN’ (guanine/adenosine) forming hydrogen bonds, ‘NI’ (adenosine) forming only van der Waals interactions, and ‘NG’ (thymine) relying on steric hindrance for specificity (Deng et al., 2012; Mak et al., 2012; Streubel et al., 2012). This suggests that TALEN specificity will be more complex than a simple number of mismatches tolerated for ‘off-target’ binding, and that high specificity may be designed without the cost of decreased activity.

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% (Doyon et al., 2011; Gabriel et al., 2011). Nucleases can also be delivered as mRNA to prevent random integration of an expression plasmid and decrease the duration of nuclease expression. 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.* (Lombardo et al., 2011) recently showed that the *AAVSI* 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. ZFN-mediated transgene insertion at the *AAVSI* 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.* (Zou et al., 2011) with the targeting of a gp91^{phox} minigene to *AAVSI* to correct the phenotype of CGD. 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.

Summary

Improvements in the design of viral vectors and the development of new tools for precise gene targeting make this an exciting time for gene therapy. The genotoxic side effects encountered in early clinical trials have often

overshadowed the clinical successes of gene therapy in our public forum. Several strategies currently under development have the potential to prevent these side effects while maintaining or improving our ability to correct disease phenotypes in patients suffering from PIDs. In particular, advances in gene targeting may allow for precise correction of disease-causing mutations or insertion of transgenes without otherwise altering the genome. Demonstrations using ZFNs and, more recently, TALENs, suggest that this approach may one day allow for a new level of sophistication for gene therapy where restoration of gene expression occurs with full endogenous regulation and without significant risk of oncogenesis.

Pathophysiology and Clinical Management of SCID-X1

Pathophysiology of SCID-X1

X-linked Severe Combined Immunodeficiency (SCID-X1) is caused by a mutation in the interleukin-2 receptor gamma chain gene (*IL2RG*) at the Xq13.1 locus (Puck et al., 1993). The *IL2RG* gene encodes the protein IL2R γ , also known as the common γ chain (γ c), which forms part of the cell-surface receptors for interleukins 2, 4, 7, 9, 15, and 21. Mutations throughout *IL2RG* are capable of causing SCID-X1, with a hotspot for mutations in exon 5 (**Figure 1.4**). Following interleukin binding, IL2R γ mediates signal transduction through a JAK-STAT signaling pathway. Without functional IL2R γ , lymphopoiesis is severely impaired

Mutations Causing SCID-X1

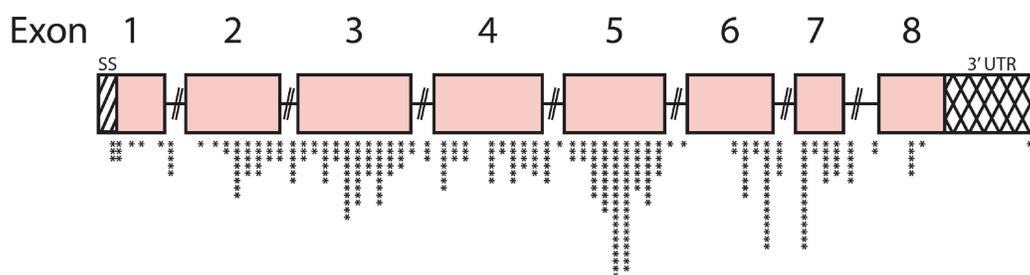


Figure 1.4: Distribution of mutations (*) in the *IL2RG* gene known to cause SCID-X1. Disease-causing mutations are spread throughout the gene with a hotspot of mutations in Exon 5. Over 340 unique mutations have been described for SCID-X1. *Image modified from nih.gov.*

and hematopoietic stem cells (HSCs) are unable to differentiate into functional T cells, B cells, or NK cells. While mature T cells and NK cells are typically absent in SCID-X1 patients, B cells are present at normal or increased levels, but are compromised by defective interleukin signaling and a lack of T cell stimulation.

Clinical Presentation and Diagnosis of SCID-X1

SCID-X1 patients often present within the first 3 months of life with severe infections including oral candidiasis and pneumonitis, with a mean presentation of ~6.5 months (Buckley et al., 1997). Other common symptoms include diarrhea, failure to thrive, and infections of the gut, with recurrent infections subsequently resulting in growth impairment and malnutrition. Without successful bone marrow transplantation, SCID-X1 is usually lethal within the first year of life.

Diagnosis of SCID can be made by measuring peripheral lymphocyte cell counts and testing lymphocyte function, in addition to genetic testing. SCID-X1 patients will typically have very low peripheral T cell and NK cell counts combined with normal or elevated B cell counts, and will have an absent or small thymus as visualized with a chest X-ray. Diagnosis of SCID-X1 is definitively established by a lack of cell-surface IL2R γ expression, which is determined by staining followed by flow cytometry, or mutational analysis of *IL2RG*.

Epidemiology of SCID-X1

The frequency of all forms of SCID is estimated to be between 1:50,000 to 1:100,000 births (Buckley et al., 1997), with SCID-X1 accounting for approximately 50% of all cases. The majority of SCID-X1 patients do not have any family history of the disease, suggesting the many cases are the result of new mutations. There is no known disparity in occurrence of SCID-X1 between different ethnic groups.

Treatment of SCID-X1

Initial management of SCID-X1 focuses on treatment and prevention of infections with a combination of medications for opportunistic infections, administration of intravenous immunoglobulin, and isolation from pathogens. The term 'bubble boy disease' resulted from isolation of patients in plastic 'bubbles', the most famous of which was David Vetter, who lived in a 'bubble' for almost his entire life until his death at the age of 12. Hematopoietic stem cell transplantation (HSCT) is curative for SCID-X1, with a 90% success rate for patients transplanted with a matched donor. For patients without a matched donor, bone marrow transplantation with a haploidentical donor has a success rate of ~70-78% with removal of T cells in donor marrow to prevent graft-versus-host disease (GVHD) (Fischer, 2000). Immunosuppression is not required for HSCT in SCID-X1 patients due to the lack of an immune system to reject engraftment.

Recently, the report of a 90% success rate for HSCT with a haploidentical donor for transplants performed within the first 3 months of life provided strong evidence in favor of newborn screening for SCIDs, and multiple states have added SCIDs to newborn screening utilizing the T-cell receptor excision circle (TREC) assay (Brown et al., 2011; Buckley, 2012).

For patients diagnosed after the first 3 months of life who do not have an available matched donor, gene therapy represents a viable option for treatment (Cavazzana-Calvo et al., 2013). Of 20 SCID-X1 patients treated with γ -retroviral-mediated gene therapy, 17 patients were disease-free with a median follow-up time of 10 years (Hacein-Bey-Abina et al., 2008) and 5 of patients developed leukemia due to insertional oncogenesis. Precise gene targeting strategies to prevent insertion oncogenesis have led to the development of *IL2RG*-specific ZFNs (Urnov et al., 2005), but this approach has not been used clinically due to the low efficiencies and high toxicities of these reagents in hematopoietic stem cells. In Chapter III of this thesis, I present a novel strategy for using TALEN-mediated gene addition of *IL2RG* cDNA to correct the functional deficit in almost all patients suffering from SCID-X1.

CHAPTER II:
DEVELOPMENT OF A METHOD FOR MEASURING GENE
EDITING OUTCOMES AT ENDOGENOUS LOCI

Abstract

Targeted genome editing with engineered nucleases increasingly enables precise sequence modifications at almost any site within the genome. A variety of reporter assays for tracking genome editing outcomes have been developed, but none has allowed for the frequency of different genome editing outcomes to be measured simultaneously at any endogenous locus. In this section I describe a method for quantifying genome editing outcomes at any site of interest using single molecule real time (SMRT) DNA sequencing. This method allows for simultaneous measurement of NHEJ and HDR modification frequencies at endogenous loci in any cell type. By directly sequencing targeted loci with long sequencing read-lengths, I further show that this method can be used to analyze gene editing of transcriptionally silent loci in primary cells and a large range of insertion and deletion mutations created by engineered nucleases.

Introduction

Genome editing with engineered nucleases is a transformative technology for efficiently modifying essentially any genomic sequence of interest (McMahon et al., 2012). This technology utilizes engineered nucleases to generate site-specific double-strand breaks (DSB) at desired genomic locations followed by resolution of DSBs using the endogenous cellular repair mechanisms of nonhomologous end-joining (NHEJ) and homology directed repair (HDR) (Porteus et al., 2005). A variety of desired genetic modifications can be achieved with this approach, including disruption of an unwanted gene through mutagenic NHEJ and precise alteration of a genomic sequence to a new sequence through HDR. There are currently four principal families of engineered nucleases used for gene editing: Zinc Finger Nucleases (ZFNs) (Porteus et al., 2005), Transcription Activator-Like Effector Nucleases (TALENs) (Bogdanove et al., 2011), Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) (Gaj et al., 2013), and meganucleases (Silva et al., 2011). The rapid development of these technologies is allowing researchers to precisely alter genomes for numerous applications, including plant engineering, generation of cell lines for basic science, human gene therapy, and industrial applications.

When a new set of gene editing reagents is developed for a custom application, the activity levels of nucleases and the frequency of the desired gene

editing event must be determined and often need to be optimized for the specific cell type and system used. This need has previously been met by a variety of methods including gel-based assays to measure mutagenic NHEJ (Guschin et al., 2010), gene addition of fluorescent reporters to measure HDR (Porteus et al., 2003; Stark et al., 2004), analysis of large numbers of single cell clones, and the use of optimization assays to measure NHEJ and HDR at engineered reporter loci (Certo et al., 2011). While each of these assays has their utility, each also has important limitations including insufficient sensitivity as required for difficult applications (gel based assays), that they are surrogate rather than direct measures of genome editing (targeted gene addition of fluorescent reporters), and the need to generate reporter cell lines (Traffic Light Reporter system). The Traffic Light Reporter system is the only one of these assays that allows simultaneous measurement of NHEJ and HDR, which is accomplished by expressing GFP in cells that undergo HDR-mediated gene correction of a GFP gene or expressing mCherry in cells that undergo frameshifts induced by NHEJ (Certo et al., 2011). While this is a very sensitive assay for measuring DSB repair pathway choice, the need to generate a fluorescent reporter locus precludes measurement at endogenous target loci and in human primary cells. High-throughput sequencing of endogenous loci overcomes these limitations, but the range of gene editing outcomes that can be measured is determined by sequencing read-lengths. Illumina (Yang et al., 2013) and 454 (Qi et al., 2013) sequencing have recently

been used to measure HDR and NHEJ outcomes using single-stranded oligodeoxynucleotides (ssODNs) or plasmids with short homology arms as donor templates, but the read-length limitations of these platforms do not allow analysis of longer arms of homology that drive more efficient HDR and the flexibility to target long gene cassettes. Here, I show that measuring genome editing outcomes at endogenous loci can be done with SMRT sequencing using an affordable approach that can be widely used.

The SMRT DNA sequencing project was performed in collaboration with Ayal Hendel and Eli Fine. For all of the experiments presented here, I developed all of the targeting reagents, performed the experiments, and generated the figures, Ayal Hendel prepared the SMRT libraries, and Eli Fine performed high-throughput sequencing analysis. The SMRT DNA sequencing technology was developed by Pacific Biosciences and allows for the determination of DNA sequence from individual DNA templates (Eid et al., 2009; Roberts et al., 2013). Single-molecule read lengths approaching 15 kb were reached in this study, with an average read length approaching 3kb. For DNA fragments shorter than the read limit of the polymerase, improved sequence accuracy (frequently reaching an average Phred QV score of 40, denoting 99.99% accuracy) is achieved by iteratively sequencing the same circular DNA template (Travers et al., 2010).

Results

Development of a high-throughput strategy for measuring targeting at endogenous loci

To develop a method for quantitatively, robustly, and rapidly measuring the different gene editing outcomes occurring at an endogenous locus of interest, I used a highly active TALEN pair to stimulate DSBs at the endogenous IL-2 receptor common γ -chain gene (*IL2RG*), mutations in which are responsible for the congenital primary immunodeficiency SCID-X1 (Shaw et al., 2011). (For more information on gene targeting for SCID-X1 see Chapter III of this dissertation.) To introduce precise sequence alterations at this locus, I designed a donor template with approximately 400 base pairs (bp) arms of homology 5' and 3' of the TALEN cut site (**Figure 2.1**). Within the 3' arm of homology I introduced seven point mutations that, upon successful HDR, would be stably integrated into the *IL2RG* gene and prevent binding and cleavage by the TALEN pair. To measure the frequency of mutagenic NHEJ and HDR at the endogenous *IL2RG* locus with this system, I developed a strategy based on SMRT DNA sequencing in collaboration with Ayal Hendel and Eli Fine, as described in the introduction to this chapter.

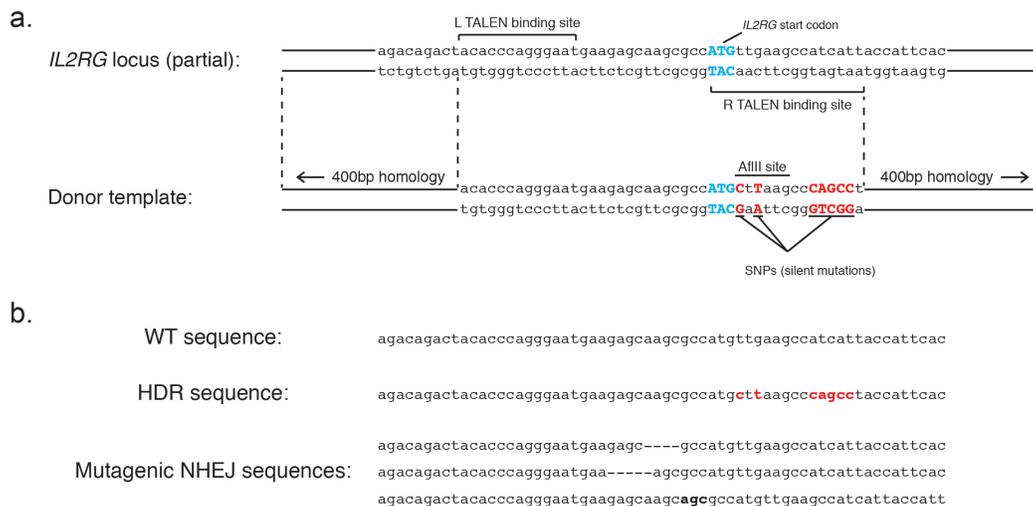


Figure 2.1: Gene targeting strategy for *IL2RG*. (a) The *IL2RG* TALENs are designed to bind a target site encompassing the start codon for *IL2RG*. When homology-directed repair (HDR) occurs between the endogenous *IL2RG* locus and the donor template, seven single-nucleotide polymorphisms (SNPs) are incorporated in *IL2RG* immediately following the start codon. These SNPs are silent mutations that provide a signature for measuring HDR frequency by SMRT sequencing. The SNPs also alter the right TALEN binding site and create a novel AflII site that can be analyzed by a restriction fragment length polymorphism (RFLP) assay. (b) Potential sequences for alleles in gene targeted populations. All alleles that have undergone HDR will have the same sequence with the seven indicated differences (red basepairs) from the WT sequence copied from the donor template. In contrast, mutagenic NHEJ will generate many different allelic sequences with different insertion and deletion mutations, three examples of which are shown. Dashes represent deleted bases; bold letters represent inserted bases.

Measurement of gene editing outcomes at IL2RG

To induce sequence alterations at *IL2RG* I first expressed the *IL2RG* TALENs from plasmid DNA in K562 cells with or without donor DNA and analyzed cell populations using the SMRT DNA sequencing method. Following gene targeting at *IL2RG*, genomic DNA from the bulk populations was collected and the *IL2RG* locus was amplified using a forward primer that is 5' and outside the start of the 5' homology arm and a reverse primer that is downstream of the TALEN pair target site (**Figure 2.2a, Figure 2.2b**). With this approach, non-integrated and randomly integrated donor templates are not amplified, removing common sources of background noise.

For cells transfected with TALENs alone, unmodified alleles and alleles with insertions or deletions indicative of mutagenic NHEJ were detected (**Figure 2.2c**). For cells transfected with both the TALENs and donor DNA unmodified alleles, alleles with insertions or deletions, and alleles with the seven point mutations precisely integrated into *IL2RG* by HDR were detected (**Figure 2.2c**). To validate the SMRT DNA sequencing analysis I first used a restriction fragment length polymorphism (RFLP) assay to measure the frequency of an AflII restriction site that is created when the seven point mutations within the donor template are incorporated into the target locus (**Figure 2.1a**). The AflII restriction site was detected at an average frequency of 14.3% by the RFLP assay compared to 16.8% by SMRT analysis (**Figure 2.3a**). The most commonly used methods for

determining the frequency of NHEJ measure any small insertion or deletion events, which is confounded by sequence alterations introduced by HDR. To confirm the actual frequency of alleles modified by NHEJ and HDR I grew single cell clones from a representative sample. Analysis of these clones with Sanger sequencing showed that 19.4% of alleles had undergone mutagenic NHEJ and 19.1% of alleles had been precisely modified by HDR, compared to 19.2% and 18.8% respectively as measured by SMRT sequencing analysis of the same population (**Figure 2.3b, Figure 2.3c**). Joseph Clark analyzed the single cell clones for this experiment. To confirm the reproducibility of SMRT sequencing analysis, a single targeted population was analyzed eight times and the standard deviations for NHEJ and HDR were found to be 0.66% and 0.79% respectively (**Figure 2.4**).

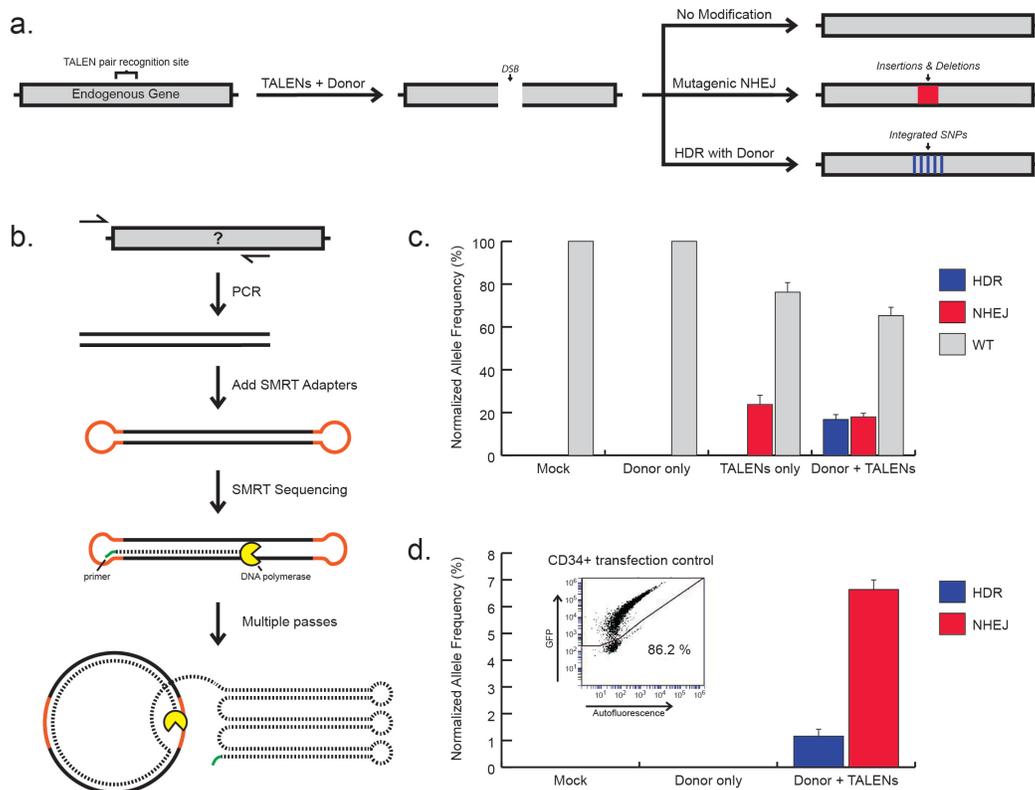


Figure 2.2: Measuring gene editing at an endogenous locus with SMRT sequencing. (a) Diagram of gene editing at an endogenous locus. TALENs create a double strand break (DSB), which can lead to no modification, insertion or deletion mutations, or integration of single nucleotide polymorphisms from a donor template (Figure 2.1). (b) Schematic of SMRT DNA sequencing analysis. The endogenous locus is amplified by PCR, with at least one primer outside the arms of homology of the donor template, and SMRT adapters are added to PCR amplicons. Individual DNA molecules are sequenced by SMRT sequencing, with read lengths averaging ~3kb in length and approaching ~15kb. (c) Measurement of gene editing outcomes at the *IL2RG* locus in K562 cells using the high-

expression TALEN plasmids. Modification frequencies are normalized to transfection efficiency. Data for graph is from Table 2.1. **(d)** Measurement of gene editing outcomes at *IL2RG* in CD34⁺ HSPCs using the high-expression TALEN plasmids. Insert shows transfection efficiency used for normalization. Data for graph is from Table 2.2. Bars represent three independent biological replicates; error bars, s.d.

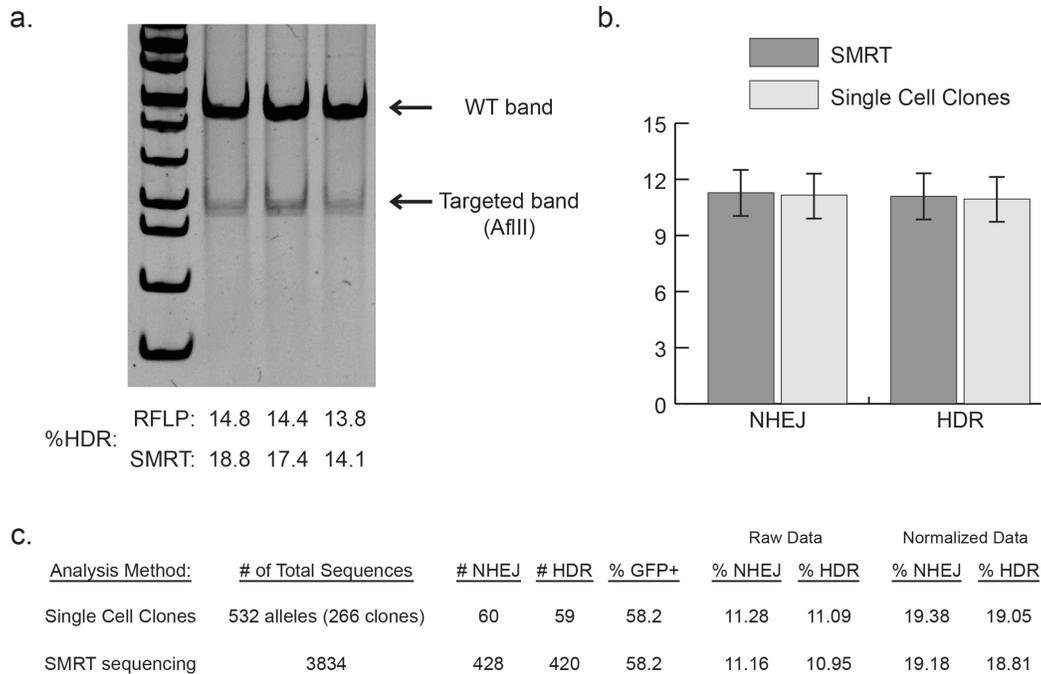


Figure 2.3: Comparison of SMRT sequencing to RFLP and single cell clone analysis. (a) RFLP analysis of K562 cells targeted with 1 μ g of each TALEN and 5 μ g donor in triplicate. The frequency of HDR in each sample as measured by RFLP and SMRT sequencing analysis is shown. (b) Quantification of NHEJ and HDR frequencies in single cell clones grown from a representative population of K562 cells. *IL2RG* alleles for each clone were amplified by PCR and analyzed with Sanger sequencing. Error bars represent 90% confidence intervals. (c) Analysis of alleles from single cell clones with Sanger sequencing showing raw and transfection normalized allelic modification frequencies compared to SMRT sequencing analysis.

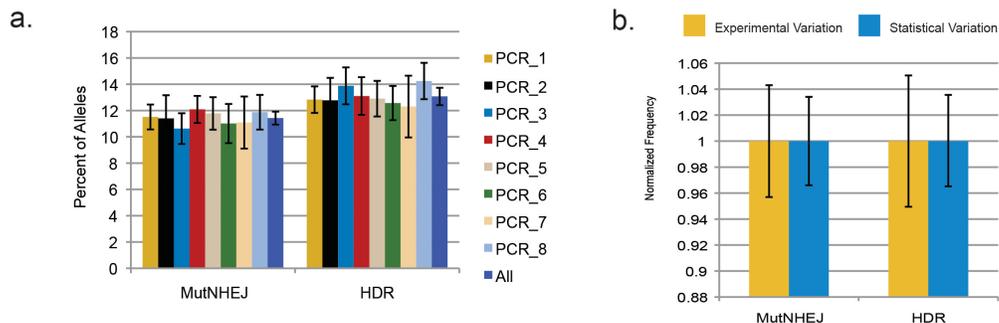


Figure 2.4: Reproducibility of SMRT sequencing for gene editing analysis.

(a) A representative sample of K562 cells was analyzed by SMRT sequencing 8 separate times to determine the variability introduced by PCR, SMRT library synthesis, and sequencing. Error bars for the individual PCRs represent standard deviation of the different SMRT sequencing runs. Error bars of the combined sample represent the standard deviation between the 8 PCRs. **(b)** Quantification of the observed experimental variation compared to the expected statistical variation for the number of sequences analyzed for the 8 replicates. Error bars for experimental variation represent the standard deviation between the 8 replicates. Error bars for the statistical variation represent 66% confidence intervals to directly compare to ± 1 standard deviation which covers 66% of the normal distribution.

Measurement of gene editing outcomes at a silent locus in human stem cells

When new gene editing tools are being developed it is important to determine optimal gene editing conditions based on the needs of a specific application, such as measuring relatively rare events in difficult to target cell types, maximizing the ratio of HDR to NHEJ, or determining the minimal effective dose to achieve a specific targeting frequency. With my gene editing tools I first tested my ability to measure gene editing events in CD34+ hematopoietic stem and progenitor cells (HSPCs) and a collaborator, Vittorio Sebastiano, Ph.D., tested our ability to measure gene targeting in human embryonic stem cells (hESCs), both of which are difficult to target, but important cell types for basic research and gene therapy.

After introducing TALENs and donor DNA into CD34+ HSPCs, SMRT sequencing analysis showed frequencies of mutagenic NHEJ and HDR of 6.6% and 1.2% respectively at the endogenous *IL2RG* locus (**Figure 2.2d**). In hESCs, which commonly require enrichment of targeted clones due to low gene editing efficiencies, addition of TALENs and donor DNA resulted in mutagenic NHEJ and HDR frequencies of 0.10% and 0.14% respectively (**Table 2.3**). Importantly, since *IL2RG* is silent in both of these cell types, these results demonstrate the ability of this approach to provide quantitative and sensitive measures of gene editing at a silent endogenous locus in important primary cells. Furthermore, the long read-lengths achieved by SMRT sequencing allowed for the measurement of

large insertion and deletion events (up to +334 bp and -355 bp in this study) and comparison of the frequency of these mutations in primary CD34+ HSPCs versus the K562 cell line (**Figure 2.5b**). HSPCs showed a strong peak in mutations ranging from -6 to -2 bp that was much more pronounced than in the K562 cells, and a corresponding drop in mutations ranging from -19 to -7 bp. Since the overall mutation rate in K562 cells was roughly two-fold higher than in HSPCs, one potential explanation is that, in K562 cells, TALENs are more frequently re-cleaving DNA that has been repaired with only small deletions (-6 to -2 bp) in the spacer region and thereby creating larger deletions that are more disruptive to the TALEN binding site and prevent further re-cleavage. TALENs using the +63 C-terminal truncation (as in this study) have been shown to be active over a wide range of spacers (~10-30 bp), supporting the theory that they are capable of re-cutting alleles that have already been disrupted. If this is the case, the use of more stringent C-terminal truncations, such as C+18 or C+28, may shift the peak mutation range in K562 cells towards that seen in HSPCs (Christian et al., 2012; Miller et al., 2011).

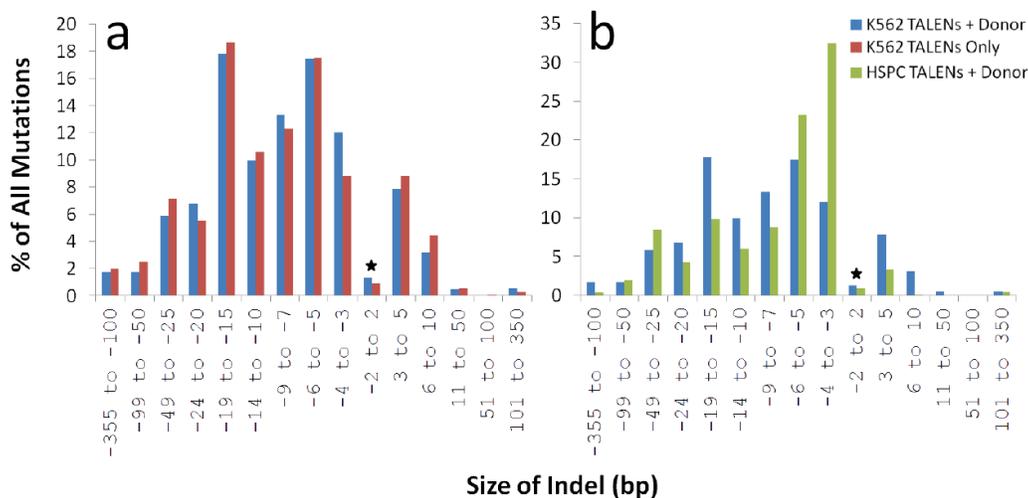


Figure 2.5: Mutation spectra caused by TALENs at *IL2RG* in the K562 cell line and primary CD34+ hematopoietic stem and progenitor cells. For all of the sequencing reads classified as mutagenic NHEJ, the length of the resulting insertion or deletion mutation (indel) was measured. For combined insertions and deletions in a single read, the overall change in the size of the fragment was measured (i.e. a 3 bp deletion and a 4 bp insertion would result in an overall +1 increase to the fragment size). **(a)** K562 cells were transfected with TALENs only or TALENs and donor DNA and the resulting mutation spectra were measured. **(b)** CD34+ HSPCs were treated with TALENs and donor DNA and the resulting mutation spectra is compared to that in K562 cells. (★) The analysis pipeline used in this study does not count 1 or 2 bp sequence alterations as indels in order to avoid false positives resulting from sequencing errors, and thus any 1 or 2 bp indels are not measured. The -2 to 2 bins therefore only represent complex indels combining insertions and deletions with a total change of ≤ 2 bp.

Optimization of gene targeting parameters at an endogenous locus

I next applied SMRT sequencing analysis to explore how varying the dose of TALENs and donor template DNA affects gene editing frequencies at the endogenous *IL2RG* gene in K562 cells. When I kept the amount of donor DNA constant and titrated down the amount of TALENs by 100-fold, I saw a progressive decrease in both mutagenic NHEJ and HDR events while their relative frequencies remained largely unchanged (**Figure 2.6a**). Using this approach, I was able to accurately detect gene editing outcomes at frequencies ranging from $> 20\%$ to $\leq 0.1\%$. When I held the amount of TALENs constant and titrated the amount of donor DNA, the overall level of modification at *IL2RG* was fairly constant while the ratio of HDR to NHEJ rose from 0.12 to 1.37 with increasing abundance of donor DNA (**Figure 2.6b, Figure 2.7**). Analysis of the mutation spectra at *IL2RG* in the presence and absence of donor DNA showed a negligible effect on the types of mutations created by NHEJ (**Figure 2.5a**).

To determine the effect of homology arm length on HDR efficiency with plasmid donors, I took advantage of the long read-lengths provided by SMRT sequencing to measure targeting efficiency with a range of homology arm lengths from 800 bp, a standard homology arm length used for gene targeting, to 100 bp (**Figure 2.6c**). Homology arms 100 bp or 200 bp in length were found to be significantly less effective for HDR, with 400 bp homology arms resulting in maximal levels of HDR. This suggests that plasmid donor templates with

homology arms less than 200 bp may suffer greatly reduced HDR efficiency, while maximal levels of HDR are achieved with relatively short 400 bp arms of homology.

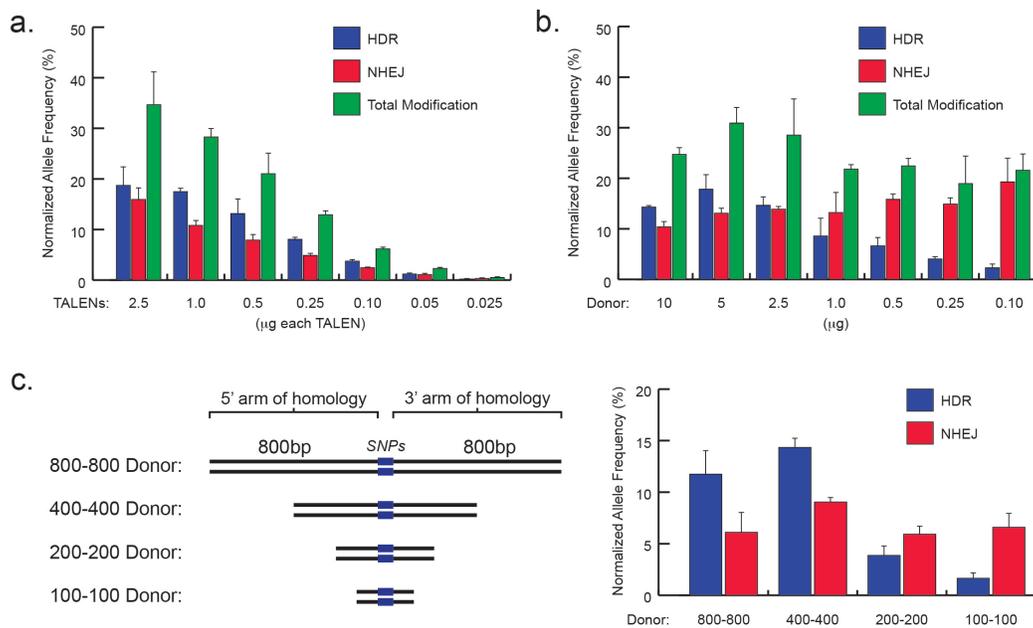


Figure 2.6: Interrogation of gene editing parameters at an endogenous locus with SMRT sequencing. (a) Titration of TALEN plasmid amount in K562 cells with donor DNA amount held constant at 5μg. Data for graph is from Table 2.4. (b) Titration of donor DNA amount in K562 cells with TALEN DNA amount held constant at 1μg for each TALEN. Data for graph is from Table 2.5. (c) Left: Schematic of donor templates with varying arm of homology lengths.

Right: Quantification of effect of homology arm length on gene editing frequencies in K562 cells. Modification frequencies are normalized to transfection efficiency. Data for graph is from Table 2.6. All experiments performed with standard TALEN expression plasmids. Bars represent three independent biological replicates; error bars, s.d.

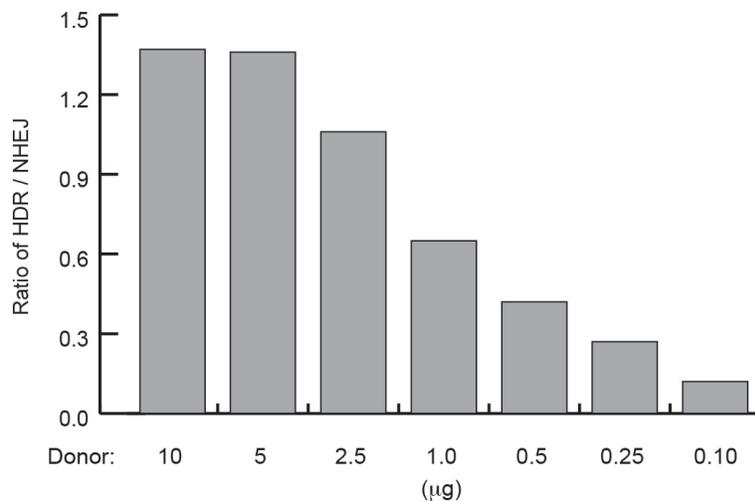


Figure 2.7: Ratio of HDR to NHEJ with varying amounts of donor DNA for samples in Figure 2.6b.

Table 2.1: Summary of SMRT sequencing data for Figure 2.2c – Targeting *IL2RG* in K562 cells

Treatment	Raw Data				Normalized Data					
	# of SMRT Sequencing Reads				%WT		%NHEJ		%HR	
	WT	NHEJ	HDR	%GFP+		AVG ± STD		AVG ± STD		AVG ± STD
Mock	6148	1	0	56.8	99.96	99.96	0.04	0.04	0.0	0.0
Donor only	5695	0	0	64.3	100.0	100.0	0.0	0.0	0.0	0.0
TALENs only	6261	1054	0	61.3	76.5	76.2 ± 4.4	23.5	23.8 ± 4.4	0.0	0.0
TALENs only	5701	1184	0	60.7	71.7		28.3		0.0	
TALENs only	8509	1001	0	54.1	80.5		19.5		0.0	
TALENs and Donor	2986	428	420	58.2	62.0	65.3 ± 4.0	19.2	18.0 ± 1.6	18.8	16.8 ± 2.4
TALENs and Donor	8707	1168	1097	57.6	64.1		18.5		17.4	
TALENs and Donor	5768	607	527	54.2	69.7		16.2		14.1	

Table 2.2: Summary of SMRT sequencing data for Figure 2.2d – Targeting *IL2RG* in CD34+ HSPCs

Treatment	Raw Data				Normalized Data					
	# of SMRT Sequencing Reads				%WT		%NHEJ		%HR	
	WT	NHEJ	HDR	%GFP+		AVG ± STD		AVG ± STD		AVG ± STD
Mock	10513	1	0	86.2	100.0	99.99	0.01	0.01	0.0	0.0
Donor only	8526	0	1	86.2	100.0	99.99 ± 0.01	0.0	0.00	0.01	0.004 ± 0.01
Donor only	8111	0	0	86.2	100.0		0.0		0.0	
Donor only	8373	0	0	86.2	100.0		0.0		0.0	
TALENs and Donor	7687	497	99	86.2	91.6	92.2 ± 0.62	7.0	6.6 ± 0.4	1.4	1.2 ± 0.3
TALENs and Donor	8461	525	94	86.2	92.1		6.7		1.2	
TALENs and Donor	8478	487	68	86.2	92.9		6.3		0.9	

Table 2.3: Summary of SMRT sequencing data for gene targeting in hESCs

	Raw Data			Normalized Data					
	# of SMRT Sequencing Reads			%WT		%NHEJ		%HR	
Treatment	WT	NHEJ	HDR		AVG ± STD		AVG ± STD		AVG ± STD
Donor only	6048	2	0	99.97	99.98 ± 0.01	0.03	0.02 ± 0.01	0	0.0
Donor only	9833	1	0	99.99		0.01		0	
TALENs and Donor	5892	5	7	99.8	99.76 ± 0.07	0.08	0.10 ± 0.06	0.12	0.14 ± 0.02
TALENs and Donor	5818	3	9	99.79		0.05		0.15	
TALENs and Donor	9307	16	14	99.68		0.17		0.15	

Table 2.4: Summary of SMRT sequencing data for Figure 2.6a – Titration of TALEN amounts

TALEN amount (each TALEN)	Raw Data				Normalized Data					
	# of SMRT Sequencing Reads				%Total Modification	%NHEJ		%HR		
	WT	NHEJ	HDR	%GFP+		AVG ± STD		AVG ± STD		AVG ± STD
2.5µg	2192	250	269	68.2	28.06	34.65 ± 6.09	13.52	15.94 ± 2.44	14.55	18.71 ± 3.71
2.5µg	1854	338	424	81.3	35.83		15.89		19.94	
2.5µg	1592	335	394	78.4	40.06		18.41		21.66	
1µg	1623	201	334	86.7	28.59	28.29 ± 1.49	10.74	10.81 ± 0.82	17.85	17.48 ± 0.73
1µg	2200	270	416	80.3	29.60		11.66		17.95	
1µg	1371	138	229	79.2	26.67		10.03		16.64	
0.5µg	3866	283	449	83.9	18.97	21.03 ± 4.09	7.33	7.90 ± 1.28	11.64	13.13 ± 2.81
0.5µg	3564	422	738	95.4	25.73		9.36		16.37	
0.5µg	4152	301	488	86.9	18.38		7.01		11.37	
0.25µg	3431	160	249	79.3	13.43	12.92 ± 0.85	5.26	4.87 ± 0.50	8.17	8.05 ± 0.37
0.25µg	4576	228	377	87.2	13.39		5.05		8.35	
0.25µg	2434	95	168	81.6	11.95		4.31		7.63	
0.10µg	5087	107	158	83.4	5.94	6.17 ± 0.30	2.40	2.44 ± 0.09	3.54	3.73 ± 0.34
0.10µg	3373	83	115	91.2	6.07		2.54		3.53	
0.10µg	4000	95	164	93.4	6.51		2.39		4.12	
0.05µg	6736	54	85	88.5	2.29	2.27 ± 0.16	0.89	1.07 ± 0.19	1.40	1.20 ± 0.18
0.05µg	7499	76	77	95.6	2.09		1.04		1.06	
0.05µg	5000	62	56	95.3	2.41		1.27		1.14	
0.025µg	7062	17	19	94.5	0.54	0.50 ± 0.13	0.25	0.30 ± 0.15	0.29	0.20 ± 0.08
0.025µg	4427	19	6	92	0.61		0.47		0.14	
0.025µg	4101	6	6	86	0.35		0.17		0.17	

Table 2.5: Summary of SMRT sequencing data for Figure 2.6b – Titration of Donor template amount

Donor amount	Raw Data				Normalized Data					
	# of SMRT Sequencing Reads			%GFP+	%Total Modification		%NHEJ		%HR	
	WT	NHEJ	HDR		AVG ± STD	AVG ± STD	AVG ± STD	AVG ± STD		
10µg	7392	773	1033	74.3	26.42	24.72 ± 1.78	11.31	10.41 ± 0.81	15.11	14.31 ± 1.03
10µg	6661	622	895	74.6	24.87		10.20		14.66	
10µg	4188	347	469	71.3	22.86		9.72		13.14	
5µg	2694	328	463	76	29.87	30.92 ± 1.01	12.38	13.07 ± 0.60	17.49	17.84 ± 0.50
5µg	5748	835	1143	80.3	31.88		13.46		18.42	
5µg	4115	579	763	79.3	31.01		13.38		17.63	
2.5µg	4484	742	821	82.7	31.26	28.53 ± 2.49	14.84	13.88 ± 0.93	16.42	14.65 ± 1.58
2.5µg	6087	783	809	78.6	26.39		12.98		13.41	
2.5µg	7212	953	972	75.4	27.94		13.83		14.11	
1µg	5813	590	347	73.6	18.86	21.80 ± 6.17	11.88	13.21 ± 4.00	6.98	8.59 ± 2.27
1µg	2594	185	140	63.1	17.65		10.05		7.61	
1µg	12234	2619	1656	89.6	28.90		17.70		11.19	
0.5µg	9574	948	463	70.1	18.32	22.44 ± 2.54	12.31	15.82 ± 3.05	6.01	6.62 ± 0.58
0.5µg	12611	2244	846	80.5	24.45		17.75		6.70	
0.5µg	14814	2713	1117	83.6	24.57		17.40		7.17	
0.25µg	18675	2582	755	75.4	20.11	18.94 ± 1.01	15.56	14.89 ± 0.58	4.55	4.05 ± 0.46
0.25µg	11605	1362	338	69.9	18.28		14.65		3.63	
0.25µg	10646	1303	358	73.2	18.44		14.47		3.98	
0.1µg	14299	3663	493	85.9	26.22	21.58 ± 4.49	23.11	19.28 ± 3.77	3.11	2.30 ± 0.73
0.1µg	12091	2050	229	74.5	21.29		19.15		2.13	
0.1µg	13246	1694	182	71.9	17.25		15.58		1.67	

Table 2.6: Summary of SMRT sequencing data for Figure 2.6c – Homology arm length comparison

Homology Arm Lengths	Raw Data			%GFP+	Normalized Data			
	# of SMRT Sequencing Reads				%NHEJ		%HR	
	WT	NHEJ	HDR			AVG ± STD		AVG ± STD
800 - 800	4731	169	364	64.4	5.0	6.10 ± 1.60	10.7	11.74 ± 2.37
800 - 800	4515	179	333	66	5.4		10.0	
800 - 800	4283	293	533	72.2	7.9		14.4	
400 - 400	9224	749	1273	76.9	8.7	9.05 ± 0.44	14.7	14.34 ± 0.99
400 - 400	7241	676	1070	78.9	9.5		15.1	
400 - 400	7906	578	851	69	9.0		13.2	
200 - 200	3911	133	101	61.7	5.2	5.93 ± 0.83	4.0	3.86 ± 0.94
200 - 200	3549	106	53	49.7	5.8		2.9	
200 - 200	4155	206	143	66.8	6.8		4.7	
100 - 100	8203	430	120	64.3	7.6	6.59 ± 1.32	2.1	1.65 ± 0.53
100 - 100	7953	253	54	59.9	5.1		1.1	
100 - 100	7199	325	80	60.8	7.0		1.7	

Discussion

The recent explosion in custom gene editing technologies is ushering in a new age of genome engineering where scientists across fields of study and using different organisms and cell types can precisely modify essentially any locus they desire. Here I have shown that SMRT DNA sequencing provides a rapid, quantitative, and sensitive strategy for measuring genome editing outcomes at any endogenous locus, including transcriptionally silent loci. In addition to initial characterization of novel nucleases and targeting constructs, I show that this method can be used to test and optimize gene targeting conditions for a specific cell type of interest. Without the need to generate reporter cell lines, this method further allows gene editing frequencies to be measured in human primary cells and other difficult to culture cell types. As the types of achievable genomic manipulations become more and more complex, this method will help scientists accurately measure the frequency of genomic alterations in different model organisms and at different target loci of their choosing. By providing this flexibility, the SMRT DNA sequencing strategy presented here can streamline the development of genome editing projects and hasten the expansion of this technology to a wider range of applications.

Materials and Methods

Construct assembly:

IL2RG TALE repeats were synthesized (Genscript) and cloned into a TALEN expression vector based on pcDNA3.1 (Invitrogen) with the $\Delta 152$ N-terminal domain and the +63 C-terminal domain fused to the FokI nuclease domain as previously described (Miller et al., 2011). The *IL2RG* targeting vectors containing the seven point mutations to prevent binding by the *IL2RG* TALEN pair were synthesized by PCR amplifying the *IL2RG* locus from genomic DNA isolated from K562 cells using the primers in Table 2.7.

Cell culture:

K562 cells (ATCC) were maintained in RPMI 1640 (Hyclone) supplemented with 10% bovine growth serum, 100 units/ml penicillin, 100 μ g/ml streptomycin and 2mM L-glutamine. Human CD34⁺ hematopoietic stem and progenitor cells (HSPCs) were purchased from Lonza (2M-101B) and thawed per the manufacturer's instructions. CD34⁺ HSPCs were maintained in X-VIVO15 (Lonza) supplemented with SCF (100ng/ml), TPO (100ng/ml), Flt3-Ligand (100ng/ml), IL-6 (100ng/ml), and StemRegenin1 (0.75 μ M). hESC line H1 (WiCell) was maintained in feeder-free culture conditions in mTeSR1 (Stem Cell Technologies) on a thin layer of Matrigel (BD). Cultures were passaged every 3-5

days enzymatically with Accutase (Innovative Cell technologies). Cells were transfected between passage 45 and 47.

Transient transfection for genome editing:

1 x 10⁶ K562 cells were transfected with 2 µg TALEN-encoding plasmid and 5 µg donor plasmid (unless otherwise indicated) by nucleofection (Lonza) using program T-016 and a nucleofection buffer containing 100mM KH₂PO₄, 15mM NaHCO₃, 12mM MgCl₂ • 6H₂O, 8mM ATP, 2mM glucose, pH 7.4. 4 x 10⁵ CD34⁺ HSPCs were nucleofected with an Amaxa 4D Nucleofector with the P3 Primary Cell Nucleofector Kit (V4XP-3032) and program EO-100 per the manufacturer's instructions. 1 x 10⁶ H1 cells were transfected with 0.5 µg or 2.5 µg of each TALEN-encoding plasmid and 4 µg donor plasmid (unless otherwise indicated) by nucleofection using an Amaxa 4D Nucleofector with the P3 Primary Cell Nucleofector Kit (V4XP-3032) and program B-105 following manufacturer's instructions.

Flow cytometry:

Samples were collected 72h post-nucleofection and analyzed for fluorescence using an Accuri C6 flow cytometer. GFP expression was measured using a 488-nm laser for excitation and a 530/30 bandpass filter for detection.

Restriction fragment length polymorphism assay:

Restriction fragment length polymorphism assay was performed as previously described in Chen, F. *et al.* (Chen et al., 2011). Briefly, genomic DNA was extracted from transfected cells with the DNeasy Blood & Tissue Kit (Qiagen). Genomic DNA was then PCR amplified with primers flanking the donor target region (see Table 2.7 for PCR primer sequences). The amplification was carried out with Accuprime polymerase (Invitrogen), using the following cycling conditions: 95 °C for 5 min for initial denaturation; 30 cycles of 95 °C for 30 s, 67 °C for 45 s and 68 °C for 120 s; and a final extension at 68 °C for 5 min. PCR products were digested with 20 U of AflIII at 37 °C for ~ 2h and resolved with PAGE.

Single cell clone analysis:

Single-cell sorting was performed by flow cytometry on a BD FACS ARIA. Genomic DNA was isolated from single clones using the DNeasy Blood & Tissue Kit (Qiagen). The *IL2RG* target region was amplified using the primers in Table 2.7 with Accuprime polymerase (Invitrogen) and the following cycling conditions: 95 °C for 5 min for initial denaturation; 30 cycles of 95 °C for 30 s, 67 °C for 45 s and 68 °C for 120 s; and a final extension at 68 °C for 5 min. PCR amplicons were sequenced using standard Sanger sequencing. Sequences were analyzed using the ApE plasmid editor by M. Wayne Davis.

SMRT sequencing:

Genomic DNA containing *IL2RG* alleles was harvested from cultured K562, CD34+ HSPC, and hESC samples using the DNeasy Blood & Tissue Kit (Qiagen). *IL2RG* alleles were amplified using the primers in Table 2.7 with Accuprime polymerase (Invitrogen) and the following cycling conditions: 95 °C for 5 min for initial denaturation; 30 cycles of 95 °C for 30 s, 67 °C for 45 s and 68 °C for 60 s; and a final extension at 68 °C for 5 min for the K562 samples and 95 °C for 5 min for initial denaturation; 30 cycles of 95 °C for 30 s, 67 °C for 45 s and 68 °C for 90 s; and a final extension at 68 °C for 5 min for the HSPC and hESC samples. Sequencing libraries were constructed, as previously described (Travers et al., 2010), using the DNA Template Prep Kit 1.0 (Pacific Biosciences, Menlo Park, CA). SMRTbell libraries contained amplicons that were pooled together, with different barcodes appended to allow multiplex analysis. Purified, closed circular SMRTbell libraries were annealed with a sequencing primer complementary to a portion of the single-stranded region of the hairpin. For all SMRTbell libraries, annealing was performed at a final template concentration between 30 and 60 nM, with a 20-fold molar excess of sequencing primer. All annealing reactions were carried out at 80°C for 2 min, with a slow cool to 25°C at a rate of 0.1°C/second. Annealed templates were stored at -20°C until polymerase binding. DNA polymerase enzymes were stably bound to the primed sites of the annealed SMRTbell templates using the DNA Polymerase Binding Kit

2.0 (Pacific Biosciences). SMRTbell templates (3 nM) were incubated with 6 nM of polymerase in the presence of phospholinked nucleotides at 30°C for 2 h. Following incubation, samples were stored at 4°C. Sequencing was performed within 72 h of binding using final on plate concentration of 0.3 nM. Each sample was sequenced as previously described (Rasko et al., 2011) using DNA Sequencing Kit 2.0 (Pacific Biosciences). Sequencing data collection was performed on the PacBio® RS (Pacific Biosciences) using movies of 55 min in each case.

SMRT Analysis Pipeline:

The SMRT Sequencing Analysis pipeline was developed by Eli Fine in Strawberry Perl and utilizes the NCBI BLAST software as well as the mEmboss Needleman-Wunsch pairwise alignment algorithm. All components of the pipeline were run on a standard Windows PC and are available for download (<https://sourceforge.net/projects/tdna-getsmart/>).

Statistical analysis:

To calculate confidence intervals, t-statistics were calculated as previously described (Pattanayak et al., 2011). 90% confidence intervals were calculated by determining the upper and lower bounds of the mutation rates that would yield P

values of 0.05. 66% confidence intervals were calculated similarly using a target *P* value of 0.32.

Table 2.7: List of primers used in Chapter II

Primer name	Primer sequence (5'-3')	Note
<i>IL2RG</i> – F1	TGCACGATCGCTCCCAAAACAGTAGAGCTTTGACAGAGATTTAAG	Used with K562 samples
<i>IL2RG</i> – R1	GATCTCGCATCGTGTTTCAGCCCCACTCCCAGC	Used with K562 samples
<i>IL2RG</i> – F2	ACTAGAGCGACCTCCCAAAACAGTAGAGCTTTGACAGAGATTTAAG	Used with K562 samples
<i>IL2RG</i> – R2	GTCACGTCTGCGTGTTTCAGCCCCACTCCCAGC	Used with K562 samples
<i>IL2RG</i> – F3	ATCTGACAGTCCTCCCAAAACAGTAGAGCTTTGACAGAGATTTAAG	Used with K562 samples
<i>IL2RG</i> – R3	CAGAGTGTCACGTGTTTCAGCCCCACTCCCAGC	Used with K562 samples
<i>IL2RG</i> – F4	ACGTGTAGTCTCTCCCAAAACAGTAGAGCTTTGACAGAGATTTAAG	Used with K562 samples
<i>IL2RG</i> – R4	TCTACTGACACGTGTTTCAGCCCCACTCCCAGC	Used with K562 samples
<i>IL2RG</i> – F5	CGCACTCGATCCTCCCAAAACAGTAGAGCTTTGACAGAGATTTAAG	Used with K562 samples
<i>IL2RG</i> – R5	TGATGTAGATCGTGTTTCAGCCCCACTCCCAGC	Used with K562 samples
<i>IL2RG</i> – F6	TCGCTATGTGCCTCCCAAAACAGTAGAGCTTTGACAGAGATTTAAG	Used with K562 samples
<i>IL2RG</i> – R6	TGTACTATATCGTGTTTCAGCCCCACTCCCAGC	Used with K562 samples
<i>IL2RG</i> – F7	TCTCTGTACTCCTCCCAAAACAGTAGAGCTTTGACAGAGATTTAAG	Used with K562 samples
<i>IL2RG</i> – R7	ATCATATGATCGTGTTTCAGCCCCACTCCCAGC	Used with K562 samples
<i>IL2RG</i> – F8	ATGACTGCGACCTCCCAAAACAGTAGAGCTTTGACAGAGATTTAAG	Used with K562 samples
<i>IL2RG</i> – R8	CAGTCTCACACGTGTTTCAGCCCCACTCCCAGC	Used with K562 samples
<i>IL2RG</i> – F9	AGCTCGAGACGCAGAGTGGCTGTGGTAATGGAAAGGAGGAAAC	Used with HSPC, and hESC samples
<i>IL2RG</i> – R9	CTGAGAGTAGCGTGTTTCAGCCCCACTCCCAGC	Used with HSPC, and hESC samples
<i>IL2RG</i> – F10	ACTAGAGCGAGCAGAGTGGCTGTGGTAATGGAAAGGAGGAAAC	Used with HSPC, and hESC samples
<i>IL2RG</i> – R10	GTCACGTCTGCGTGTTTCAGCCCCACTCCCAGC	Used with HSPC, and hESC samples
<i>IL2RG</i> – F11	TGAGCACGTAGCAGAGTGGCTGTGGTAATGGAAAGGAGGAAAC	Used with HSPC, and hESC samples
<i>IL2RG</i> – R11	CAGACTCGTACGTGTTTCAGCCCCACTCCCAGC	Used with HSPC, and hESC samples
<i>IL2RG</i> – F12	ACGTGTAGCTGCAGAGTGGCTGTGGTAATGGAAAGGAGGAAAC	Used with HSPC, and hESC samples
<i>IL2RG</i> – R12	TCTACTGACACGTGTTTCAGCCCCACTCCCAGC	Used with HSPC, and hESC samples
<i>IL2RG</i> – F13	TCTCTGTACTGCAGAGTGGCTGTGGTAATGGAAAGGAGGAAAC	Used with HSPC, and hESC samples
<i>IL2RG</i> – R13	ATCATATGATCGTGTTTCAGCCCCACTCCCAGC	Used with HSPC, and hESC samples
<i>IL2RG</i> – F14	AGCATCAGACGCAGAGTGGCTGTGGTAATGGAAAGGAGGAAAC	Used with HSPC, and hESC samples
<i>IL2RG</i> – R14	GTATGTCTCACGTGTTTCAGCCCCACTCCCAGC	Used with HSPC, and hESC samples
<i>IL2RG</i> – F15	ACGTCTCGTGCAGAGTGGCTGTGGTAATGGAAAGGAGGAAAC	Used with HSPC, and hESC samples
<i>IL2RG</i> – R15	CTGTGATACTCGTGTTTCAGCCCCACTCCCAGC	Used with HSPC, and hESC samples
<i>IL2RG</i> – F16	TACACGCACTGCAGAGTGGCTGTGGTAATGGAAAGGAGGAAAC	Used with HSPC, and hESC samples

<i>IL2RG</i> – R16	AGATCGATGTCGTGTTTCAGCCCCACTCCCAGC	Used with HSPC, and hESC samples
<i>IL2RG</i> – F17	CTAACTATAACGGTCCTAAGGTAGCGATTAATTAAGACCAAGTCAAGGAAGAGGCATGG	Used for 400-400 donor template
<i>IL2RG</i> – R17	GCTTAAGCATGGCGCTTGCTCTTCATTCC	Used for all <i>IL2RG</i> donor templates
<i>IL2RG</i> – F18	CATGCTTAAGCCCAGCCTACCAT	Used for all <i>IL2RG</i> donor templates
<i>IL2RG</i> – R18	CTGGAGCTCGGCCGGCCGGCTCATGTCTGTAATCCTGGTGC	Used for 400-400 donor template
<i>IL2RG</i> – F19	GGAAATAACTATAACGGTCCTAAGGTAGCGATTAATTAATGGGAGAAACACCACAGAAGC	Used for 800-800 donor template
<i>IL2RG</i> – R19	CTGGAGCTCGGCCGGCCCCGACTTATGACTTACCCAGGAG	Used for 800-800 donor template
<i>IL2RG</i> – F20	ACCTAACTATAACGGTCCTAAGGTAGCGATTAATTAAGTTTACCACCTACA GCAGCACC	Used for 200-200 donor template
<i>IL2RG</i> – R20	CTGGAGCTCGGCCGGCCGCCACATGATTGTAATGGCCAGTG	Used for 200-200 donor template
<i>IL2RG</i> – F21	CTAACTATAACGGTCCTAAGGTAGCGATTAATTAATCTTCCACCGGAAGCTATGACAG	Used for 100-100 donor template
<i>IL2RG</i> – R21	CTGGAGCTCGGCCGGCCACCAGCTGTGGTGTCTTCATTCC	Used for 100-100 donor template
<i>IL2RG</i> – F22	GCAGAGTGGCTGTGGTAATGGAAAGGAGGAAAC	Used for RFLP assay and single cell clone analysis
<i>IL2RG</i> – R22	CTGGGTACTGCAGATATCCAGAGCCTAGCCTC	Used for RFLP assay and single cell clone analysis

CHAPTER III:
DEVELOPMENT OF A NOVEL GENE THERAPY
STRATEGY FOR SCID-X1

Abstract

Gene therapy trials for SCID-X1 have demonstrated clinical efficacy in a majority of patients, providing a viable option for patients without a suitable donor for bone marrow transplantation. Here I describe a novel gene therapy strategy for SCID-X1 that utilizes precise integration of a functional *IL2RG* cDNA through gene targeting with TALENs. Using highly active TALENs targeting *IL2RG* exon 1, I show that insertion of cDNA under the control of the endogenous promoter produces IL2R γ activity levels on par with WT cells. These gene targeting reagents are able to stimulate significant levels of precise *IL2RG* cDNA gene addition in CD34⁺ HSPCs, the critical cell type for clinical translation of this approach. By utilizing precise gene targeting at *IL2RG*, the gene therapy strategy described here has the potential to correct the functional deficit in almost all SCID-X1 patients without the risk of insertional oncogenesis from pseudo-random viral vector integration.

Introduction

X-Linked Severe Combined Immunodeficiency (SCID-X1) is a genetic disorder caused by mutations in the gene encoding interleukin 2 receptor gamma chain (IL2R γ), which forms part of the receptor for interleukins IL-2, IL-4, IL-7, IL-9, IL-15, & IL-21 (Mukherjee et al., 2013; Shaw et al., 2011). A non-functional *IL2RG* gene results in extensive defects in interleukin signaling that cripple the ability of lymphocytes to differentiate into functional T-cells, B-cells, and natural killer cells, resulting in a devastating lack of an adaptive immune system. Without successful bone marrow transplantation patients usually die in the first year of life as a result of severe infections (Hacein-Bey-Abina et al., 2002).

While allogeneic bone marrow transplantation of hematopoietic stem cells (HSCs) from an HLA-identical donor has a high success rate, only about one-third of patients have an available matched donor (Qasim et al., 2009). The past two decades have seen a revolution in the way primary immunodeficiencies (PIDs), including SCID-X1, are approached clinically as gene therapy trials have sought, for the first time, to directly correct the genetic defects that underlie these diseases (Aiuti et al., 2002; Blaese et al., 1995; Bordignon et al., 1995; Boztug et al., 2010; Chinen et al., 2007; Gaspar et al., 2004; Hacein-Bey-Abina et al., 2002; Kang et al., 2010; Ott et al., 2006). The goal of gene therapy is to correct the

functional deficit caused by a genetic mutation in a patient's own cells by providing a functional copy of the mutated gene. The earliest gene therapy trials for SCID-X1 in the 1990s proved that addition of a functional copy of *IL2RG*, in this case delivered as a γ -retrovirus, provides clinical benefit to patients, with 17 of 19 patients experiencing improvement in lymphopoiesis and immune function (Gaspar et al., 2004; Hacein-Bey-Abina et al., 2002). Despite their clinical success, these trials were halted when five of the patients developed leukemia secondary to insertion of the γ -retroviral vector near proto-oncogenes. In addition to these SCID-X1 trials, the danger of insertional oncogenesis was seen in multiple other early gene therapy trials including those for X-CGD and WAS (Boztug et al., 2010; Gaspar et al., 2011; Hacein-Bey-Abina et al., 2008; Hacein-Bey-Abina et al., 2003; Howe et al., 2008; Stein et al., 2010).

In response to the tumorigenic risks associated with pseudo-random viral insertion, gene targeting emerged as an exciting alternative for gene therapy where a disease-causing mutation could be corrected directly without altering the rest of the genome (Kim et al., 1996; Porteus et al., 2003; Urnov et al., 2005). Gene targeting relies on engineered nucleases to create double-strand breaks at specific sites in the genome, which can be repaired with exogenous donor templates to precisely introduce new genetic sequences. To realize this potential, the first pair of zinc-finger nucleases successfully made to target an endogenous human gene were reported for *IL2RG* in 2005 (Urnov et al., 2005). These ZFNs

target a mutational hot-spot for SCID-X1 in exon 5 of *IL2RG*, which creates the possibility of correcting many SCID-X1-causing mutations directly. For clinical translation, correction of *IL2RG* needs to be accomplished in a patient's hematopoietic stem and progenitor cells, and the potential of these *IL2RG* ZFNs to stimulate gene targeting in primary cells was demonstrated by Urnov et al. and Lombardo et al., who showed gene targeting frequencies of 5% and 0.1% in primary T cells and CD34+ HSPCs respectively (Lombardo et al., 2007; Urnov et al., 2005). In addition to the *IL2RG* ZFNs, a pair of ZFNs developed to target the *CCR5* gene has been shown to stimulate targeted gene disruption in T cells and CD34+ HSPCs. Disruption of *CCR5* by these ZFNs knocks out expression of the *CCR5* co-receptor used during infection by HIV, creating a population of T cells that are resistant to infection by *CCR5*-tropic HIV (Holt et al., 2010; Perez et al., 2008).

While targeted gene disruption of *CCR5* has advanced to phase I clinical trials, the use of engineered nucleases to stimulate therapeutically-relevant levels of gene targeting in CD34+ HSPCs has proven more difficult. Two critical obstacles for the translation of this technology to gene targeting-based clinical trials have been the relative toxicity of highly-active ZFNs in primary cells and the difficulty of obtaining high-frequencies of gene targeting in primary cells. To overcome these obstacles to gene targeting in human CD34+ HSPCs I first sought to develop low-toxicity ZFNs to *IL2RG* utilizing the OPEN platform (Pruett-

Miller et al., 2008). Following the development of a new class of engineered nucleases, TALENs, in 2010 and lackluster results with my *IL2RG* ZFNs, I next developed a novel gene therapy strategy for SCID-X1 utilizing TALENs targeting exon 1 of the *IL2RG* gene.

Phenotypic correction of SCID-X1 with gene targeting can be achieved through three different strategies: (1) direct correction of a disease-causing mutation, (2) gene addition of full *IL2RG* cDNA to a safe-harbor locus, or (3) full or partial *IL2RG* cDNA addition within the endogenous *IL2RG* gene. Strategy 1 is perhaps the most elegant as it requires minimal alteration to the genome and restores full endogenous expression of IL2R γ , but clinical translation would be hampered by the requirement for different sets of nucleases and donor templates capable of targeting the many disease-causing mutations throughout *IL2RG*, each of which would likely have different efficacy and toxicity profiles. Strategy 2 offers a general approach that could be used for all SCID-X1 patients, but sacrifices endogenous regulation of IL2R γ expression by targeting the gene to a different genomic locus. Strategy 3 offers a compromise approach where gene addition of cDNA within *IL2RG* can phenotypically correct all downstream mutations, while still driving IL2R γ expression with its endogenous promoter. It is important to note that while this strategy uses the endogenous *IL2RG* promoter, regulation of expression by genetic elements downstream of the gene addition site, including introns, can be compromised.

Here I report my initial efforts to develop low-toxicity ZFNs for SCID-X1 followed by the development of TALENs to *IL2RG* exon 1. These TALENs are shown to be capable of high-efficiency gene targeting of *IL2RG* cDNA in cell lines and primary CD34+ HSPCs with lower toxicity than commonly-used ZFNs. Furthermore, I demonstrate that gene addition of codon-optimized *IL2RG* cDNA with an artificial intron provides physiologically-relevant levels of IL2R γ activity. By targeting full *IL2RG* cDNA to *IL2RG* exon 1, this strategy has the potential to restore endogenously-regulated IL2R γ expression in over 98% of patients suffering from SCID-X1, including patients with missense, nonsense, or frameshift mutations in exons 1-8 or RNA processing mutations in *IL2RG* introns.

Results

Development of OPEN ZFNs targeting IL2RG exon 5

The emergence of zinc-finger nucleases as tools for targeted gene editing of human genes led to the first successes in the gene targeting field (Bibikova et al., 2003; Kim et al., 1996; Porteus et al., 2003). While the potential for ZFNs to be used in clinical applications has been readily apparent, translation of the technology has not progressed rapidly, with only one pair of ZFNs advancing to clinical trials after a decade of development (Cannon et al., 2011). Two challenges that have slowed the path to clinical application for ZFNs are the

difficulty of making highly active ZFN pairs and the toxicity of these nucleases in human cells (Ramirez et al., 2008; Sander et al., 2011). In order to engineer highly active ZFNs to novel target sites, researchers have principally used either a ‘modular assembly’ approach where individual zinc-fingers with well-characterized specificities are joined together or context-dependent approaches where zinc-fingers are selected in the context of the rest of the zinc-finger array for the desired target site (Beerli et al., 2002; Greisman et al., 1997; Hurt et al., 2003; Isalan et al., 1997; Segal et al., 2003; Urnov et al., 2005). While modular assembly is a straight-forward and rapid approach to generate zinc-finger arrays, the overall success rate of this technique is low and the high binding affinities of the individual zinc-fingers regardless of adjacent zinc-fingers can create highly toxic ZFNs (Cornu et al., 2008; Pruetz-Miller et al., 2008).

To develop low-toxicity ZFNs for *IL2RG* we utilized the Oligomerized Pool Engineering (OPEN) strategy, which has been shown to produce lower toxicity ZFNs with a higher success rate than modular assembly approaches (Maeder et al., 2008; Pruetz-Miller et al., 2008), to design novel ZFNs targeting *IL2RG* exon 5 (**Figure 3.1**). The OPEN selections for these ZFNs were performed by Jiuli Zhang. Following selection of potential zinc-finger protein (ZFP) arrays for each target half-site, I synthesized ZFNs and tested the ability of these ZFN pairs to recognize and cut the target DNA sequence using the single-strand annealing assay (**Figure 3.2a**). Multiple pairs of ZFNs had high levels of activity

for this target sequence, including pairs that showed higher levels of activity in this assay than previously reported *IL2RG* ZFNs designed by modular assembly (Urnov et al., 2005) (**Figure 3.2b**). To confirm the ability of these ZFNs to stimulate DSBs and gene editing at the endogenous *IL2RG* gene I next tested these ZFNs using a gene targeting assay (**Figure 3.2c**). When the *IL2RG* ZFNs were introduced into K562 cells with a donor template harboring a Ubc-eGFP expression cassette, the ZFNs stimulated increased integration of the donor sequence at frequencies from 1-4%. While these results were initially encouraging, there was a high degree of variability in the efficiency with which these ZFNs stimulated targeted gene addition, with no increase in gene addition relative to donor only controls being seen in multiple experiments.

To confirm that these ZFNs were reproducibly creating targeted DSBs at *IL2RG* exon 5 I used the Cell assay, which allows for more direct measurement of nuclease activity by determining the percentage of alleles that have been disrupted by mutagenic NHEJ. Following expression of the OPEN ZFNs in K562 cells, 3-8% of endogenous *IL2RG* alleles were mutated compared to 14% by the previously reported modular assembly (MA) ZFNs (**Figure 3.3a**). These Cell assay results mirrored the gene addition results by showing increased activity of the MA ZFNs compared to the OPEN ZFNs, in contrast to the SSA assay results which showed higher levels of activity with the OPEN ZFNs (**Figure 3.3b**). One possible explanation for this disparity is that the relative abundance of the

exogenous SSA reporter plasmid in the SSA assay is less sensitive to ZFN expression levels than gene targeting, which occurs at endogenous loci present at only 1-2 copies per cell depending on the specific chromosomal target. Taken together, these results suggest that while the SSA assay is a viable method to screen large sets of ZFNs for activity, the relative activities shown in this assay are not necessarily predictive for which ZFNs will stimulate the highest levels of gene targeting (**Figure 3.3b**). To increase the protein expression level of the OPEN ZFNs, I cloned a codon-optimized FokI nuclease domain (Fn) in place of the standard Fn domain. Codon-optimization of this domain, which accounts for approximately 2/3rds of the total size of the ZFN protein in three finger ZFNs, significantly increased the expression of levels of ZFNs to levels similar to those of the MA ZFNs as determined by Western blot (**Figure 3.3c**). Comparison of the activity levels of these codon-optimized ZFNs with standard OPEN ZFNs showed increased activity at the endogenous *IL2RG* target site (**Figure 3.3d**). Despite this increase in protein expression and gene editing efficiency, comparison of codon-optimized OPEN ZFNs with the previously reported MA ZFNs revealed that these OPEN ZFNs were still only 10-40% as active at the endogenous *IL2RG* gene as the MA ZFNs (**Figure 3.3e**).

Despite reproducible stimulation of mutagenic NHEJ with my OPEN ZFNs at *IL2RG* exon 5, robust and reproducible gene addition was not observed with these ZFNs. This combination of results was highly surprising given that the

generation of DSBs necessary for mutagenic NHEJ is precisely the same event that is responsible for stimulating targeted HR. The most likely explanation for these results is that gene addition is indeed occurring in these samples, but the percentage of cells harboring stable integration of the Ubc-eGFP donor is not consistently different between the donor plus ZFN and donor alone samples. This could result from a combination of a) low targeting frequencies that are obscured by random integration frequencies of the donor plasmid in K562 cells, and b) ZFN toxicities that are known to decrease cell survival of highly-transfected cells which in turn decreases the number of cells harboring both random and targeted integrations over time. While further optimization of these OPEN ZFNs was possible, I chose instead to develop a new gene therapy strategy for SCID-X1 using transcription activator-like effector nucleases (TALENs), which have been shown to have similar efficacy as highly-active ZFNs with lower levels of toxicity (Mussolino et al., 2011).

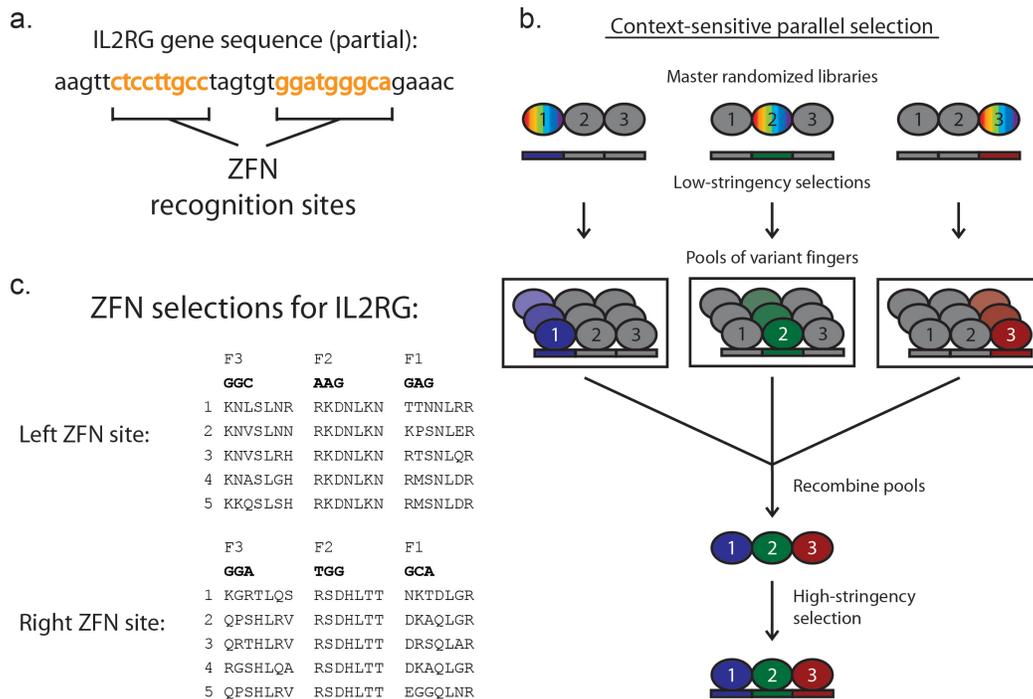


Figure 3.1: Context-dependent selection of ZFNs for *IL2RG* exon 5 using the OPEN platform. (a) Partial sequence of the WT *IL2RG* gene including the ZFN target sequences. (b) Schematic of the OPEN platform modified from (Pruett-Miller et al., 2008). Randomized libraries of ZFP sequences are put through two rounds of context-sensitive selection to determine the combination of 3 ZFPs with the most activity at the target sequence. (c) ZFN DNA binding domain sequences selected to bind to the left and right target sequences in *IL2RG* exon 5.

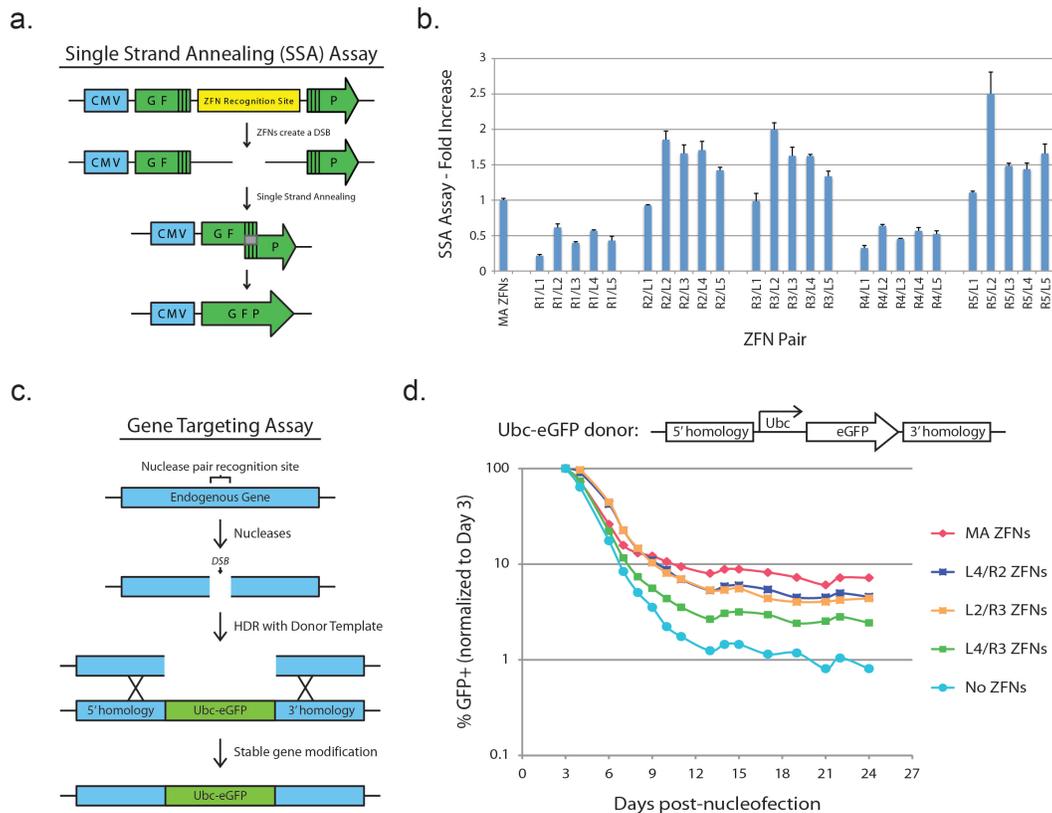


Figure 3.2: Measuring the activity of OPEN ZFNs with fluorescent reporter assays. (a) Schematic of single strand annealing (SSA) assay. A plasmid harboring ZFN recognition sites within a GFP gene is co-transfected with ZFNs in HEK 293T cells. When ZFNs bind to the ZFN recognition site and create a DSB, short regions of homology on either side of the DSB (vertical black bars) allow for the GFP gene to be repaired by single strand annealing, resulting in GFP expression. (b) SSA assay data showing the fold increase in %GFP+ cells with co-transfection of the SSA reporter plasmid and ZFNs relative to transfection of SSA reporter plasmid alone. Data is normalized relative to the MA ZFNs. Error bars, s.d. (c) Schematic of gene targeting assay. A donor template with arms of homology flanking the ZFN target site and a Ubc-eGFP insert is co-transfected

with ZFNs in K562 cells. When ZFNs create DSBs at the genomic target locus, the DSB can be repaired by homology-directed repair (HDR) with the donor template, resulting in stable integration of the GFP expression cassette. **(d)** Gene targeting assay data showing the stimulation of stable GFP integration with ZFNs. The frequency of Ubc-eGFP donor template random integration is seen in the no ZFNs control.

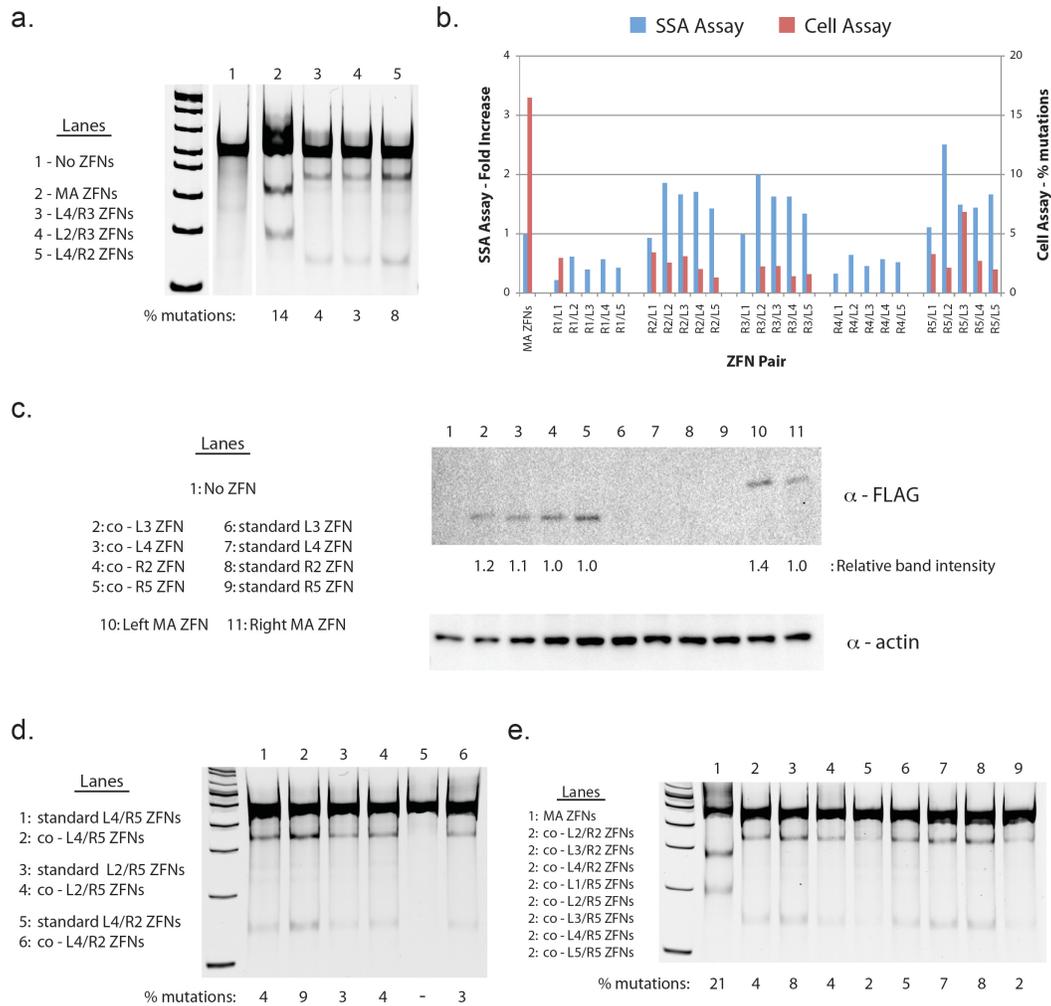


Figure 3.3: Measurement of standard OPEN ZFN expression levels and activity at the endogenous target site compared to codon-optimized OPEN ZFNs and MA ZFNs. (a) Cell assay showing site-specific induction of mutations at *IL2RG* exon 5 by OPEN ZFNs and MA ZFNs. The allelic mutation frequency is proportional to the intensity of the two bottom bands relative to the intensity of the top band for each sample. The MA ZFNs and OPEN ZFNs generate different sized bands in this assay since they induce mutations at different positions within

IL2RG exon 5. **(b)** Comparison of Cell assay results with SSA assay results for OPEN ZFNs and MA ZFNs. SSA assay data is normalized to the activity of the MA ZFNs and Cell assay data represents the actual allelic mutation frequency at the endogenous target site in K562 cells. **(c)** Western blot showing relative expression levels for OPEN ZFNs with and without a codon-optimized Fn domain. α -FLAG antibody is used to detect expression levels of each FLAG-tagged ZFN. The larger size of the 4-finger MA ZFNs compared to the 3-finger OPEN ZFNs is clearly shown on this blot. Relative intensities for each band are shown for each lane with detectable protein levels. Detection of actin with α -actin antibody is used as a loading control. **(d)** Cell assay showing activity levels for three OPEN-ZFN pairs with either a standard or codon-optimized Fn domain. **(e)** Cell assay showing activity levels for the MA ZFNs and 8 of the most active OPEN ZFNs expressed with a codon-optimized Fn domain. MA – modular assembly; Fn – FokI nuclease domain; co-ZFN – ZFN expressed with codon-optimized Fn domain.

Development of highly-active TALENs for gene targeting at IL2RG exon 1

The introduction of TALENs as gene editing tools in 2010 represents an important milestone in the development of gene targeting strategies for gene therapy as the ease of designing new TALENs and the high success rates for generating highly-active TALENs vastly expanded the potential genomic sites that can be targeted (Christian et al., 2010; Miller et al., 2011; Mussolino et al., 2012; Reyon et al., 2012). To develop a new gene targeting reagents for SCID-X1, I first designed TALENs targeting the start codon of *IL2RG* in exon 1 (**Figure 3.4a**). Targeted gene addition at the start codon creates a unique opportunity for gene therapy as the full cDNA for a gene can be targeted to its endogenous promoter, potentially restoring endogenously-regulated expression regardless of the downstream mutation in an individual patient.

To test the ability of my *IL2RG* exon 1 TALENs to create targeted DSBs at the endogenous site, I measured each TALEN pairs ability to mutate the target locus using the Cell assay (**Figure 3.4b**). Remarkably, all TALEN pairs designed with an optimal spacer length showed very high levels of activity with up to 44% of *IL2RG* alleles being mutated. These activity levels are as high as the most active pairs of ZFNs and TALENs reported in the literature (Perez et al., 2008; Reyon et al., 2012; Urnov et al., 2005). The importance of the spacer length for the activity of TALEN pairs was further highlighted by the result that both TALEN pairs with a suboptimal spacer length of 27 bp showed an intermediate

level of activity while the two TALEN pairs with non-functional spacer lengths less than 10 bp showed a complete absence of activity at the target locus (**Figure 3.4b**). I next tested the capacity for these TALEN pairs to stimulate targeted gene addition by transfecting K562 cells with TALENs and a CMV-eGFP donor template. Each TALEN pair was able to stimulate targeted gene addition in proportion to its activity at the target locus as seen in the Cell assay, with TALEN pair L3/R3 showing the highest level of a targeting with 16% of cells stably expressing eGFP (**Figure 3.4c**).

While the observed increases in Cell assay band intensity and stable gene integration seen with the *IL2RG* exon 1 TALENs are indicative of precise gene targeting, I next sought to confirm that these results were indeed the result of ‘on-target’ activity by directly sequencing TALEN-induced mutations and measuring targeted integration of a restriction site at *IL2RG* exon 1 (**Figure 3.5, Figure 3.6**). Analysis of the specific mutations created by TALEN pair L1/R1 showed a typical pattern of insertions and deletions with mutations occurring within the spacer region between the two TALEN binding sites (**Figure 3.5**). The frequency of mutations as determined by sequencing of PCR products was found to be ~38%, which validates the accuracy of the Cell assay that reported a mutation frequency of ~44% for the same targeted population. By creating donor templates with point mutations that disrupt the TALEN binding sequence and introduce novel restriction sites, I was able to directly analyze the frequency of *IL2RG* allele

modification by HR using a restriction fragment length polymorphism (RFLP) assay (**Figure 3.6a**). Targeting of a PmeI site with TALEN pair L3/R3 resulted in precise modification of ~10% of *IL2RG* alleles, corresponding to a cellular modification frequency of ~10-20% depending on the frequency of mono- and bi-allelic targeting (**Figure 3.6b**). Targeting frequencies using donor templates that altered either the spacer sequence, the left TALEN binding site, or both TALEN binding sites were not significantly different, supporting the surprising conclusion that under these conditions disruption of TALEN binding sites to prevent cutting of the donor template or re-cutting of targeted alleles does not improve gene targeting frequencies (**Figure 3.6b**). To further determine the frequency of mono- and bi-allelic modification I grew single cell clones from a K562 population targeted with a PmeI restriction site. RFLP analysis of these clones revealed a high frequency of biallelic modification, with 4 of 19 clones (21%) harboring mutations on both alleles compared to 3 of 19 clones (16%) with only one modified allele (**Figure 3.6c**).

Ideal gene therapy reagents combine high efficacy with low toxicity. To determine the relative toxicity of my *IL2RG* exon 1 TALENs compared to the *CCR5*-specific ZFNs that are currently in phase I clinical trials (clinicaltrials.gov NCT00842634) and the previously reported *IL2RG* exon 5-specific ZFNs, I used a Cell assay time-course that measures the change in the allelic modification frequency in a population over time. This assay improves on previous toxicity

assays that measure the difference in cell survival of transfected cells with and without nucleases. By directly measuring the toxicity of nucleases in the population of cells that are successfully targeted, this assay is not obscured by cells that are transfected at a low level and experience relatively low toxicity while also not expressing high enough levels of nucleases for successful gene targeting. Over the course of two weeks from Day 3 to Day 17 post-transfection, cells targeted with the *IL2RG* exon 1-specific TALENs showed little decrease in allelic modification frequency, with the L1/R1 and L2/R2 TALEN pairs not stimulating any loss of signal and the L3/R3 TALEN pair stimulating a moderate signal decrease of 17% (**Figure 3.7a**). In stark contrast, the highly-active ZFNs to both *CCR5* and *IL2RG* exon 5 stimulated significant loss of signal with decreases of 56% and 36% respectively (**Figure 3.7b**). The *IL2RG* exon 1-specific TALENs also stimulated absolute levels of modification as high or higher than the ZFNs at both Day 3 and Day 17, suggesting that the observed increase in toxicity from the ZFNs was not due to relative over-expression of the ZFNs with saturation of mutagenic NHEJ.

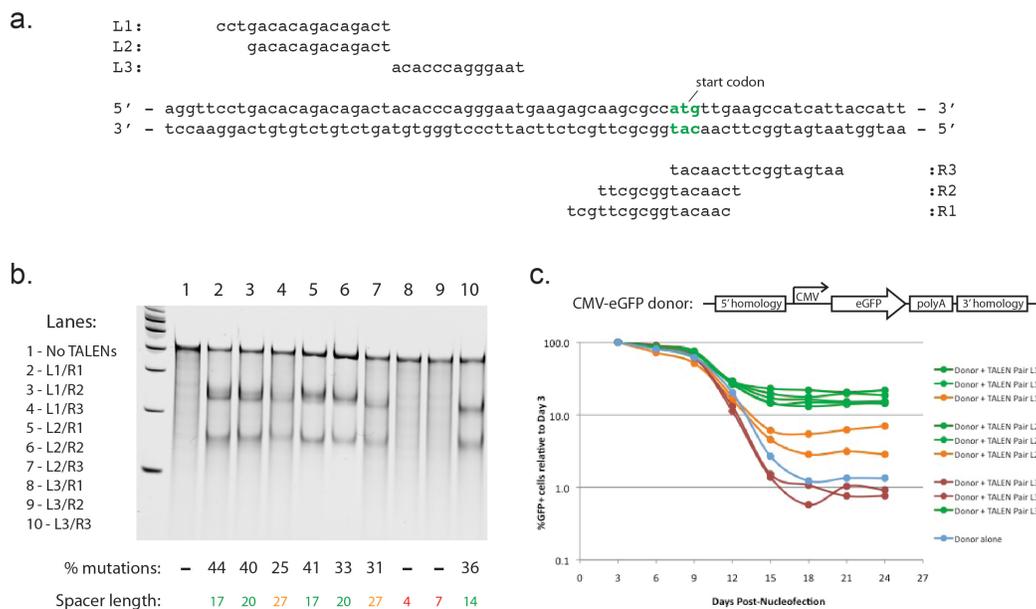


Figure 3.4: TALENs designed to target *IL2RG* exon 1. (a) Schematic of TALEN target sites surrounding the *IL2RG* start codon (denoted in green). (b) Cell assay showing TALEN activity levels at the target site. K562 cells were transfected with each of the 9 possible TALEN pairs and a no TALEN control and the resulting allelic mutation frequencies are shown. The spacer length between each TALEN pair is listed below the allelic mutation frequency, with the number color-coded to represent optimal (green), suboptimal (orange), and non-functional (red) spacer lengths. (c) Gene addition of a CMV-eGFP fluorescent reporter to *IL2RG* in K562 cells. TALEN pairs with optimal (green), suboptimal (orange), and non-functional (red) spacer lengths are color-coded as in panel (b), with the donor alone sample shown in blue.

WT *IL2RG* exon 1 sequence (partial):

L1 TALEN binding site
R1 TALEN binding site

aggttcctgacacagacagactacacccagggatgaagagcaagcgcc**atg**ttgaagccatcattaccatt
start codon

Mutated sequences:

```

aggttcctgacacagacagactaca----ggaatgaagagcaagcgccatggtgaagccatcattaccatt
aggttcctgacacagacagactaca-----ggaatgaagagcaagcgccatggtgaagccatcattaccatt
aggttcctgacacagacagactac-----ggaatgaagagcaagcgccatggtgaagccatcattaccatt
aggttcctgacacagacagactac-----ggaatgaagagcaagcgccatggtgaagccatcattaccatt
aggttcctgacacagacagactacacccag-----gaagagcaagcgccatggtgaagccatcattaccatt
aggttcctgacacagacagactacacc-----atgaagagcaagcgccatggtgaagccatcattaccatt
aggttcctgacacagacagactacacc-----gaagagcaagcgccatggtgaagccatcattaccatt
aggttcctgacacagacagactacacc-----aagagcaagcgccatggtgaagccatcattaccatt
aggttcctgacacagacagactacacc-----caagcgccatggtgaagccatcattaccatt
aggttcctgacacagacagactaca-----ccatggtgaagccatcattaccatt
aggttcctgacacagacag-----tgttgaagccatcattaccatt
aggttcctgacacag-----ccatcattaccatt
aggttcctgacacag-----ccatcattaccatt
aggttcctgacacagacagactacacccagggAAatgaagagcaagcgccatggtgaagccatcattaccatt
aggttcctgacacagacagactacacccagggGGAatgaagagcaagcgccatggtgaagccatcattaccatt

```

% Mutations = 37.5% (15/40); 13 deletions, 2 insertions

Figure 3.5: Specific mutations created by *IL2RG* exon 1 TALENs. The *IL2RG* locus was PCR amplified and individual PCR products were sequenced with Sanger sequencing. This sequencing analysis, which showed an allelic mutation frequency of 37.5% (15/40), is in close agreement with the Cell assay analysis for this sample, which showed an allelic mutation frequency of 44% (Figure 3.4b, lane 2). Deleted bases are indicated by dashes. Inserted bases are in bold and underlined.

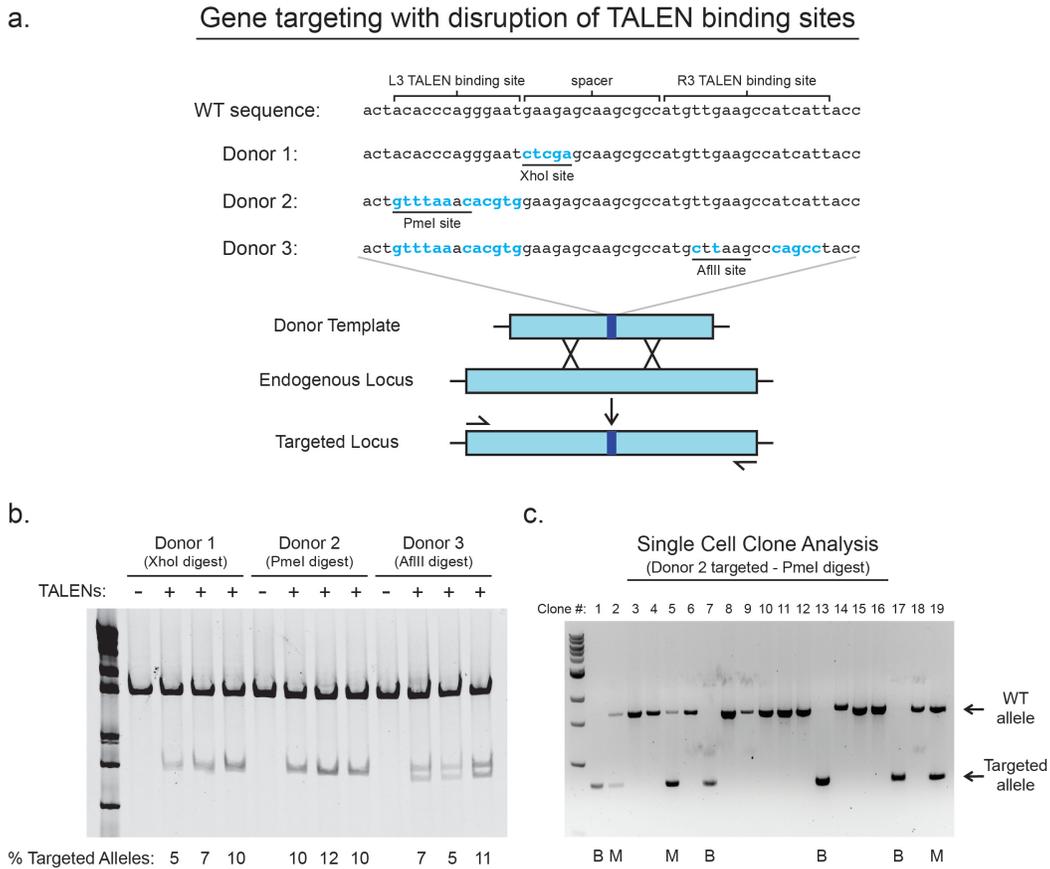


Figure 3.6: Gene targeting of restriction sites to *IL2RG* exon 1. (a) Schematic showing gene targeting of different restriction site-containing donor templates to *IL2RG* exon 1. The point mutations in the donor templates alter the TALEN binding site by changing the spacer sequence (Donor 1), the left TALEN binding site (Donor 2), or both TALEN binding sites (Donor 3). Following targeting, the *IL2RG* locus is amplified with PCR primers outside the donor template arms of homology for analysis with the RFLP assay. Point mutations (blue bases) and restriction sites within each donor template are indicated. (b) RFLP assay measuring the gene targeting frequency in K562 cells targeted with the donor templates shown in panel (a) and TALEN pair L3/R3. The percentage of targeted

alleles is calculated as the intensity of the cut (bottom) band(s) relative to the total intensity of the cut and uncut bands in each lane. (c) Single cell clone analysis of a K562 cell population targeted with Donor 2 and TALEN pair L3/R3 was performed by Joseph Clark. Single cells were sorted by FACS and *IL2RG* alleles were analyzed with the RFLP assay. Clones with monoallelic or biallelic targeting are indicated with an M or B respectively.

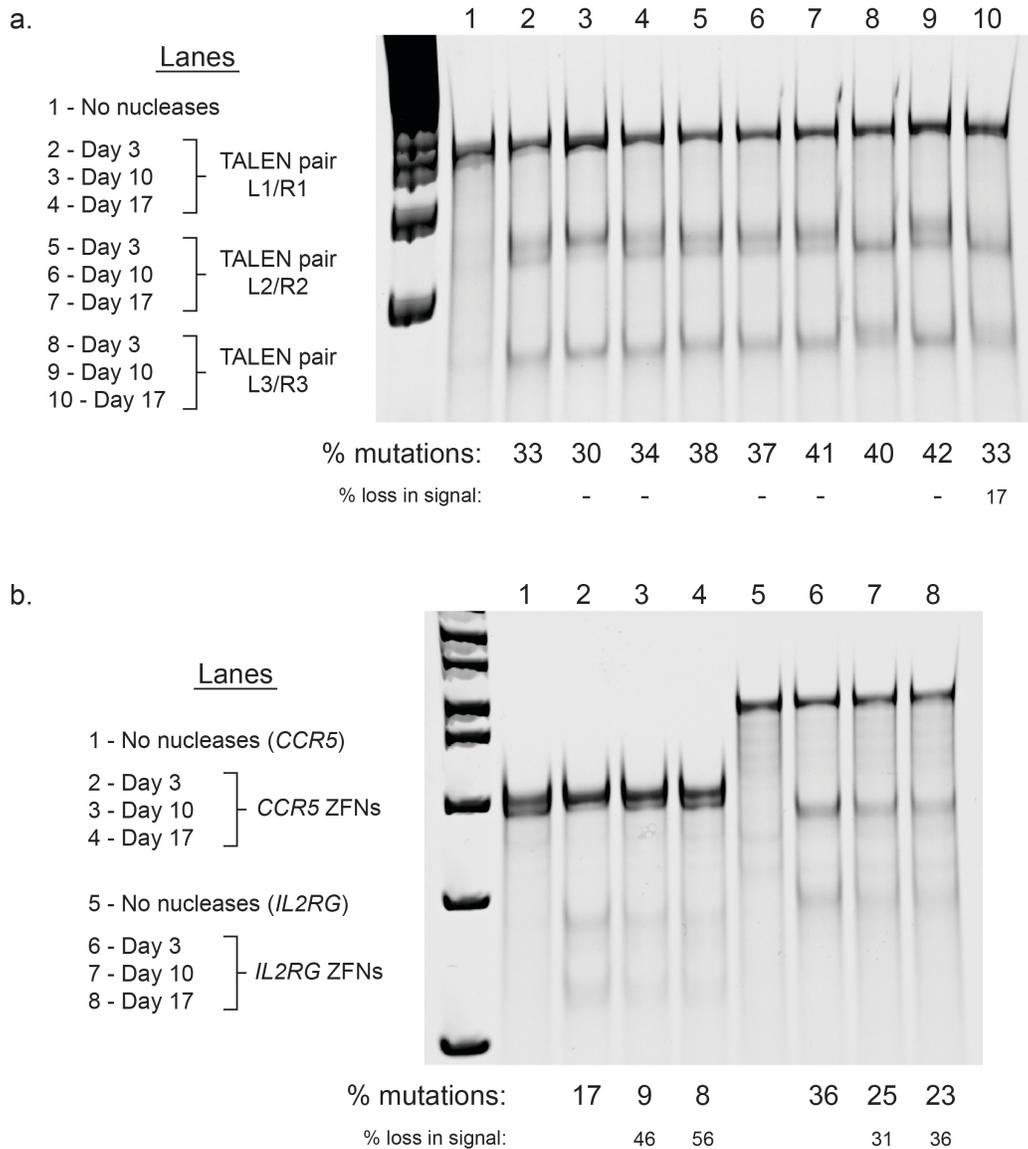


Figure 3.7: Comparison of toxicity between *IL2RG* exon 1 TALENs and commonly-used ZFNs. (a) Cell assay time course showing the allelic mutation frequency with three pairs of *IL2RG* exon 1 TALENs in K562 cells at day 3, day 10, and day 17 post-nucleofection. The decrease in mutation frequency over time is indicated at day 10 and day 17 relative to the initial frequency at day 3. K562

cells were grown at 37°C post-nucleofection. **(b)** Cell assay time course showing the allelic mutation frequency with the commonly-used *CCR5* and *IL2RG*-specific ZFNs in K562 cells. The decrease in mutation frequency over time is indicated at day 10 and day 17 relative to the initial frequency at day 3. K562 cells were grown at 37°C post-nucleofection. In this assay, the frequency of allelic modification in each sample is proportional to the intensity of the two lower bands in each lane relative to the higher WT band. Nuclease-mediated toxicity is measured as a relative survival disadvantage of nuclease-modified cells over time.

Development of a novel gene therapy strategy for SCID-X1

The clinical paradigm suggested here for SCID-X1 gene therapy involves *ex vivo* gene targeting of allogenic CD34+ HSPCs with integration of full *IL2RG* cDNA at *IL2RG* exon 1. For this strategy to be curative, expression levels of functional IL2R γ protein from the inserted transgene must be capable of restoring the physiological capacity of CD34+ HSPCs to respond to interleukin signaling and participate in lymphopoiesis. To test the ability of my TALENs to stimulate targeted integration of full *IL2RG* cDNA, I transfected K562 cells with TALEN pair L3/R3 and a donor template including *IL2RG* cDNA directly following the *IL2RG* promoter and a Ubc-eGFP fluorescent reporter (**Figure 3.8a**). High levels of TALEN activity and stable gene integration were detected using the Cell assay and FACS analysis (**Figure 3.8b**, **Figure 3.8c**). The ability of the endogenous *IL2RG* promoter to drive expression from a targeted transgene was tested by targeting eGFP directly following the *IL2RG* promoter without other exogenous promoter elements (**Figure 3.8c**). Following gene targeting, eGFP expression was successfully detected by FACS in a percentage of cells similar to that seen with targeting of the Ubc-eGFP reporter, suggesting that expression of eGFP from the endogenous *IL2RG* promoter produces detectable protein levels in essentially all targeted cells. The relatively large size of the cDNA donor insert, 3.7kb, did not decrease gene targeting efficiency relative to the smaller, 1.2kb, insert of the eGFP donor. Interestingly, eGFP+ cells in the donor alone sample were observed

at a very low frequency of 0.2% for the eGFP donor compared to 1.5% for the Ubc-eGFP reporter in the cDNA donor, reflecting the requirement of the eGFP donor to be integrated following an endogenous promoter to express significant levels of eGFP (**Figure 3.8c**). High-throughput sequencing to measure the percentage of *IL2RG* alleles harboring the targeted cDNA sequence confirmed that ~17% of alleles had been precisely modified in the cDNA donor-targeted population, confirming the on-target gene addition frequency suggested by the stable integration of the Ubc-eGFP reporter (**Figure 3.9**).

The benefit of targeting exon 1 of *IL2RG* is that a single gene therapy protocol could potentially be used to treat all SCID-X1 patients with mutations downstream of the *IL2RG* start codon, which are responsible for more than 98% of SCID-X1 cases. A potential drawback of this approach, however, is that regulation of gene expression levels by downstream elements such as introns can be lost. To determine the activity levels of IL2R γ following targeted insertion of cDNA I targeted WT cDNA (with the exception of minor codon usage mutations in exon 1 used for targeting analysis), codon-optimized cDNA, and codon-optimized cDNA with an artificial intron to *IL2RG* in RAMOS cells (**Figure 3.10a**). The level of functional interleukin-4 receptor (IL4R) on RAMOS cells is limited by the availability of the IL2R γ subunit due to relatively low expression levels compared to other receptor subunits, making these cells ideal for measuring IL2R γ activity levels (Junttila et al., 2012). Following targeting of RAMOS cell

populations with TALEN pair L3/R3 and donor templates, I sorted individual eGFP⁺ cell clones and confirmed targeting of *IL2RG* with PCR (**Figure 3.10b**). I next tested the level of IL2R γ activity in targeted clones and WT cells using a phosphoflow assay that measures IL2R γ -dependant phosphorylation of STAT6 following stimulation of cells with IL-4 (**Figure 3.10c**). Since RAMOS cells are derived from a male, they have only one copy of *IL2RG*, ensuring that IL2R γ in targeted clones is only expressed from the targeted allele. In RAMOS cells with IL2R γ expression knocked-out by insertion of a Ubc-eGFP cassette (KO clones), IL2R γ activity dropped to very low levels compared to WT cells (**Figure 3.10d**). IL2R γ activity in RAMOS clones targeted with WT cDNA (cDNA clones) confirmed the ability of IL2R γ to be expressed from a targeted transgene, though the level of IL2R γ activity was significantly lower than in WT cells. Codon-optimization and the addition of an upstream artificial promoter represent two potential strategies to increase protein expression levels from targeted *IL2RG* cDNA. Analysis of IL2R γ activity in RAMOS clones targeted with codon-optimized cDNA (*co*-cDNA clones) alone and codon-optimized cDNA with an artificial intron (intron-*co*-cDNA clones) showed increased activity relative to clones expressing WT cDNA, with intron-*co*-cDNA clones showing IL2R γ activity levels as high or higher than WT RAMOS cells (**Figure 3.10d**). These results show that physiologically-relevant levels of IL2R γ expression can be

achieved by gene addition of modified *IL2RG* cDNA, supporting the potential of this gene therapy strategy as a generally-effective approach for SCID-X1 patients.

Having confirmed the ability of these gene targeting reagents to stimulate gene addition at *IL2RG* exon 1 and that expression from a targeted *IL2RG* cDNA produces physiologically-relevant levels of IL2R γ activity, I next tested my ability to apply this strategy to CD34⁺ HSPCs. Initial experiments in CD34⁺ HSPCs revealed that these primary cells are highly sensitive to toxicity resulting from nucleofection of plasmid DNA and are transfected at significantly lower efficiencies than commonly-used cell lines. To decrease the amount of total DNA necessary for efficient gene targeting, I cloned my *IL2RG* exon 1 TALENs into a ‘high-expression’ backbone based on the pmaxCloning vector (**Figure 3.11a**). Compared to standard TALEN expression plasmids, expression of TALENs using the high-expression constructs generated significantly increased activity at both high and low total plasmid amounts (**Figure 3.11b**). To test the capacity of these gene targeting reagents to function at the significantly decreased amounts of DNA and protein expression levels suitable for nucleofection of CD34⁺ HSPCs (**Figure 3.12a**), I titrated the amount of donor template and TALEN expression DNA in K562 cells. These experiments demonstrated the robustness of these reagents, as efficient gene addition at *IL2RG* was seen with donor template amounts as low as 50ng and TALEN expression plasmid amounts as low as 10ng, both of which represent at least 100-fold decreases from optimal levels (**Figure 3.12c**, **Figure**

3.12d). As expected, the high-expression TALEN vectors stimulated higher levels of gene addition than the standard TALEN expression vectors at low DNA amounts (**Figure 3.12b**).

To test the ability of TALENs expressed from the high-expression constructs to target *IL2RG* in HSPCs, I transfected CD34⁺ cells purified from umbilical cord blood with TALEN pair L3/R3. Analysis of TALEN-treated HSPCs revealed significant levels of gene modification at *IL2RG* up to 10% as measured by the Cell assay (**Figure 3.13a**). When an *IL2RG* cDNA donor was co-transfected with TALENs in CD34⁺ cells, stable expression of the Ubc-eGFP fluorescent reporter was seen in 0.7% of cells compared to < 0.1% of cells treated with the cDNA donor alone (**Figure 3.13b**). SMRT sequencing analysis of the targeted population revealed that ~1.4% of alleles were precisely targeted, which represents a significant increase from the fluorescent read-out of 0.7%. A possible explanation for this discrepancy is that not all CD34⁺ cells actively express integrated Ubc-eGFP either due to silencing or lack of sufficient expression with this promoter. These results represent a significant improvement in gene targeting efficiency in CD34⁺ HSPCs, which has not been previously reported at frequencies above 0.1% (Lombardo et al., 2007). While the high-activity and relatively low toxicity of the *IL2RG* TALENs presented here make these gene targeting reagents potentially ideal for translation to gene therapy in patient

CD34+ HSPCs, the overall toxicity of the gene targeting process in CD34+ cells must be further improved to make this a viable clinical strategy.

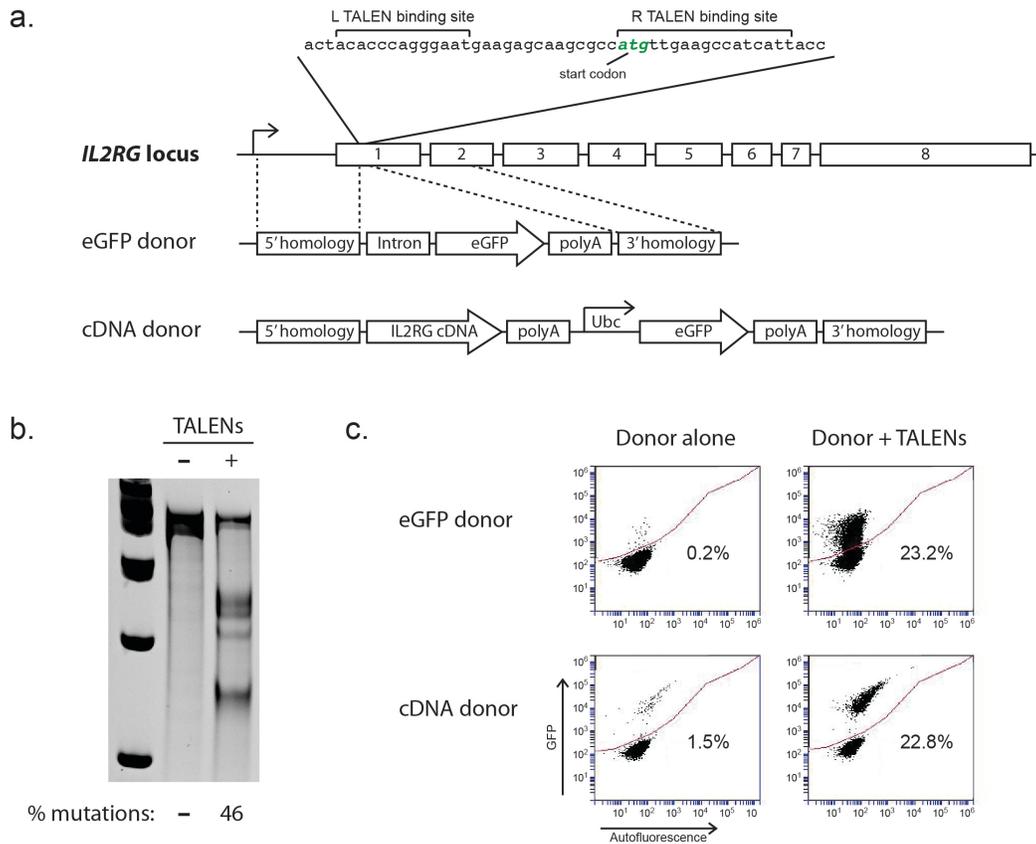


Figure 3.8: Novel gene therapy strategy for SCID-X1. (a) Schematic showing the TALEN pair L3/R3 target site and donor template elements in relation to the endogenous *IL2RG* locus. The inserts for the eGFP and cDNA donors that are integrated at *IL2RG* exon 1 following HR are illustrated. (b) Cell assay showing targeted disruption of the *IL2RG* exon 1 target site by TALEN pair L3/R3 in K562 cells. (c) Representative FACS plots showing stable integration of the eGFP and cDNA donor templates in K562 cells.

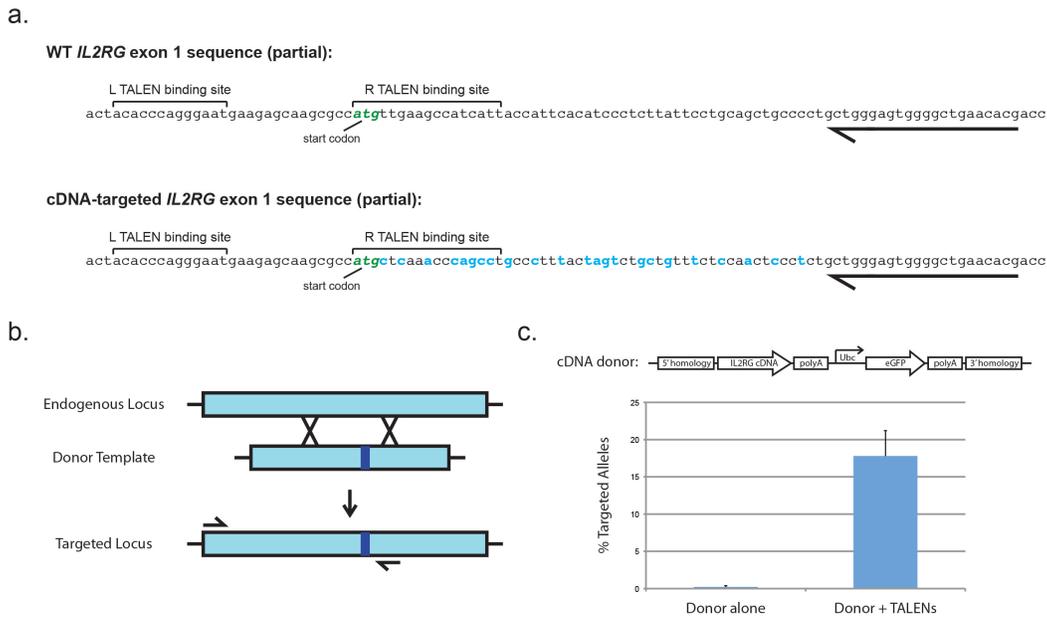


Figure 3.9: High throughput sequencing approach using SMRT sequencing to confirm targeted gene addition. (a) Schematic comparing the sequence of part of *IL2RG* exon 1 before targeting (WT) and after targeting with a cDNA donor (cDNA-targeted). TALEN binding sites, the start codon (green bases), codon-usage mutations integrated following gene targeting (blue bases), and the reverse PCR primer used to amplify the locus (partial arrows) are indicated. The SNPs integrated following HR with the donor template do not alter the amino acid sequence of IL2R γ , but provide a signature used to calculate the relative frequency of ‘WT’ and ‘targeted’ alleles. The PCR amplicons created from the WT and targeted alleles are the same size, which prevents bias for either allele during SMRT sequencing. (b) Schematic illustrating the relative positions of the donor template arms of homology and PCR primers used for SMRT sequencing analysis. The vertical blue bar represents the insert sequence in the donor template, and is not to scale. (c) SMRT sequencing analysis of K562 cells

measuring the percentage of alleles harboring the targeted cDNA signature following gene targeting with the indicated donor template. Error bars, s.d. For more information on SMRT sequencing and the reliability of this high throughput sequencing approach, see Figures 2.2 – 2.4.

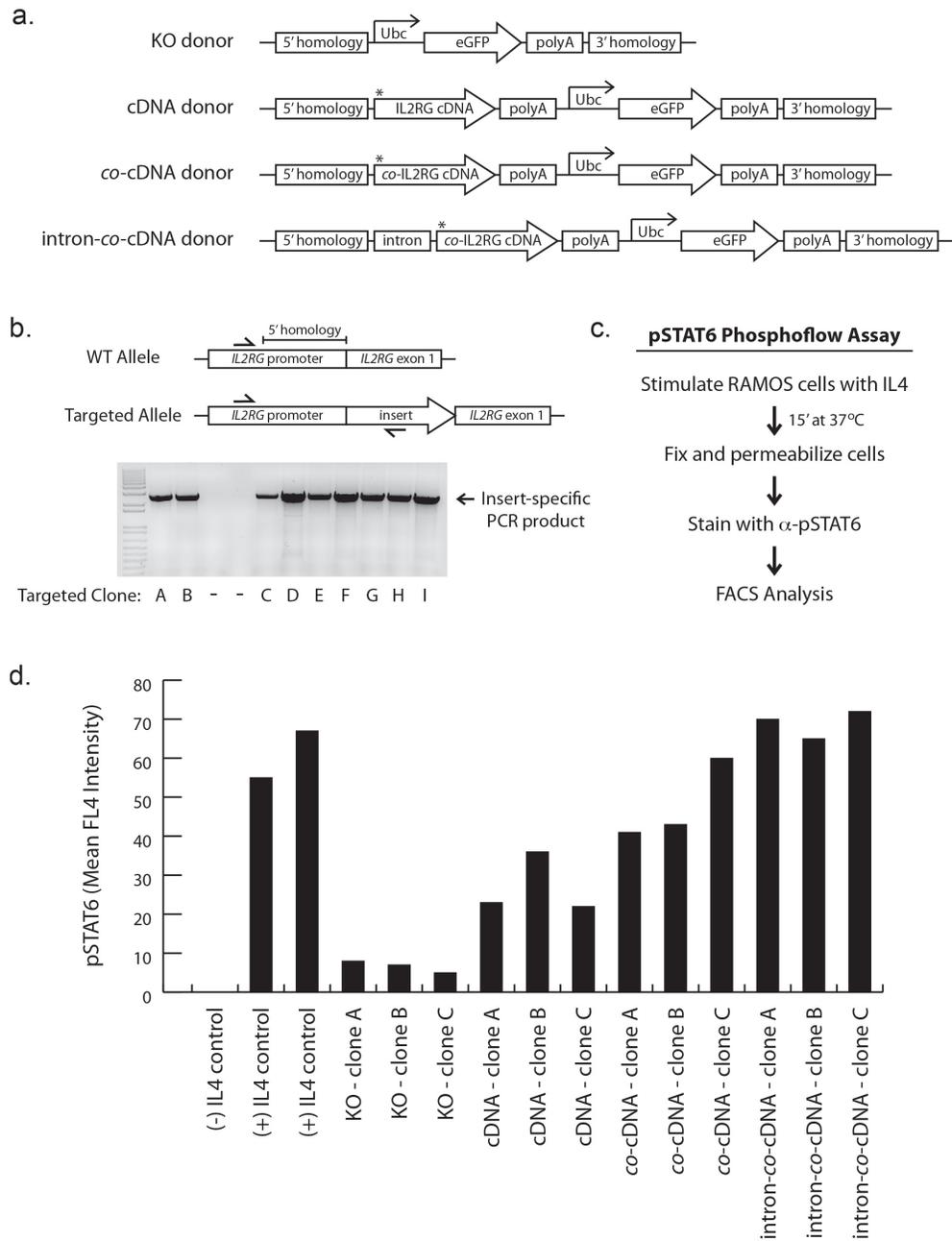


Figure 3.10: Measurement of IL2R γ activity in gene targeted RAMOS cells.
(a) Schematics illustrating the donor templates used to target either precise knockout of *IL2RG* expression (KO) or integration of *IL2RG* cDNA at *IL2RG*

exon 1. (*) indicates 23 codon-usage mutations, illustrated in Figure 3.9a, used for cDNA targeting analysis by high-throughput sequencing. **(b)** PCR confirmation of targeted integration at *IL2RG* in RAMOS single cell clones. **(c)** Phosphoflow assay workflow. RAMOS cells are first stimulated with 10ng/ml IL4 for 15'. Cells are then fixed, permeabilized, and stained with an antibody specific for phosphorylated STAT6 as an indicator of successful signal transduction through the IL2R γ -dependant IL4 receptor. Data is normalized to the (-) IL4 negative control. **(d)** Phosphoflow assay showing relative IL2R γ activity in WT RAMOS cells (controls) and clonal populations with targeted insertion of the sequences illustrated in panel (a). Data from three separate clones is shown for each gene targeting modification at *IL2RG*.

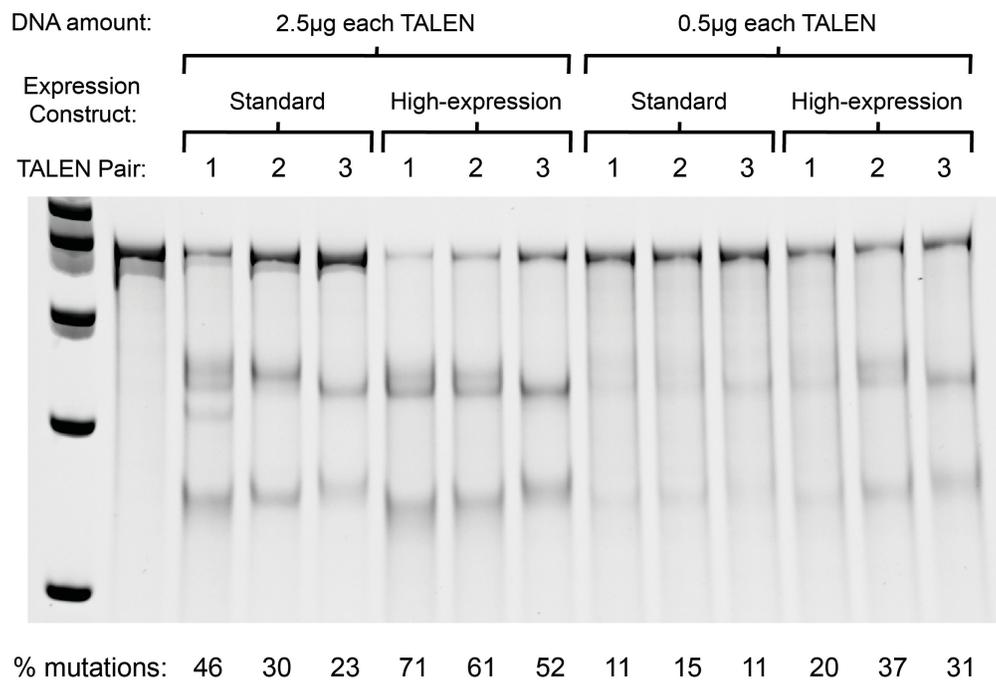


Figure 3.11: Comparison of activity levels with ‘high-expression’ and standard TALEN expression constructs. Cell assay showing the activity at *IL2RG* exon 1 for three TALEN pairs expressed from the ‘high-expression’ vector and the standard expression vector. K562 cells were nucleofected with the indicated amount of each TALEN and allelic mutation frequencies were measured on Day 3 post-nucleofection. TALEN pair 1 – L1/R1; TALEN pair 2 – L2/R2, TALEN pair 3 – L3/R3.

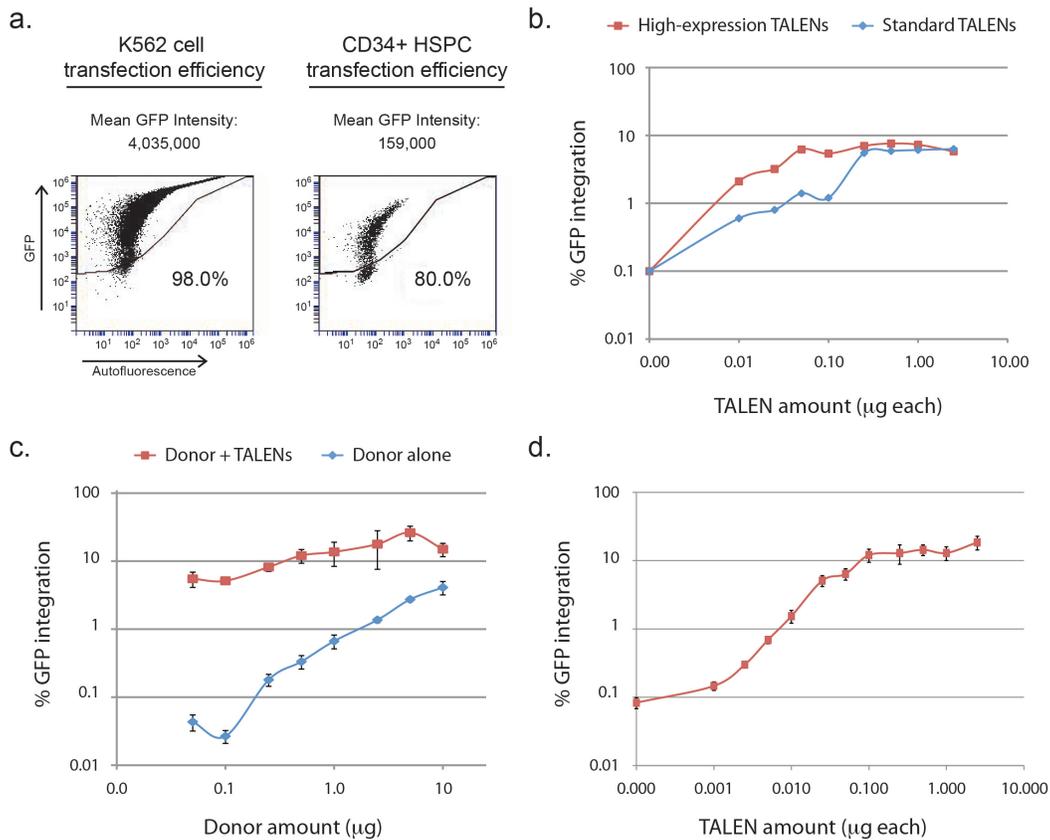


Figure 3.12: Measurement of gene targeting efficiencies with DNA and protein expression levels relevant to CD34+ hematopoietic stem and progenitor cells. (a) Representative FACS plots showing transfection efficiency and fluorescent intensity of K562 cells and CD34+ HSPCs 72 hours post-nucleofection with a GFP expression plasmid. (b) Gene targeting efficiency with different amounts of TALEN constructs in K562 cells. This experiment was performed by Joseph Clark. (c) Gene targeting efficiency with different amounts of donor template in K562 cells. (d) Gene targeting efficiency with different amounts of high-expression TALEN constructs in K562 cells. Error bars, s.d.

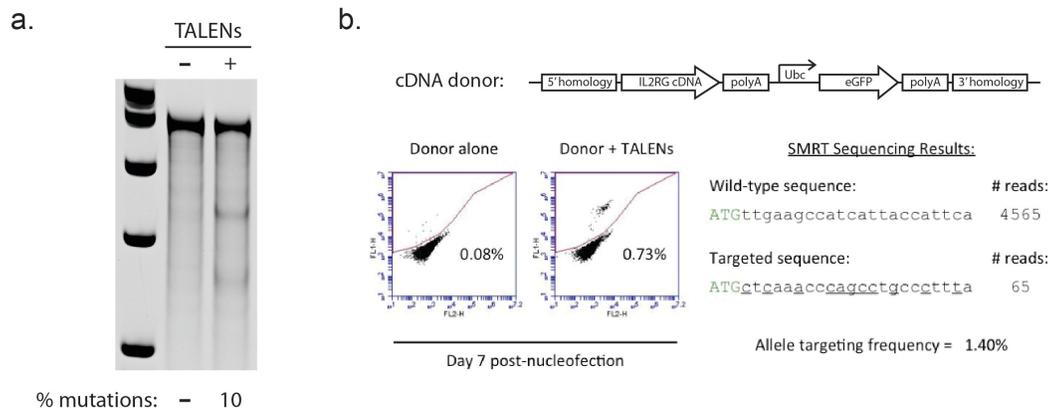


Figure 3.13: Gene targeting in CD34⁺ hematopoietic stem and progenitor cells. (a) Cell assay showing targeted gene disruption of the *IL2RG* target site by TALEN pair L3/R3 in CD34⁺ HSPCs. (b) Targeted gene addition of a cDNA donor with a fluorescent reporter at *IL2RG* exon 1 in CD34⁺ HSPCs. FACS plots showing stable integration of the cDNA donor and SMRT sequencing results measuring on-target gene addition frequencies are shown.

Discussion

I show here a general strategy for correcting the functional deficit causing SCID-X1 by precisely targeting full *IL2RG* cDNA under the control of the endogenous promoter at *IL2RG* exon 1. Previous observations of spontaneous reversions of SCID-causing mutations in T-cell precursors (Bousso et al., 2000; Hirschhorn et al., 1996; Stephan et al., 1996) and gene therapy trials with virus-mediated transgene insertion have clearly demonstrated the potential for small numbers of gene corrected cells to restore immune system function in patients suffering from SCID-X1 (Gaspar et al., 2004; Hacein-Bey-Abina et al., 2002). Gene targeting with ZFNs to insert partial cDNA at *IL2RG* exon 5 presented a more precise approach for restoring *IL2RG* expression, but had low efficiency in clinically-relevant hematopoietic stem cells (Lombardo et al., 2007). Clinical application of these strategies has been hampered by insertional oncogenesis, low gene targeting efficiencies in hematopoietic stem cells, and nuclease toxicity (Fischer et al., 2010). The *IL2RG* exon 1-specific TALENs described here are significantly less toxic than previously described ZFNs, including ZFNs currently being used in clinical trials, and stimulate 10-fold higher levels of gene targeting in CD34+ HSPCs than the highest levels previously achieved.

The observation that expression of IL2R γ from codon-optimized *IL2RG* cDNA with an artificial intron produces IL2R γ activity levels as high or higher than those in WT cells illustrates the potential for cDNA addition at exon 1 as a

powerful strategy for gene therapy. Due to the variety and complexity of the different mutations capable of disrupting gene expression or protein function, the ability to overcome a majority of disease-causing mutations with a single gene targeting strategy will expedite the development of gene therapy tools for many monogenic diseases. Highlighting the utility of this approach, the gene targeting strategy presented here represents a viable option for over 98% of known SCID-X1-causing mutations, excluding only those in upstream promoter elements and those that disrupt the TALEN binding sites surrounding the *IL2RG* start codon.

Despite the improvements in activity levels and toxicity with this system, *in vivo* demonstrations of the efficacy of this approach have proven challenging. The most likely obstacle facing successful engraftment and *in vivo* functionality of targeted HSCs is the overt toxicity seen in CD34+ HSPC populations following successful gene targeting. The desire to reduce toxicity without reducing efficacy has led to the development of numerous toxicity-minimizing techniques. The majority of these techniques center on reducing the creation of DSBs by engineered nucleases, including improving nuclease specificity with obligate heterodimer domains and reducing total nuclease expression through fusion to degradation domains or delivery as mRNA instead of DNA (Doyon et al., 2011; Pruett-Miller et al., 2009; Zou et al., 2011). Strategies to improve other elements of gene targeting by altering the reagent delivery method or optimizing donor

template design also hold promise for reducing cytotoxicity in stem cells (Jia et al., 2010; Liu et al., 2002).

The novel gene targeting strategy at *IL2RG* exon 1 presented here has the potential to form the foundation of a gene therapy protocol capable of restoring *IL2R γ* activity in a patient's own hematopoietic stem cells for the vast majority of patients suffering from SCID-X1. Furthermore, the ability of *IL2RG* expression to be driven by the endogenous promoter in precisely-targeted cells, combined with the potential to design new TALENs to almost any site in the genome, suggests that this approach could be generally useful for many monogenic diseases, particularly those where endogenously-regulated gene expression is critical. With further optimization of gene targeting in CD34⁺ HSPCs, this strategy may provide the same restoration of immune system function seen in SCID-X1 trials without the oncogenic potential that has prevented wide application of gene therapy in the clinic.

Materials and Methods

Cell culture and transfection:

HEK 293T cells were maintained at 37°C with 5% CO₂ in DMEM (Cellgro) supplemented with 10% bovine growth serum, 100 units/ml penicillin, 100µg/ml streptomycin, and 2mM L-glutamine. For transfection, HEK 293T cells were grown in 24 wells until reaching ~70% confluency and transfected with calcium phosphate as previously described (Porteus et al., 2003).

K562 cells (ATCC) and **RAMOS cells** (generous gift from Aaron Ring and Dr. K. Christopher Garcia) were maintained at 37°C with 5% CO₂ in RPMI 1640 (Hyclone) supplemented with 10% bovine growth serum, 100 units/ml penicillin, 100µg/ml streptomycin and 2mM L-glutamine. K562 cells were transfected by nucleofection (Lonza) of 1 x 10⁶ cells using program T-016 and a nucleofection buffer containing 100mM KH₂PO₄, 15mM NaHCO₃, 12mM MgCl₂ • 6H₂O, 8mM ATP, 2mM glucose, pH 7.4. RAMOS cells were transfected by nucleofection (Lonza) of 2 x 10⁶ cells in the nucleofection buffer described above using program O-006.

Human CD34+ HSPCs were purchased from Lonza (2M-101B) and thawed per the manufacturer's instructions. CD34+ HSPCs were maintained at 37°C with 5% CO₂ in X-VIVO15 (Lonza) supplemented with SCF (100ng/ml), TPO (100ng/ml), Flt3-Ligand (100ng/ml), IL-6 (100ng/ml), and StemRegenin1

(0.75 μ M). For transfection, 4×10^5 CD34+ HSPCs were nucleofected with an Amaxa 4D Nucleofector with the P3 Primary Cell Nucleofector Kit (V4XP-3032) and program EO-100 per the manufacturer's instructions.

ZFN, TALEN, and Donor Template Construct assembly:

ZFNs: *IL2RG* exon 5-specific ZFPs were selected by Jiuli Zhang using the OPEN B2H selection strategy as previously described (Pruett-Miller et al., 2008). ZFPs were subsequently cloned upstream of the FokI nuclease domain in a FLAG-ZFN expression vector based on pcDNA6 (Invitrogen). MA ZFNs specific for *CCR5* (Perez et al., 2008) and *IL2RG* (Urnov et al., 2005) were synthesized with a codon-optimized Fn domain and cloned into the pcDNA6 expression vector. Codon-optimized OPEN ZFNs were generated by cloning ZFPs upstream of the codon-optimized Fn domain in the MA ZFN expression vector.

TALENs: *IL2RG* TALE repeats were synthesized (Genscript) and cloned into a TALEN expression vector based on pcDNA3.1 (Invitrogen) with the $\Delta 152$ N-terminal domain and the +63 C-terminal domain fused to the FokI nuclease domain as previously described (Miller et al., 2011). High-expression TALENs were cloned into pmaxCloning (Lonza) downstream of a CMV promoter and artificial intron and upstream of an SV40 polyA.

Donor Templates: The *IL2RG* targeting vectors were synthesized by PCR amplifying the *IL2RG* locus from genomic DNA isolated from K562 cells using

the primers in Figure 3.14. Codon-usage mutations were added to WT *IL2RG* cDNA through PCR. Codon-optimized *IL2RG* cDNA was synthesized (Genescript) with the same exon 1 codon-usage signature as WT cDNA and a short WT sequence at the end of exon 1 for PCR amplification. The artificial intron was PCR amplified from pmaxGFP expression plasmid (Lonza) and cloned upstream of *IL2RG* cDNA. Restriction site donor templates for *IL2RG* exon 1 were generated by PCR and cloned using a naturally-occurring HaeII site directly 5' to the *IL2RG* start codon. CMV-eGFP and Ubc-eGFP were inserted between the exon 1 and exon 5 arms of homology using standard cloning techniques. All cDNA donor templates were cloned with the cDNA integrated in-frame with the *IL2RG* start codon while maintaining the *IL2RG* 5' UTR and Kozak sequence with the exception of the artificial intron donor, where the intervening sequence between the artificial intron and start codon in pmaxGFP was maintained. A BGH polyA sequence following the stop codon and a downstream Ubc-eGFP-BGH polyA expression cassette was added to all cDNA constructs using standard cloning techniques. For the promoterless eGFP donor, eGFP was amplified by PCR and cloned with eGFP integrated in-frame with the *IL2RG* start codon.

Single-strand annealing (SSA) assay:

The SSA reporter was generated by inserting the *IL2RG* exon 5 target sites for both the OPEN and MA ZFNs inside a disrupted GFP gene downstream of a

CMV promoter. The insert is flanked on both sides by a duplicated 42 bp region of GFP. The SSA reporter and ZFNs are delivered into HEK 293T cells by calcium phosphate transfection. GFP expression in transfected cells was measured on a BD FACS Aria (Becton Dickinson) at 48 hours post-transfection.

Cell assay and time-course toxicity assay:

The Cell assay was performed as previously described (Guschin et al., 2010). Briefly, K562 cells or CD34+ HSPCs were nucleofected with either 2.5µg or 500ng of each nuclease respectively. Genomic DNA was harvested with the DNeasy Blood & Tissue Kit (Qiagen) on day 3 post-nucleofection for standard Cell assay analysis and additionally on day 10 and day 17 post-nucleofection for the Cell time-course toxicity assay. The *IL2RG* exon 1, *IL2RG* exon 5, and *CCR5* loci were amplified by PCR using the primers listed in Figure 3.14 with Accuprime polymerase (Invitrogen). 200-400 ng of PCR product was analyzed for cutting by the Cell nuclease per the manufacturer's instructions (Transgenomic).

Flow cytometry:

Samples were analyzed for fluorescence using either a BD FACS Aria (Becton Dickinson) or an Accuri C6 flow cytometer (Becton Dickinson). GFP

expression was measured in the FL1 channel and α -pSTAT6-647 was detected in the FL4 channel.

Western blot:

All ZFN proteins were cloned with an N-terminal FLAG tag. K562 cells were nucleofected with 2.5 μ g ZFN expression plasmid and cells were collected at 24 hours post-nucleofection. Expression of ZFNs was detected by Western blot as described in (Pruett-Miller et al., 2009) using 1:10,000 mouse α -Flag M2 primary antibody (Sigma-Aldrich) followed by 1:10,000 goat α -mouse-HRP secondary antibody (Sigma-Aldrich). β -action expression was detected using 1:2,000 rabbit α -actin primary antibody (Sigma-Aldrich) followed by 1:10,000 goat α -rabbit-HRP secondary antibody (Sigma-Aldrich). Western Blotting Luminal Reagent (Santa Cruz Biotechnology) was used for visualization of HRP-conjugated antibody.

Restriction fragment length polymorphism assay:

Restriction fragment length polymorphism assay was performed as previously described in Chen, F. *et al.* (Chen et al., 2011). Briefly, genomic DNA was extracted from cells with the DNeasy Blood & Tissue Kit (Qiagen). Genomic DNA was then PCR amplified using Accuprime polymerase (Invitrogen) with primers outside the donor template arms of homology (see Figure 3.14 for primer

sequences). PCR products were digested with 20 U of either XhoI, PmeI, or AflII at 37 °C for 2h and resolved with PAGE.

Single cell clone generation and targeting analysis:

Single-cell sorting was performed by flow cytometry on a BD FACS Aria with single cells sorted into 96 wells. Clonal populations were expanded with standard growth conditions for each cell type. Genomic DNA was isolated from clonal populations using the DNeasy Blood & Tissue Kit (Qiagen). In **K562 cell clones**, allelic modification was analyzed with the RFLP assay by Joseph Clark. To generate targeted **RAMOS cell clones**, RAMOS cells were first nucleofected with 5µg donor template and 1µg of each TALEN. Following gene targeting, individual GFP+ cells were sorted into 96 wells. Targeting for each clone was determined by PCR screening with the primers in Figure 3.14. The successful amplification of a targeting-dependant PCR product confirmed clones that were precisely modified at the *IL2RG* locus.

High-throughput (SMRT) sequencing:

Genomic DNA containing *IL2RG* alleles was harvested from cultured K562 cells and CD34+ HSPCs using the DNeasy Blood & Tissue Kit (Qiagen). *IL2RG* alleles were amplified using the primers in Figure 3.14 with Accuprime polymerase (Invitrogen). SMRT DNA library construction from PCR products

was performed per the manufacturer's instructions and sequencing data collection was performed on the PacBio® RS (Pacific Biosciences). Analysis of SMRT sequencing results was done using the SMRT analysis pipeline developed by Eli Fine in Strawberry Perl, which utilizes the NCBI BLAST software as well as the mEmboss Needleman-Wunsch pairwise alignment algorithm. All components of the pipeline were run on a standard Windows PC and are available for download (<https://sourceforge.net/projects/tdna-getsmart/>).

Statistical analysis:

Data from three or more biological replicates was used to determine statistical significance, with the mean \pm standard deviation reported. Statistical analysis was performed using the Student's T-test. P-values < 0.05 were considered significant.

Phosphoflow assay:

RAMOS cells were analyzed using the phosphoflow assay as previously described (Junttila et al., 2012; Krutzik et al., 2003). Briefly, 5×10^5 RAMOS cells were either unstimulated or stimulated with 10ng/ml IL4 for 15 minutes at 37°C in growth media. Cells were then fixed in 1.5% paraformaldehyde for 10 minutes at room temperature and permeabilized with ice-cold methanol (100% v/v) for 30 minutes on ice. Cells were stained with α -pSTAT6 Ax647 (1:50

dilution, BD Biosciences #612601) for 1 hour in the dark at room temperature, followed by analysis by FACS. pSTAT6 levels were calculated by subtracting the mean FL4 intensity of the stimulated samples from that of the unstimulated sample.

Cell assay *IL2RG* exon 1 and sequencing of mutations

F: tcacacagcacatattgccacaccctctg

R: tgcccacatgattgtaatggccagtggc

Cell assay *IL2RG* exon 5

F: tagcagagatgacactgggtgggtgttcagg

R: taacaacacgctaaccaaccctacacaga

Cell assay *CCR5*

F: aagatgattatcaagtgtcaagtcc

R: caaagtcccactgggcg

RFLP assay and single cell clone analysis

F: gcagagtggctgtggtaatggaaggaggaaac

R: ctgggtactgcagatatccagaccctagcctc

SMRT sequencing analysis

F: gcagagtggctgtggtaatggaaggaggaaac

R: cgtgttcagccccactcccagc

***IL2RG* exon 1 donor constructs**

5' homology arm F: taactataacggctctaaggtagcgattaattaatgggagaaacaccacagaagc

5' homology arm R: tagggataacagggtaattctagaggcgcttgctctcattcct

3' homology arm F: ctcgagggcgcgccgtttaaactagtgaaagcaagcgccatgttga

3' homology arm R: atctatgtcgggtcggagaaaagaggtaatgaaatggcaccgacttatgacttaccccagga

***IL2RG* exon 5 donor construct**

5' homology arm F: taactataacggctctaaggtagcgaagctctgttctctgctcc

5' homology arm R: tagggataacagggtaatgttaaagcggctccgaacag

3' homology arm F: tggcaaacagctattatgggtattatgggtccactctgtggaagtgtcagc

3' homology arm R: atctatgtcgggtcggagaaaagaggtaatgaaatggcaagaccctgcaaacctcctc

Targeting confirmation for RAMOS single cell clones

F: gcagagtggctgtggtaatggaaggaggaaac

R: gctgcgcccttctgtgacgt

Figure 3.14: List of primers used in Chapter III (5' – 3')

CHAPTER IV:

CONCLUSIONS AND FUTURE DIRECTIONS

As our knowledge of the biological underpinnings of genetic diseases has grown, so too has our ability to create ever more effective treatments for the patients who suffer from them. The emergence of gene targeting technologies creates, for the first time, the opportunity to attack these diseases at their source by precisely correcting disease-causing mutations and restoring function in a patient's own cells. The potential for these technologies to alter the way that we approach many diseases is vast. Gene targeting-based strategies for primary immunodeficiencies like SCID-X1, hemoglobinopathies like sickle-cell disease, infectious agents like HIV, bleeding disorders like hemophilia B, and other diseases have already been presented, and many more are under development. In addition to providing new cures, these therapies could replace expensive and complicated life-long treatment strategies with a single intervention, decreasing the burden of treatment-related side effects and the total cost of treatment by hundreds of thousands of dollars over the life of a patient. With further development of these strategies to make them safer and more effective, gene targeting may one day allow patients to experience the ultimate goal of gene therapy, a life without the burden of genetic disease.

Medical application of gene targeting: A paradigm shift for genetic therapies

Application of gene targeting technologies to the treatment of human disease

Gene therapy trials over the past two decades have focused on providing a functional copy of a gene through imprecise introduction of gene expression vectors throughout the genome. Fundamentally, these strategies seek to answer a question of immense importance when a critical gene is dysfunctional: “How can we express this gene?” While this is a logical first question in the development of genetic therapies, the capacities of gene targeting are extending far beyond this basic problem to more and more complex modifications. Driven by applications in basic science as well as clinical translation, gene targeting technologies have advanced at a truly extraordinary pace.

From the first demonstration of an engineered nuclease for a human gene in 2005, advances in ZFN design and the introduction of TALENs and CRISPR-Cas9 nucleases have seen targeted genome modification achieved in hundreds of genes in more than 15 organisms (Gaj et al., 2013; Jinek et al., 2012; Joung et al., 2013; Sun et al., 2013). TALENs targeting novel genomic sites can be made and characterized in only a few weeks, and sets of CRISPRs capable of simultaneously inducing modifications at multiple genes can be synthesized in a matter of days (Cong et al., 2013; Mali et al., 2013; Reyon et al., 2012). While

CRISPRs were initially shown to have low specificities, the rapid adaptation of technologies developed for ZFNs and TALENs allowed multiple groups to improve the specificity of Cas9-mediated editing in only matter of months (Fu et al., 2013; Mali et al., 2013; Ran et al., 2013). To determine if CRISPR-Cas9 nucleases provided a benefit in efficacy or toxicity I tested CRISPRs targeting *IL2RG* in the K562 cell line and CD34+ hematopoietic stem and progenitor cells (HSPCs) (**Figure 4.1**). The utility of the CRISPRs for basic research was illustrated by the fact that they stimulated higher frequencies of gene targeting than TALENs in K562 cells, but for unknown reasons the CRISPRs showed almost no activity in CD34+ HSPCs.

While much of this work is designed for application in basic science, the complexity and range of these genomic alterations holds great promise for the clinic. Far from simply inserting a functional replacement of a gene, genomic engineers are now capable of targeting non-monogenic diseases with techniques such as the reprogramming of T cells to generate tumor-specific lymphocytes (Provasti et al., 2012) and the stacking of genetic resistance to protect against HIV infection (Voit et al., 2013). While increased complexity will undoubtedly create new hurdles for clinical translation, the scope and precision of these genetic modifications is shifting the paradigm for gene therapy. Through gene targeting with engineered nucleases, scientists are now beginning to ask a new and incredibly powerful question: “What do we want the genome to look like?”

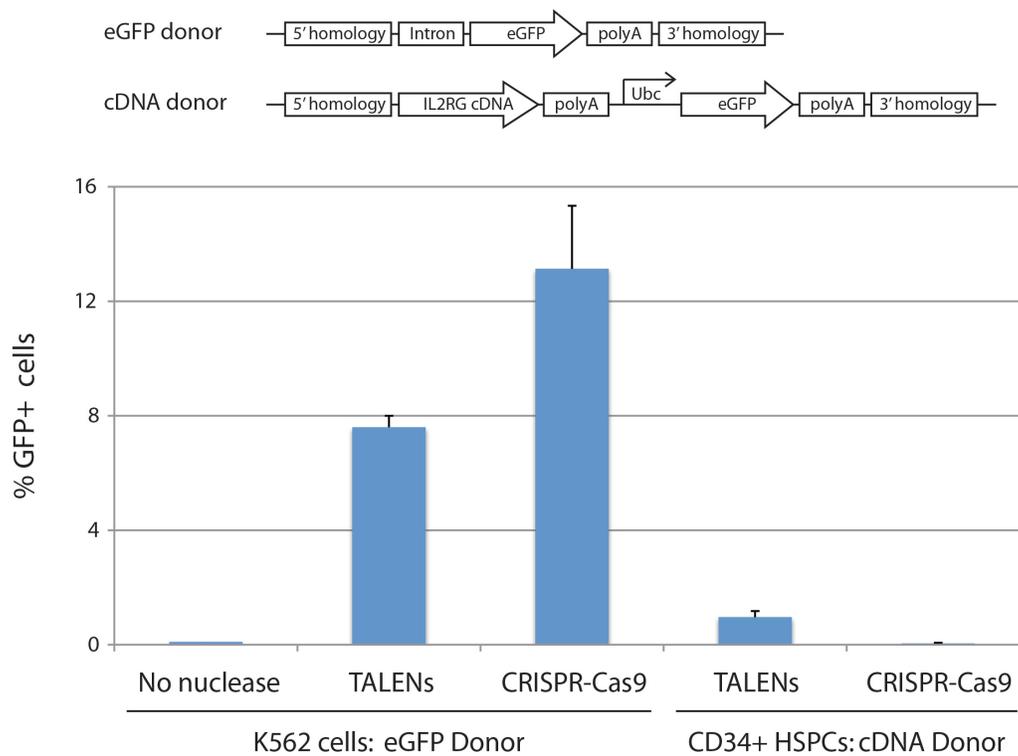


Figure 4.1: Gene targeting at *IL2RG* exon 1 with CRISPR-Cas9 nucleases. A CRISPR-Cas9 nuclease targeting a site near the TALEN target site in *IL2RG* exon 1 was synthesized and tested for activity in K562 cells and CD34+ HSPCs. This CRISPR-Cas9 nuclease was designed and synthesized by Niraj Punjya.

Development of a safe and effective cure for SCID-X1

In this thesis, I report my contributions to the field of gene therapy with the development of a novel strategy for the correction of SCID-X1 and a valuable method for measuring gene editing events at endogenous loci. As one of the first diseases successfully treated by gene therapy and the very first target of an engineered nuclease to a human gene, SCID-X1 has a unique place in the history of gene therapy. Despite the fact that highly active ZFNs targeting *IL2RG* were reported eight years ago, functional correction in SCID-X1 patient hematopoietic stem cells has not been demonstrated. The gene targeting strategy presented in Chapter III of this thesis advances the potential for gene therapy of SCID-X1 in three important ways.

First, I describe a pair of highly active TALENs that precisely target the start codon for *IL2RG* in exon 1. This is the first demonstration of TALEN-mediated targeting of the *IL2RG* gene. These TALENs are capable of modifying over 70% of *IL2RG* alleles in K562 cell populations and stimulate precise gene addition at *IL2RG* exon 1 in over 20% of cells. This level of activity is on par with the most active engineered nucleases ever reported. In addition, these TALENs are significantly less toxic than the previously reported *IL2RG*-specific ZFNs and *CCR5*-specific ZFNs that are currently being used in clinical trials.

Second, I present a novel strategy for targeting full *IL2RG* cDNA to the *IL2RG* start codon. This strategy combines the general applicability of gene

addition at 'safe-harbor' loci with restoration of endogenously-regulated gene expression, and has the potential to be effective for over 98% of SCID-X1 patients. A potential pitfall of this approach is the loss of downstream regulatory elements, such as introns, and indeed the addition of WT cDNA using this strategy resulted in IL2R γ activity 40-60% below normal levels. Utilizing the addition of an artificial intron and cDNA codon-optimization, I was able to overcome this hurdle and restore IL2R γ activity levels to those seen in WT cells.

Third, I report a significant improvement in the efficacy of gene targeting in CD34⁺ hematopoietic stem and progenitor cells (HSPCs). While high frequency gene modification in cell lines has become relatively common in recent years, the ability to efficiently modify CD34⁺ HSPCs remains one of the most significant obstacles to clinical translation of gene targeting. Using the *IL2RG* exon 1-specific TALENs, I was able to modify 5-10% of *IL2RG* alleles in CD34⁺ HSPC populations and stimulate precise addition of *IL2RG* cDNA to *IL2RG* exon 1 in over 1% of cells. This represents an over ten-fold improvement to the highest frequencies of gene targeting previously reported in CD34⁺ HSPCs of ~0.1%.

The high activity and relatively low toxicity of this TALEN-mediated gene therapy strategy represent significant improvements in gene targeting for SCID-X1, but *in vivo* demonstration of this approach has proven challenging. With the finding that physiologically-relevant levels of IL2R γ activity are achieved following targeting with this strategy, the generation of targeted CD34⁺

HSPCs capable of *in vivo* functionality stands as the final preclinical hurdle for the translation of this approach.

Overcoming toxicity in hematopoietic stem cells: The path to precise gene therapy for SCID-X1

Despite all of the advances and improvements in gene targeting technology over the past decade, *in vivo* demonstration of gene correction in hematopoietic stem cells is yet to be achieved. The ability of the *IL2RG* exon 1-specific TALENs presented here to stimulate gene targeting in > 1% of CD34+ HSPCs has proven robust, but experiments attempting to engraft targeted stem cells in mice have been unsuccessful. Interestingly, control samples of CD34+ HSPCs transfected with only a GFP expression plasmid have successfully engrafted in NSG mice (collaboration with Adam Hartigan, data not shown). This suggests that it is the gene targeting process itself, or the transfection conditions necessary for efficient gene targeting, that is responsible for unacceptably high toxicity, and not the general process of nucleofection with plasmid DNA.

To overcome this challenge, I have begun testing multiple strategies for decreasing the toxicity of gene targeting. One method that has been demonstrated to reduce total nuclease expression levels and provide efficient targeting in hiPSCs is the delivery of TALENs with mRNA instead of plasmid DNA. mRNA delivery of my *IL2RG* exon 1-specific TALENs was found to successfully

stimulate similar levels of targeted gene addition to those achieved with plasmid DNA, but also produced similar levels of toxicity in CD34+ HSPCs (**Figure 4.2**). The most likely explanation for this finding is that the majority of the toxicity in this approach is generated by the sheer amount of donor template plasmid DNA that is necessary for optimal levels of gene targeting. In support of this conclusion, samples of CD34+ HSPCs transfected with only donor template were found to have similar overt toxicity levels as samples transfected with donor template and TALENs.

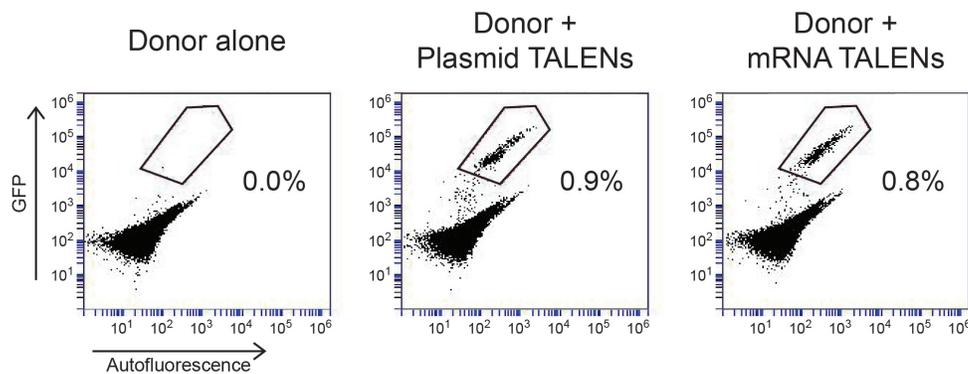


Figure 4.2: Gene targeting in CD34+ HSPCs with mRNA delivery of TALENs. CD34+ HSPCs were nucleofected with donor template and either 250ng of each high-expression TALEN plasmid or 2 μ g of mRNA encoding each TALEN. The mRNA used in this experiment was synthesized by Joseph Clark.

A potential strategy for reducing plasmid-related toxicity is to deliver the donor template as miniVector DNA. MiniVectors are double-stranded DNA molecules produced by removing the bacterial backbone elements from plasmids. MiniVectors have been shown to provide improved expression and stability in human stem cells compared to plasmids due to reduced gene silencing (Jia et al., 2010; Lu et al., 2013). For gene targeting, miniVector DNA could provide an additional advantage over plasmids by reducing the total size of the donor template molecule and therefore allowing more copies of the donor template to be delivered for a given amount of transfected DNA. To determine the potential for miniVectors to function as donor templates, I compared gene targeting frequencies in K562 cells transfected with *IL2RG* exon 1-specific TALENs and either miniVector or plasmid donor template DNA (**Figure 4.3**). This experiment, which is the first demonstration of miniVector DNA being used for gene targeting, revealed that miniVector DNA is capable of functioning as a donor template in gene targeting at efficiencies slightly below that of plasmid DNA. Despite a small decrease in gene targeting efficiency at maximal levels of donor template with miniVector DNA, a more than 3-fold decrease in the total amount of miniVector DNA did not decrease targeting efficiency and provided ~70% of the modification attained with the molar equivalent of plasmid DNA.

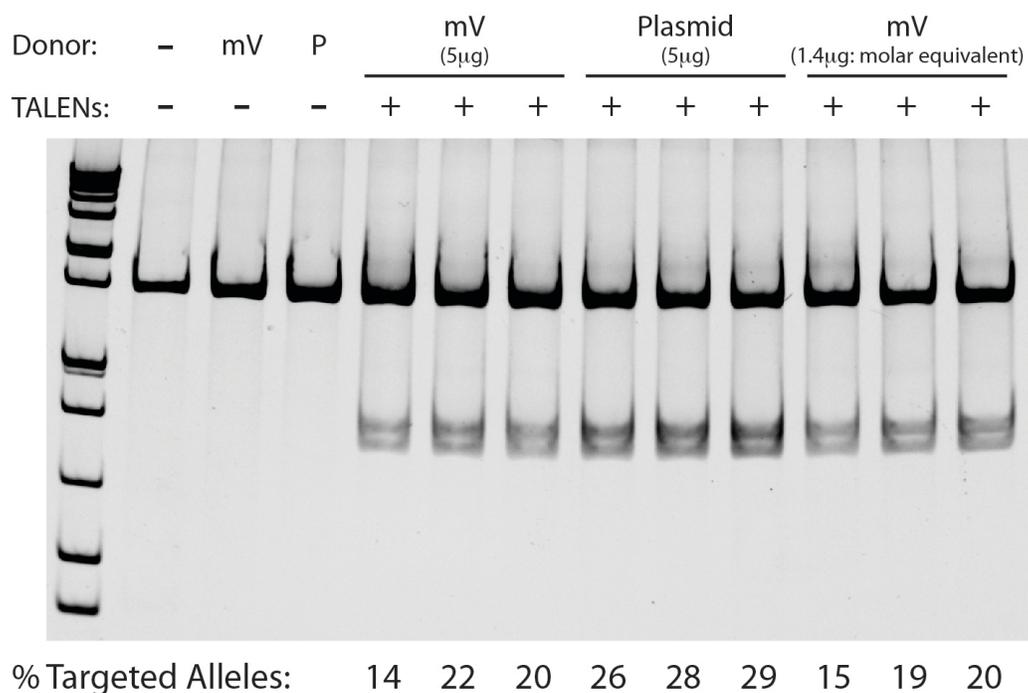


Figure 4.3: Gene targeting using miniVector DNA as a donor template. K562 cells were nucleofected with either miniVector DNA or plasmid DNA donor templates and *IL2RG* exon 1-specific TALENs. Precise integration of a restriction site present in the donor was measured with the RFLP assay. The miniVector DNA used in this experiment was provided by Twister Biotech, Inc.

While these strategies have yet to decrease the toxicity of gene targeting in CD34+ HSPCs below the threshold for *in vivo* engraftment, further development of these and other toxicity-reducing strategies could allow for *in vivo* demonstration of gene targeting in hematopoietic stem cells in the near future. Interestingly, observations from these experiments suggest that for this TALEN-mediated *IL2RG* targeting strategy it is the amount of donor template, and not the engineered nucleases, that is the primary contributor to high levels of toxicity. The fact that CD34+ HSPCs modified with more toxic *CCR5* ZFNs are capable of engraftment supports the conclusion that it is the addition of the donor template for gene targeting, and not nuclease-mediated DSBs, that is the primary hurdle for clinical translation.

Gene targeting: A new paradigm for gene therapy

The purposeful generation of DSBs with engineered nucleases seemingly creates a requirement for genotoxicity with gene targeting, but it is precisely the potential for safety with this approach that has driven this field forward. Precise genetic modification with nucleases optimized for specificity has the potential to change the principal risk of gene therapy from the permanent insertion of viral vectors throughout the genome to the generation of a small number of double-strand breaks over a period of a few days. Even the newest SIN-lentiviral vectors

that hold great promise for improving the safety of gene therapy do so while integrating at over 35,000 sites throughout the genome (Aiuti et al., 2013). In light of this, the fact that using a nuclease to prevent genotoxicity is ironic does nothing to diminish its appeal.

Gene targeting strategies are shifting the paradigm for gene therapy by replacing pseudo-random addition of transgenes with precise and potentially complex genomic modifications. Therapeutic strategies like the one presented here for SCID-X1 have the potential to revolutionize the way monogenic disorders and other diseases are approached clinically. As a demonstration of the elegance of potential gene targeting therapeutics, a mouse model of hemophilia B was recently corrected with one injection of a hepatotropic virus encoding nucleases and a donor template specific for the gene encoding blood coagulation factor IX (Li et al., 2011). If a safe CD34-tropic delivery method becomes available, the *IL2RG*-specific targeting system presented here could one day provide a life-long cure for patients suffering from SCID-X1 with as little clinical intervention as a single injection.

Overall conclusions:

50 years ago, as of the presentation of this thesis, Marshall Nirenberg did not yet know the entirety of the genetic code and Har Gobind Khorana was busy describing the first synthesis of an oligonucleotide. Today, precise modification of the genome to correct disease phenotypes can be achieved in human stem cells and 85% of SCID-X1 patients treated in gene therapy trials have experienced correction of their immunodeficiency. In this thesis, I have presented a strategy for precisely correcting the underlying deficit causing SCID-X1 in the same CD34+ hematopoietic stem cells whose dysfunction causes this disease. As the translation of the precise tools of gene targeting to the clinical world of gene therapy continues, the work presented here could form the foundation for a safe and effective cure for SCID-X1. 50 years from now, when a child is diagnosed with SCID-X1, a new standard of care based on basic scientific findings including those presented here may allow the doctor to inform the parents simply that their child will need one extra injection, and that the rest of their immunizations will have to be delayed for 3-6 months.

BIBLIOGRAPHY

- Aiuti A, Biasco L, Scaramuzza S, Ferrua F, Cicalese MP, Baricordi C, Dionisio F, Calabria A, Giannelli S, Castiello MC *et al.* Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome. *Science* 2013; 341:1233-151.
- Aiuti A, Cattaneo F, Galimberti S, Benninghoff U, Cassani B, Callegaro L, Scaramuzza S, Andolfi G, Mirolo M, Brigida I *et al.* Gene therapy for immunodeficiency due to adenosine deaminase deficiency. *N Engl J Med* 2009; 360:447-58.
- Aiuti A, Slavin S, Aker M, Ficara F, Deola S, Mortellaro A, Morecki S, Andolfi G, Tabucchi A, Carlucci F *et al.* Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. *Science* 2002; 296:2410-3.
- Astrakhan A, Sather BD, Ryu BY, Khim S, Singh S, Humblet-Baron S, Ochs HD, Miao CH, Rawlings DJ. Ubiquitous high-level gene expression in hematopoietic lineages provides effective lentiviral gene therapy of murine Wiskott-Aldrich syndrome. *Blood* 2012; 119:4395-407.
- Avedillo Diez I, Zychlinski D, Coci EG, Galla M, Modlich U, Dewey RA, Schwarzer A, Maetzig T, Mpofu N, Jaeckel E *et al.* Development of novel efficient SIN vectors with improved safety features for Wiskott-Aldrich syndrome stem cell based gene therapy. *Mol Pharm* 2011; 8:1525-37.
- Beerli RR, Barbas CF, 3rd. Engineering polydactyl zinc-finger transcription factors. *Nat Biotechnol* 2002; 20:135-41.
- Bibikova M, Beumer K, Trautman JK, Carroll D. Enhancing gene targeting with designed zinc finger nucleases. *Science* 2003; 300:764.
- Blaese RM, Culver KW, Miller AD, Carter CS, Fleisher T, Clerici M, Shearer G, Chang L, Chiang Y, Tolstoshev P *et al.* T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years. *Science* 1995; 270:475-80.
- Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 2009; 326:1509-12.

- Bogdanove AJ, Voytas DF. TAL effectors: customizable proteins for DNA targeting. *Science* 2011; 333:1843-6.
- Booth C, Gaspar HB, Thrasher AJ. Gene therapy for primary immunodeficiency. *Curr Opin Pediatr* 2011; 23:659-66.
- Bordignon C, Notarangelo LD, Nobili N, Ferrari G, Casorati G, Panina P, Mazzolari E, Maggioni D, Rossi C, Servida P *et al*. Gene therapy in peripheral blood lymphocytes and bone marrow for ADA-immunodeficient patients. *Science* 1995; 270:470-5.
- Bouso P, Wahn V, Douagi I, Horneff G, Pannetier C, Le Deist F, Zepp F, Niehues T, Kourilsky P, Fischer A *et al*. Diversity, functionality, and stability of the T cell repertoire derived in vivo from a single human T cell precursor. *Proc Natl Acad Sci U S A* 2000; 97:274-8.
- Boztug K, Schmidt M, Schwarzer A, Banerjee PP, Diez IA, Dewey RA, Bohm M, Nowrouzi A, Ball CR, Glimm H *et al*. Stem-cell gene therapy for the Wiskott-Aldrich syndrome. *N Engl J Med* 2010; 363:1918-27.
- Briggs AW, Rios X, Chari R, Yang L, Zhang F, Mali P, Church GM. Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Res* 2012.
- Brown L, Xu-Bayford J, Allwood Z, Slatter M, Cant A, Davies EG, Veys P, Gennery AR, Gaspar HB. Neonatal diagnosis of severe combined immunodeficiency leads to significantly improved survival outcome: the case for newborn screening. *Blood* 2011; 117:3243-6.
- Brunet E, Simsek D, Tomishima M, DeKolver R, Choi VM, Gregory P, Urnov F, Weinstock DM, Jasin M. Chromosomal translocations induced at specified loci in human stem cells. *Proc Natl Acad Sci U S A* 2009; 106:10620-5.
- Buckley RH. The long quest for neonatal screening for severe combined immunodeficiency. *J Allergy Clin Immunol* 2012; 129:597-604; quiz 605-6.
- Buckley RH, Schiff RI, Schiff SE, Markert ML, Williams LW, Harville TO, Roberts JL, Puck JM. Human severe combined immunodeficiency:

genetic, phenotypic, and functional diversity in one hundred eight infants. *J Pediatr* 1997; 130:378-87.

Cannon P, June C. Chemokine receptor 5 knockout strategies. *Curr Opin HIV AIDS* 2011; 6:74-9.

Cattoglio C, Pellin D, Rizzi E, Maruggi G, Corti G, Miselli F, Sartori D, Guffanti A, Di Serio C, Ambrosi A *et al.* High-definition mapping of retroviral integration sites identifies active regulatory elements in human multipotent hematopoietic progenitors. *Blood* 2010; 116:5507-17.

Cavazzana-Calvo M, Andre-Schmutz I, Fischer A. Haematopoietic stem cell transplantation for SCID patients: where do we stand? *Br J Haematol* 2013; 160:146-52.

Cavazzana-Calvo M, Payen E, Negre O, Wang G, Hehir K, Fusil F, Down J, Denaro M, Brady T, Westerman K *et al.* Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia. *Nature* 2010; 467:318-22.

Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* 2011; 39:e82.

Certo MT, Ryu BY, Annis JE, Garibov M, Jarjour J, Rawlings DJ, Scharenberg AM. Tracking genome engineering outcome at individual DNA breakpoints. *Nat Methods* 2011; 8:671-6.

Chen F, Pruett-Miller SM, Huang Y, Gjoka M, Duda K, Taunton J, Collingwood TN, Frodin M, Davis GD. High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat Methods* 2011; 8:753-5.

Chiarle R, Zhang Y, Frock RL, Lewis SM, Molinie B, Ho YJ, Myers DR, Choi VW, Compagno M, Malkin DJ *et al.* Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* 2011; 147:107-19.

Chinen J, Davis J, De Ravin SS, Hay BN, Hsu AP, Linton GF, Naumann N, Nomicos EY, Silvin C, Ulrick J *et al.* Gene therapy improves immune

function in preadolescents with X-linked severe combined immunodeficiency. *Blood* 2007; 110:67-73.

Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 2010; 186:757-61.

Christian ML, Demorest ZL, Starker CG, Osborn MJ, Nyquist MD, Zhang Y, Carlson DF, Bradley P, Bogdanove AJ, Voytas DF. Targeting G with TAL effectors: a comparison of activities of TALENs constructed with NN and NK repeat variable di-residues. *PLoS One* 2012; 7:e45383.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013; 339:819-23.

Connelly JP, Barker JC, Pruett-Miller S, Porteus MH. 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; 18:1103-10.

Cooray S, Howe SJ, Thrasher AJ. Retrovirus and lentivirus vector design and methods of cell conditioning. *Methods Enzymol* 2012; 507:29-57.

Cornu TI, Thibodeau-Beganny S, Guhl E, Alwin S, Eichtinger M, Joung JK, Cathomen T. DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Mol Ther* 2008; 16:352-8.

Deng D, Yan C, Pan X, Mahfouz M, Wang J, Zhu JK, Shi Y, Yan N. Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* 2012; 335:720-3.

Doyon Y, Vo TD, Mendel MC, Greenberg SG, Wang J, Xia DF, Miller JC, Urnov FD, Gregory PD, Holmes MC. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat Methods* 2011; 8:74-9.

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* 2009; 323:133-8.

- Fischer A. Severe combined immunodeficiencies (SCID). *Clin Exp Immunol* 2000; 122:143-9.
- Fischer A, Hacein-Bey-Abina S, Cavazzana-Calvo M. 20 years of gene therapy for SCID. *Nat Immunol* 2010; 11:457-60.
- Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* 2013; 31:822-6.
- Gabriel R, Lombardo A, Arens A, Miller JC, Genovese P, Kaepffel C, Nowrouzi A, Bartholomae CC, Wang J, Friedman G *et al.* An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol* 2011; 29:816-23.
- Gaj T, Gersbach CA, Barbas CF, 3rd. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* 2013; 31:397-405.
- Gaspar HB, Cooray S, Gilmour KC, Parsley KL, Adams S, Howe SJ, Al Ghonaium A, Bayford J, Brown L, Davies EG *et al.* Long-term persistence of a polyclonal T cell repertoire after gene therapy for X-linked severe combined immunodeficiency. *Sci Transl Med* 2011; 3:97ra79.
- Gaspar HB, Cooray S, Gilmour KC, Parsley KL, Zhang F, Adams S, Bjorkegren E, Bayford J, Brown L, Davies EG *et al.* Hematopoietic stem cell gene therapy for adenosine deaminase-deficient severe combined immunodeficiency leads to long-term immunological recovery and metabolic correction. *Sci Transl Med* 2011; 3:97ra80.
- Gaspar HB, Parsley KL, Howe S, King D, Gilmour KC, Sinclair J, Brouns G, Schmidt M, Von Kalle C, Barington T *et al.* Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. *Lancet* 2004; 364:2181-7.
- Greisman HA, Pabo CO. A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* 1997; 275:657-61.
- Guschin DY, Waite AJ, Katibah GE, Miller JC, Holmes MC, Rebar EJ. A rapid and general assay for monitoring endogenous gene modification. *Methods Mol Biol* 2010; 649:247-56.

- Hacein-Bey-Abina S, Garrigue A, Wang GP, Soulier J, Lim A, Morillon E, Clappier E, Caccavelli L, Delabesse E, Beldjord K *et al.* Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest* 2008; 118:3132-42.
- Hacein-Bey-Abina S, Hauer J, Lim A, Picard C, Wang GP, Berry CC, Martinache C, Rieux-Laucat F, Latour S, Belohradsky BH *et al.* Efficacy of gene therapy for X-linked severe combined immunodeficiency. *N Engl J Med* 2010; 363:355-64.
- Hacein-Bey-Abina S, Le Deist F, Carlier F, Bouneaud C, Hue C, De Villartay JP, Thrasher AJ, Wulffraat N, Sorensen R, Dupuis-Girod S *et al.* Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy. *N Engl J Med* 2002; 346:1185-93.
- Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Leboulch P, Lim A, Osborne CS, Pawliuk R, Morillon E *et al.* LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* 2003; 302:415-9.
- Hirschhorn R, Yang DR, Puck JM, Huie ML, Jiang CK, Kurlandsky LE. Spontaneous in vivo reversion to normal of an inherited mutation in a patient with adenosine deaminase deficiency. *Nat Genet* 1996; 13:290-5.
- Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, Cost GJ, Zhang L, Santiago Y, Miller JC *et al.* Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol* 2011; 29:731-4.
- Holt N, Wang J, Kim K, Friedman G, Wang X, Taupin V, Crooks GM, Kohn DB, Gregory PD, Holmes MC *et al.* Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. *Nat Biotechnol* 2010; 28:839-47.
- Howe SJ, Mansour MR, Schwarzwaelder K, Bartholomae C, Hubank M, Kempinski H, Brugman MH, Pike-Overzet K, Chatters SJ, de Ridder D *et al.* Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J Clin Invest* 2008; 118:3143-50.

- Hunter MJ, Tuschong LM, Fowler CJ, Bauer TR, Jr., Burkholder TH, Hickstein DD. Gene therapy of canine leukocyte adhesion deficiency using lentiviral vectors with human CD11b and CD18 promoters driving canine CD18 expression. *Mol Ther* 2011; 19:113-21.
- Hurt JA, Thibodeau SA, Hirsh AS, Pabo CO, Joung JK. Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc Natl Acad Sci U S A* 2003; 100:12271-6.
- Huston MW, van Til NP, Visser TP, Arshad S, Brugman MH, Cattoglio C, Nowrouzi A, Li Y, Schambach A, Schmidt M *et al.* Correction of murine SCID-X1 by lentiviral gene therapy using a codon-optimized IL2RG gene and minimal pretransplant conditioning. *Mol Ther* 2011; 19:1867-77.
- Isalan M, Choo Y, Klug A. Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proc Natl Acad Sci U S A* 1997; 94:5617-21.
- Jia F, Wilson KD, Sun N, Gupta DM, Huang M, Li Z, Panetta NJ, Chen ZY, Robbins RC, Kay MA *et al.* A nonviral minicircle vector for deriving human iPS cells. *Nat Methods* 2010; 7:197-9.
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012; 337:816-21.
- Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* 2013; 14:49-55.
- Junttila IS, Creusot RJ, Moraga I, Bates DL, Wong MT, Alonso MN, Suhoski MM, Lupardus P, Meier-Schellersheim M, Engleman EG *et al.* Redirecting cell-type specific cytokine responses with engineered interleukin-4 superkines. *Nat Chem Biol* 2012; 8:990-8.
- Kang EM, Choi U, Theobald N, Linton G, Long Priel DA, Kuhns D, Malech HL. Retrovirus gene therapy for X-linked chronic granulomatous disease can achieve stable long-term correction of oxidase activity in peripheral blood neutrophils. *Blood* 2010; 115:783-91.
- Kang HJ, Bartholomae CC, Paruzynski A, Arens A, Kim S, Yu SS, Hong Y, Joo CW, Yoon NK, Rhim JW *et al.* Retroviral gene therapy for X-linked

chronic granulomatous disease: results from phase I/II trial. *Mol Ther* 2011; 19:2092-101.

Kim YG, Cha J, Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A* 1996; 93:1156-60.

Krutzik PO, Nolan GP. Intracellular phospho-protein staining techniques for flow cytometry: monitoring single cell signaling events. *Cytometry A* 2003; 55:61-70.

Li H, Haurigot V, Doyon Y, Li T, Wong SY, Bhagwat AS, Malani N, Anguela XM, Sharma R, Ivanciu L *et al.* In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature* 2011; 475:217-21.

Liu H, Agarwal S, Kmiec E, Davis BR. Targeted beta-globin gene conversion in human hematopoietic CD34(+) and Lin(-)CD38(-) cells. *Gene Ther* 2002; 9:118-26.

Lombardo A, Cesana D, Genovese P, Di Stefano B, Provasi E, Colombo DF, Neri M, Magnani Z, Cantore A, Lo Riso P *et al.* Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat Methods* 2011; 8:861-9.

Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, Ando D, Urnov FD, Galli C, Gregory PD *et al.* Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol* 2007; 25:1298-306.

Lu J, Zhang F, Kay MA. A mini-intronic plasmid (MIP): a novel robust transgene expression vector in vivo and in vitro. *Mol Ther* 2013; 21:954-63.

Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, Eichinger M, Jiang T, Foley JE, Winfrey RJ, Townsend JA *et al.* Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell* 2008; 31:294-301.

Mak AN, Bradley P, Cernadas RA, Bogdanove AJ, Stoddard BL. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* 2012; 335:716-9.

- Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, Yang L, Church GM. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* 2013; 31:833-8.
- Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. RNA-guided human genome engineering via Cas9. *Science* 2013; 339:823-6.
- McMahon MA, Rahdar M, Porteus M. Gene editing: not just for translation anymore. *Nat Methods* 2012; 9:28-31.
- Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ *et al.* A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* 2011; 29:143-8.
- Moreno-Carranza B, Gentsch M, Stein S, Schambach A, Santilli G, Rudolf E, Ryser MF, Haria S, Thrasher AJ, Baum C *et al.* Transgene optimization significantly improves SIN vector titers, gp91phox expression and reconstitution of superoxide production in X-CGD cells. *Gene Ther* 2009; 16:111-8.
- Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. *Science* 2009; 326:1501.
- Mukherjee S, Thrasher AJ. Gene therapy for PIDs: progress, pitfalls and prospects. *Gene* 2013; 525:174-81.
- Munoz IG, Prieto J, Subramanian S, Coloma J, Redondo P, Villate M, Merino N, Marenchino M, D'Abramo M, Gervasio FL *et al.* Molecular basis of engineered meganuclease targeting of the endogenous human RAG1 locus. *Nucleic Acids Res* 2011; 39:729-43.
- Mussolino C, Cathomen T. TALE nucleases: tailored genome engineering made easy. *Curr Opin Biotechnol* 2012.
- Mussolino C, Morbitzer R, Lutge F, Dannemann N, Lahaye T, Cathomen T. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res* 2011; 39:9283-93.

- Naldini L. Ex vivo gene transfer and correction for cell-based therapies. *Nat Rev Genet* 2011; 12:301-15.
- Ng YY, Baert MR, Pike-Overzet K, Rodijk M, Brugman MH, Schambach A, Baum C, Hendriks RW, van Dongen JJ, Staal FJ. Correction of B-cell development in Btk-deficient mice using lentiviral vectors with codon-optimized human BTK. *Leukemia* 2010; 24:1617-30.
- Ott MG, Schmidt M, Schwarzwaelder K, Stein S, Siler U, Koehl U, Glimm H, Kuhlcke K, Schilz A, Kunkel H *et al.* Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nat Med* 2006; 12:401-9.
- Pattanayak V, Ramirez CL, Joung JK, Liu DR. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Methods* 2011; 8:765-70.
- Perez EE, Wang J, Miller JC, Jouvenot Y, Kim KA, Liu O, Wang N, Lee G, Bartsevich VV, Lee YL *et al.* Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol* 2008; 26:808-16.
- Pike-Overzet K, Rodijk M, Ng YY, Baert MR, Lagresle-Peyrou C, Schambach A, Zhang F, Hoeben RC, Hacein-Bey-Abina S, Lankester AC *et al.* Correction of murine Rag1 deficiency by self-inactivating lentiviral vector-mediated gene transfer. *Leukemia* 2011; 25:1471-83.
- Porteus MH, Baltimore D. Chimeric nucleases stimulate gene targeting in human cells. *Science* 2003; 300:763.
- Porteus MH, Carroll D. Gene targeting using zinc finger nucleases. *Nat Biotechnol* 2005; 23:967-73.
- Porteus MH, Cathomen T, Weitzman MD, Baltimore D. Efficient gene targeting mediated by adeno-associated virus and DNA double-strand breaks. *Mol Cell Biol* 2003; 23:3558-65.
- Provasi E, Genovese P, Lombardo A, Magnani Z, Liu PQ, Reik A, Chu V, Paschon DE, Zhang L, Kuball J *et al.* Editing T cell specificity towards

leukemia by zinc finger nucleases and lentiviral gene transfer. *Nat Med* 2012; 18:807-15.

Pruett-Miller SM, Connelly JP, Maeder ML, Joung JK, Porteus MH. Comparison of zinc finger nucleases for use in gene targeting in mammalian cells. *Mol Ther* 2008; 16:707-17.

Pruett-Miller SM, Reading DW, Porter SN, Porteus MH. Attenuation of zinc finger nuclease toxicity by small-molecule regulation of protein levels. *PLoS Genet* 2009; 5:e1000376.

Puck JM, Deschenes SM, Porter JC, Dutra AS, Brown CJ, Willard HF, Henthorn PS. The interleukin-2 receptor gamma chain maps to Xq13.1 and is mutated in X-linked severe combined immunodeficiency, SCIDX1. *Hum Mol Genet* 1993; 2:1099-104.

Qasim W, Gaspar HB, Thrasher AJ. Progress and prospects: gene therapy for inherited immunodeficiencies. *Gene Ther* 2009; 16:1285-91.

Qi Y, Zhang Y, Zhang F, Baller JA, Cleland SC, Ryu Y, Starker CG, Voytas DF. Increasing frequencies of site-specific mutagenesis and gene targeting in *Arabidopsis* by manipulating DNA repair pathways. *Genome Res* 2013; 23:547-54.

Ramirez CL, Foley JE, Wright DA, Muller-Lerch F, Rahman SH, Cornu TI, Winfrey RJ, Sander JD, Fu F, Townsend JA *et al.* Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat Methods* 2008; 5:374-5.

Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y *et al.* Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* 2013.

Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, Paxinos EE, Sebra R, Chin CS, Iliopoulos D *et al.* Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N Engl J Med* 2011; 365:709-17.

Reyon D, Tsai SQ, Khayter C, Foden JA, Sander JD, Joung JK. FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* 2012; 30:460-5.

- Rivat C, Santilli G, Gaspar HB, Thrasher AJ. Gene therapy for primary immunodeficiencies. *Hum Gene Ther* 2012; 23:668-75.
- Roberts RJ, Carneiro MO, Schatz MC. The advantages of SMRT sequencing. *Genome Biol* 2013; 14:405.
- Rouet P, Smih F, Jasin M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc Natl Acad Sci U S A* 1994; 91:6064-8.
- Sander JD, Dahlborg EJ, Goodwin MJ, Cade L, Zhang F, Cifuentes D, Curtin SJ, Blackburn JS, Thibodeau-Beganny S, Qi Y *et al.* Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat Methods* 2011; 8:67-9.
- Santilli G, Almarza E, Brendel C, Choi U, Beilin C, Blundell MP, Haria S, Parsley KL, Kinnon C, Malech HL *et al.* Biochemical correction of X-CGD by a novel chimeric promoter regulating high levels of transgene expression in myeloid cells. *Mol Ther* 2011; 19:122-32.
- Scaramuzza S, Biasco L, Ripamonti A, Castiello MC, Loperfido M, Draghici E, Hernandez RJ, Benedicenti F, Radrizzani M, Salomoni M *et al.* Preclinical safety and efficacy of human CD34(+) cells transduced with lentiviral vector for the treatment of Wiskott-Aldrich syndrome. *Mol Ther* 2013; 21:175-84.
- Segal DJ, Beerli RR, Blancafort P, Dreier B, Effertz K, Huber A, Koksich B, Lund CV, Maguenat L, Valente D *et al.* Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry* 2003; 42:2137-48.
- Shaw KL, Kohn DB. A tale of two SCIDs. *Sci Transl Med* 2011; 3:97ps36.
- Silva G, Poirot L, Galetto R, Smith J, Montoya G, Duchateau P, Paques F. Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Curr Gene Ther* 2011; 11:11-27.

- Stark JM, Pierce AJ, Oh J, Pastink A, Jasin M. Genetic steps of mammalian homologous repair with distinct mutagenic consequences. *Mol Cell Biol* 2004; 24:9305-16.
- Stein S, Ott MG, Schultze-Strasser S, Jauch A, Burwinkel B, Kinner A, Schmidt M, Kramer A, Schwable J, Glimm H *et al.* Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. *Nat Med* 2010; 16:198-204.
- Stephan V, Wahn V, Le Deist F, Dirksen U, Broker B, Muller-Fleckenstein I, Horneff G, Schroten H, Fischer A, de Saint Basile G. Atypical X-linked severe combined immunodeficiency due to possible spontaneous reversion of the genetic defect in T cells. *N Engl J Med* 1996; 335:1563-7.
- Streubel J, Blucher C, Landgraf A, Boch J. TAL effector RVD specificities and efficiencies. *Nat Biotechnol* 2012; 30:593-5.
- Sun N, Liang J, Abil Z, Zhao H. Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. *Mol Biosyst* 2012; 8:1255-63.
- Sun N, Zhao H. Transcription activator-like effector nucleases (TALENs): a highly efficient and versatile tool for genome editing. *Biotechnol Bioeng* 2013; 110:1811-21.
- Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res* 2010; 38:e159.
- Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S, Jamieson AC, Porteus MH, Gregory PD, Holmes MC. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 2005; 435:646-51.
- van Rensburg R, Beyer I, Yao XY, Wang H, Denisenko O, Li ZY, Russell DW, Miller DG, Gregory P, Holmes M *et al.* Chromatin structure of two genomic sites for targeted transgene integration in induced pluripotent stem cells and hematopoietic stem cells. *Gene Ther* 2012.
- van Til NP, de Boer H, Mashamba N, Wabik A, Huston M, Visser TP, Fontana E, Poliani PL, Cassani B, Zhang F *et al.* Correction of murine Rag2 severe

combined immunodeficiency by lentiviral gene therapy using a codon-optimized RAG2 therapeutic transgene. *Mol Ther* 2012; 20:1968-80.

Voit RA, McMahon MA, Sawyer SL, Porteus MH. Generation of an HIV resistant T-cell line by targeted "stacking" of restriction factors. *Mol Ther* 2013; 21:786-95.

Wood AJ, Lo TW, Zeitler B, Pickle CS, Ralston EJ, Lee AH, Amora R, Miller JC, Leung E, Meng X *et al.* Targeted genome editing across species using ZFNs and TALENs. *Science* 2011; 333:307.

Yang L, Guell M, Byrne S, Yang JL, De Los Angeles A, Mali P, Aach J, Kim-Kiselak C, Briggs AW, Rios X *et al.* Optimization of scarless human stem cell genome editing. *Nucleic Acids Res* 2013.

Zhang F, Frost AR, Blundell MP, Bales O, Antoniou MN, Thrasher AJ. A ubiquitous chromatin opening element (UCOE) confers resistance to DNA methylation-mediated silencing of lentiviral vectors. *Mol Ther* 2010; 18:1640-9.

Zou J, Maeder ML, Mali P, Pruett-Miller SM, Thibodeau-Beganny S, Chou BK, Chen G, Ye Z, Park IH, Daley GQ *et al.* Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell* 2009; 5:97-110.

Zou J, Sweeney CL, Chou BK, Choi U, Pan J, Wang H, Dowey SN, Cheng L, Malech HL. Oxidase-deficient neutrophils from X-linked chronic granulomatous disease iPS cells: functional correction by zinc finger nuclease-mediated safe harbor targeting. *Blood* 2011; 117:5561-72.