GENERATION OF HIV-RESISTANT T-CELLS AND CORRECTION OF THE SICKLE CELL MUTATION BY TARGETED GENOME ENGINEERING

APPROVED BY SUPERVISORY COMMITTEE

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

_____
Joel Goodman, Ph.D. (Chair)

_____
Paul Sternweis, Ph.D.

_____
Jon Graff, M.D.,Ph.D.

DEDICATION

To Sijy, my inspiration, my love

GENERATION OF HIV-RESISTANT T-CELLS AND CORRECTION OF THE SICKLE CELL MUTATION BY TARGETED GENOME ENGINEERING

by

RICHARD ALEXANDER VOIT

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

May 2014

GENERATION OF HIV-RESISTANT T-CELLS AND CORRECTION OF THE SICKLE CELL MUTATION BY TARGETED GENOME ENGINEERING

RICHARD ALEXANDER VOIT, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2014

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

Targeted genome engineering is a powerful method to create specific modifications at chromosomal loci. This technique makes it feasible to precisely alter DNA sequences by introducing a specific DNA double-strand break, which is repaired by the natural cellular machinery. These double strand breaks are induced by engineered chimeric nucleases – either zinc finger nucleases (ZFNs)

or Tal effector nucleases (TALENs) – and depending on the experimental design, can result in gene disruption, gene correction or targeted transgene integration. In this thesis, I present two applications of this approach in the context of two prevalent human diseases, HIV infection and sickle cell disease.

HIV infects CD4+ T-cells by binding to the CD4 receptor and either the CCR5 or CXCR4 co-receptor on the surface of those cells. Previously, ZFNs were described that create gene specific knockouts of *CCR5*, protecting cells against CCR5-tropic (R5) HIV, but not against CXCR4-tropic (X4) HIV. I hypothesized that combining ZFN-mediated *CCR5* disruption with targeted integration of a cassette of anti-HIV genes would confer higher levels of resistance against R5-tropic virus and also be protective against X4-tropic virus. In a T-cell reporter line, I showed that *CCR5* disruption alone conferred 16-fold protection against R5-tropic virus but had no effect against X4-tropic HIV. In contrast, *CCR5* disruption, combined with targeted gene integration into that locus, of the anti-HIV restriction factors human-rhesus hybrid TRIM5α, APOBEC3G D128K and Rev M10 was completely protective against both viral tropisms.

Sickle cell disease is caused by a point mutation in the β-globin gene, and I sought to correct this mutation by synthesizing TALENs specific for that site. The β-globin TALENs stimulated integration of therapeutic β-globin cDNA in approximately 20% of cells prior to selection. Using FDA-approved drugs to select for modified cells, I showed virtually complete enrichment of targeted cells.

Furthermore, I used the β-globin TALENs to target GFP to the β-globin start codon and designed TALENs to target tdTomato to the start codon of the γ-globin gene, upregulation of which is a goal of sickle cell disease pharmacotherapy. In this way, I developed an endogenous dual promoter reporter system and screened for drugs that preferentially upregulated γ-globin.

TABLE OF CONTENTS

PRIOR PUBLICATIONS

Voit RA, Barker JC, Porteus MH. (2013) Nuclease-mediated targeted genome modification in mammalian cells. *Site-directed insertion of transgenes.* Springer. 327-52

Voit RA, McMahon MA, Sawyer S, Porteus MH. (2013) HIV-resistant T-cells generated by targeted "stacking" of restriction factors. *Molecular Therapy.* 21(4): 786-95.

Voit RA, Hendel A, Pruett-Miller SM, Porteus MH. (2013) Therapeutic gene editing and the generation of endogenous fluorescent reporters at the human globin loci. (Submitted).

Lin Y, Fine E, **Voit RA**, Porteus MH, Craddick TJ, Bao, G. (2013) A new design tool for predicting TALE Nuclease activity. (Submitted).

Bauer DE, Kamran SC, Lessard S, Xu J, Fujiwara Y, Lin C, Shao Z, Canver MC, Smith EC, Pinello L, Sabo P, Vierstra J, **Voit RA**, Yuan GC, Porteus MH, Stamatoyannopoulos JA, Lettre G, Orkin S. (2013) An adult-stage erythroid enhancer of BCL11A determines fetal hemoglobin level. (Submitted)

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

5mC – 5-methylcytosine

AAV – adeno-associated virus

AAVS1 – AAV integration site 1

ADA – adenosine deaminase

AIDS - acquired immunodeficiency syndrome

APO – APOBEC3G D128K

BCNU - carmustine

CGD - chronic granulomatous disease

CMV – cytomegalovirus promoter

DSB – double strand break

EPO - erythropoietin

FIs - fusion inhibitors

GFP – green fluorescent protein

HAART – highly-active anti-retroviral therapy

HbA – adult hemoglobin

HbF – fetal hemoglobin

HbS – sickle hemoglobin

hESC – human embryonic stem cell

HIV - human immunodeficiency virus

HR – homologous recombination

hrhTRIM5α - human-rhesus hybrid TRIM5α

HSC – hematopoietic stem cell

HSCT – hematopoietic stem cell transplant

hTRIM5α - human TRIM5α

IDLV – integration defective lentivirus

IL2Rγ – interleukin 2 receptor γ

INSTIs - integrase strand transfer inhibitors

iPSC – induced pluripotent stem cell

IRES – internal ribosomal entry site

MGMT – methylguanine methyltransferase

MSC - mesenchymal stromal cells

NHEJ – non-homologous end joining

NNRTIs - non-nucleoside reverse transcriptase inhibitors

NRTIs - nucleoside/nucleotide reverse transcriptase inhibitors

O6BG – O6 benzylguanine

OPEN – Oligomerized pool engineering

PIs - protease inhibitors

PNH – paroxysmal nocturnal hemoglobinuria

Psi – pseudo-uridine

Rev – Rev M10

rhTRIM5α – rhesus TRIM5α

ROS - reactive oxygen species

RTCN – ratio to cell negative

RVD – repeat variable domain

SCD – sickle cell disease

SCID – severe combined immunodeficiency

SMRT – single molecule, real-time

SSA – single strand annealing

TALEN – Tal effector nuclease

Ubc – ubiquitin C promoter

ZFN – zinc finger nuclease

CHAPTER I: INTRODUCTION AND REVIEW OF THE LITERATURE

The role of targeted genome engineering in the treatment of human disease

The ultimate goal of biomedical research is to elucidate the mechanisms of disease pathogenesis and develop therapies to ameliorate symptoms and provide cures. Often when successful, decades of basic science and translational research culminates in the synthesis of a small molecule with pharmaceutical efficacy that alters disease progression. Sometimes, as in the case of antibiotics for infections, an acute drug regimen can eliminate the underlying cause of the infirmity. In other cases, for example statins in cholesterol regulation, chronic drug therapy provides ongoing benefit as long as the patient continues to take his medicine. Because high cholesterol is most commonly a manifestation of the complex interplay between many environmental and genetic factors, it can be controlled using medicine, but the underlying pathologies are so varied and interdependent that a permanent "cure" is not feasible. Another set of illnesses, the monogenic diseases arise from a known deficiency in a single gene, and small molecule based therapy provides treatment for the consequences of the underlying genetic defect but does not directly address the root cause. In the last several decades, however,

new classes of therapeutics have either been developed or are being developed that attempt to treat genetic diseases at their root cause including the use of protein replacement therapy, bone marrow transplantation, and gene therapy.

Enzyme replacement therapy for the lysosomal storage diseases and severe combined immunodeficiency caused by mutations in the adenosine deaminase (ADA) gene, as well as clotting factor replacement for hemophilia are examples of protein therapeutics in which protein is administered directly to replace the missing protein. Protein replacement therapy provides life-altering treatment for patients for whom it is available, but does not cure the disease as the underlying genetic defect remains and is limited in both cost and effectiveness. The high cost of protein therapy (hundreds of thousands of dollars per year per patient) means this type of therapy is not available to patients from resource poor parts of the world. Finally, despite attempts at prolonging the half-life of therapeutic proteins, protein replacement therapy requires life-long repeated infusions at daily, weekly, or monthly intervals.

Bone marrow transplantation in which the patient's bone marrow is replaced with the bone marrow from a person who does not have the disease is a highly effective therapy for patients with genetic diseases of the blood such as sickle cell disease, thalassemia and sever combined immunodeficiency (SCID) and is even effective in non-blood genetic diseases such as adrenoleukodystrophy and Hurler's syndrome. Bone marrow transplantation replaces the hematopoietic

system containing the disease-causing mutation with a hematopoietic system free of the genetic defect. Bone marrow transplantation remains the only broadly used stem cell therapy to treat patients. However, the limitations of stem cell therapy by bone marrow transplantation are several-fold. First, like protein therapeutics, bone marrow transplantation is a highly technical and expensive procedure and so is not available in many parts of the world. Second, many patients do not have a suitable immunologically matched donor. Third, there are long-term side effects from bone marrow transplantation including graft versus host disease and side effects of high doses of myeloablative conditioning therapy that can result in conditions worse than the original disease.

Gene therapy is an alternative to protein replacement therapy or allogenic stem cell therapy in which the genomes of a patient's own cells are modified to correct the underlying pathologic mutation, and those gene-modified cells are used as treatment for the disease. The modification of the patient's cells can be done *in vivo* in which the modification takes place without the cells leaving the body or *ex vivo* in which the cellular modification takes place outside the body and then the cells are transplanted back into the patient. Unlike small molecule therapy or protein replacement therapy, this strategy is directed at the root cause of the disease, meaning that a total cure could be achieved with a single intervention. Unlike allogeneic bone marrow transplantation, this approach uses

the patient's own cells and so the problems of finding matched donors and graft versus host disease are avoided.

There have been several gene therapy clinical trials for genetic diseases of the blood including for SCID-X1, ADA-SCID, Wiskot-Aldrich syndrome, chronic granulomatous disease, and thalassemia (Abuljadayel et al., 2006; Aiuti et al., 2009; Bank et al., 2005; Boztug et al., 2010; Cavazzana-Calvo et al., 2000; Cavazzana-Calvo et al., 2005; Gaspar et al., 2006; Kang et al., 2009; Malech et al., 1997). In each of these trials, the patient's bone marrow cells were removed, transduced with a retroviral or lentiviral vector that carries a wild type copy of the mutated gene and then transplanted back into the patient. Tens of patients of worldwide have benefited from this therapy. A fundamental feature of this viral based strategy is that the transgene integrates into the genome in an uncontrolled fashion. A detrimental consequence of this uncontrolled integration is that it can activate a latent oncogene, as was the case in a number of gene therapy patients, causing frank leukemia, myelodysplasia, or a clonal proliferation (Hacein-Bey-Abina et al., 2008; Hacein-Bey-Abina et al., 2003; Ott et al., 2006). Although there are vector and disease specific differences, the overall frequency of this event seems to be greater than 10% which may make this approach unsustainable for the treatment of genetic diseases of the blood.

Targeted genome engineering, a technology that enables site-specific, controlled insertion of a transgene, allows for the benefits of transgene technology

while avoiding the complications of random integration. Targeted genome engineering strategies in human cells have been developed to stimulate sequence-specific DNA modifications to achieve one of three therapeutic endpoints. First, the targeted disruption of an endogenous locus can provide a rapid, efficient means to achieve permanent gene knockouts in human cells. Precise gene knockout has therapeutic implications in preventing HIV infection for example, by eliminating a major route of entry of the virus into host cells, as will be discussed at length. Additionally, gene knockouts in human cells provide a powerful research tool that has heretofore been restricted to model organisms. The second use of targeted genome engineering is the direct correction of disease-causing mutations at endogenous loci. In this way, the mutated disease-causing DNA sequence is replaced by a non-mutated version of that gene. Third, targeted genome modification can be used to insert transgenes to precise locations in the genome, including sites deemed safe harbors, which are distant enough from known oncogenes that the risk of insertional oncogenesis is minimized. With such potentially precise methods of genomic manipulation available, researchers now have the ability to begin to address the root cause of many of the more than 10,000 monogenic diseases.

Specific DNA double-strand breaks stimulate targeted genome engineering

Central to achieving high frequencies of targeted genome engineering in human cells is the ability to induce DNA double-strand breaks (DSBs) in a precise, sequence-specific manner. To do so, engineered chimeric nucleases have been developed and can be classified into one of three general classes: 1) zinc finger nucleases (ZFNs), 2) transcription activator-like effector nucleases (TALENs), and 3) homing endonucleases (Figure 1.1). ZFNs and TALENs share a common general structure in which an engineered DNA binding domain is fused to the nuclease domain derived from the FokI restriction enzyme. The genome specificity of ZFNs and TALENs derives from both the sequence specificity of the DNA binding domain and the need for the FokI nuclease domain to dimerize (Bitinaite et al., 1998) in order to create a DNA double-strand break at a specific target site. Homing endonucleases have less well-defined domains, but nonetheless can be engineered to induce site-specific DSBs. These three classes of engineered nucleases will be detailed in the next sections, followed by a discussion of the cellular response to DSBs, and how that response can result in specific, targeted genome modifications.

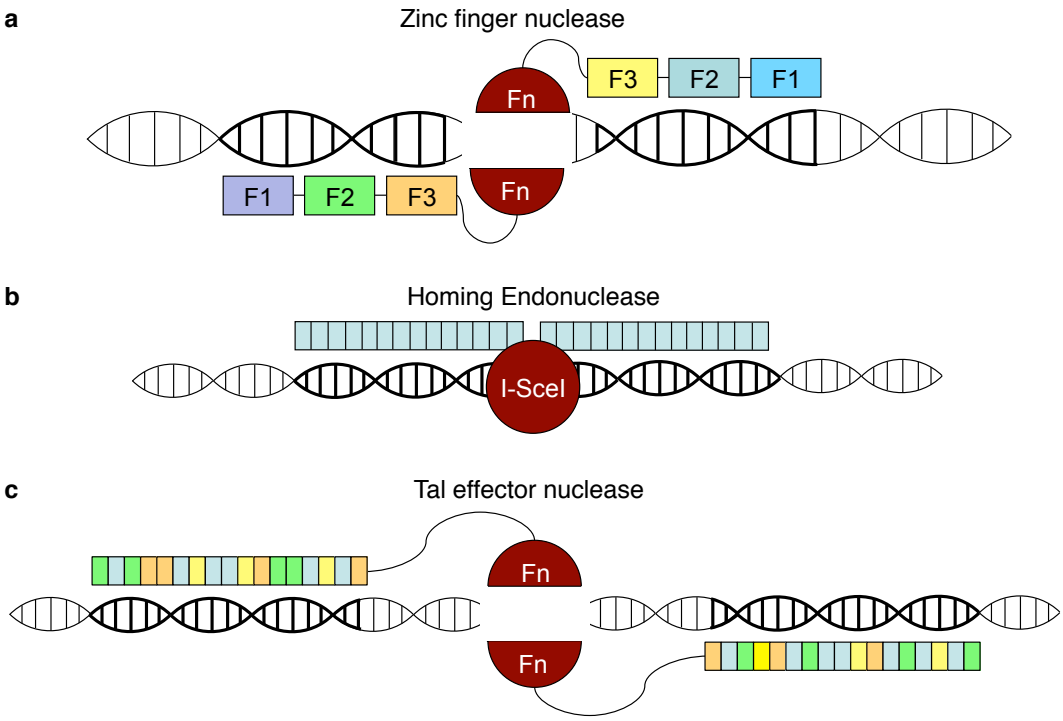


Figure 1.1: Induction of DNA double strand breaks by three classes of chimeric nucleases. (a) The DNA binding domain of zinc finger nucleases (ZFNs) is discrete from the nuclease domain and consists of arrays of three or four zinc fingers, each of which binds three nucleotides. When two ZFNs bind recognition sequences separated by 5-6 nucleotides, the nuclease domains dimerize to form an active enzyme and stimulate DNA cleavage. **(b)** Meganucleses such as I-SceI have long (12-40 base pair) DNA recognition sequences and bind and cleave DNA as monomers. **(c)** Tal effector nucleases (TALENs) bind DNA using a series of 34 amino acid repeats, each containing two key residues that mediate binding to one nucleotide. Arrays of 11-17 (or more) repeats are fused to a nuclease domain, which must dimerize with another TALEN separated by a spacer of 12-23 nucleotides.

Zinc Finger Nucleases

ZFNs consist of a zinc finger DNA binding domain fused to the non-specific nuclease domain derived from the type IIS FokI restriction enzyme (Bitinaite et al., 1998; Kim and Chandrasegaran, 1994). Zinc fingers have a conserved ββα structure and the α helix fits into the major groove of DNA, interacting with three nucleotides. The first six amino acids of the α helix, along with the immediately preceding amino acid, give the zinc finger its nucleotide specificity. By stringing three or four zinc fingers in series, proteins can be designed with 9- or 12-base pair (or longer) recognition sequences. When these zinc finger motifs are fused at their C-termini to the FokI nuclease domain, the resulting chimeric protein is a ZFN. Study of the activity of the wild type FokI endonuclease revealed that efficient cleavage is mediated by the dimerization of nuclease domains (Bitinaite et al., 1998). The dimerization requirement for efficient activity is maintained for chimeric nucleases including ZFNs. For ZFNs, site-specific cleavage is achieved by having two different ZFNs bind in an inverted orientation on opposite DNA strands thereby allowing the nuclease domain to dimerize and cut the DNA in the spacer region between the two ZFN binding sites (Bibikova et al., 2001). As mentioned, each individual ZFN has a 9- or 12- base pair binding site, a length not sufficient to confer genomic specificity, but a pair of ZFNs has an 18- or 24-base pair binding site, making a ZFN binding site theoretically unique in the human genome.

There are several approaches to the design of ZFNs. The simplest is to use modular-assembly in which individual fingers are strung together in an array of three or more. This approach is relatively straightforward and has been useful in several different studies (Beerli et al., 2000; Dreier et al., 2001; Dreier et al., 2005). The major disadvantage is that it has a low success rate when target sites do not fit a certain type of sequence and even with potentially ideal target sites, the zinc finger nucleases made using this strategy have lower activity and higher off-target effects than proteins made using other strategies. A similar strategy is to use modular-assembly to string together two-finger modules into an array (Perez et al., 2008; Urnov et al., 2005). This strategy was first developed by Gendaq, which was bought by Sangamo Biosciences, and is now sold to the research community through Sigma-Aldrich Pharmaceuticals. Because the method has not been publicly described, the overall success rate has not been published but the strategy has had significant success targeting multiple genes. This method is based on a proprietary archive of two-finger modules and thus is not freely available to researchers. A derivative of this approach is to assemble a ZFN using two-finger modules and then to subsequently refine the original protein for improved properties ("Sangamo approach"). It is through this method that the most highly active ZFNs from Sangamo Biosciences have been made but is also not an approach that is freely available to the scientific community at large. An alternative approach to assembling pre-determined modules, is to use selection

strategies, either by phage display or using bacterial based methods (including the Oligomerized Pool ENgineering (OPEN) method), to select for zinc finger domains that bind the desired target sequence out of a randomized pool of proteins. These selection approaches are difficult to use but do result in nucleases with properties of high activity and lower off-target effects (Pruett-Miller et al., 2008). Finally, there are two derivatives of the OPEN approach to making ZFNs. In context-dependent assembly, modular assembly is used to assemble ZFNs using two-finger modules that have been previously identified using the OPEN system (Sander et al., 2010). In a hybrid approach, individual modules can be combined with the OPEN system to generate proteins that have high activity and relatively low toxicity (Wilson and Porteus, unpublished data).

Homing Endonucleases

Zinc finger nucleases have had the widest application in targeted genome modifications, but other classes of nucleases are also being developed to stimulate targeted DNA double strand breaks. The first of these classes includes meganucleases, also called homing endonucleases. The major family of meganucleases, called LAGLIDADG endonucleases because of the presence of a conserved amino acid motif, are derived from the mitochondria and chloroplasts of eukaryotic unicellular organisms such as yeast. As their name suggests, meganucleases have very long DNA recognition sites (between 12 and 40 base

pairs), allowing them to bind DNA with very high specificity. The first example of a meganuclease used in gene targeting is I-SceI. In 1994 Maria Jasin and colleagues stably integrated two I-SceI sites into mouse chromosomes and demonstrated cleavage at one or both of those sites by transient expression of I-SceI (Rouet et al., 1994). To use meganucleases in targeted genome modification, however, the nucleases must be re-engineered to recognize a new target site. Several groups in academia and industry (including Cellectis and Precision Biosciences) have had some success with the re-engineering of meganucleases to recognize new target sites for the purposes of genome engineering using targeted genome modification (Arnould et al., 2006; Arnould et al., 2010; Grizot et al., 2009a; Smith et al., 2006). However, the re-engineering of meganucleases to new target sites is a challenging endeavor and has not been widely adopted.

TAL Effector Nucleases

Recently, another DNA-binding domain has been identified as a potential motif for use in chimeric nucleases. Transcription activator-like (TAL) effectors are virulence factors in the phytopathogenic bacteria *Xanthomonas* and bind to DNA through a series of nearly identical repeats. Each repeat of 34 amino acids has two key residues that recognize and bind one nucleotide, and proteins optimally with 11 to 17 repeats can be designed *de novo* to bind very specific DNA sequences (Boch et al., 2009; Moscou and Bogdanove, 2009). TAL effector

repeats, like zinc finger proteins, can be fused to the FokI nuclease domain to generate TALENs which dimerize and effectively cleave cognate DNA sequences. A significant advantage of TALENs compared to ZFNs is their seemingly total modularity; the same repeat recognizes and binds the same nucleotide regardless of the context of its neighboring repeats. This modularity eliminates most of the screening steps required in the synthesis of other chimeric nucleases. In fact, at least five groups have established rapid methods to design and synthesize novel TALENs (Cermak et al., 2011; Li et al., 2011; Reyon et al., 2012; Sanjana et al., 2012; Zhang et al., 2011). Notably the Voytas lab has made their protocol readily accessible and the necessary reagents publicly available (Cermak et al., 2011). In early studies, the activity of best ZFNs and the best TALENs seems comparable, but making highly active TALENs is much easier and more efficient than making highly active ZFNs. Also TALENS seem to have fewer off-target effects and lower toxicity than ZFNs (Mussolino et al., 2011). Thus, it is likely that ZFNs and TALENs will both be useful nuclease platforms for targeted genome modifications in the future.

Targeted gene knockout by mutagenic non-homologous end joining

Cells are constantly bombarded with DNA damaging agents, such that each nucleus is subject to between 10 and 100 DSBs per day on average (Burma

et al., 2006). Most natural DSBs are the result of ongoing environmental stresses such as ionizing radiation and metabolically derived reactive oxygen species. Unrepaired DSBs lead to cell death, and thus multiple, redundant cellular DNA repair pathways have evolved to repair these important genetic lesions. The two most important DSB repair mechanisms are non-homologous end joining (NHEJ) and homologous recombination (HR) (Figure 1.2). In NHEJ, the DNA ends created by a DSB are recognized and bound by the Ku70/Ku80 heterodimer, which along with the catalytic subunit of DNA-dependent protein kinase (DNA-PKcs), recruits the DNA ligase IV-XRCC4 complex to the site of the DSB. The DNA ligase IV-XRCC4 complex catalyzes the rejoining of the two DNA ends (Burma et al., 2006). When the DNA ends are compatible, as is the case with nuclease induced DSBs, repair by NHEJ most often proceeds without altering the DNA sequence near the break site. However, occasionally through the 5' to 3' exonuclease activity of the accessory factor Artemis and the gap-filling activity of DNA polymerases λ and μ, small insertions or deletions are introduced at the site of the nuclease-induced break (Ogiwara and Kohno, 2011), resulting in repair by mutagenic NHEJ.

When a DSB created by an engineered nuclease is repaired by mutagenic NHEJ, the resulting frameshift mutations often lead to truncated or non-functional protein products. In this method of targeted genome modification, the site of the mutation is controlled by the specificity of the nucleases but the specific genomic

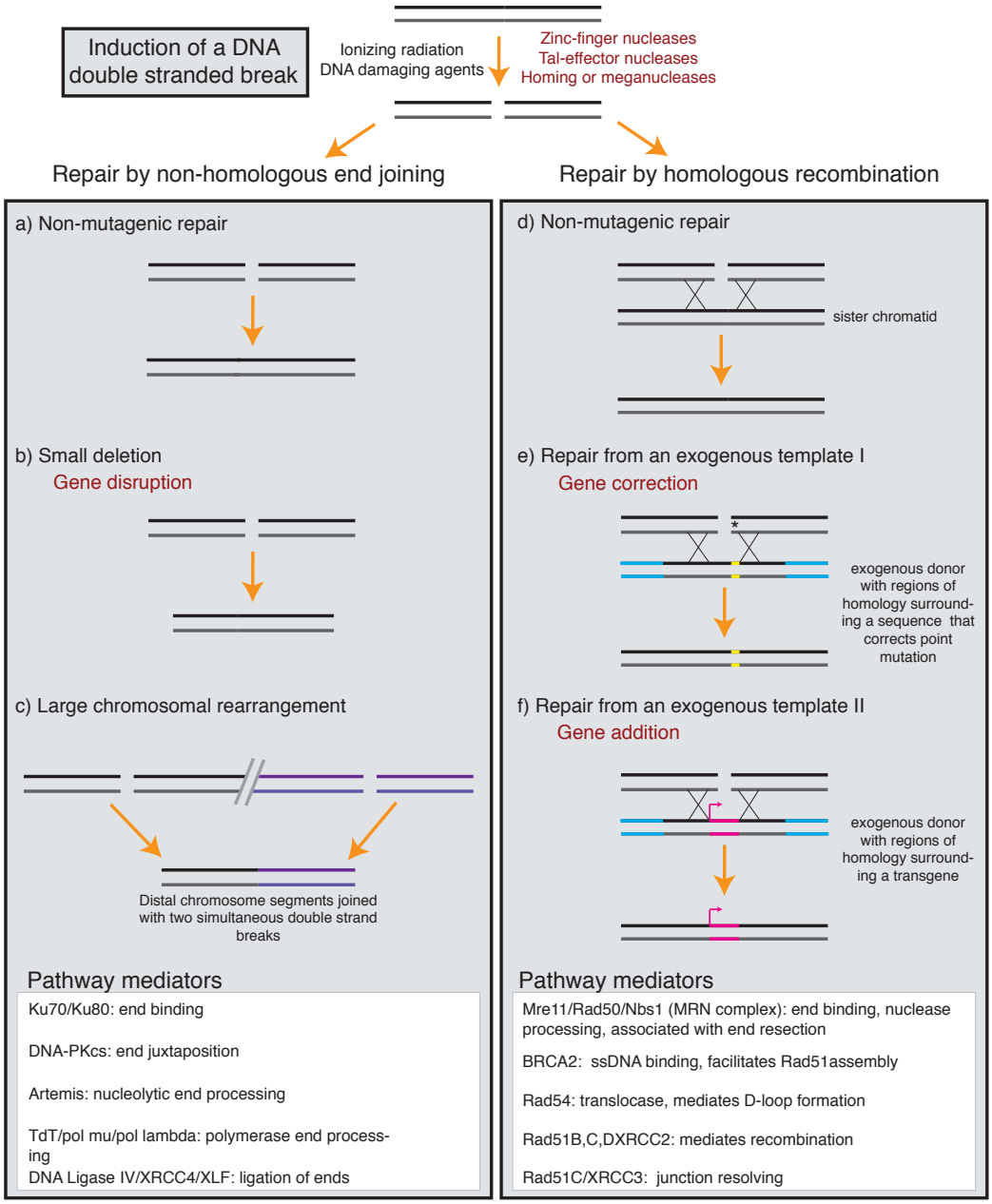


Figure 1.2: NHEJ and HR repair pathways in mammalian cells. Repair by NHEJ can result in **(a)** no sequence modifications, **(b)** small insertions or deletions, or **(c)** large chromosomal rearrangements. Repair by HR can result in **(d)** no sequence modifications, **(e)** gene correction using an exogenous donor, or **(f)** gene addition using an exogenous donor.

modification at that site is uncontrolled. Thus, by combining the engineering of gene specific nucleases with the intrinsic mutagenic property of NHEJ, researchers can create targeted gene knockouts. Depending on the activity and expression level of the nucleases, targeted gene mutation frequencies of greater than 50% can be generated (Perez et al., 2008).

Targeted gene insertion by homologous recombination

Homologous recombination is a high-fidelity repair pathway in which regions of homologous DNA serve as a template to accurately repair DNA damage. In a normal setting in dividing cells, repair of a DSB proceeds using the sister chromatid as the homologous template, in a process that fixes the break without altering the genomic sequence. This process is initiated when the MRN complex, consisting of meiotic recombination 11 (Mre11), Rad50 and Nijimegen breakage syndrome 1 (Nbs1), senses DNA ends. The MRN complex activates ataxia telangiectasia mutated (ATM) protein, which in turn phosphorylates histone protein H2AX to serve as a scaffold for other repair proteins. ATM also initiates cell cycle arrest by phosphorylation of p53, and the accumulation of p53-binding parter 1 (53BP1) is an important marker of DSBs. In a series of steps, the MRN complex creates 3'-overhanging ssDNA by 5' to 3' end resection, in conjunction with CtBP-interacting protein (CtIP). The creation of this 3' ssDNA overhang of 50-100 bases prevents the binding of the Ku70/Ku80 complex and

therefore is the critical step in repair pathway choice between NHEJ and HR (Stracker and Petrini, 2011). Replication protein A (RPA) binds to and stabilizes the ssDNA overhang, but RPA binding also blocks access of the critical HR protein Rad51. In a breast cancer 2 (BRCA2)-dependent process, Rad51 replaces RPA on the ssDNA, catalyzing strand invasion of a homologous DNA template, which in the normal setting is often the sister chromatid (Holloman, 2011).

In the experimental setting, if an extra-chromosomal piece of donor DNA with homology to the break site is supplied, the same HR machinery can use the exogenous DNA to repair the DSB. Through careful design of the exogenous template, a disease-causing point mutation can be changed to the wild type base, leading to reversion of the mutation when the donor template is used in HR ("gene correction"). Using an alternative donor template design in which a single gene or even cassette of genes is flanked by homology arms on each side of the DSB, gene targeting by HR can occur in which the entire cassette between the arms is inserted precisely into the genome at the site of the nuclease induced DSB ("targeted gene addition"). Including either drug resistance markers or genes encoding fluorescent proteins in between the homology arms, allows for the selection or enrichment of targeted cells. In gene targeting by HR, not only is the site of the genome modification determined, but by harnessing HR to repair the induced DSB using a provided donor template, the exact sequence of the genome modification is also determined. Depending on the activity of the engineered

nuclease, gene correction rates and targeted gene addition rates of greater than 20% can be achieved (Benabdallah et al., 2010; Hockemeyer et al., 2009; Urnov et al., 2005), and these frequencies approach 100% following drug selection or fluorescence enrichment.

Specific genome modification using non-nuclease approaches

In addition to nuclease-mediated genome editing strategies, adeno-associated virus (AAV) has been used for targeted genome modifications (Khan et al., 2011). Through mechanisms still not clear, at high multiplicities of infection, AAV can stimulate targeted transgene integration or gene correction at frequencies up to 1% under certain conditions. While this frequency is a log lower than what can be achieved using nuclease-mediated genome modification it does have the advantage of not needing to engineer site-specific nucleases.

Papapetrou *et al* (2010) have described another site-specific viral targeting approach in induced pluripotent stem (iPS) cells. This lentiviral-mediated approach depends upon retrospective analysis and selection of optimal integration events at a "safe harbor" locus. The authors define a safe harbor locus as a locus that meets the following citeria: a locus that is 50kb from the 5' end of any gene, 300kb from any known cancer-related gene, 300kb from any microRNA and one that is located outside transcription units or "ultraconserved" regions. The iPS cells for this experiment were derived from patients with β-thalassemia and

consequently did not express functional β-globin when differentiated into the erythroid lineage. The iPS cells were infected at a low multiplicity of infection with a β-globin expressing lentivirus. Single copy integrant clones were then selected and screened for the genomic location of integration. After screening, the authors found that they could isolate an iPS clone expressing β-globin efficiently from a locus that met the "safe harbor" criteria. Although this retrospective strategy differs from many of the prospective targeting approaches, the result is similar and perhaps provides a plausible alternative for targeted genome engineering.

Applications of targeted gene disruption in human disease

The following sections will describe ZFN and TALEN pairs that have been shown to stimulate targeted genome modifications in human cells. These modifications can be classified into three major categories: 1) endogenous gene disruption, 2) endogenous gene correction, and 3) safe harbor gene addition (Table 1.1).

The simplest method of targeted genome modification is gene disruption by mutagenic NHEJ after the induction of a DSB by the engineered nucleases. ZFN-induced mutagenic NHEJ repair has used for a wide variety of experimental purposes including the creation of knockout human cell lines as well as primary

Human genes modified by targeted genome engineering			
Cell type	Gene	Frequency of modification	References
Gene disruption			
CD4+/HSC	CCR5	54%/17%	Perez *et al*/Holt *et al*
HT1080	CAG repeats	>0.006%	Mittleman *et al*
293	HOXB13	9.60%	Maeder *et al*
293	CFTR	1.20%	Maeder *et al*
Gene correction			
K562/CD4+	IL2Ry	20%/5%	Urnov *et al*/Lombardo *et al*
K562	VEGF	7.7%	Maeder *et al*
CHO	XPC	0.20%	Arnould *et al*
293	RAG1	6%	Grizot *et al*
iPSC	β-globin	10%/0.3% (selected)	Sebastiano *et al*/Zou *et al*
iPSC	α_1-antitrypsin	54% (selected)	Yusa *et al*
Gene addition			
K562	IL2Ry	15%	Moehle *et al*/Lombardo *et al*
hESC/iPSC	AAVS1	56%/75% (selected)	Hockemeyer *et al*/Zou *et al*
MSC/hESC	CCR5	40%/5.3%	Benabdallah *et al*/Lombardo *et al*
hESC/iPSC	PIGA	0.24%	Zou *et al*
hESC	OCT4	94% (selected)	Hockemeyer *et al*
hESC	PITX3	11% (selected)	Hockemeyer *et al*

Table 1.1 Human genes modified by targeted genome engineering. HSC, hematopoietic stem cell; HT1080, human fibrosarcoma cell line; 293, human embryonic kidney cell line; K562, human erythroleukemia cell line; CHO, chinese hamster ovary cell line; hESC, human embryonic stem cell; iPS, induced pluripotent stem cell; MSC, mesenchymal stem cell

cells including stem cells (Alwin et al., 2005; DeKelver et al., 2010; Hockemeyer et al., 2009; Kim et al., 2009; Perez et al., 2008; Porteus and Baltimore, 2003; Urnov et al., 2005; Zou et al., 2009). Further, whole animal knockout has been achieved by this method in model organisms that include flies (Beumer et al., 2006; Beumer et al., 2008; Bibikova et al., 2002; Bozas et al., 2009), zebrafish (Doyon et al., 2008; Foley et al., 2009; Meng et al., 2008), mice (Carbery et al., 2010; Cui et al., 2010; Meyer et al., 2010), rats (Geurts et al., 2009; Mashimo et al., 2010), rabbits (Flisikowska et al., 2011), frogs (Young et al., 2011), sea urchins (Ochiai et al., 2010), plants (Lloyd et al., 2005; Wright et al., 2005), nematodes (Morton et al., 2006) and silkworms (Takasu et al., 2010).

More recently, TALENs have been reported to achieve efficient gene disruption in yeast (Christian et al., 2010; Li et al., 2010), plants (Cermak et al., 2011), nematodes (Wood et al., 2011), zebrafish (Huang et al., 2011; Sander et al., 2011), rats (Tesson et al., 2011) and human cells (Hockemeyer et al., 2011; Miller et al., 2010; Reyon et al., 2012). To create gene specific knockouts, the ZFNs or TALENs have been introduced as either DNA expression plasmids or as mRNA through a variety of delivery techniques including standard transfection, nucleofection, adenoviral infection, integration-defective lentiviral infection, or by glass-needle microinjection.

HIV resistance generated by CCR5 knockout

From a clinical gene therapy perspective the most important use of this gene disruption approach has been to create HIV-resistant cells through the ZFN-induced mutation in the *CCR5* gene. CCR5 is a necessary co-receptor for infection of CD4+ T-cells by HIV. People with homozygous Δ32 mutations in *CCR5*, which results in non-functional CCR5 protein, are resistant to HIV infection but are otherwise phenotypically normal. The hypothesis based on these findings is that by creating an immune system that has *CCR5* mutated in a patient who was already infected with HIV, one could alter the course of the disease. In an exciting proof-of-principle experiment, an HIV infected patient underwent a bone marrow transplant for acute myelogenous leukemia using donor marrow that was homozygous for the *CCR5 Δ32* allele. Remarkably, after the transplant the patient's HIV viral load became undetectable and his donor CD4 count, derived from the CCR5 Δ32 donor, rose (Allers et al., 2010; Hutter et al., 2009). The results of this patient's treatment demonstrated that it might be possible to engineer the immune system to become HIV resistant with a beneficial clinical outcome. Unfortunately, bone marrow transplantation using homozygous Δ32 donors is not a broadly feasible strategy for the treatment of HIV because most patients will not have an HLA matched homozygous Δ32 donor and because of the general toxicity from allogeneic bone marrow transplantation.

An alternative strategy would be to engineer the patient's own immune system to become HIV resistant by creating a ZFN mediated knockout of the *CCR5* gene in autologous cells, either CD4+ T-cells or hematopoietic stem cells (HSCs). Both ZFNs (Holt et al., 2010; Perez et al., 2008) and TALENs (Miller et al., 2010) have been designed to efficiently disrupt CCR5, but the ZFNs have been the most well-studied and characterized. In pre-clinical studies, Carl June and co-workers have demonstrated that using CCR5-directed ZFNs, primary T-cell populations can be generated in which >50% of the CCR5 alleles have been mutated and that these T-cells are resistant to a CCR5-tropic strain of HIV (Perez et al., 2008). Paula Cannon and her co-workers have demonstrated that the same CCR5 ZFNs can cause inactivation of the CCR5 gene in human CD34+ HSCs and that T-cells derived from these gene modified cells are resistant to HIV (Holt et al., 2010).

Based on these pre-clinical findings, clinical trials have begun using on this strategy. In the first clinical trials to open, patient derived CD4+ T-cells were infected with an adenovirus that expressed the CCR5 ZFNs causing CCR5 gene modification. The population of autologous T-cells was then expanded *in vitro* and transplanted back into the patient. These trials have enrolled a number of patients and the early results look very promising. First, the autologous transplant of ex vivo modified CD4+ T-cells is safe and well-tolerated and leads to the persistent increase in total CD4 count. Also during a controlled drug interruption

after an initial spike in HIV RNA levels, 3 of 6 patients displayed a reduction in viral titer. Most notably, one patient was naturally heterozygous for the *CCR5 Δ32* allele and then received the ZFN treatment to knock out the wild type *CCR5* allele. In this patient, HIV RNA levels became undetectable during the drug interruption. More broadly, the clinicians observed that a higher percentage of bi-allelically CCR5-modified CD4+ T-cells was a significant predictor of HIV RNA suppression during drug interruption. This is not surprising because one functional *CCR5* allele produces sufficient CCR5 protein to mediate viral entry into T-cells. This is one significant limitation to the long-term effectiveness of this approach. The authors of this study report fewer than 10% of all of the re-infused CD4+ T-cells had been biallelically modified (Ando et al., 2011).

A second major limitation to generating HIV resistance by *CCR5* knockout is that instead of using the CCR5 co-receptor to infect T-cells, HIV is also capable of using the CXCR4 co-receptor (See HIV pathogenesis, below). To address this, two groups have recently reported the synthesis of ZFNs that efficiently disrupt CXCR4 (Wilen et al., 2011; Yuan et al., 2012), generating resistance to CXCR4-tropic HIV. To achieve resistance to both CCR5- and CXCR4-tropic HIV, Wilen et al stimulated CXCR4 knockout in cells derived from a CCR5 Δ32 patient. For a continued discussion of this work and my alternative approach to generating dual-tropic resistance, see Chapter 2.

Other nuclease-induced knockouts relevant to human diseases

Of the handful of published ZFN pairs designed to disrupt human genes, the CCR5 ZFNs have had the most pre-clinical and clinical success. Others, however, are being developed that are perhaps promising therapeutic tools. One such pair of ZFNs has been developed to recognize CAG triplet repeat tracts. CAG tracts are physiologically normal in multiple genes when the number of repeats is fewer than 30, however expansion of repeat number leads to debilitating diseases such as myotonic dystrophy, Huntington's disease and spinocerebellar ataxia. ZFNs that bind these repeats were designed by modular assembly, and demonstrate the ability to contract CAG repeats in a reporter assay in human cells (Mittelman et al., 2009).

A third pair of ZFNs used to disrupt an endogenous human gene has been developed against the *PIG-A* gene, located on the X-chromosome Xp22.1. Mutations in this gene cause paroxysmal nocturnal hemoglobinuria (PNH). The *PIG-A* gene encodes a phosphatidylinositol glycan class A protein that functions in a complex required for production of many glycosyl-phosphatidyl-inositol anchored proteins (GPI-APs). Among these anchor proteins are a few involved in complement regulation, without which cells are inappropriately recognized and lysed by the complement system. For patients with PNH, a somatic mutation within the hematopoietic progenitor population results in pools of abnormal red blood cells. Their lysis by the complement system results in a hemolytic anemia

and hemoglobinuria. Zou *et al* (2009) demonstrated that the PIG-A ZFNs could target and knock out the *PIG-A* gene in both human embryonic stem (hES) cells and iPS cells by either mutagenic NHEJ-mediated repair of the DSB or by targeting a hygromycin resistance cassette flanked by homology arms to the locus. PIG-A knock out was phenotypically characterized by resistance to aerolysin-mediated cell killing and by cell surface absence of GPI-APs. The authors confirmed that restoration of a PIG-A transgene could rescue these phenotypes. Though this work did not yet demonstrate targeting of the PIG-A gene in a therapeutically relevant manner for PNH patients, the PIG-A ZFNs could hold potential for a ZFN-mediated gene correction strategy in HSCs.

Because TALENs were only first described at the end of 2010, studies describing their potential in human disease applications are not as complete as those using ZFNs. Nonetheless, the rapid, streamlined method of synthesizing TALENs has resulted in the report of many pairs with the potential to affect human disease. The best example is the recent report (Reyon et al., 2012) that describes the facile synthesis of 96 pairs of TALENs against genes implicated in cancer or epigenetic regulation using a novel method called FLASH assembly. Of the 96 tested TALEN pairs, 84 of them induced significant gene modification by mutagenic NHEJ.

Though endogenous gene disruption is clinically relevant in perhaps only a few scenarios, it could have vast impact on the way diseases are modeled and

studied. As demonstrated, highly efficient targeting strategies now allow for robust and precise genetic manipulation in previously inaccessible cells types such as hES and iPS cells. Moreover, ZFNs and TALENs have been used to create targeted gene mutations to create genetically modified organisms in a wide variety of species for which it previously was not possible to perform targeted gene modification. The ability to create precise genetically modified lines in species other than mouse has the potential to result in improved models of human disease, improved species for agricultural purposes, and in the long-term general improvements in human health.

Examples of therapeutically relevant gene correction

By harnessing the capability of nucleases to stimulate gene targeting through homologous recombination, small, precise changes can be introduced into the genome to correct disease causing mutations. Along with an active pair of nucleases, a donor DNA template must also be designed that is homologous to the endogenous gene and contains the small changes that are to be introduced into the genome. The nucleases and the donor DNA template must then be delivered to a cell-type that is capable of reconstituting the diseased tissue or organ. Translation of this therapeutic paradigm to the clinic has not yet been accomplished, but progress has been made *in vitro* towards this end, and work is currently underway to demonstrate the feasibility of the strategy in animal models.

Proof of principle GFP correction

The first examples of nuclease mediated gene correction of a chromosomally integrated gene in human cells were described in 2003 and 2005 (Porteus and Baltimore, 2003; Urnov et al., 2005). In these first examples of ZFN mediated gene correction, a mutated GFP gene was introduced as a single copy into the genome of cells to a create reporter line. The frequency of gene correction was measured by determining the frequency that cells became GFP positive. When the donor DNA template was introduced without nucleases, the frequency of gene correction was on the order of 10^{-6}. However, when ZFNs that target the integrated reporter gene were co-transfected with the donor DNA template into HEK293T cells, the frequency of gene correction increases to 0.5-2.2% (Porteus and Baltimore, 2003; Urnov et al., 2005). This frequency can increase to ~10% if the cells are transiently arrested in G2/M by vinblastine (Urnov et al., 2005). A similar stimulation in gene correction using the GFP reporter system was also found in hES cells and iPS cells (Zou et al., 2009). Importantly, the gene corrected hESCs and hIPSCs retained their pluripotency and did not develop any overt karyotypic abnormalities. Finally, gene correction rates of ~2% were obtained in primary mouse fibroblasts using a transgenic gene targeting mouse reporter line in which the mutated GFP gene was knocked-in to the ROSA26 locus (Connelly et al., 2010).

Gene correction in SCID

Urnov *et al* (2005) published the first example of gene correction of an endogenous human gene. Mutations in the interleukin 2 receptor γ (IL2Rγ) gene cause SCID-X1, the most common form of severe combined immunodeficiency. ZFNs to exon 5 were engineered using the Sangamo strategy. Transfection of these ZFNs along with a gene correction DNA donor template, demonstrated gene correction efficiencies of ~20% in K562 cells and ~5% in primary T-cells. Lombardo *et al* (2007) built on this work and demonstrated that using integration defective lentivirus (IDLV) rather than standard transfection techniques, similar rates of gene targeting could be obtained in HEK293T cells, K562 cells, and a transformed lymphoblastoid cell line.

Other endogenous human genes targeted by gene correction

Other endogenous human genes that have been successfully targeted through gene correction by ZFNs include VEGF (Maeder et al., 2008) and α1 antitrypsin (Yusa et al., 2011). Also, two concurrent works described the gene correction of β-globin by ZFNs constructed using the Sangamo method (Zou et al., 2011a) and the OPEN method (Sebastiano et al., 2011). Both groups showed gene correction in a sickle cell iPS line, but at relatively low frequencies after drug selection (0.33% Zou, and 9.7% Sebastiano). These works provide the basis for the improved β-globin gene correction strategy detailed in Chapter 3.

In addition to ZFNs, TALENs and homing endonucleases have also been designed to stimulate HR at chromosomal loci in the human genome. TALENs directed against OCT4 and PITX3 stimulated integration of GFP and a drug resistance marker in human ES and iPS cells (Hockemeyer et al., 2011). Also, an I-CreI derived homing endonuclease was designed to target XPC, a gene involved in the skin disorder Xeroderma Pigmentosum (Arnould et al., 2007). The endonuclease-binding site from XPC was cloned into the lacZ gene, disrupting expression of lacZ. Then the mutated lacZ construct was integrated into the genome of CHO-K1 cells, transfected with the designed homing endonuclease and a lacZ repair template, and shown to successfully stimulate HR.

Another gene relevant to human disease that has been targeted with a homing endonuclease is RAG1. RAG1 is essential for VDJ recombination in B- and T-cells, and mutations in RAG1 cause Omenn syndrome, a SCID variant. RAG1 homing endonucleases stimulate HR at the endogenous locus in HEK-293 cells at rates of up to 6% (Grizot et al., 2009b; Munoz et al., 2010). Although the results of these two studies with homing endonucleases show promising initial targeting results, there have not yet been reports of the correction of disease-causing mutations in primary human cells.

Advantages of targeted gene correction

Endogenous gene correction has several advantageous features. First, with modification of the endogenous gene, the resultant transgene product remains under the control of endogenous regulatory elements at its correct chromosomal location. This is crucial for maintaining the biological optimum of protein expression that may be lost with ectopically located transgene integrations. Second, direct gene correction rather than virally mediated transgene addition minimizes the risk of insertional oncogenesis from uncontrolled viral integration sites. Third, gene correction, in contrast to gene addition, could be used to treat dominant diseases or diseases in which the mutant protein has dominant features. In sickle cell disease for example, the point mutation in β-globin causes the pathological polymerization of hemoglobin molecules. Simply inserting a β-globin transgene in a random or a safe-harbor locus, would not address the problems caused by the mutant protein. In contrast, after gene correction, one (or both) of the mutated alleles is converted to the wild type sequence, leading to reversion of the sickle cell phenotype. The major challenges to translating nuclease mediated gene correction into therapy include developing high quality nucleases to disease causing genes, demonstrating that clinically relevant levels of gene correction can occur in the appropriate cell type and finally demonstration that gene corrected cells can correct the disease phenotype.

Transgene integration by safe harbor gene addition

An alternative to direct gene correction is to use homologous recombination to target transgene addition to a safe harbor in the genome. While the defining features of a safe genetic safe harbor are still being established, in principle a safe-harbor locus can be generally defined as one in which the insertion of a transgene has no aberrant physiologic consequences. Another feature of a safe harbor locus, in addition to its safety, is it must be a location in which a transgene can be expressed at the levels necessary to achieve therapeutic efficacy. A robust example of safe harbor utilization is the frequent targeting of the murine ROSA26 locus for knock-in mouse technology. One significant benefit of safe harbor gene addition is that only one pair of ZFNs or TALENs needs to be designed and optimized to achieve targeting of any number of therapeutic transgenes (as opposed to the individual tailoring that is required for endogenous gene correction).

Gene targeting at the IL2Rγ locus

Targeted gene addition by homologous recombination can be achieved using nucleases engineered to recognize the specific safe harbor site. In addition, a donor DNA plasmid is engineered by inserting a transgene cassette (driving the expression of one or several genes) between arms of homology. The arms of homology can be as short as a few hundred basepairs in length and flank the site

of the nuclease-induced break. Targeted gene addition occurs when the homologous recombination machinery uses the donor DNA template to repair the nuclease induced DSB. Moehle *et al* (2007) first described this technology using the previously described IL2Rγ ZFNs. While IL2Rγ is not a true safe harbor locus, important proof-of-principle targeting experiments were conducted using these ZFNs. The authors began by inserting a very small, in-frame tag of only 4 amino acids (RAKR, furin cleavage site) into the IL2R-γ gene, and they achieved targeting frequencies of up to 15% by this strategy in K562 cells. Then, to investigate the full potential of gene addition, the authors targeted a 900 base pair GFP cassette, an 1100 base pair GFP-PolyA cassette and a full transcriptionally active unit of a promoter-GFP that measured 1500 base pairs. The authors demonstrated targeting frequencies of 6%, 3% and 6% for these three constructs, respectively. Finally, the authors generated a donor plasmid with a very large insert (8kb) between homology arms that contained three independent, transcriptionally active units. Remarkably, the authors found by Southern blot analysis that 5-8% of the IL2Rγ alleles had been targeted. These data highlight the potential for nuclease-mediated gene addition strategies by suggesting that even very large payloads can be efficiently and precisely targeted to safe harbor loci.

Lombardo *et al* (2007) described gene addition using the IL2Rγ ZFNs in a panel of human cell types using delivery of targeting components. First, the authors attempted targeting with a donor construct where GFP was flanked by

IL2Rγ homology arms. They reported targeting efficiencies of up to 6.6% and up to 2.4% in K562s and in an Epstein Barr virus transformed lymphoblastoid cell line, respectively. Because SCID can be caused by many mutations in the IL2R-γ gene, the authors reasoned that a potential treatment for the disease could be to target normal IL2R-γ cDNA to the endogenous locus, as a one shot approach to cover all of the different mutations (instead of designing novel ZFNs to target each mutated exon separately). They showed targeting of IL2Rγ cDNA to the endogenous IL2Rγ locus at a frequency of up to 6% in their lymphoblastoid cell line and that functional IL2Rγ expression was restored.

Genome modification at the AAVS1 locus

One site being developed as a potential safe harbor is the AAVS1 integration site on human chromosome 19. The AAVS1 locus encodes the ubiquitously expressed *PPP1R12C* gene. Because adeno-associated virus integration at this site does not seem to be deleterious and because the AAVS1 site is ubiquitously expressed, it has characteristics that may make it a good safe harbor. To this end, Sangamo Biosciences developed a pair of zinc-finger nucleases that recognized the AAVS1 locus and targeted several transgenes to the AAVS1 locus in various cell types, including hESCs (DeKelver et al., 2010; Hockemeyer et al., 2009). In Hockemeyer *et al*, the authors demonstrated three

methods of transgene expression within this locus. First, they targeted a puromycin resistance cassette driven by the endogenous *PPP1R12C* promoter through a splice acceptor. Second, they achieved constitutively expressed eGFP by targeting an eGFP cassette with cytomegalovirus (CMV) promoter with a chicken β-actin enhancer. Finally, the authors targeted the eGFP gene with a tetracycline response element and demonstrated that eGFP expression could be induced by administration of doxycycline (Hockemeyer et al., 2009).

Recently, Zou *et al* (2011b) described an AAVS1 ZFN-mediated targeting strategy in iPS cells derived from a patient with X-linked chronic granulomatous disease (CGD). Patients with CGD have a deficiency of the p91-phox gene product and lack reactive oxygen species (ROS)-mediated microbicidal activity in neutrophils. As a result, these patients present with severe, recurrent, and atypical infections. In this paper, the authors described iPS cell derivation from patients with CGD, that, when differentiated into neutrophils, expressed no functional p91-phox protein and showed reduced ROS production compared to wild type iPS cells. The authors targeted a p91-phox minigene ($gp91^{phox}$) to the AAVS1 locus in the iPS cells using mRNA delivery of the AAVS1 ZFNs and achieved targeted transgene insertion in 75% (15/20) of the iPS clones they examined. Importantly, when the targeted iPS cells were differentiated into neutrophils, oxidase activity was restored, correcting the original disease phenotype. The authors then compared the AAVS1- $gp91^{phox}$ cells with $gp91^{phox}$ expressing cells generated by

random lentiviral insertion. In the untargeted cells, $gp91^{phox}$ was initially expressed at very high levels, but only 4% of the resultant differentiated neutrophils were weakly oxidase positive at the end of the experiment. The authors suggested that this was because "a significant proportion of lentivector genomic inserts are subject to accelerated silencing" when the iPS cells were differentiated into neutrophils (Zou et al., 2011b). These data highlight one of the advantages of using precise, site-specific genome engineering and instead of relying on random integration of virally-introduced transgenes. Notably, TALENs to the same AAVS1 target site have recently been described (Hockemeyer et al., 2011) that stimulate the integration of a drug resistance cassette specifically into the AAVS1 locus in 50% of drug-resistant ES and iPS cells.

Use of CCR5 as a safe harbor locus

Another example of a potential safe harbor locus is *CCR5*. In addition to using ZFNs to create mutations in the *CCR5* gene, there have also been reports using those ZFNs to target transgenes to that locus. One example shows the successful gene addition of erythropoietin (EPO) to *CCR5* in up to 40% of human mesenchymal stromal cells (MSCs). EPO is a cytokine that drives the maturation of red blood cells in the bone marrow, and loss of EPO leads to severe anemia. Injection of the MSCs into immunodeficient mice resulted in increased levels of circulating EPO and higher hematocrit compared to controls (Benabdallah et al.,

2010). These results indicate that ZFN-mediated targeting of plasma-soluble factors to a safe harbor locus may be a viable therapeutic option.

Non-clinical applications of targeted genome engineering

Generation of isogenic cell lines and endogenously-tagged proteins

Site-specific genome engineering holds promise for use as a clinical therapeutic in monogenic diseases. However, precise, genetic modification of human cells can also be developed as a tool to better understand more basic biological questions. One example (DeKelver et al., 2010) utilized ZFNs directed to the *AAVS1* locus to generate a panel of targeted stable human cell lines. The authors postulated that targeting trangenes to a safe harbor locus would circumvent the confounding factors associated with random integration and clonal selection in classic stable cell line protocols, and would allow for the generation of more predictable and "isogenic" stable cell lines. To this end, the authors demonstrated targeting of a variety of constructs to the AAVS1 locus that included 1) a GFP reporter in K562, Hep3B, and HEK293 cells where targeting frequencies ranged from 3-10% 2) glucocorticoid receptor response element luciferase reporter constructs in U2OS and 3) shRNA cassettes targeted to CD58 and to components in the mTOR pathway in K562, HEK293 and hESC. Though the scientific potential for targeted, isogenic cell lines is evident, and the ability to target a variety of constructs to many cell types was thoroughly demonstrated, a

comparison between isogenic stable cells and stable cell lines generated by conventional means was lacking in this report.

Similar to the generation of targeted stable cells lines is the application of genome targeting for endogenous protein tagging. Classically, cellular trafficking studies have been accomplished by overexpression of a tagged fusion protein from plasmid DNA. Often, however, protein overexpression can perturb the cell's normal physiology and confound experimental results. A recent example of targeted protein tagging demonstrated that clathrin-mediated endocytosis could better be studied by the ZFN-mediated generation of fluorescent fusion proteins at the endogenous loci of two proteins required for this process (Doyon et al., 2011). ZFN-mediated gene targeting can also be used to create tagged proteins that allow for lineage tracing. One example tagged OCT4 with eGFP as a reporter to monitor the pluripotent state of hESCs. In this same publication, the authors also targeted a non-hESC expressed locus (PITX3) with eGFP in hESC and iPS cells. They suggested that this strategy could be used as a reporter when the hESC or iPS cells were differentiated into neurons (where PITX3 is expressed), though these data were not reported (Hockemeyer et al., 2009). Similar results were obtained at OCT4 and PITX3 using TALEN-mediated genome editing strategies (Hockemeyer et al., 2011).

Induction of gross chromosomal changes using two pairs of nucleases

Another non-therapeutic role of ZFNs is the demonstration that two distinct pairs of nucleases can stimulate chromosomal translocations when simultaneously administered to cells. The combination of the IL2Rγ and AAVS1 ZFNs induced reciprocal translocations between the X-chromosome (site of IL2Rγ) and chromosome 19 (AAVS1 site) at a frequency of 10^{-4} in HEK-293 cells and $2x10^{-6}$ in hES cells (Brunet et al., 2009). Although these rates are several orders of magnitude lower than rates of gene targeting at a single locus, this strategy provides a framework to study the mechanisms of chromosomal rearrangement and the factors required for this process. Additionally, developing ZFNs to chromosomal sites that have been implicated in oncogenic translocations would provide a way to study the expression and regulation of the resulting fusion proteins from their endogenous loci.

Another application of ZFNs for chromosomal manipulation is the use of two distinct pairs that recognize sites on the same chromosome to induce deletions. Using ZFNs designed to the adjacent genes *CCR2* and *CCR5*, Lee *et al* (2009) used a simple PCR approach to detect deletions of up to 15 kbp within the gene cluster. Further, when the authors utilized ZFNs upstream to the *CCR5* locus, they were able to induce extremely large chromosomal deletions of up to 15 Mbp. Similarly, Sollu *et al* (2010) described chromosomal deletions at a frequency of 10% using two pairs of nucleases designed to different sites within

the HOXB13 locus. In this way, it may now be possible to achieve the targeted deletion of entire gene clusters, introns or even specific exons in human cells. While the ability to create specific chromosomal rearrangements by ZFNs has important research applications, it also has critical implications for the translation of ZFNs and TALENs into clinical tools since the induction of unintended gross chromosomal rearrangements would be a serious adverse event.

Safety concerns for targeted genome engineering in human cells

ZFN expression in mammalian cells results in reduced cell viability (Porteus and Baltimore, 2003), which is directly correlated with the expression level of the ZFN as transfecting more DNA results in a higher degree of cell toxicity (Pruett-Miller et al., 2008). The mechanism of reduced cell viability is directly related to the creation of off-target DNA double-strand breaks. Staining for 53BP1 foci, a marker for double-strand breaks, demonstrates that increased toxicity is directly correlated with increased numbers of 53BP1 foci (Pruett-Miller et al., 2008). Elimination of increased 53BP1 foci also eliminates the reduced cell viability associated with ZFN cell expression (Pruett-Miller et al., 2009). Intriguingly, TALENs seem to have less cytotoxicity than ZFNs when expressed in cells (Mussolino et al., 2011).

The identification of the sites of the ZFN and TALEN off-target double-strand breaks has been a challenging problem. Gabriel *et al* (2011) used the

integration of known adaptors into sites of double-strand breaks in a non-homologous fashion to identify such sites, and uncovered a number of genomic off-target sites that had not been previously known for the CCR5 ZFNs. The nature of their approach, however, was biased against identifying rare off-target sites and underestimated the number of off-target sites. A second group used an entirely different strategy to identify off-target sites for the same CCR5 ZFNs in the same cell type (Pattanayak et al., 2011). Interestingly, while both groups confirmed a number of new off-target loci, the set of sites from the two different approaches were completely non-overlapping. Thus, the full set of off-target sites, even for this highly studied nuclease pair, has not been described and it is likely that further improvements will be needed to generate a comprehensive or near comprehensive list.

The reduced cell viability after ZFN exposure has prompted an investigation into strategies to reduce this effect. Two major strategies have been developed. The first is to modify the nuclease domain such that a nuclease pair can only cleave DNA as heterodimers ("obligate heterodimer" nuclease domains). Use of obligate heterodimer nucleases results in a marked improvement in cell viability (Doyon et al., 2010; Miller et al., 2007; Pruett-Miller et al., 2008; Szczepek et al., 2007). In some cases, however, the use of the obligate heterodimer nuclease domains also results in reduced on-target activity (Wilson and Porteus unpublished data). An alternative strategy is to regulate the

expression level of the ZFNs. Since the targeted genome modification is a "hit and run" strategy, sustained nuclease expression is not needed and, in fact, sustained nuclease expression can result in increased cell toxicity. Pruett-Miller *et al* (2009) demonstrated that using small molecule regulation of ZFN protein level that on-target activity can be maintained while decreasing the cell toxicity and off-target foci formation to background levels. Similarly, delivery of nucleases as mRNA prevents sustained expression of the proteins and yet was more effective in stimulating targeted genome modifications (Chen et al., 2011). The importance of regulating ZFN protein level both in amount and over time was confirmed in the off-target studies of the CCR5 ZFNs (Pattanayak et al., 2011).

Epidemiology, pathogenesis and treatment of HIV

Epidemiology

The human immunodeficiency virus (HIV) is an enveloped, single-stranded RNA lentivirus in the class *retroviridae* that probably first infected humans by zoonotic transmission from SIV-infected (simian immunodeficiency virus) chimpanzees in the 1950s (Gao et al., 1999). Infection by HIV leads to the progressive destruction of the immune system and the development of acquired immunodeficiency syndrome (AIDS). According to the World Health Organization statistics, HIV has claimed the lives of more than 25 million people worldwide and currently infects 34 million more people, including almost 4

million children (www.who.int). Two-thirds of all HIV infected individuals live in sub-Saharan Africa, as do three-fourths of all newly infected people. Overall, the prevalence of HIV in the region is greater than 5%. In Dallas County, 1 in 200 individuals is infected with HIV, including 1 in 90 African-Americans (www.dallascounty.org).

Pathogenesis

The entry of HIV into its primary target cells, CD4+ T-cells and macrophages, is mediated by the binding of the viral envelope gp120 protein to the CD4 receptor and either the CCR5 or CXCR4 co-receptor. Mutations in the third variable (V3) region of gp120 dictate co-receptor specificity (Saracino et al., 2009), and viral tropisms exist that utilize one or both co-receptors (CCR5- or R5-tropic, CXCR4- or X4-tropic, and R5/X4 dual tropic). Initial HIV infection is virtually always mediated by R5-tropic virus but during the course of infection viral clones can undergo co-receptor switching to a T-cell specific X4-tropic or R5/X4 dual tropic virus (Schuitemaker et al., 1992). After HIV entry, uncoating of the viral capsid occurs followed by reverse transcription of the RNA genome into double stranded DNA by the error-prone reverse transcriptase enzyme. In fact, on average one base is mutated during every cycle of replication (Abram et al., 2010), allowing the virus to rapidly evolve and develop drug resistance. The double strand DNA genome is then integrated into the host cell genome from

which viral transcripts and genomes are produced. Organized packaging of viral RNA and protein precedes budding of viral particles from the host cell surface.

Untreated, the disease can progress rapidly, destroying CD4 cells and limiting the ability of the immune system to control opportunistic infections and malignancies. Accumulation of multiple or severe co-morbid conditions is the clinical hallmark of AIDS. Most common among these conditions are oral thrush, pneumocystis pneumonia, tuberculosis, esophagitis, chronic diarrhea, toxoplasmosis, various forms of encephalopathy, Kaposi's sarcoma, Hodgkin's lymphoma and rectal and anal carcinomas (Holmes et al., 2003).

Treatment

The standard of care for patients infected with HIV is treatment with a combination of drugs that inhibit HIV replication by different mechanisms or at different stages of the viral lifecycle. This drug regimen is broadly known as highly active anti-retroviral therapy (HAART). The National Institutes of Health publishes and frequently updates guidelines for the appropriate treatment of HIV patients and the following data are culled from those recommendations (www.aidsinfo.nih.gov). The two primary measurements of the severity of infection in a patient are the level of HIV RNA in the plasma, called viral load, and the concentration of CD4+ T-cells in the blood, called the CD4 count. Initiation of treatment is strongly recommended in patients with CD4 counts of

<500 cells/mm^3, based on data from randomized, controlled trials and observational cohort studies.

There are six mechanistic classes of anti-HIV drugs: nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs), protease inhibitors (PIs), fusion inhibitors (FIs), CCR5 antagonists, and integrase strand transfer inhibitors (INSTIs). The standard initial HAART regimen for treatment-naïve patients consists of two NRTIs in combination with an NNRTI, PI, INSTI or CCR5 antagonist. Most commonly, two NRTIs (emtracitabine and tenofovir) are prescribed along with the NNRTI efavirenz. In some patients drug resistance develops, due in large part to suboptimal adherence to drug and dosing regimens or drug intolerance, which has led to the formulation of a one-pill, once-a-day combination of emtracitabine, tenofovir and efavirenz called ATRIPLA. Despite this simplification, there is significant interest in developing a one-shot, effective therapy to inhibit initial infection or prevent immunological collapse by protecting a subset of CD4+ T-cells from becoming infected. The former refers to the continuing development of an HIV vaccine, which will be discussed in the following paragraph, and the latter refers to gene therapy approaches, which were previously mentioned and provide the rationale for the work presented in Chapter 2.

Despite decades of research and vaccine development, no HIV vaccine showed any clinical significance until the RV144 Thailand trial results were

released in December of 2009. Prior trials, including the VAX003 and VAX004 studies, used HIV gp120 proteins as an antigen in an attempt to elicit an antibody-based response to provide immunological protection, but they failed because the antibodies were ineffective against all the varied strains of HIV (Girard and Plotkin, 2011). The RV144 trial used a prime-boost vaccine regimen in which participants were given four priming injections of a canarypox vector carrying HIV *env* followed by two boosts with gp120 (the same that was used as a single agent in the VAX004 trial). The results of this trial showed a modest 31% decrease in HIV infection in low-risk individuals who received the vaccine compared to placebo controls, but the overall numbers were low (29 new HIV cases in the placebo group, compared to 17 in the vaccine group). In the high-risk cohort (individuals who reported high risk behavior such as multiple sexual partners, commercial sex work, previous sexually transmitted infection, or needle sharing), there was no difference between the vaccine and placebo groups (Rerks-Ngarm et al., 2009). These data show the first evidence of statistical clinical success of any kind for HIV vaccines, but also demonstrate that there are considerable improvements that must be made for this to become a viable solution to curbing the global HIV pandemic.

Epidemiology, pathogenesis and treatment of sickle cell disease

Epidemiology

Sickle cell disease (SCD) is an autosomal recessive disease caused by a point mutation in the β-globin gene. It is the most common monogenic disease worldwide, and in the United States it occurs in 1 of 500 African-American births. One in twelve African-Americans has the sickle cell trait, or one mutated allele. SCD is even more prevalent in Sub-Saharan Africa and other endemic malaria locations because sickle cell trait is protective against developing severe complications including death from malaria.

Pathogenesis

Hemoglobin is the oxygen-carrying protein of red-blood cells and it is a heterotetramer, consisting in adults of two α-globin chains and two β-globin chains (HbA). In the fetus and newborn, instead of β-globin, structurally similar γ-globin chains associate with α-globin to form fetal hemoglobin (HbF). Sickle cell disease is caused by the substitution of the hydrophobic valine for the hydrophilic glutamine in the sixth position of β-globin, and the mutated β-globin molecules associate with wild type α-chains to form sickle hemoglobin (HbS). Deoxygenated HbS molecules can polymerize in the microvasculature, causing red blood cells to assume a sickle shape, blocking normal blood flow and leading

to painful vaso-occlusive crises (Odievre et al., 2011). Aplastic crises, in which many red blood cell precursor reticulocytes are destroyed, and hemolytic crises, in which the destruction of mature red blood cells is accelerated, both lead to dramatic and often life-threatening reductions in hemoglobin levels. Other severe complications of SCD include anemia, stroke, acute chest syndrome and increased susceptibility to infection (www.mayoclinic.com/health/sickle-cell-anemia).

Treatment

The only current cure for sickle cell disease is an allogenic hematopoietic stem cell transplant (HSCT) from a non-diseased, matched donor, most often a matched sibling. However, the immunosuppression required for the transplant and the risk of developing graft versus host disease, combined with the fact that fewer than 15% of sickle cell disease patients have matched siblings, prevent HSCT from being a widespread therapeutic option. In fact, although there are more than 70,000 individuals in the United States with SCD, fewer than 500 HSCTs have been performed (Shenoy, 2011). *Ex vivo* correction of the mutated β-globin gene in HSCs from sickle cell patients, followed by autologous HSCT would achieve the same phenotypic cure without the limitations of an allogenic transplant. This gene correction strategy is described in Chapter 3 of this work.

Symptomatic treatment of SCD involves pain management, careful monitoring and treatment of infections and red blood cell transfusions. The first

drug approved to alter the regulation of globins in sickle cell disease is hydroxyurea, which reactivates the expression of γ-globin. In this way, HbF is produced and protects red blood cells from assuming a sickle shape (Steinberg and Rodgers, 2001). A multicenter, clinical study of 299 adults with severe SCD found that hydroxyurea reduced by half the number of hospitalizations and the frequency of pain, acute chest syndrome and blood transfusions (Charache et al., 1995) however there was no significant difference in 28-month mortality (Steinberg and Rodgers, 2001). Understanding the molecular mechanism of γ-globin reactivation and discovering other small molecule compounds that can upregulate γ-globin expression are avenues of study with direct clinical relevance. In Chapter 3, I present a novel high-throughput method to screen compounds for their relative effects on the endogenous expression of β-globin and γ-globin by using precise genome editing at those loci.

CHAPTER II: HIV RESISTANT T-CELLS GENERATED BY TARGETED "STACKING" OF RESTRICTION FACTORS

Abstract

Zinc finger nucleases are site-specific genome editing proteins that have been developed, among other things, to combat HIV infection. Using ZFN-meditated homologous recombination, we targeted a cocktail of anti-HIV restriction factors to the *CCR5* locus in a T-cell reporter line, and identified combinations of factors that provide robust resistance to infection by CCR5-tropic (R5-tropic) and CXCR4-tropic (X4-tropic) HIV. *CCR5* disruption alone, which mimics the strategy being used in clinical trials, confers 16-fold protection against R5-tropic HIV, but it has no effect against an X4-tropic virus. Rhesus TRIM5α, chimeric human-rhesus TRIM5α, APOBEC3G D128K or Rev M10 alone targeted to *CCR5* confers significantly improved resistance to infection by both variants compared to *CCR5* disruption alone. The combination of three factors targeted to *CCR5* blocks infection at multiple stages, providing complete protection against infection by R5-tropic and X4-tropic HIV. This is the first

demonstration of a targeted gene stacking strategy effective in protecting against both R5-tropic and X4- tropic HIV infection.

Introduction

One of the major challenges to eradicating HIV infection is the virus's ability to mutate and evade therapy (Boutwell et al., 2010; Cohen, 2003; Richman et al., 2004). This has led to a broad interest in developing alternative treatment strategies including cell-based gene therapy approaches to restrict infection (Malim et al., 1989; Podsakoff et al., 2005; Ranga et al., 1998; Rossi et al., 2007; Sawyer et al., 2005; Xu et al., 2004). For example, a recent report describes the lifelong immunity conferred to humanized mice by the delivery of a gene encoding a broadly neutralizing antibody against HIV (Balazs et al., 2011). HIV infection is mediated through binding to the CD4 receptor and either the CCR5 (CCR5-tropic virus) or CXCR4 (CXCR4-tropic virus) co-receptor on the surface of $CD4^+$ T cells, the primary target cells *in vivo*. In patients, early infection is typically established by CCR5-tropic (R5-tropic) virus, while CXCR4-tropic (X4-tropic) or dual-tropic variants predominate in late stage disease (Schuitemaker et al.). Interestingly, individuals who are homozygous for the Δ32 variant of the *CCR5* gene are resistant to HIV infection but are otherwise healthy (Liu et al., 1996), making *CCR5* an intriguing target for HIV therapy. This has been done

both by small molecule approaches to inhibit binding of HIV to the CCR5 receptor (Chen et al., 2012) and by genetic manipulation to create HIV resistant cells that don't express CCR5 on the cell surface (Holt et al., 2010; Perez et al., 2008). Moreover, the demonstration of an apparent cure of an HIV infected patient by allogenic bone marrow transplantation from a matched *CCR5* Δ32 donor was recently reported (Allers et al., 2010; Hutter et al., 2009). Although it is not known whether it was the donor cells alone or a combination of ablative therapy and transplantation with HIV resistant cells that led to the apparent cure, it strongly supports the idea that using genetically modified cells is a promising approach for altering the course of HIV infection.

Specific genome modification can be achieved with engineered proteins called zinc finger nucleases (Bitinaite et al., 1998; Kim and Chandrasegaran, 1994; Porteus and Baltimore, 2003). ZFNs are composed of a zinc finger DNA binding domain fused to a FokI endonuclease domain. Each zinc finger recognizes and binds to a 3-nucleotide sequence, such that a 4-fingered protein recognizes 12 base pairs. Anti-parallel binding of two ZFNs to contiguous sites separated by a short DNA spacer leads to dimerization of the endonuclease domain and creation of a site-specific DNA double-strand break which can be repaired either by potentially mutagenic NHEJ or high-fidelity homologous recombination with a homologous DNA donor template. ZFNs have recently been developed that target the *CCR5* gene, and upon induction of a site-specific double

strand break and mutagenic repair by NHEJ, populations of HIV resistant T-cells (Perez et al., 2008) and HSCs (Holt et al., 2010) have been created which phenotypically mimic the *CCR5* Δ32 allele. The potential limitation of this approach is that in a patient infected with both X4- and R5-tropic virus, simply mutating *CCR5* in a fraction of T-cells or HSCs may not be sufficient to alter the course of the disease. Instead, cells that are genetically resistant to both HIV tropisms need to be created.

One way to generate cells that are resistant to both R5-tropic and X4-tropic HIV is to simultaneously knock out expression of CCR5 and CXCR4. In fact, recent reports (Wilen et al., 2011; Yuan et al., 2012) have described a ZFN-mediated *CXCR4* disruption strategy effective in protecting human $CD4^+$ T-cells against X4-tropic but not R5-tropic infection. To achieve dual tropic resistance, Wilen *et al* (2011) disrupted *CXCR4* in T-cells from *CCR5* Δ32 patients, suggesting a potential double knockout strategy using two pairs of ZFNs against *CXCR4* and *CCR5*.

Besides being used for targeted gene disruption, ZFNs can also be used to stimulate precise targeting of a gene therapy payload to a specific genomic locus by homologous recombination, while minimizing the risk of uncontrolled genomic insertion, which has caused serious adverse events including leukemia in several clinical gene therapy trials (Cavazzana-Calvo et al.; Hacein-Bey-Abina et al., 2003; Ott et al., 2006; Pike-Overzet et al., 2007). Taking advantage of the

benefits of this system, we have developed a novel method for creating multiple genetic resistances to HIV infection with a single gene targeting event. Our approach mimics the current treatment strategy of highly active anti-retroviral therapy. A HAART regimen typically consists of three or more drugs that inhibit HIV infection at multiple stages of the virus lifecycle. It has been shown that treatment with two antiviral drugs is better than one and that a three-drug regimen is superior (Schmit and Weber, 1997). By analogy, we hypothesize that by creating multiple layers of genetic resistance to HIV ("stacking" genetic traits), we can generate broader and more robust inhibition of both R5- and X4-tropic HIV. The ultimate goal would be to develop an approach that, when translated into humans, would create a reservoir of protected T-cells that would stave off immune collapse and the onset of AIDS, either alone or in combination with anti-retroviral drugs.

Genetic studies have recently revealed several cellular proteins that confer effective resistance to HIV. TRIM5α and APOBEC3G are restriction factors that constitute a newly appreciated arm of the mammalian innate immune system (Baumann, 2006). Interestingly, rhesus macaque TRIM5α (Stremlau et al., 2004) and an engineered human-rhesus hybrid TRIM5α (Sawyer et al., 2005) inhibit effective HIV capsid disassembly in the cytoplasm of infected human cells, while the human version of TRIM5α is significantly less effective. In fact, a variant of rhesus TRIM5α was recently used to generate a transgenic cat that was protected

from infection by feline immunodeficiency virus (Wongsrikeao et al., 2011). APOBEC3G is a cytidine deaminase that is packaged with newly formed viral particles and causes hypermutation of the viral genome (Santa-Marta et al., 2005). The APOBEC3G D128K mutant escapes depletion by the viral protein vif (Xu et

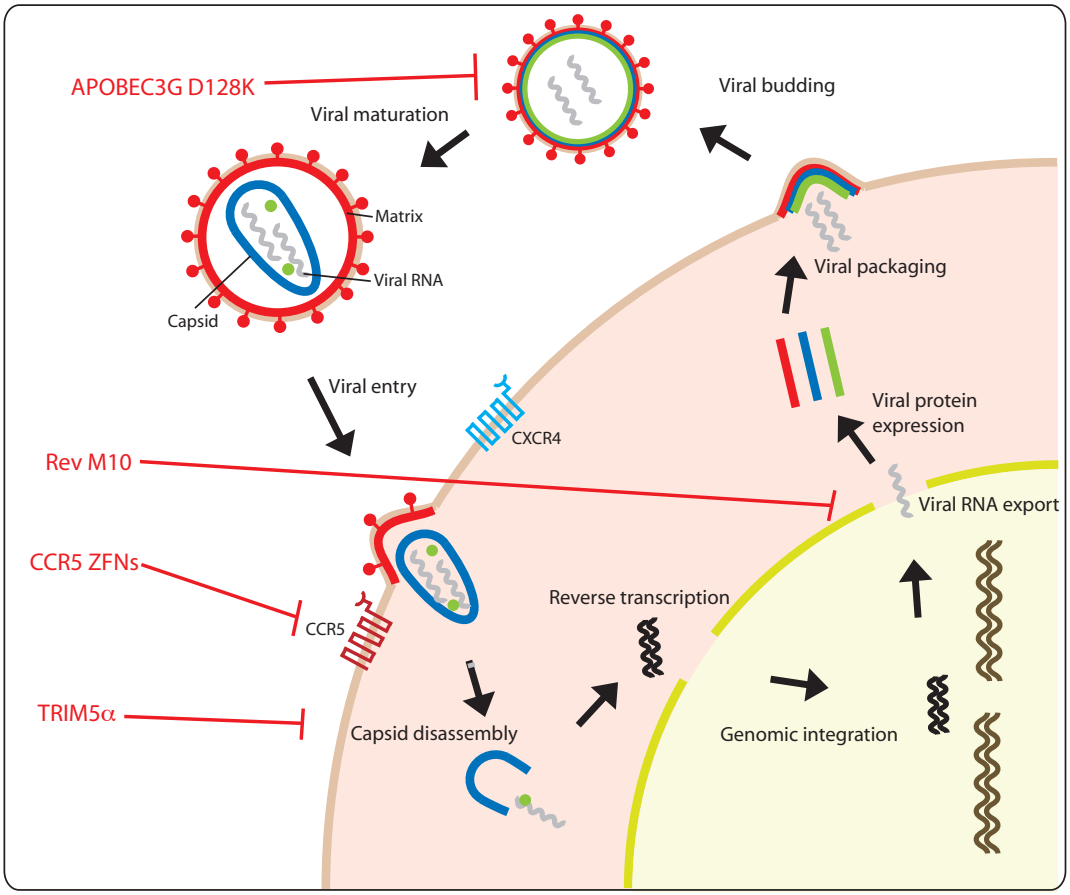


Figure 2.1: Disruption of the HIV lifecycle by targeted gene therapy. Current pharmacological therapies simultaneously attack multiple stages of the HIV lifecycle. Similarly, we have designed a ZFN-based targeting strategy to insert into the *CCR5* locus the genes APOBEC3G D128K, Rev M10, and human-rhesus TRIM5α. In doing so, we stack genetic resistance to R5-tropic and X4-tropic HIV.

al., 2004) and thus successfully interferes with HIV replication. Rev M10 is a dominant negative form of a viral protein that prevents the export of early viral RNAs from the nucleus (Malim et al., 1989; Malim et al., 1991) (Figure 2.1). Rev M10 differs from the other two proteins in that it is not a naturally occurring molecule, although it effectively behaves as a restriction factor when expressed in cells.

Here, the strengths of previous transgenic approaches have been combined with recent knowledge of HIV cellular restriction to generate the first example of T-cells that are broadly resistant to HIV infection. Our strategy involves creating multi-layered genetic resistance to HIV (schematized in Figure 2.2) by disrupting *CCR5* while simultaneously integrating a cassette of anti-HIV restriction factors. We demonstrate that this approach conveys complete suppression of viral replication as well as full-spectrum resistance to both R5-tropic and X4-tropic HIV forms.

Results

ZFN-mediated gene targeting at the CCR5 locus

To determine the efficiency of targeting to the *CCR5* locus by homologous recombination (Figure 2.3a), we constructed a GFP expression cassette with one kilobase arms of homology from the *CCR5* gene centered at the CCR5 ZFN cut site (Perez et al., 2008), and included a negative selectable herpes simplex virus

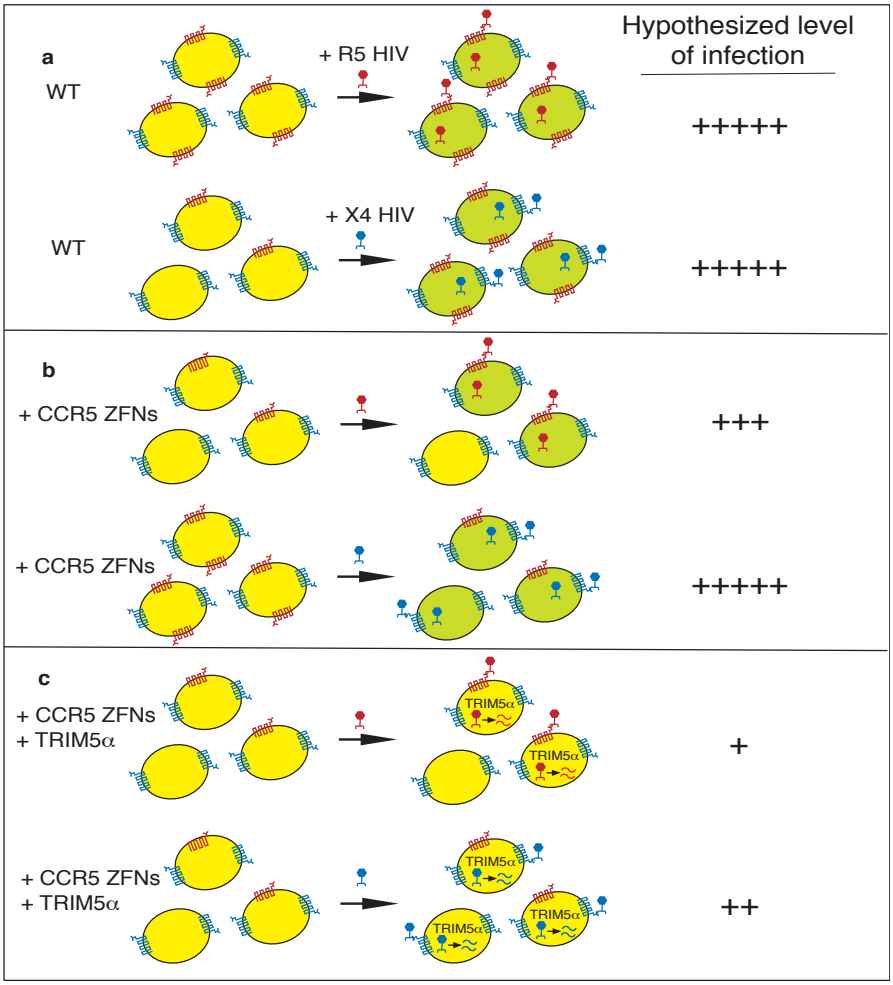


Figure 2.2: Rationale for targeting anti-HIV genes to the *CCR5* locus. **(a)** Wild type JLTRG-R5 cells are susceptible to infection by R5- (red) or X4-tropic (blue) HIV because they express both CCR5 (red receptor) and CXCR4 (blue receptor). Since JLTRG-R5 cells express GFP upon infection, we hypothesize that wild type cells will express high levels of GFP after infection by R5- and X4-tropic HIV (level of infection: "+++++"). **(b)** Treatment with CCR5 ZFNs alone will confer protection against R5-tropic HIV in the cells that have biallelic disruption of *CCR5*, leading to fewer GFP positive cells ("+++"), but ZFN treatment alone will not protect cells from infection by X4 HIV. **(c)** Cells targeted with TRIM5α or other anti-HIV genes at one or both *CCR5* alleles will have multiple genetic layers of protection against R5 infection ("+") and will also be protective against X4 infection ("++") because the anti-HIV factors are post-entry inhibitors of the HIV lifecycle, resulting in very few GFP positive cells. Targeting multiple anti-HIV genes simultaneously serves to stack genetic resistance against both viral tropisms.

thymidine kinase (HSV-TK) domain outside the homology ("CCR5-GFP" targeting vector, Figure 2.3b). Delivery of the CCR5-GFP targeting vector along with CCR5 ZFN DNA or mRNA into human erythroleukemic K562 cells by nucleofection resulted in stable GFP expression in up to 30% of the transfected cells (20% overall) compared to 2% (1.4% overall) in the absence of ZFNs (Figure 2.3c). Negative selection with ganciclovir led to a 2.5 fold enrichment of targeted cells compared to random integrants, consistent with targeting at the

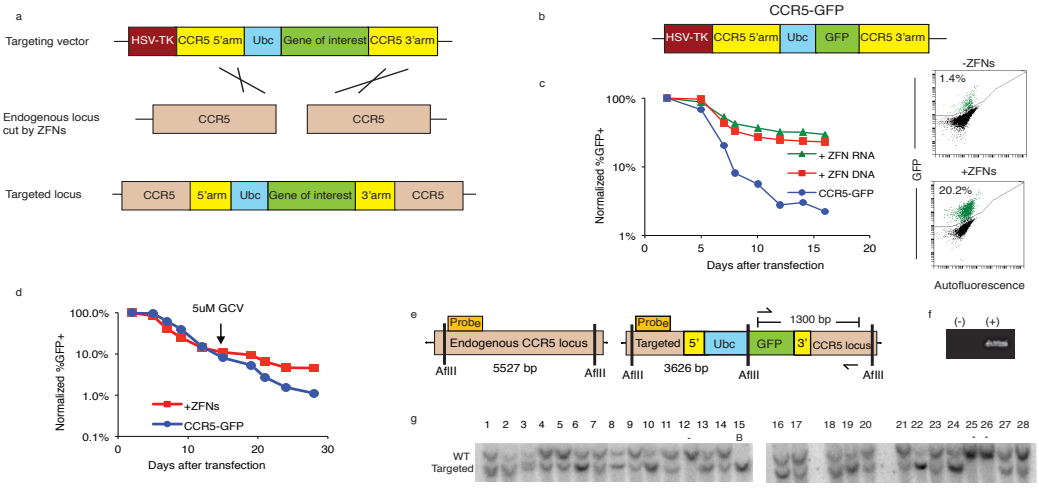


Figure 2.3: Targeting GFP to the endogenous *CCR5* locus using ZFN-mediated homologous recombination. (**a**) Schematic of ZFN-mediated gene targeting by homologous recombination. **(b)** CCR5-GFP targeting vector. For complete list and abbreviations see Figure 2.4. **(c)** Targeting of Ubc-GFP to *CCR5* **(d)** Negative selection with ganciclovir enriched for targeted cells. **(e)** Diagram of genomic PCR and Southern blot strategies used to show targeting at *CCR5*. Probe – DNA probe used for Southern blot, arrows – PCR primers. **(f)** Genomic PCR demonstrating ZFN-mediated targeting at *CCR5.* **(g)** Southern blot analysis of GFP positive clones following targeting. "-" untargeted, B – biallelically targeted.

same genomic locus (Figure 2.3d).

To confirm that targeting occurred at the *CCR5* locus, we designed a PCR strategy with a forward primer that binds a sequence in the CCR5-GFP targeting vector and a reverse primer recognizing a site in the CCR5 locus (Figure 2.3e). A band indicative of targeting was detected in ZFN-treated samples but not in control samples (Figure 2.3f). Clonal molecular analysis demonstrated that 89% of the GFP positive cells (25/28 clones) were the result of targeting the GFP expression cassette to the *CCR5* locus (Figures 2.3g), thus showing an overall targeting rate of 27% in the unsorted population.

Generation of HIV-resistant cells by targeting restriction factors to CCR5

We next investigated whether targeting an anti-HIV restriction factor to the *CCR5* locus by homologous recombination would confer greater restriction to HIV infection compared to *CCR5* disruption alone (Figure 2.2). To do so, we modified the CCR5-GFP targeting vector to replace the GFP expression cassette with a TRIM5α-IRES-puro cassette ("CCR5-TRIM5α targeting vector", Figure 2.4 and 2.5a). Three versions of this targeting vector were constructed in which the TRIM5α gene is either human (h), which does not restrict HIV infection; rhesus (rh), which confers robust resistance (Stremlau et al., 2004); or a human-rhesus hybrid (hrh), which has been shown to confer intermediate resistance to HIV (Sawyer et al., 2005). A fourth version of the targeting vector was created to

include only the puromycin resistance gene, allowing for the selection of *CCR5* disrupted cells ("CCR5-hTRIM5α," "CCR5-rhTRIM5α," "CCR5-hrhTRIM5α," "CCR5-IRES-puro" targeting vectors, Figure 2.4). ZFN-mediated integration of the CCR5-rhTRIM5α targeting vector at the *CCR5* locus was detected at greater than 50% of the alleles in K562 cells following selection with puromycin (Figure 2.5b).

For use in HIV challenge experiments, we used JLTRG-R5, a human Jurkat T-cell reporter line (Ochsenbauer-Jambor et al., 2006). JLTRG-R5 cells express CD4 and CXCR4 at levels similar to primary $CD4^+$ T-cells and express

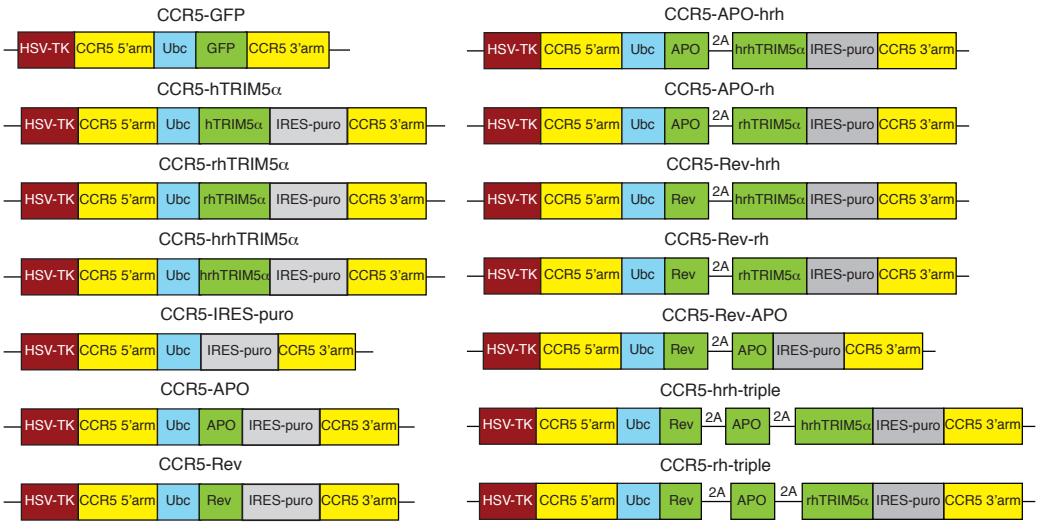


Figure 2.4: CCR5 targeting vectors. Schematic representations of CCR5 targeting vectors used in this study. (HSV-TK – herpes simplex virus thymidine kinase, Ubc – ubiquitin C promoter, hTRIM5α – human TRIM5α, rhTRIM5α – rhesus TRIM5α, hrhTRIM5α – human-rhesus hybrid TRIM5α, Rev – Rev M10, APO – APOBEC3G D128K, IRES-puro – internal ribosome entry site – puromycin N-acetyltransferase).

CCR5 at levels similar to peripheral blood mononuclear cells (Ochsenbauer-Jambor et al., 2006). Additionally, they have an integrated LTR-GFP reporter that expresses GFP upon HIV infection (Figure 2.5c). We used the CCR5 ZFNs to integrate the CCR5-hrhTRIM5α and CCR5-rhTRIM5α targeting vectors into the *CCR5* locus in JLTRG-R5 cells. Again, we showed targeting of up to 50% of the alleles in the puromycin-selected population (Figure 2.5e). Similarly, we targeted CCR5-hTRIM5α and CCR5-IRES-puro to *CCR5* in JLTRG-R5 (data not shown). Protein levels of each of the targeted TRIM5α variants were detected (Figure 2.5e, lanes 2,3,4), although the expression of the rhTRIM5α was approximately two-fold lower than hrhTRIM5α or hTRIM5α as quantified by band intensities using ImageJ software (Figure 2.5f). To check the status of the untargeted *CCR5* allele, we stained the targeted cells for CCR5 and analyzed by FACS. Notably, in the puromycin-selected population, greater than 99% of the untargeted alleles were cut by the ZFNs and mutagenically repaired by NHEJ, eliminating virtually all expression of functional CCR5 (Figure 2.5g). In this way, we have efficiently created a population of cells that are disrupted at both *CCR5* alleles, one by gene targeting and the second by mutagenic NHEJ.

Targeted single-factor cell lines are significantly protected against R5-tropic and X4-tropic HIV

To quantitate the level of resistance conferred by targeting these TRIM5α variants to *CCR5*, we challenged the targeted cells with R5-tropic HIV-1_{BaL-1}. Samples were analyzed by flow cytometry for GFP expression from the integrated LTR-GFP reporter, and ratio to cell negative (RTCN) values were calculated as previously described (Vodros and Fenyo, 2005). Briefly, RTCN is the cross

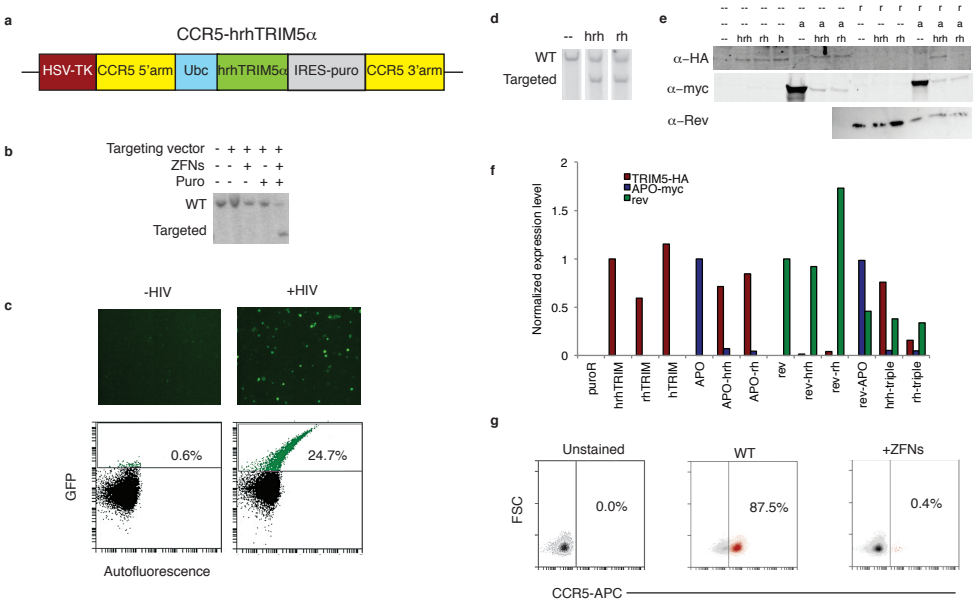


Figure 2.5: Establishment of JLTRG-R5 T-cell line derived cells expressing anti-HIV genes. (a) CCR5-hrhTRIM5α targeting vector. **(b)** Southern blot showing targeting of hrhTRIM5α in K562s and enrichment with puromycin. **(c)** JLTRG-R5 cells express GFP upon infection by HIV. **(d)** Southern blot showing targeting of anti-HIV genes in JLTRG-R5 cells. **(e)** Protein expression from targeted cassettes. **(f)** Quantification of bands from Western blots in **(e)**, normalized to hrhTRIM, APO and rev lanes, respectively. **(g)** FACS plots showing ZFN-mediated disruption of the untargeted *CCR5* allele in CCR5-IRES-puro cells. (r – Rev M10, a – APOBEC3G D128K-myc, hrh – human-rhesus hybrid TRIM5α-HA, rh – rhesus TRIM5α-HA, h – human TRIM5α HA).

product of percent GFP positive cells times mean fluorescence intensity, normalized to uninfected samples and was used to quantify level of infection (the higher the RTCN value, the greater the level of infection). Infection was followed over the course of 14 days, and as expected, wild type JLTRG-R5 cells were susceptible to infection by R5-tropic virus with a maximum RTCN value of 660 (Figure 2.6a, blue line, Day 11). CCR5-IRES-puro cells, which mimic previously reported *CCR5* disruption, reached a maximum infection of 42 with R5-tropic

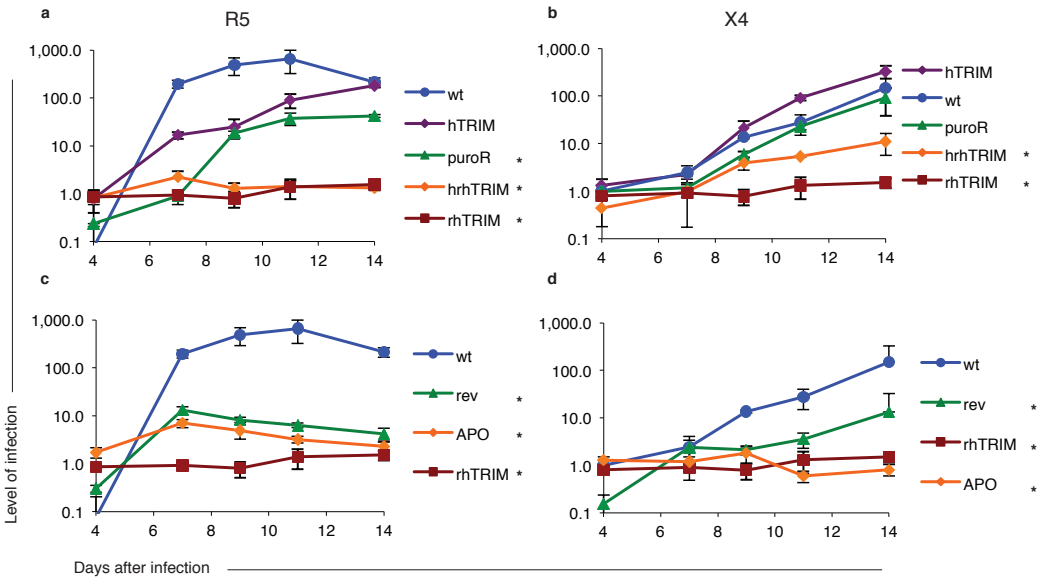


Figure 2.6: Targeting a single anti-HIV restriction factor to CCR5 confers significant resistance to R5-tropic and X4-tropic HIV. Infection timecourse of CCR5-TRIM5α cells and CCR5-IRES-puro cells with R5-tropic **(a)** and X4-tropic **(b)** HIV. Infection timecourse of CCR5-APO and CCR5-rev cells with R5-tropic **(c)** and X4-tropic **(d)** HIV. (*p < .05 compared to wild type at maximum infection., Error bars represent standard deviation, small error bars are obscured by markers.)

HIV, a protection of 16-fold (Figure 2.6a and Table 2.1). Similarly, CCR5-hTRIM5α cells showed modest but not statistically significant ($p < .07$) restriction of R5-tropic HIV infection, consistent with simply knocking-out *CCR5*. In contrast, CCR5-hrhTRIM5α cells and CCR5-rhTRIM5α cells demonstrated 426-fold (RTCN = 1.6) and 494-fold (RTCN = 1.3) resistance to R5-tropic HIV, respectively as measured through day 14 (Figure 2.6a and Table 2.1). Thus compared to the level of protection achieved by *CCR5* disruption alone through the targeting of IRES-puro, combining *CCR5* disruption with the targeted

Anti-HIV genes			Fold resistance to R5 HIV	Fold resistance to X4 HIV
WT cells			1	1
--	--	--	16	2
		+ Rev	80	14
	+ APO	--	134	104
		+ Rev	226	68
hrhTRIM	--	--	426	17
		+ Rev	39	3
	+ APO	--	100	6
		+ Rev	521	177
rhTRIM	--	--	494	142
		+ Rev	215	54
	+ APO	--	215	28
		+ Rev	92	66
hTRIM	--	--	4	1

Table 2.1: Fold protection of *CCR5* targeted cells after 14 days of infection. Fold protection was calculated as the maximum RTCN value of wild type cells divided by the maximum RTCN value of each cell line through the first 14 days of infection by R5-tropic or X4-tropic HIV. "-- -- --" has no additional anti-HIV genes besides *CCR5* disruption achieved by CCR5-IRES-puro targeting.

integration of an anti-HIV factor increases the resistance of these cells an additional 30-fold, providing virtually complete inhibition of R5-tropic infection by day 14.

Since TRIM5α acts as a post-entry restriction factor, we hypothesized that the CCR5-hrhTRIM5α and CCR5-rhTRIM5α cells would also be resistant to infection by X4-tropic HIV, while the CCR5-IRES-puro cells would not. Each of the targeted cell lines was challenged with X4-tropic HIV-1_{NL4-3}. Wild type JLTRG-R5 cells were infected to a maximum RTCN value of 187, and the CCR5-IRES-puro cells and the CCR5-hTRIM5α cells were not significantly protected from X4-tropic infection (Figure 2.6b and Table 2.1), supporting the hypothesis that the resistance these cells displayed against the R5-tropic virus was due solely to disruption of *CCR5*. Importantly, the CCR5-hrhTRIM5α cells and the CCR5-rhTRIM5α cells displayed 17-fold (RTCN = 11) and 142-fold (RTCN = 1.3) resistance to X4-tropic HIV, respectively (Figure 2.6b and Table 2.1). In this way, we showed that compared to *CCR5* disruption alone, targeting rhesus- or human-rhesus hybrid TRIM5α to *CCR5* both increases the protection of a T-cell line against R5-tropic infection and confers significant resistance to X4-tropic infection.

To investigate whether other post-entry restriction factors would also serve to protect against both viral tropisms, we modified the targeting vectors to

include the vif-resistant D128K mutant of APOBEC3G or dominant negative Rev M10 ("CCR5-APO" and "CCR5-Rev" targeting vectors, Figure 2.4). Following targeting, we detected expression of the myc-tagged APOBEC3G D128K and Rev M10 (Figure 2.5e, lanes 5 and 8, and Figure 2.5f). We showed that the CCR5-APO cells and the CCR5-Rev cells were protected 134-fold (RTCN =4.9) and 80-fold (RTCN = 8.2) against R5-tropic HIV, respectively (Figure 2.6c and Table 2.1). Similarly, against X4-tropic HIV, these restriction factors provided 104-fold (RTCN = 1.8) and 14-fold (RTCN = 13.4) protection through day 14, respectively (Figure 2.6d and Table 2.1).

Targeted stacking of restriction factors confers complete resistance to R5-tropic and X4-tropic infection

Since the combination of *CCR5* disruption with TRIM5α expression was more protective than *CCR5* disruption alone against R5-tropic HIV, we sought to determine if stacking resistance factors together in combination as a cassette of anti-HIV genes would confer greater resistance to both X4-tropic and R5-tropic variants. To do this we again modified the targeting vectors to include all combinations of Rev M10, APOBEC3G D128K and either hrh-TRIM5α or rh-TRIM5α ("CCR5-Rev-hrh," "CCR5-Rev-rh," "CCR5-APO-hrh," "CCR5-APO-rh," "CCR5-Rev-APO," "CCR5-hrh-triple," and "CCR5-rh-triple," Figure 2.4) and targeted them to *CCR5* in JLTRG-R5 cells. Protein expression of the anti-

HIV genes was markedly lower from many of the two- and three-factor combination cassettes compared to the single-factor cell lines (Figures 2.5e and 2.5f). Because of the robust protection provided by each of the factors alone (RTCN values between 1.3 and 13.4), there was no significant benefit of adding additional factors through day 14 of infection (Table 2.1). However by day 23 of

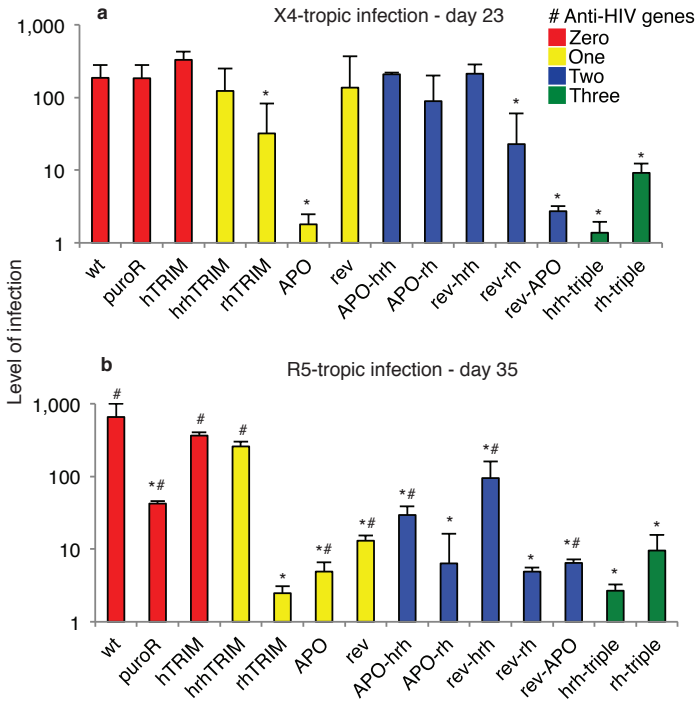


Figure 2.7: Stacking genetic resistance provides complete protection from infection by R5-tropic and X4-tropic HIV. Maximum infection level of cell lines targeted at *CCR5* with zero (red bars), one (yellow), two (blue) or three (green) restriction factors by day 23 of infection with X4-tropic HIV **(a)** or by day 35 of infection with R5-tropic HIV **(b)**. (Error bars represent standard deviation, *p < .05 compared to wt, in **(b)** # p < .05 compared to CCR5-hrh-triple).

continuous infection, the resistance to X4-tropic infection was either completely or partially abrogated in the CCR5-hrhTRIM5α cells and the CCR5-rhTRIM5α cells, with RTCN values of 123 and 32, respectively (Figure 2.7a and Table 2.2) similar to a previous report of decreased TRIM5α restriction upon prolonged infection (Pacheco et al., 2010). Similarly, the CCR5-Rev cells showed a maximum RTCN value of 137 by day 23. By day 23, many of the two-factor targeted cell lines had also become infected by X4-tropic virus (Figure 2.7a and Table 2.2).

Of the one- and two-factor targeted lines (Figure 2.7a, yellow and blue bars, respectively and Table 2.2), the CCR5-rhTRIM5α, CCR5-APO, CCR5-Rev-rh, and CCR5-rev-APO cells were significantly less infected than wild type cells. Interestingly, the CCR5-APO and CCR5-Rev-APO cells remained virtually uninfected at day 23 (Table 2.2). One explanation for this observation is that since the protein level of APOBEC3G D128K in these two lines was at least 20-fold higher than in any of the other lines (Figure 2.5f, blue bars), it provided enough restriction that additional factors were not necessary to further inhibit infection. However, such overexpression of a cytidine deaminase could have detrimental cellular genomic consequences (see Discussion). For example, we were able to generate HEK 293T cell lines with targeted integration into the *CCR5* locus of all of the anti-HIV restriction factors and combination of factors except with the CCR5-APO and CCR5-Rev-APO targeting vectors which express the highest

levels of APOBEC3G D128K (data not shown). We believe this inability to create 293T cell lines with targeted integration of a highly expressed *APOBEC3G D128K* gene is a reflection of the potential toxicity of its overexpression. Importantly, the modest expression of all three anti-HIV factors in the CCR5-hrh-triple and CCR5-rh-triple cell lines (Figures 2.5e and 2.5f) conferred significant protection to these cells through 23 days of infection (Figure 2.7a, green bars and Table 2.2).

Anti-HIV genes			Fold resistance to R5 HIV	Fold resistance to X4 HIV
WT cells			1	1
--	--	--	16	1
		+ Rev	50	1
	+ APO	--	134	104
		+ Rev	102	68
hrhTRIM	--	--	3	2
		+ Rev	7	1
	+ APO	--	22	1
		+ Rev	246	135
rhTRIM	--	--	266	6
		+ Rev	134	8
	+ APO	--	104	2
		+ Rev	69	20
hTRIM	--	--	2	1

Table 2.2: Fold protection of CCR5 targeted cells after extended infection. Some previously resistant cell lines **(Table 2.1)** became infected after extended infection times of 35 days (R5) or 23 days (X4).

Because of the additional layer of protection that *CCR5* disruption confers against R5-tropic HIV (compared to X4-tropic HIV), the single- and double-factor targeted cells remained protected from infection by R5-tropic virus through day 23, but some became moderately infected by day 35 (Figure 2.7b, yellow and blue bars, respectively and Table 2.2). With the exception of the CCR5-hTRIM5α and CCR5-hrhTRIM5α cells, all other lines still have a significantly lower maximum infection level than wild type cells by day 35 (Figure 2.7b, *). However the CCR5-hrh-triple and CCR5-rh-triple cells are significantly less infected than most of the single- and double-factor cell lines (Figure 2.7b, #), demonstrating that the combination of all three factors targeted to *CCR5* is most protective against R5-tropic infection.

Expression of rhTRIM5α and hrhTRIM5α inhibits the initial round of infection

Finally, to further investigate the mechanism of HIV resistance in each of these cell lines, we used a single round infectivity assay (McMahon et al., 2007; McMahon et al., 2011). First, we sought to determine whether this assay was sensitive enough to detect differences in the susceptibility of our targeted cell lines. To do this, we generated replication incompetent CCR5-tropic pseudovirions containing red fluorescent protein in place of the viral envelope protein by co-transfection of HEK 293T cells with the NL4-3-ΔE-RFP plasmid

and the CCR5 envelope SF162 plasmid. Culture supernatants were collected and used to infect wild type JLTRG-R5 cells. Infection was assessed by quantifying RFP positive cells two days post-infection. Importantly, wild type cells had measurable infection at day 2, while cell lines targeted at *CCR5* were significantly protected against infection likely due to the disruption of *CCR5* (Figure 2.8a).

To quantify the contribution the targeted anti-HIV restriction factors have independent of the effects from *CCR5* disruption, we made VSV-G pseudotyped single-round RFP virus (HIV-$1_{NL4\text{-}3\text{-}\Delta E\text{-}RFP\text{-}VSVG}$) and infected cells as above. Because TRIM5α is a post-entry, pre-integration restriction factor, CCR5-hrhTRIM5α cells and CCR5-rhTRIM5α cells should still be protected against single round VSV-G pseudotyped HIV infection. In fact, when challenged in the

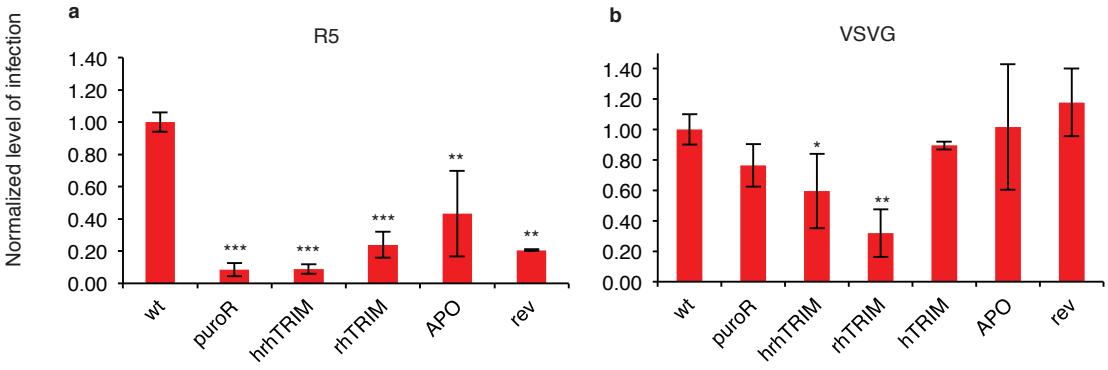


Figure 2.8: TRIM5α blocks the initial round of infection by HIV. Infection of *CCR5*-targeted cell lines with single round HIV pseudotyped with (**a**) R5 envelope or (**b**) VSV-G envelope (Error bars represent standard deviation * $p < .05$, ** $p < .005$, *** $p < .0005$).

single round assay, CCR5-hrhTRIM5α cells became significantly less infected, reaching only 59% the level of infection compared to wild type cells. Similarly, CCR5-rhTRIM5α cells were infected at 32% of wild type (Figure 2.8b). Furthermore, because Rev M10 and APOBEC3G D128K are post-integration restriction factors, they should not inhibit the initial cycle of infection, but only subsequent rounds. As predicted, the cells targeted with CCR5-hTRIM5α, CCR5-APO, and CCR5-Rev showed no protection in the single round infectivity assay (Figure 2.8b).

Discussion

Using ZFN-mediated homologous recombination to target a cassette of anti-HIV restriction factors (rh or hrh-TRIM5α, APOBEC3G D128K and Rev M10) to the *CCR5* locus, we have created robust resistance to infection by both X4- and R5-tropic HIV and have established a framework to test the relative fold protection genetic restriction factors provide against HIV infection (Table 2.1 and Table 2.2). This is the first demonstration of targeted trait stacking against HIV infection using homologous recombination to precisely integrate the restriction factors into the genome. A recent report (Lei et al., 2011) describes ZFN-mediated insertion of GFP into *CCR5* at targeting rates of up to 10% using a baculoviral delivery method in Ghost-CCR5 cells. A second study (Lombardo et

al., 2007) used integration-defective lentivirus to deliver CCR5 ZFNs and targeting vector and achieved targeting rates of 0.1% in human CD34+ hematopoietic stem/progenitor cells and 5% in human embryonic stem cells, but neither group infected the targeted cells with HIV. Here, we achieve targeting rates of 27% with GFP, and further demonstrate targeting of anti-HIV factors that confer effective resistance to infection by HIV.

Since targeting a puromycin resistance gene alone to *CCR5* is functionally equivalent to *CCR5* disruption by ZFN-induced NHEJ, we were able to directly compare the efficiency of our ZFN-mediated homologous recombination trait-stacking strategy with the published ZFN-mediated *CCR5* disruption approach currently in clinical trials (Ando et al., 2011; Mitsuyasu et al., 2011). In fact, our strategy overestimates the effectiveness of *CCR5* disruption alone because we selected for targeted cells, most of which have undergone mutagenic NHEJ at the untargeted CCR5 allele (Figure 2.8d), in effect resulting in a population of cells with bi-allelic CCR5 disruption. Even in this context, against R5-tropic HIV, targeted integration of an anti-HIV gene to *CCR5* provides up to 30-fold increased protection compared to *CCR5* disruption alone.

In an effort to restrict infection by X4-tropic virus, CXCR4 ZFNs have been recently described (Wilen et al., 2011; Yuan et al., 2012), and when used in combination with CCR5 ZFNs could confer resistance to both viral tropisms. However, the possible implementation of this approach is limited to the mature

post-thymic CD4+ T-cell compartment and would not be feasible in HSCs because of the requirement of functional CXCR4 in B-cell development (Dar et al., 2006; Nagasawa et al., 1996) and HSC homing. The use of ZFNs against *CXCR4* and *CCR5* to confer dual-tropic HIV resistance would also require bi-allelic gene disruption at two distinct sites in the same cell, a relatively rare occurrence even with highly active ZFN pairs. Finally, the simultaneous creation of two double-strand breaks would likely lead to increased genomic instabilities as simultaneous double-strand breaks can lead to chromosomal translocations (Brunet et al., 2009; Chiarle et al., 2011) or deletions (Lee et al., 2009).

Our results suggest an alternative strategy to generate cells with dual-tropic resistance. In addition to robust protection against R5-tropic HIV, our single factor cell lines showed significant protection through the first two weeks of infection by X4-tropic HIV. By the third week, the cells targeted with CCR5-hrhTRIM5α, CCR5-rhTRIM5α and CCR5-Rev became infected, but the CCR5-hrh-triple and CCR5-rh-triple lines remained uninfected. Interestingly, the CCR5-APO cell line also maintained resistance to both R5-tropic and X4-tropic infection through 3 weeks. Because of this, one explanation of the complete resistance of the CCR5-triple cell lines is that the protection was due to APOBEC3G D128K alone with no additional benefit from Rev M10 or TRIM5α. We have evidence to suggest, however, that the complete protection was due to the effects of all three factors. The expression of the myc-tagged APOBEC3G D128K was at least 20-

fold higher in the CCR5-APO cells than in the CCR5-hrh-triple and CCR5-rh-triple lines. Furthermore, in the cell lines expressing two out of the three anti-HIV factors (ie, CCR5-APO-rh), expression of APOBEC3G D128K was similar to the CCR5-rh-triple cells, but the CCR5-APO-rh cells became infected by X4-tropic HIV (RTCN = 92) by the third week (Figure 2.7a). Therefore, when APOBEC3G D128K was expressed at low levels, it was only in combination with Rev M10 and rh- or hrh-TRIM5α that it provided complete protection against R5- and X4-tropic HIV. Because APOBEC3G is a cytidine deaminase, limiting its overexpression is preferred to avoid unintended host cell genomic DNA modifications (Albin and Harris, 2010; Harris et al., 2002; Yamanaka et al., 1995) or other cellular events that have been linked to hepatic metastases in colorectal cancer (Ding et al.). Similarly, the seemingly contradictory observation that the double-factor CCR5-rev-hrh cells were less resistant to infection than the single-factor CCR5-rev and CCR5-hrhTRIM5α cells can be explained by the lower level of protein expression in the CCR5-rev-hrh cells. Also, higher expression of TRIM5α in the CCR5-hrh-triple cells compared to the CCR5-rh-triple cells also explains the slightly higher resistance in the triple line expressing the human-rhesus hybrid protein (Figures 2.5e and 2.5f).

In summary, targeted trait stacking is an improvement over previous studies that solely relied on *CCR5* (Holt et al., 2010; Kim et al., 2009; Perez et al., 2008) or *CXCR4* (Wilen et al., 2011; Yuan et al., 2012) disruption to prevent HIV

infection for three reasons. First, unique to our system, mono-allelically modified cells confer effective resistance to infection (Figure 2.2). Second, just as pharmacological therapies against HIV infection inhibit multiple stages of the HIV lifecycle, our system provides multiple genetic blockades against infection (Figure 2.1). In the first 14 days following infection, all combinations of anti-HIV factors provided resistance and there were only slight differences in the effectiveness of some of the single-, double- and triple-factor cassettes. However, by day 35 of infection with R5-tropic HIV, many of the cell lines targeted with single factor cassettes had become infected. Notably the CCR5-hrh-triple cells remained uninfected throughout the course of the experiment. The third major improvement of this strategy is that targeted trait stacking at the *CCR5* locus is effective in preventing infection by both R5- and X4-tropic HIV. RTCN values for CCR5-hrh-triple cells were less than 2 for both viral tropisms. The large difference in the "fold protection" against X4-tropic (135-fold) and R5-tropic (522-fold) can be attributed primarily to the variable susceptibility of wild type cells to X4- and R5-tropic infection. Combining the targeted trait stacking strategy described here with recent advances in ZFN modification of human hematopoietic stem and progenitor cells is the next pre-clinical hurdle in developing robust, long-term genetic protection against infection by HIV.

Materials and Methods

Cell lines and cell culture

K562s (ATCC) and JLTRG-R5 (AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH from Dr. Olaf Kutsch) were maintained in RPMI 1640 (Hyclone) supplemented with 10% bovine growth serum, 100units/ml penicillin, 100μg/ml streptomycin and 2mM L-glutamine.

Plasmid construction

Flag-tagged CCR5 ZFNs (Perez et al., 2008) were synthesized and cloned into pcDNA6 (Invitrogen). Targeting vectors were constructed by PCR amplifying CCR5 arms of homology centered on the ZFN cut site using the following primers: 5'fwd, 5'-TTCCTGCCTCATAAGGTTGC-3', 5' rev 5'-AGGATGAGGATGACCAGCAT-3', 3'fwd 5'-GATAAACTGCAAAAGGCTGAAGAG-3', 3' rev 5'-AGACCCTCTATAACAGTAACTTCCT-3'. pcRev (AIDS Research and Reference Reagent Program from Dr. Bryan R. Cullen) and pcDNA-APO3G (AIDS Research and Reference Reagent Program from Drs. Klaus Strebel and Sandra Kao) were subjected to site-directed mutagenesis to create the rev M10 (Malim et al., 1991) and APOBEC3G D128K (Xu et al., 2004) mutants. HA-tagged TRIM5α constructs were used as previously described (Sawyer et al.,

2005). The three anti-HIV genes were cloned in all combinations and followed by IRES-puromycin acetyltransferase between the CCR5 arms of homology. Included outside the homology arms was HSV-TK domain for negative selection.

Virus preparation and titer

Plasmid pNL4-3 (generous gift of Dr. Beth Levine) was transfected into HEK293T cells to make replication competent CXCR4-tropic HIV. Supernatant was collected on day 2 and viral titer was measured by p24 ELISA (Cell Biolabs) following the manufacturer's protocol. Cell-free viral supernatant of CCR5-tropic HIV-$1_{Ba\text{-}L}$ (AIDS Research and Reference Reagent Program from Dr. Suzanne Gartner, Dr. Mikulas Popovic and Dr. Robert Gallo) was propagated in wild type JLTRG-R5 cells and viral supernatant was collected, titered, and used in subsequent experiments. Single round HIV was produced as previously described (McMahon et al., 2007). Briefly pNL4-$3_{\Delta E\text{-}RFP}$ was co-transfected into HEK293T cells with either pVSVG (VSVG envelope) or pSF162 (R5 envelope) and viral supernatant was collected on day 2 after transfection.

Cell transfection and selection

K562 cells were nucleofected (Lonza) with 10μg targeting vector and 1μg of each ZFN plasmid or mRNA using program T-016 and nuclefection buffer containing 100mM KH_2PO_4, 15mM $NaHCO_3$, 12mM $MgCl_2 \bullet 6H_20$, 8mM ATP, 2mM

glucose, pH 7.4. JLTRG-R5 cells were nucleofected with the same conditions except with program I-010. Targeted cells were selected with 0.5μg/ml puromycin. Negative selection was performed with 5uM ganciclovir.

Gene targeting

Genomic PCR was performed to show targeting to *CCR5* using a forward primer within GFP (5'-TTCAAGATCCGCCACAACATCG-3') and a reverse primer outside the 3' arm of homology (5'-ACAGATGCCAAATAAATGGATG-3'). A Southern blot probe was generated upstream of the 5' arm of homology by PCR amplification of genomic DNA using the following primers: fwd 5'-GGCCAGAAGAGCTGAGACATCCG-3', rev 5'-CGTCTGCCACCACAGATGAATGTC-3'. Radioactive Southern blotting was performed using standard techniques.

Western blot

Expression of targeted anti-HIV genes was detected by Western blot using the following primary antibodies: 1:500 mouse α-rev (Thermo), 1:5,000 mouse α-myc (Roche), 1:300 rabbit α-HA (Santa Cruz); and secondary antibodies: 1:10,000 goat α-mouse-HRP (Santa Cruz) and 1:10,000 goat α-rabbit-HRP (Santa Cruz).

Immunostaining for CCR5

Surface expression of CCR5 was measured by staining 0.5 million JLTRG-R5 cells with 10μl APC-conjugated α-CCR5 antibody (BD Biosciences). Incubation was performed in 100μl PBS/2% serum for 30 minutes at 4^{o}C and flow cytometry was performed using an Accuri C6 cytometer (Accuri).

Quantitation of infection

In multi-round infection experiments, 10^5 wild type or targeted JLTRG-R5 cells were incubated with either 1ng p24 X4-tropic or 10ng p24 R5-tropic HIV. Because JLTRG-R5 cells contain an integrated LTR-GFP (Ochsenbauer-Jambor et al., 2006), the level of infection was determined using GFP fluorescence and was calculated as the ratio to cell negative ($RTCN = (\%GFP)_{sample} * (MFI)_{sample} / (\%GFP)_{wt} * (MFI)_{wt}$) (Vodros and Fenyo, 2005). Fluorescence was measured by flow cytometry using an Accuri C6 every 2-3 days. At every time point an aliquot of each sample was pelleted and fixed by resuspension in either 4% paraformaldehyde or 2% formaldehyde, and incubated 30 minutes at 4^{o}C prior to FACS. In single-round infection experiments, 10^5 wild type or targeted cells were spin infected with 100μl viral supernatant at 1200xg for 2 hours. On day 2, cells were prepared for FACS analysis as above.

CHAPTER III: THERAPEUTIC GENE EDITING AND THE GENERATION OF ENDOGENOUS FLUORESCENT REPORTERS AT THE HUMAN GLOBIN LOCI

Abstract

Tal-effector nucleases are engineered proteins that induce DNA double-strand breaks in a sequence specific manner and can stimulate precise genome editing. Sickle cell disease and β-thalassemia are common genetic disorders caused by mutations in β-globin so we engineered a pair of highly active TALENs that induce modification of 54% of human β-globin alleles near the site of the sickle mutation. These TALENS stimulate targeted integration of therapeutic, full-length β-globin cDNA to the endogenous β-globin locus in 19% of cells prior to selection as quantified by single molecule real time sequencing. We also developed highly active TALENs to human γ-globin, a pharmacologic target in sickle cell disease therapy. Using the β-globin and γ-globin TALENs, we generated cell lines that express GFP under the control of the endogenous β-globin promoter and tdTomato under the control of the endogenous γ-globin promoter. With these fluorescent reporter cell lines, we screened a library of small molecule compounds for their differential effect on the transcriptional

activity of the endogenous β- and γ-globin loci and identified several that preferentially upregulate the γ-globin gene in comparison to the β-globin gene.

Introduction

Sickle cell disease is the most common monogenic disease worldwide and is caused by a single point mutation in the β-globin gene. Painful clinical symptoms begin shortly after birth as mutated β-globin subunits replace non-defective γ-globin chains in the predominant form of hemoglobin. Current pharmacological treatment with hydroxyurea partially reverses this globin switching by increasing the production of γ-globin (Lavelle, 2004; Steinberg and Rodgers, 2001). This has led to broad interest in developing other compounds and discovering new mechanisms that preferentially upregulate γ-globin (Bauer and Orkin, 2010; Liu et al., 2010; Steinberg and Rodgers, 2001; Xu et al., 2011), and also in developing methods to study globin regulation (Chan et al., 2012; Howden et al., 2008). Analyses of differential expression of β- and γ-globin generally have been limited to hemoglobin electrophoresis or qRT-PCR, but recent reports have described a method of using the expression of fluorescent molecules driven by the β- and γ-globin promoters as a readout of differential globin regulation. In those studies, the authors integrated into the genome a bacterial artificial chromosome containing the entire 200kb β-globin locus (which includes both β-

globin and γ-globin among other genes), modified such that the β- and γ-globin promoters drive expression of fluorescent proteins (Chan et al., 2012; Howden et al., 2008). The integration of the complete genomic locus presumably maintains much of the physiologically relevant regulation of expression, but it does not allow for the direct analysis of the endogenous locus. Instead, direct modification of the endogenous β- and γ-globin loci is required.

Endogenous genomic loci can be precisely altered using engineered zinc finger nucleases (Hockemeyer et al., 2009; Perez et al., 2008; Urnov et al., 2005; Yusa et al., 2011) and Tal effector nucleases (Hockemeyer et al., 2011; Miller et al., 2010; Reyon et al., 2012). ZFNs and TALENs are comprised of a specifically engineered DNA binding domain, which is fused to the FokI endonuclease domain. Anti-parallel binding of a pair of ZFNs or TALENs to contiguous sites leads to the dimerization of the FokI domain, resulting in a targeted DNA double-strand break. Repair of the break can proceed by mutagenic non-homologous end joining or by high-fidelity homologous recombination with a homologous DNA donor template. Compared to TALENs, ZFNs have shorter DNA recognition domains and induce more off-target breaks, leading to higher levels of cytotoxicity (Mussolino et al., 2011). The DNA recognition domain of TALENs is characterized by repeated arrays of 34 identical amino acids, except in positions 12 and 13. These two amino acids comprise the repeat variable domain (RVD), which contacts the DNA and provides the nucleotide recognition specificity of

each repeat array (Boch et al., 2009; Moscou and Bogdanove, 2009). Unlike the other DNA bases which each show strong preference for a single RVD, guanine can be recognized by one of two RVDs with different binding characteristics. The asparagine-asparagine (NN) RVD can form a high-affinity hydrogen bond with guanine, but is not specific because it can also hydrogen bond with adenine (Deng et al., 2012; Mak et al., 2012). Conversely, the asparagine-lysine (NK) RVD seems to be more specific for guanine (Miller et al., 2010) but is less commonly found in naturally occurring TAL effector proteins (Moscou and Bogdanove, 2009).

Recent reports have described the development and use of β-globin ZFNs to correct the sickle mutation in human iPS cells. The low rates of confirmed targeting described in these studies (1 out of 300 (Zou et al., 2011a) and 28 out of 286 (Sebastiano et al., 2011) drug resistant clones were targeted) could be increased by improving the efficiency and toxicity profile of the engineered nucleases. Here, we used highly active and minimally toxic β-globin TALENs to stimulate homologous recombination of therapeutic β-globin cDNA to the endogenous β-globin locus in 19% of cells prior to selection. To analyze the efficiency of both the cutting by the TALENs and the rate of targeted integration, we employed a rapid, accurate and economical deep sequencing method known as single molecule, real time (SMRT) sequencing (Eid et al., 2009). Then, we used TALENs to target GFP to the endogenous β-globin locus and tdTomato to the

endogenous γ-globin locus to generate cell lines that report on the activity of each of these genes. We used these fluorescent reporter cell lines to rapidly and robustly screen a panel of small molecule compounds for differential regulation of the of the endogenous β- and γ-globin genes.

Results

Design and characterization of β- and γ-globin TALENs

To develop a system that robustly and rapidly reports on the activity of both the β-globin and γ-globin loci, we designed a gene targeting strategy using engineered nucleases. Recent reports have described low but significant levels of genome modification at the endogenous β-globin locus using ZFNs (Sebastiano et al., 2011; Zou et al., 2011a), and we first sought to improve the rate of gene targeting at the β-globin locus by designing custom TALENs to that site. First, we identified four putative left (βL1-βL4) and four right (βR1-βR4) TALEN binding sites near the sickle mutation in β-globin (Fig. 3.1a), and synthesized the eight individual TALENs using the NK RVD to bind each guanine. Notably, we made slight modifications of the final TALEN expression vector to include the N- and C-terminal TALEN truncations that have been shown to be sufficient for optimal TALEN activity (Miller et al., 2010). In an extrachromosomal single strand annealing (SSA) assay, we identified six TALEN pairs that stimulated SSA at

least 10-fold above background (Figure 3.1b). We then re-constructed the most active TALEN pair (βL4-NK/βR4-NK) to contain the NN RVD (βL4-NN/βR4-NN) (Figure 3.2a) using the Golden Gate cloning strategy previously described (Cermak et al., 2011). To investigate their activities at the endogenous

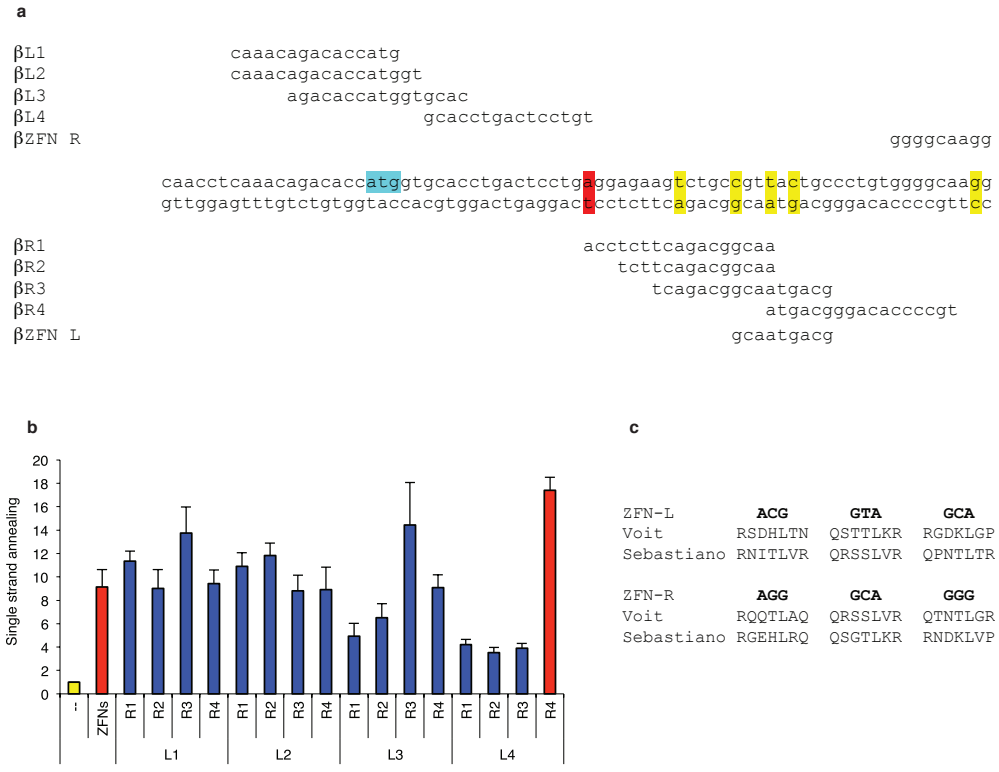


Figure 3.1: Design and activity of β-globin TALENs and ZFNs. (a) DNA binding sites of the β-globin TALENs and ZFNs. Blue – ATG start codon, red – sickle mutation, yellow - different between β-globin and δ-globin. **(b)** Fold increase in single strand annealing compared to control (yellow). Red bars – nucleases used in subsequent experiments. **(c)** Comparison of the DNA-binding a-helices of the ZFNs used in this study compared to previously published ZFNs (Perez et al., 2008).

chromosomal β-globin locus, we used the Surveyor nuclease assay in HEK293T cells. The NK versions modified up to 18% of alleles and the NN TALENs modified 48% of alleles (Figure 3.2b). As a comparison, we also used a modification of the OPEN method to generate ZFNs to the β-globin locus (Maeder et al., 2008; Pruett-Miller et al., 2008). These ZFNs were made

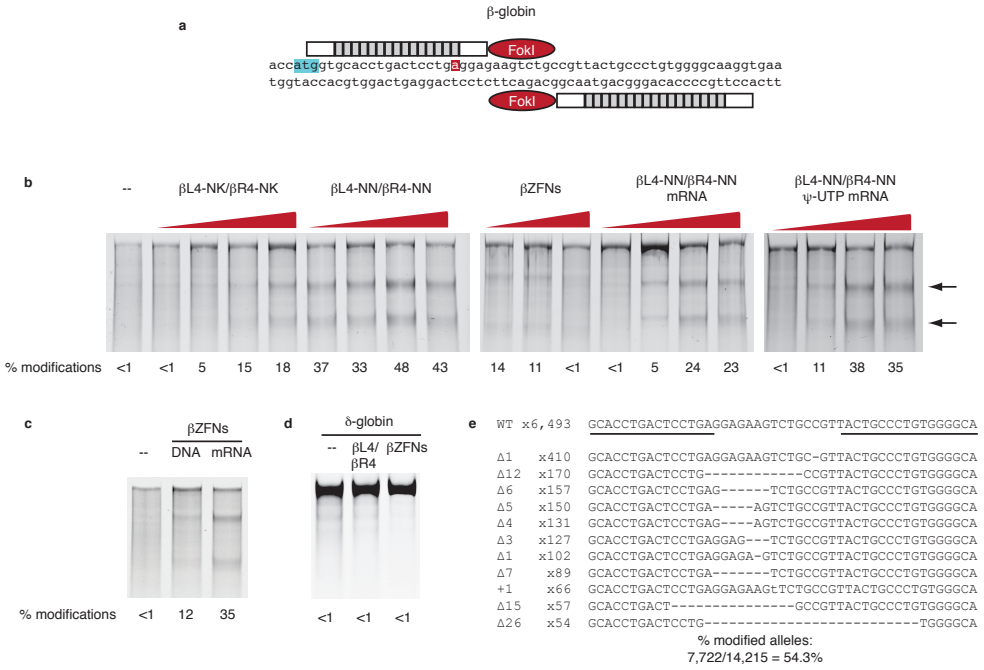


Figure 3.2. TALEN and ZFN activity at the endogenous β-globin locus. **(a)** Schematic of βL4/βR4 TALEN. **(b)** Titration of each β-globin nuclease pair in the Surveyor nuclease assay. **(c)** Comparison of the activity of β-globin ZFNs delivered as DNA or pseudo-UTP modified mRNA. **(d)** Activity of the β-globin nucleases at the δ-globin locus. **(e)** SMRT sequencing of β-globin alleles mutated by treatment with βL4-NN/βR4-NN TALENs. The 11 most abundant mutated alleles are shown, and the frequency of each is indicated. TALEN binding sites are underlined. (Δ represents deletions, + represents insertions).

independently from the ones reported by Sebastiano *et al* (Sebastiano et al., 2011) but are designed to the same target sequence and are very similar in the amino acid sequence of the alpha-helices that mediate DNA binding (Figure 3.1c). Although the ZFNs were much more cytotoxic than were the TALENs (Figure 3.3), the ZFNs were also very active, modifying up to 35% of β-globin alleles in the Survyeor nuclease assay (Figure 3.2b). Interestingly, delivery of TALENs as mRNA did not increase the already high frequency of cutting, but delivery of the ZFNs as mRNA increased the signal from 12% to 35% (Figure 3.2c). Importantly, neither the TALENs nor the ZFNs showed any modification at the δ-globin locus (Figure 3.2d), which has high sequence homology with β-globin (Figure 3.1a).

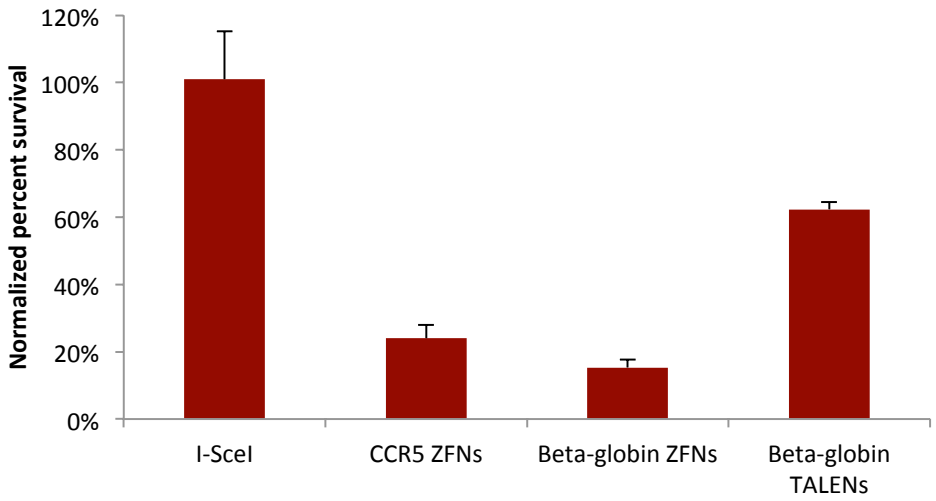


Figure 3.3: Toxicity assay of β-globin nucleases. The ratio of GFP positive cells on day 6 to day 2 for each sample was normalized to samples transfected with the non-toxic nuclease I-SceI. CCR5 ZFNs (Perez et al., 2008) were used as an example of toxic nucleases.

To confirm the frequency of genome modification by βL4-NN/βR4-NN, we used single molecule real time (SMRT) sequencing, a rapid, high-throughput method for sequencing of the β-globin locus following TALEN treatment(Eid et al., 2009). SMRT sequencing allows for simultaneous analysis of up to 30,000 sequences, as well as multiplexing various samples at once. Analysis of 14,215 β-globin sequences revealed TALEN modification of 54% (Figure 3.2e).

Next, to modify the endogenous γ-globin locus, we designed and constructed three left (γL1- γL3) and two right (γR2- γR3) NN TALENs that bind sequences near the ATG start codon of γ-globin (Figure 3.4). Because of the sequence identity between Aγ-globin and Gγ-globin these TALEN pairs do not distinguish the two loci. To measure the activity of the γ-globin TALENs, we again used the Surveyor nuclease assay, which resulted in modification of up to 44% of γ-globin alleles with the γL3/ γR2 pair (Figure 3.5a and 3.5b). Two other TALEN pairs modified greater than 30% of γ-globin alleles (Figure 3.5b).

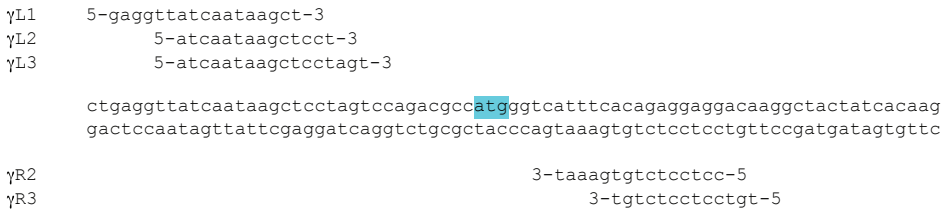


Figure 3.4: DNA binding sites of the γ-globin TALENs. The ATG start site is highlighted in blue.

Analysis of 14,790 γ-globin SMRT sequences revealed a modification rate of 53% with γL3/γR2 (Figure 3.5c).

TALEN-mediated β-globin targeting by homologous recombination

We then sought to determine at what frequency these highly active TALENs stimulated gene targeting by homologous recombination (Figure 3.6a). First to target the β-globin locus, we designed a targeting vector with approximately 1 kilobase arms of homology 5' and 3' of the TALEN cut site. In between the homology arms, we included a Ubc-GFP expression cassette that, upon successful homologous recombination, would be stably integrated into the

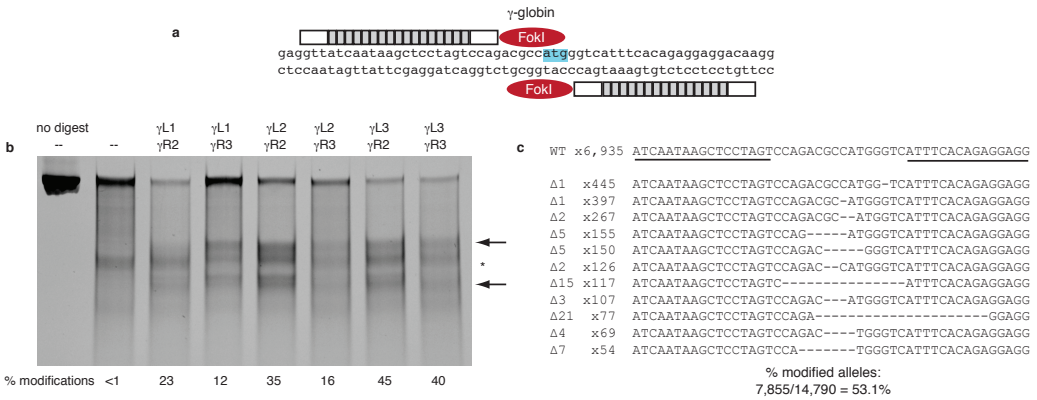


Figure 3.5: TALEN activity at the endogenous γ-globin locus. (a) Schematic of the γL3/γR2 TALEN binding site in the human γ-globin gene. Blue - ATG start codon. **(b)** γ-globin gene disruption in K562 cells. Arrows – specific cleavage products, * non-specific cleavage product. **(c)** SMRT sequencing of γ-globin alleles mutated by treatment with γL3/γR2 TALENs. The 11 most abundant mutated alleles are shown, and the frequency of each is indicated. TALEN binding sites are underlined.

β-globin locus (Figure 3.6b "β-Ubc-GFP" targeting vector). Gene targeting by nucleofection of β-Ubc-GFP with βL4-NN and βR4-NN TALEN expression plasmids into erthyroleukemic K562 cells, resulted in stable integration of Ubc-GFP in 19% of transfected cells (13% overall) compared to less than 1% in the absence of TALENs (Figure 3.6c). We then compared the activities of the NK and NN β-globin TALENs in the gene targeting assay. In confirmation of the

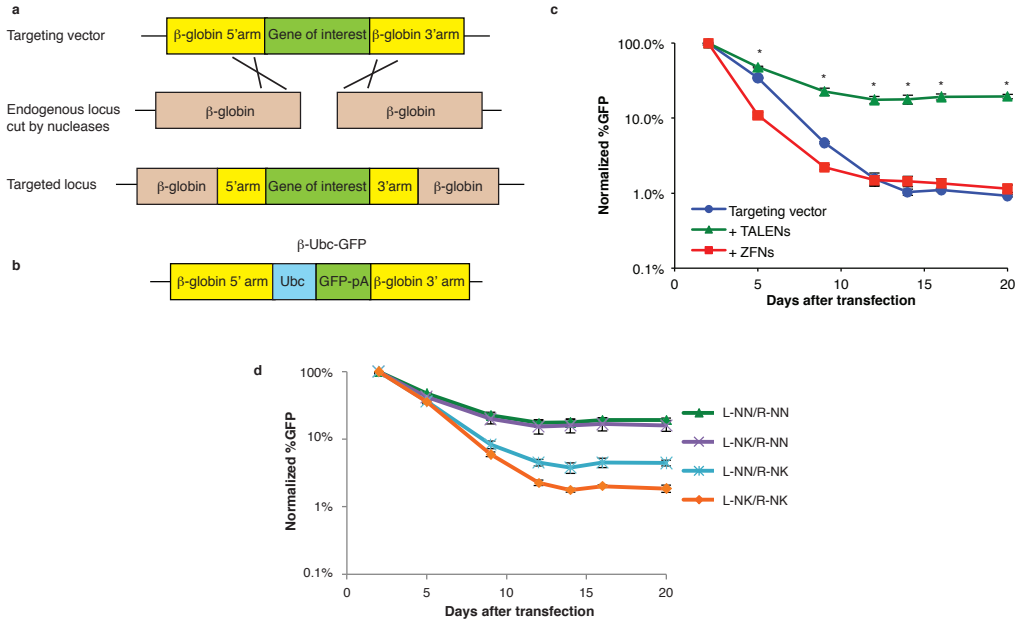


Figure 3.6 High-frequency gene targeting using β-globin TALENs. **(a)** Schematic of nuclease-mediated gene targeting to the endogenous β-globin locus. **(b)** β-Ubc-GFP targeting vector. Ubc- ubiquitin C promoter, pA- BGH polyadenylation signal sequence. **(c)** Gene targeting of β-Ubc-GFP to the endogenous β-globin locus in K562 cells using βL4/βR4 TALENs and ZFNs (* $p < .005$ compared to targeting vector alone). **(d)** Comparison of NN and NK versions of βL4/βR4 TALENs in the gene targeting assay.

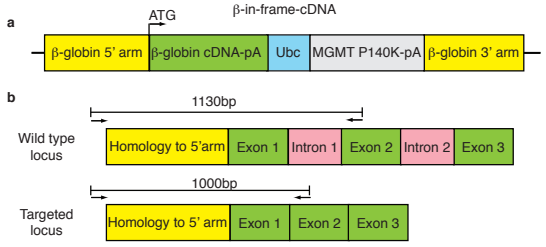
Surveyor assay data, the NN versions stimulated a significantly higher rate of targeted integration compared to the NK TALENs. Interestingly, when paired with βR4-NN, both βL4-NK and βL4-NN stimulate high rates of targeting (~20%). However, when paired with βR4-NK, βL4-NK resulted in 1.8% stable GFP expression, while βL4-NN led to 4.5% stable GFP expression (Figure 3.6d). Despite high rates of modification in the Surveyor assay (Figure 3.2b), the ZFNs did not stimulate targeting of the β-Ubc-GFP at levels that could be discriminated from background random integrants (Figure 3.6c). In this direct comparison of ZFNs and TALENs designed to target nearly the same sequence (Figure 3.1), we found that the TALENs were significantly better because of their greater cutting activity, significantly greater stimulation of targeting, and their lower toxicity. These data also demonstrate better activity with TALENs using NN as the RVD to recognize guanine compared to NK but that NK TALENs can have excellent activity in the correct context.

Targeting β-globin cDNA to the endogenous β-globin locus

We next sought to target full-length β-globin cDNA to the endogenous β-globin ATG start site. In this way, endogenous β-globin regulatory elements would express β-globin from the cDNA instead of from the wild type genomic sequence, a strategy that would be clinically relevant for both sickle cell disease

and β-thalassemia. We modified the β-Ubc-GFP targeting vector, replacing the Ubc-GFP cassette with β-globin cDNA fused in-frame to the natural β-globin ATG start codon, already present in the 5' arm of homology (Figure 3.7a, "β-in-frame-cDNA" targeting vector). Also included in the β-in-frame-cDNA targeting vector was a drug selection cassette encoding a mutant form of methylguanine methyltransferase (MGMT P140K), which allowed for enrichment of targeted cells by treatment with the combination of O6-benzylguanine (O6BG) and carmustine, (BCNU).

To determine the frequency of targeting and the efficiency of drug selection, we again employed SMRT sequencing. First, we targeted K562s with the β-in-frame-cDNA targeting vector using βL4/βR4 TALENs. Then we pulsed the samples three times with O6BG and BCNU and harvested gDNA after each



c

Drug pulses	-TALENs		+TALENs	
	# Sequences	% Targeting	# Sequences	% Targeting
0	941	0%	1100	8%
1	1101	0%	1319	69%
2	1119	0%	1100	61%
3	1100	0%	1493	61%

Figure 3.7: Targeting therapeutic β-globin cDNA to the endogenous β-globin locus. (a.) β–in-frame-cDNA targeting vector. When targeted, the cDNA is expressed from the endogenous ATG start site. pA- BGH polyadenylation signal sequence, Ubc- Ubiquitin C promoter, MGMT – Methylguanine methyltransferase **(b.)** Schematic of the wt (top) and targeted (bottom) β-globin locus. The absence of intron 1 in the cDNA-targeted locus results in a shorter PCR product. **(c.)** Results from SMRT sequencing of targeted β-globin alleles without selection (top row) and with up to three rounds of drug selection.

pulse. To amplify the β-globin locus, we used a forward primer that is 5' and outside the start of the 5' homology arm and a reverse primer in exon 2 of β-globin (Figure 3.7b). In this way, random integrants were not amplified. The presence of intron 1 in the wild type genomic DNA sequence of this locus, and its absence in the targeted β-globin cDNA, allowed us to determine the ratio of targeted alleles to wild type alleles after each pulse based on the length of the sequence, which could then be confirmed by the sequence content (Figure 3.7b). In the absence of drug selection, 8% of the alleles were targeted as determined by analyzing the sequence of 1,100 alleles. The targeting frequency of 8% of alleles is consistent with the observed rate of β-Ubc-GFP targeting in 19% of cells (Figure 3.6c). Pulsing the targeted cells with O6BG/BCNU up to three times resulted in the enrichment of targeted alleles such that they accounted for greater than 60% of all sequenced alleles (Figure 3.7c).

Generation of fluorescent β- and γ-globin reporters by endogenous locus tagging

Next, we redesigned the β-Ubc-GFP targeting vector such that a promoterless GFP was fused in frame to the β-globin ATG start codon (Figure 3.8a, "β-in-frame-GFP" targeting vector). In this way, upon targeting to the endogenous β-globin locus, GFP would be driven by the endogenous β-globin

promoter and would be subject to the regulatory elements controlling β-globin expression. We targeted the β-in-frame-GFP targeting vector to the β-globin locus, using either βL4/βR4 TALENs or ZFNs. Because of the naturally low level of β-globin expression in K562s (Fordis et al., 1984; Zein et al., 2010), the overall percentage of GFP positive cells underestimated the targeting frequency compared to that seen with the β-Ubc-GFP targeting vector. Nonetheless, in the presence of βL4 and βR4 TALENs, there was a significantly higher percentage of

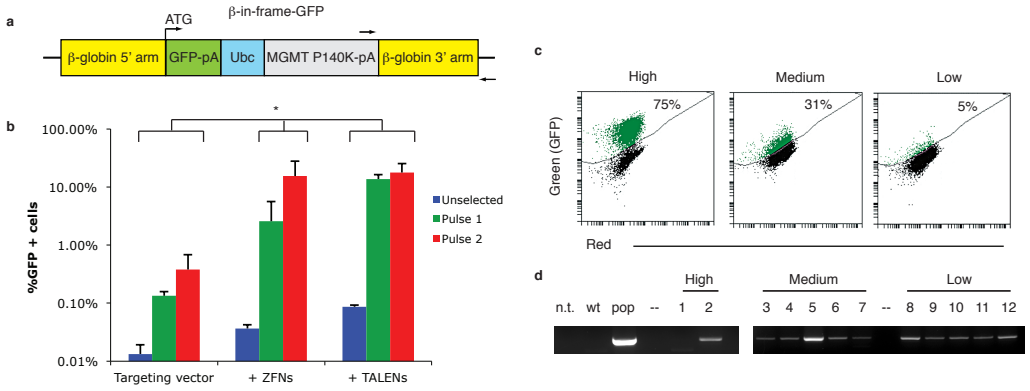


Figure 3.8: Generation of β-globin reporter cells by in-frame GFP targeting to the endogenous β-globin locus. (a) β–in-frame-GFP targeting vector. When targeted, GFP is expressed from the endogenous ATG start site. Arrows indicate PCR primers used in **(d)** to confirm targeting. pA- BGH polyadenylation signal sequence, Ubc- Ubiquitin C promoter, MGMT – Methylguanine methyltransferase. **(b)** Targeting of β-in-frame-GFP to the endogenous β-globin locus using βL4/βR4 TALENs and ZFNs (* $p < .05$ compared to targeting vector alone). **(c)** FACS plots of targeted clones, classified by level of GFP expression. **(d)** Genomic PCR using primers in **(a)** to detect presence of a targeted β-globin locus in 11 of 12 GFP positive clones. n.t.- no template control, wt- genomic DNA from wild type cells, pop- genomic PCR from targeted population, -- no sample.

GFP positive cells than in control samples (Figure 3.8b, blue bars). Selection with two pulses of O6BG and BCNU resulted in significant enrichment of GFP positive cells in the TALEN and ZFN samples compared to the targeting vector alone (Figure 3.8b). Notably, with up to four pulses with O6BG and BCNU, the overall percentage of GFP positive cells never increased above 20% (data not shown). We believe this is due to the low activity of the β-globin promoter in K562s. When we sorted for GFP positive cells from the TALEN sample, over the course of two weeks in culture, the population went from being greater than 95% GFP positive to approximately 15% (data not shown), reflecting the dynamic nature of β-globin expression in K562s. When we analyzed individual clones from the drug selected TALEN sample, we observed three distinct patterns of GFP expression that we designated "high," "medium" and "low" (Figure 3.8c).

To determine whether these clonal populations expressed GFP because of targeting to the β-globin locus, we used a genomic PCR assay spanning the junction of integration (Figure 3.8a, arrows). Indeed, 11 of 12 clones showed targeted integration (Figure 3.8d). Interestingly, the one clone that was not targeted (clone #1) was a "high" GFP expressing clone that had undergone random integration into a genomic site near strong promoter elements. These data confirm previous results that targeted cells show low levels of GFP expression because of the low activity of the β-globin promoter in these cells.

To develop a fluorescence-based reporter of the endogenous γ-globin locus, we targeted tdTomato in-frame to the ATG start codon of γ-globin, using a homologous targeting vector containing in-frame tdTomato followed by a neomycin drug resistance cassette (Figure 3.9a, "γ-in-frame-tdTomato"). Unlike β-globin, γ-globin is highly expressed in K562 cells so the fluorescent readout from the targeted γ-in-frame-tdTomato accurately reflected the overall integration rate despite the lack of an exogenous promoter. Co-transfection of γ-in-frame-tdTomato with γL3/γR2 TALENs resulted in stable tdTomato expression in 34% of transfected cells (23% overall), compared to less than 1% in samples without TALENs (Figure 3.9b and 3.9c). Genomic PCR spanning the integration junction (Figure 3.9a, arrows) revealed the presence of a targeted band in samples treated with any of the three most active pairs of γ-globin TALENs (Figure 3.9e, left).

To create a dual-fluorescent reporter that expresses GFP from the endogenous β-globin locus and tdTomato from the endogenous γ-globin locus (Figure 3.9d), we used the γ-globin TALENs to target the γ-in-frame-tdTomato vector to the γ-globin locus in a previously targeted β-globin-GFP clone (Figure 3.9e, right). In this way, we generated three cell lines that report on the activity of endogenous globin promoters, the β-globin-GFP reporter, γ-globin-tdTomato reporter and the β-globin-GFP/ γ-globin-tdTomato dual reporter (Figure 3.9f).

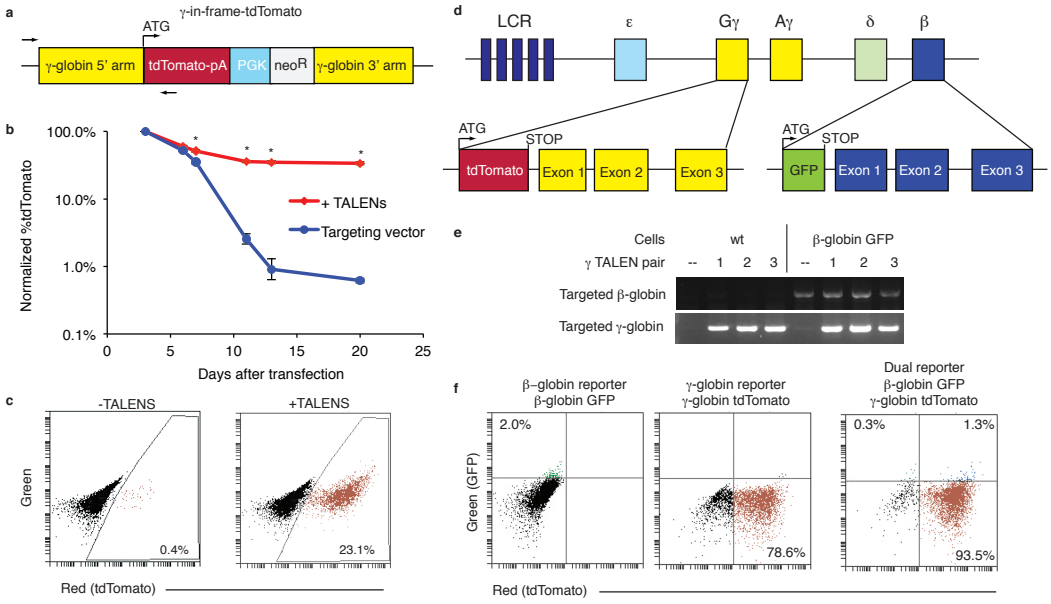


Figure 3.9: Generation of fluorescent γ-globin and dual globin reporter cell lines. (a) γ–in-frame-tdTomato targeting vector. When targeted, tdTomato is expressed from the endogenous ATG start site. Arrows indicate PCR primers used in **(e)** to confirm targeting. pA- BGH polyadenylation signal sequence, PGK- Phosphoglycerate kinase promoter, neo^R – Neomycin phosphotransferase. **(b)** Targeting of γ-in-frame-tdTomato to the endogenous γ-globin locus using γL3/γR2 TALENs (* $p < .005$ compared to targeting vector alone). **(c)** FACS plots showing stable integration of tdTomato on day 20 in the presence of TALENs. **(d)** Schematic of the targeted globin loci showing tdTomato being expressed from the endogenous g-globin ATG start site and GFP from the endogenous b-globin ATG start site. The wild type γ- and β-globin gene sequences are still present after targeting but are not expressed because of the stop codons that follow the targeted tdTomato and GFP sequences. LCR – locus control region. **(e)** Genomic PCR using primers in **(a)** to detect presence of a targeted γ-globin locus in samples treated with γ-globin TALENs (Pair 1: γL3/γR2, pair 2: γL3/γR3, pair 3: γL2/γR2). Wild type cells (left) were targeted to generate the γ-globin tdTomato reporter line, and β-globin-GFP cells (right) were targeted to generate the dual reporter cell line. **(f)** FACS plots showing the fluorescent profile of the β-globin-GFP reporter (left) the γ-globin-tdTomato reporter (center), and β-globin-GFP/γ-globin-tdTomato dual reporter (right).

Using endogenous fluorescent reporter cells to screen globin-modulating compounds

Next, we sought to establish these fluorescent reporter lines as tools that can be used to compare the globin-modulating activities of small molecule compounds. Hydroxyurea, is used clinically to increase the production of γ-globin and it has been shown to upregulate γ-globin in K562s (Banan et al., 2012; Erard et al., 1981). Indeed, K562s treated for four days with 400 μM hydroxyurea

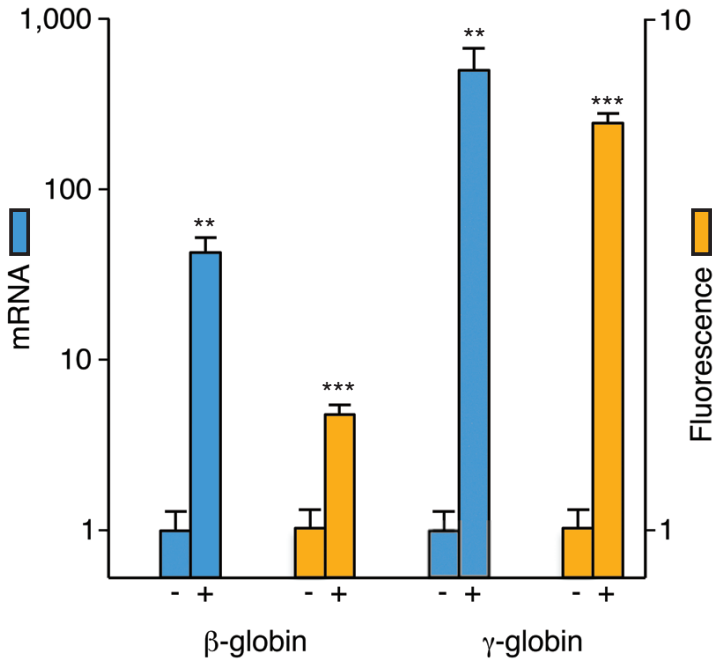


Figure 3.10: Validation of fluorescent globin reporter cells. Effect of 400μM hydroxyurea on β- and γ-globin transcript levels (blue bars) and on GFP and tdTomato expression in targeted cell lines (orange bars).

showed a significant 62-fold increase in β-globin expression. γ-globin mRNA levels were even more elevated than β-globin transcripts after treatment with hydroxyurea, increasing 932 fold (Figure 3.10). We treated the β-globin-GFP reporter cells and the γ-globin-tdTomato reporter cells with hydroxyurea and measured mean fluorescence intensity on day four. GFP and tdTomato intensities were significantly higher compared to untreated cells, and the increase in tdTomato was significantly greater than the increase in GFP, mirroring the changes in β- and γ-globin expression levels. These results show that the reporter cell lines can be used to rapidly, accurately and robustly measure the activity of the endogenous globin loci.

To expand our analysis, we treated cells from the β-globin GFP, γ-globin-tdTomato, and β-globin-GFP/ γ-globin-tdTomato cell lines with 5 concentrations of 17 different compounds shown previously to modulate globin expression (Figure 3.11). Of these, ten significantly increased the expression of the endogenous γ-globin locus, the most striking of which were guanine, guanosine, apicidin and hydroxyurea (Figure 3.12a). Similarly ten compounds increased the expression of endogenous β-globin, with the best inducers being guanosine, guanine and GMP (Figure 3.12b). The ideal pharmacological therapy for sickle cell disease is a drug that preferentially induces the production of γ-globin compared to β-globin. Therefore the most relevant analysis was of the ratio of

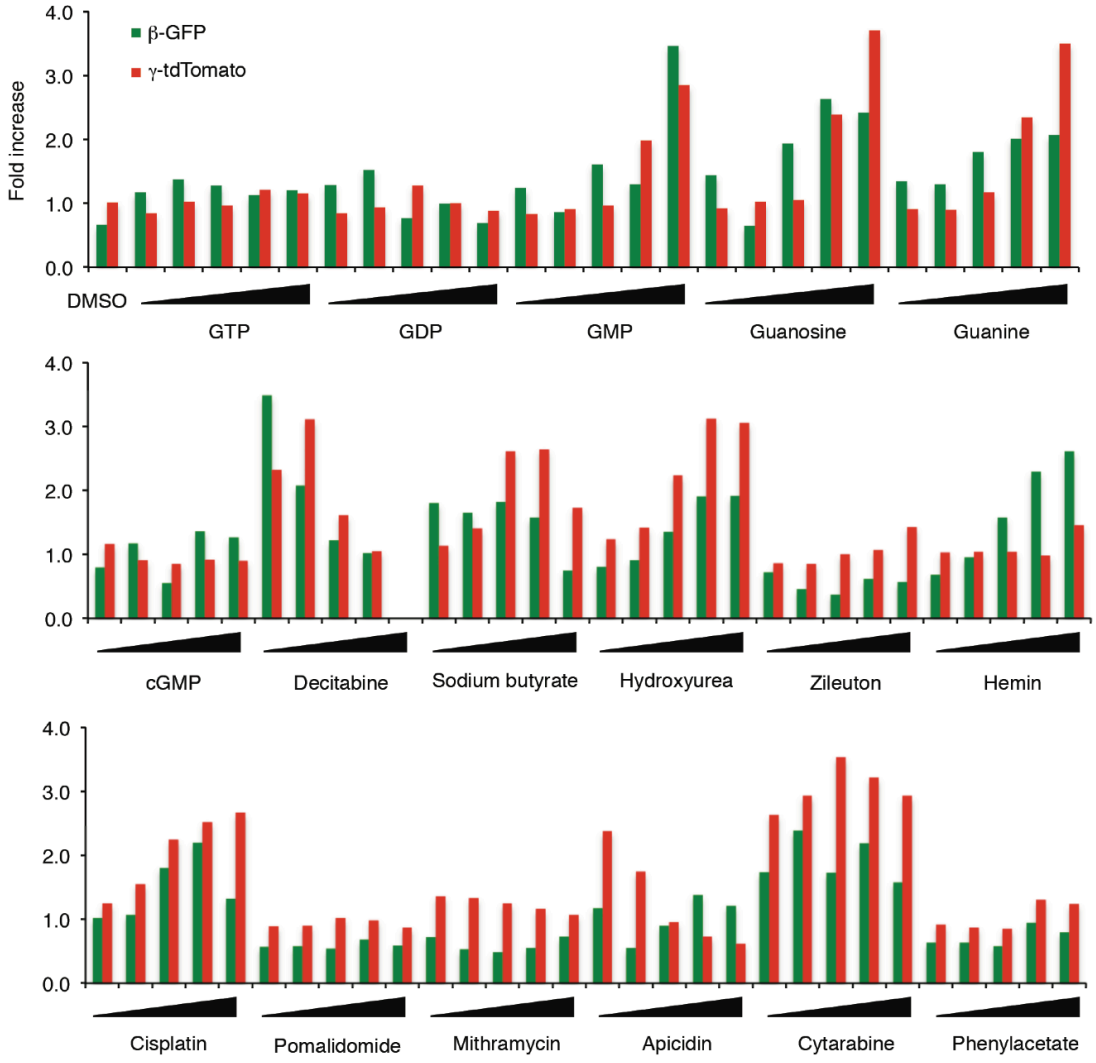


Figure 3.11: Titration of globin-modulating compounds. Increase in GFP (green bars) and tdTomato (red bars) expression from the dual reporter cell line after treatment for four days with the indicated compound. The five drug concentrations used were 1x, 2x, 5x, 10x, and 20x and the 1x concentration for each is as follows: GTP, GDP, GMP, guanosine, guanine, cGMP, hydroxyurea, zileuton- 20μM; decitabine, cytarabine- 1μM; sodium butyrate- 60μM; hemin- 5μM; cisplatin- 400nM; pomalidomide- 2μM; mithramycin- 10μM; apicidin- 200nM; phenylacetate- 400μM.

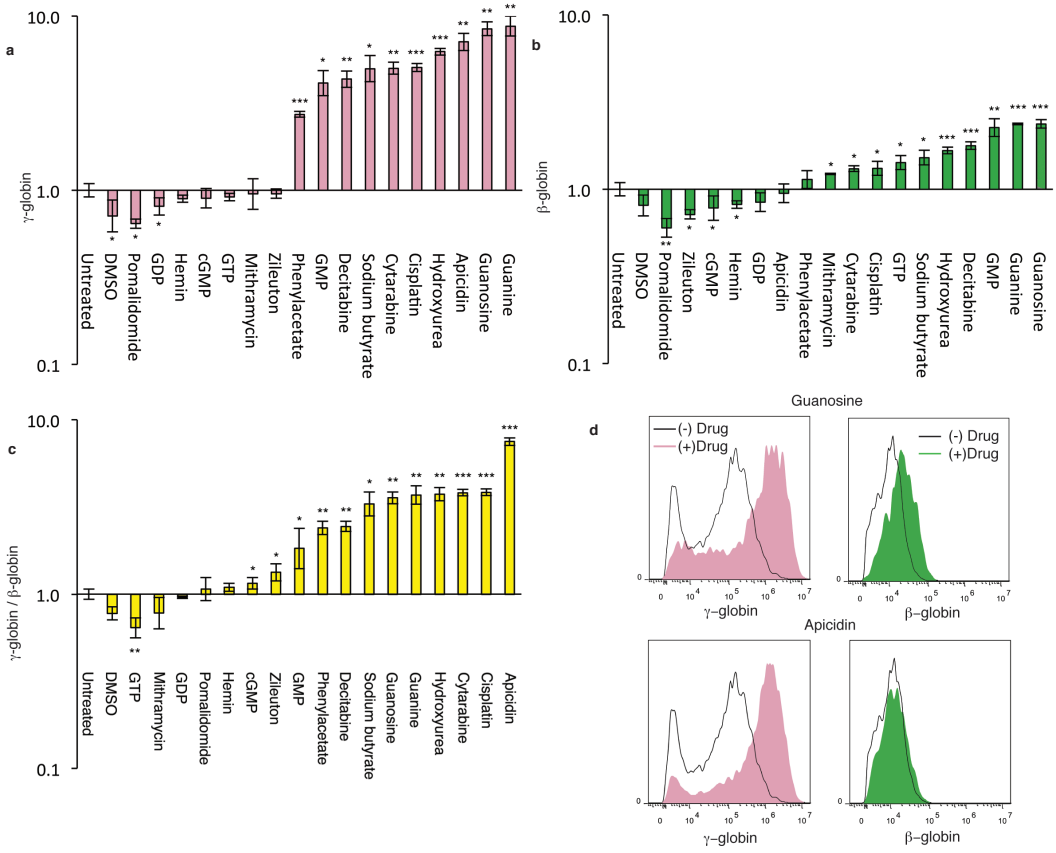


Figure 3.12: Using fluorescent reporter cell lines to screen globin-modulating compounds. Effect of drug treatment on **(a)** tdTomato expression in g-globin tdTomato cells, **(b)** GFP expression in β-globin-GFP cells, and **(c)** the ratio of tdTomato/GFP expression in γ-globin-tdTomato and β-globin-GFP cells. FACS plots showing the effect on the expression of tdTomato and GFP from the γ-globin and β-globin loci of **(d)** guanosine, top; and apicidin, bottom. Drug concentrations: DMSO- 0.2%, pomalidomide- 10μM, GDP- 400μM, hemin- 5μM, cGMP- 400μM, GTP- 200μM, mithramycin 200nM, zileuton- 200μM, phenylacetate- 4mM, GMP- 400μM, decitabine- 10μM, sodium butyrate- 1200μM, cytarabine- 5μM, cisplatin- 4μM, hydroxyurea- 400μM, apicidin- 400nM, guanosine- 400μM, guanine- 400μM.

induction of γ- to β- globin (Figure 3.12c). Compounds such as guanosine increased the expression of both γ- and β-globin (Figure 3.12d, top). However, apicidin was a strong inducer of γ-globin but had no activity at the β-globin promoter (Figure 3.12d, bottom). Importantly hydroxyurea, the clinical standard of care for induction of γ-globin had one of the highest γ/β induction ratios of all the screened compounds. In this way, we have established a system to robustly, rapidly and simultaneously report on the activity of the endogenous β- and γ-globin promoters.

Discussion

The emergence of the TALEN platform for engineering nucleases has made possible the rapid, open-source generation of highly active genome editing proteins. TALENs have been used to cause site-specific gene disruption and gene targeting in yeast (Christian et al., 2010), plants (Cermak et al., 2011), nematodes (Wood et al., 2011), zebrafish (Huang et al., 2011; Sander et al., 2011), rats (Tesson et al., 2011) and human cells (Hockemeyer et al., 2011; Miller et al., 2010; Reyon et al., 2012). A recent report described TALENs designed to human β-globin and showed gene correction of a mutated GFP gene, which had been disrupted by the insertion of the β-globin sequence recognized by the TALENs. The authors report 5% gene correction of GFP by homologous recombination but

do not report on the activity of the TALENs on the endogenous β-globin locus (Sun et al., 2012).

Here, we synthesized and compared the activities of NN-TALENs, NK-TALENs and ZFNs designed to the same genomic region. Using the most active TALENs, we targeted β-globin cDNA to the ATG start codon of the endogenous β-globin locus in human cells and introduced a deep sequencing method to precisely detect rates of targeting. Similarly, we used TALENs to generate β- and γ-globin reporter cell lines by targeting the start codons of the endogenous loci with GFP and tdTomato. Finally, we showed that our endogenously tagged reporter cells provide a rapid and facile method to analyze the globin modulating activities of small molecule compounds.

Our strategy of using SMRT sequencing to validate the activity of engineered nucleases as determined by the Surveyor nuclease assay allows for the analysis of many more sequences as compared to standard Sanger sequencing methods at a fraction of the cost of other deep sequencing platforms such as Illumina. We believe that using deep sequencing to determine cutting and targeting frequencies will be especially beneficial in primary cells such as CD34+ hematopoietic stem cells in which these rates are considerably lower compared to cell lines.

Using TALENs to target full-length β-globin cDNA to the endogenous β-globin locus provides an alternate method to gene conversion of the sickle mutation using ZFNs as recently described (Sebastiano et al., 2011; Zou et al., 2011a). First, we showed considerably higher nuclease activity, using a TALEN platform that is less toxic than ZFNs (Figure 3.3 and (Mussolino et al., 2011)). In the β-in-frame-GFP targeting experiments, which have low background signal because of the lack of exogenous promoter, we achieved targeted integration with the ZFNs after drug selection, showing that the ZFNs are capable of stimulating gene targeting at the β-globin locus. However, we were unable to detect targeting of the β-Ubc-GFP cassette with ZFNs at levels above background random integration, presumably due to the toxicity of the ZFNs.

Another improvement in our strategy is that cDNA targeting would be therapeutic in both sickle cell disease, in which the causative mutation is at codon 6 of the β-globin gene, and β-thalassemia, in which causative mutations can occur throughout the length of the β-globin gene. The co-conversion of the sickle mutation with the downstream integration of a drug resistance cassette in the first intron as described (Sebastiano et al., 2011; Zou et al., 2011a) has been demonstrated to be less efficient in cases when there is homologous sequence in between the site of the conversion and the insertion of the selectable marker (Porteus, 2006) such as the first exon of β-globin. Therefore, when we designed

the β-in-frame-cDNA targeting vector we introduced silent mutations in every sixth nucleotide of the cDNA sequence between the nuclease cut site and the end of the first exon. By reducing the homology between the genomic locus and the cDNA, we shunted the repair to proceed via homologous recombination with the 3' arm of homology (instead of with the short stretch of homology in exon 1 of the cDNA), ensuring that the drug selection cassette is also targeted to the locus. We chose to use the MGMT P140K-based drug selection strategy because it is effective *in vitro* (Davis et al., 1999) and relies on the FDA-approved compounds O6BG and BCNU, which can enrich for targeted cells *in vivo* (Neff et al., 2003).

Dozens of reports have analyzed the effect of drugs on globin expression, primarily by analyzing transcript levels by qRT-PCR, hemoglobin electrophoresis or benzidine staining (Bianchi et al., 2001; Chan et al., 2012; Cortesi et al., 1999; Fibach et al., 1995; Haynes et al., 2004; Howden et al., 2008; Keefer et al., 2006; Liu et al., 2010; Meiler et al., 2011; Mischiati et al., 2004; Moutouh-de Parseval et al., 2008; Osti et al., 1997; Witt et al., 2003) .We established fluorescent cell lines as accurate reporters of differential globin expression by comparing the induction of β- and γ-globin mRNA transcripts with the increase in GFP and tdTomato signal following treatment with hydroxyurea. Then we used the fluorescent globin reporters in a mini drug screen to demonstrate their utility as tools to rapidly and accurately measure modulations in globin expression. This is the first proof-of-principle example of using precise genome engineering to

generate cell lines with endogenous promoter reporters, validating the output by direct comparison to mRNA transcript levels, and then using the dual reporter cell line to screen for small molecules that differentially regulate two genes. In this way, we introduce a novel method to analyze endogenous promoter activity in the context of the most prevalent monogenic disease. These reporter cells could be used in an unbiased high-throughput drug screen to identify novel γ-globin specific inducers to be the next generation of pharmacologic therapy for patients with sickle cell disease. More generally, this strategy could be broadly applied to generate multi-color reporter cell lines to allow rapid screening for conditions and compounds that promote the activity of a particular pathway or determine cellular fate.

Materials and Methods

Cell lines and transfections

K562 cells (ATCC) were maintained in RPMI 1640 (Hyclone) supplemented with 10% bovine growth serum, 100 units/ml penicillin, 100μg/ml streptomycin and 2mM L-glutamine. K562s were transfected by nucleofection (Lonza) using program T-016 and a nucleofection buffer containing 100mM KH_2PO_4, 15mM $NaHCO_3$, 12mM $MgCl_2 \bullet 6H_20$, 8mM ATP, 2mM glucose, pH 7.4. HEK293T cells were maintained in DMEM (Cellgro) supplemented with 10% bovine growth serum, 100 units/ml penicillin, 100μg/ml streptomycin and

2mM L-glutamine. HEK293T cells were transfected either by calcium phosphate or Lipofectamine 2000 (Invitrogen).

Nuclease and targeting vector construction

β-globin NK TALENs were synthesized (Genscript) using the Δ152 N-terminal domain and the +63 C terminal domain previously described(Miller et al., 2010) and fused to the FokI nuclease domain and cloned into pcDNA3.1 (Invitrogen). β-globin NN TALENs and γ-globin NN TALENs were synthesized using a Golden Gate cloning strategy(Cermak et al., 2011) and cloned with the same N- and C- termini and nuclease domain into pcDNA3.1. The β-globin ZFNs were synthesized using the B2H selection strategy previously described(Pruett-Miller et al., 2008). The β-Ubc-GFP targeting vector was synthesized by PCR amplifying arms of homology from genomic DNA isolated from K562 cells using the primers in Figure 3.13 and cloning a Ubc-GFP expression cassette in between the arms. The β-in-frame-cDNA and β-in-frame-GFP targeting vectors were synthesized by overlap PCR to insert β-globin cDNA (OriGene) or GFP directly in frame to the β-globin ATG start codon using the primers in Figure 3.13. Silent mutations were introduced into the β-globin cDNA sequence at every sixth base pair between the nuclease cut site and the end of exon 1. The MGMT P140K drug selection cassette (generous gift from Dr. Stan Gerson) was cloned into the

targeting vector inside the arms of homology. The γ-in-frame-tdTomato targeting vector was generated by genomic PCR of the 5' and 3' arms of homology using primers in Figure 3.13. TdTomato was fused in frame to the γ-globin ATG start codon by overlap PCR. A neomycin phosphotransferase cassette was cloned in between the arms of homology.

In vitro transcription of nucleases

TALEN and ZFN mRNA was synthesized *in vitro* with the MEGAscript T7 kit (Ambion), polyadenylated *in vitro* with the poly(A) tailing kit (Ambion) and purified with the MEGAclear kit (Ambion) following the manufacturer's protocols. Two versions of mRNA were synthesized, using unmodified nucleotides or using pseudouridine-5'-triphosphate (Trilink) in place of UTP (Warren et al., 2010).

Single strand annealing (SSA) and toxicity assays

An SSA reporter was generated by disrupting a GFP gene by duplicating an internal 42 base pair region and separating the duplicated region with a 72 base pair fragment from the β-globin region containing the nuclease recognition sites. The SSA reporter and each nuclease were transfected by calcium phosphate into HEK293T cells and analyzed on an Accuri C6 flow cytometer (Accuri) after two days. The toxicity assay was performed as previously described (Pruett-Miller et

al., 2008). Briefly, HEK293T cells were co-transfected by calcium phosphate with a pair of nucleases and a GFP expression plasmid. The cells were analyzed by FACS for percent GFP positive on day 2 and day 6. The day2/day6 ratio was normalized to a non-toxic nuclease sample.

Surveyor nuclease assay

The Surveyor nuclease assay was performed as previously described(Guschin et al., 2010). Briefly, 6 x 10^5 HEK293T cells were lipofected with 1.5μg of each nuclease or 10^6 K562s were nucleofected with 2.5μg of each nuclease unless otherwise indicated. After 3 days genomic DNA was isolated using the DNeasy kit (Qiagen) and the locus of interest was PCR amplified using the primers in Figure 3.13 using Accuprime polymerase (Invitrogen). 200ng of the PCR product was treated with the Surveyor nuclease (Transgenomic) following the manufacturer's protocol.

SMRT sequencing and cDNA targeting

PCR products prior to cutting by the Surveyor nuclease were prepared for SMRT sequencing following the manufacturer's protocol (Pacific Biosciences). For the SMRT sequencing of the β-globin cDNA targeting events, 10^6 K562s were nucleofected with 10μg b-in-frame-cDNA and 1μg each of βL4 and βR4 TALENs. Aliquots were removed after three days when the first round of

selection was begun by adding 50μM O6BG (Sigma) for one hour and then adding 40μM BCNU (Sigma) for one hour before changing the media. Cells were allowed to recover for 7-10 days at which time another aliquot was harvested and another round of selection started. Genomic DNA was isolated (Qiagen) and the β-globin region was PCR amplified using primers in Figure 3.13, which did not amplify random integrants. Primers with unique three base pair tags were used in the PCR reactions from each time point, such that the samples could be combined and analyzed in one SMRT sequencing reaction. Data were analyzed using CLC Genomics Workbench software.

Generation of fluorescent reporter cell lines

10^6 K562 cells were nucleofected with 10μg of the targeting vector and 1μg of each TALEN. β-globin-GFP cells were enriched by 4 rounds of selection with O6BG and BCNU and clones were established by limiting dilution. γ-globin-tdTomato cells were enriched by treatment with 500μg/ml G418 and clones were established by limiting dilution. Targeting was confirmed by genomic PCR spanning the integration junctions using primers in Figure 3.13.

Quantitative real-time PCR

Wild type K562 cells were treated for 4 days with 400μM hydroxyurea and total mRNA was harvested by Trizol/chloroform extraction and purified on RNeasy columns (Qiagen). 1μg total RNA was used to synthesize cDNA with the iScript cDNA kit (Bio-Rad) following the manufacturer's protocol. Biological triplicates were each assayed in triplicate by qRT-PCR using SYBR green (Applied Biosystems) on a CFX384 real-time thermocycler (Bio-Rad) using the primers in Figure 3.13 using the following conditions: initial denaturation (3 minutes at 95^{o}C), 3-step PCR cycle (10 seconds at 95^{o}C, 30 seconds at 55^{o}C, 5 seconds at 65^{o}C, 40 cycles). PCR efficiency (between 91-119%) was calculated using serial dilutions of template for each primer set. mRNA expression was quantified using the $2^{-\Delta\Delta Ct}$ method (Livak Methods 2001) as compared to the housekeeping gene GAPDH.

Screening globin-modulating compounds

β-globin-GFP cells and γ-globin-tdTomato cells were treated for four days with the indicated concentrations of GTP (Sigma), GDP (Sigma), GMP (Sigma), guanosine (Sigma), guanine (Sigma), cGMP (Sigma), Decitabine (Sigma), Sodium butyrate (Sigma), hydroxyurea (Sigma), zileuton (Sigma), hemin (Sigma), cisplatin (Santa Cruz Biotech), pomalidomide (Sigma), mithramycin (Fisher), apicidin (Sigma), cytarabine (Sigma), or phenylacetate (Sigma).

Fluorescence was measured using an Accuri C6 cytometer (Accuri), and was reported as the fold change in fluorescence intensity after 4 days.

Statistical analysis

Data from at least three samples were used to determine significance by statistical analysis. Mean ± s.d. is reported. Statistical significance was determined by Student's *t*-test and p-values < 0.05 were considered significant.

β-globin targeting vector
5' arm
F: AGGCAGAAACAGTTAGATGTCC
R: TAGCAACCTCAAACAGACACCATG
3' arm
F: GGGGCAAGGTGAACGTGG
R: CTCTTTGCACCATTCTAAAGAAT
Site directed mutagenesis to introduce silent mutations (lower case) in β-globin cDNA
F: GCCCTcTGGGGaAAGGTcAACGTcGATGAgGTTGGcGGTGAaGCCCTcGGCAGGCTGCTGGTGGTCTACCC
R: CCATAGAGCCCACCGCATCC (sequence in BGH polyadenylation signal)
cDNA overlap PCR with ATG in 5' arm
F1: AGGCAGAAACAGTTAGATGTCC
R1: TTgACCTTtCCCCAgAGGGCgGTAACaGCAGAtTTCTCCTCAGGAGTCAGATGCACCAT
F2: GCCCTcTGGGGaAAGGTcAACGTcGATGAgGTTGGcGGTGAaGCCCTcGGCAGGCTGCTGGTGGTCTACCC
R2: CCATAGAGCCCACCGCATCC (sequence in BGH polyadenylation signal)
GFP overlap PCR with ATG in 5' arm
F1: AGGCAGAAACAGTTAGATGTCC
R1: TCCTCGCCCTTGCTCACCATGGTGTCTGTTTGAGGTTGCTA
F2: ATGGTGAGCAAGGGCGAGGA
R2: CCATAGAGCCCACCGCATCC (sequence in BGH polyadenylation signal)

γ-globin targeting vector
5'arm overlap with tdTomato
F1: TGTACACGCACATCTTATGTCT
R1: CCTCGCCCTTGCTCACCATGGCGTCTGGACTAGGAG
F2: ATGGTGAGCAAGGGCGAGG
R2: CCATAGAGCCCACCGCATCC (sequence in BGH polyadenylation signal)
3'arm
F: ATGGGTCATTTCACAGAGGA
R: GACTTTCAAAATCTACTCCAGC

Surveyor nuclease assay
β-globin
F: CCAACTCCTAAGCCAGTGCCAGAAGAG
R: CTCTTGGGTTTCTGATAGGCACTGACT
γ-globin
F: AAACGGTCCCTGGCTAAACT
R: TGAGAAGCGACCTGGACTTT
δ-globin
F: TCGACTGTTGCTTACACTTT
R: TAATCTGAGGGTAGGAAAAC

Unique tags for SMRT sequencing of cDNA targeting
F: CCAGACACTCTTGCAGATTAGTC
R: nnnTAGACCACCAGCAGCCT
nnn represents the following unique tags: AAA, TTT, ACT, TGA, AGC, TCG, ATC, TAG

Genomic PCR to confirm targeting
β-globin
F: CCTGCTGGCCGAAACTGA
R: ATGCAGAGATATTGCTATTGCCTTAAC
γ-globin
F: AGTGTGTGGACTATTAGTCAATAA
R: ATGAACTCTTTGATGACCTCC

qRT-PCR
β-globin
F: AACTGTGTTCACTAGCAACCTCAA
R: GAGTGGACAGATCCCCAAAGGA
γ-globin
F: ACTCGCTTCTGGAACGTCTGA
R: GTATCTGGAGGACAGGGCACT
GAPDH
F: GAAGGCTGGGGCTCATTT
R: CAGGAGGCATTGCTGATGAT

Figure 3.13: List of primers used in this study

CHAPTER IV: CONCLUSIONS AND FUTURE DIRECTIONS

The ability to precisely modify the genome of human cells has the potential to shift the paradigm of how monogenic and other human diseases are treated. Instead of repeated infusions of deficient proteins like Factor VIII to treat hemophilia or ADA to treat one form of SCID, which are very laborious and expensive regimens, gene therapy holds the promise of a robust, long-term cure. Additionally, gene therapy can address the root cause of other monogenic diseases like sickle cell disease and X-linked SCID, in which the deficiency is not a secreted protein, cannot be treated by protein infusion, and can currently only be cured through allogeneic stem cell transplants. Targeted modification of a patient's own stem cells avoids the side-effects and reliance on immunosuppressive drugs associated with these allogeneic transplants and is now under investigation as a way to cure HIV infection. As the specificity of these methods continues to be improved and more clinical data from patient trials becomes available, targeted genome engineering will perhaps one day fulfill the promise of a cure for monogenic diseases.

From bench to bedside: curing patients of AIDS and sickle cell disease

Development of lifelong cell-based therapy for HIV

In this work, I present the contributions I have made to this field, using targeted genome engineering in pre-clinical studies to generate HIV-resistant cells and correct the causative mutation of the most prevalent monogenic disorder, sickle cell disease. To confer HIV resistance, I used CCR5 ZFNs that are currently in clinical trials to target a cassette of anti-HIV genes to the *CCR5* locus. Unlike the CCR5 disruption strategy used in those trials, my novel approach generated cells that were highly resistant to both R5-tropic and X4-tropic HIV. This is a significant advance because it confers protection to both mono-allelically and bi-allelically modified cells against infection by both viral tropisms. Furthermore, by combining the disruption of CCR5 with the addition of up to three anti-HIV genes, I stacked genetic resistance much in the same way that HAART regimens stack pharmacological interventions simultaneously at multiple stages of the HIV lifecycle.

The next step in translating this effective approach into a viable therapy is to achieve targeting in primary human CD4+ T-cells or CD34+ hematopoietic stem cells. While seemingly not a significantly different experimental approach, getting targeting in T-cells proved to be a considerable hurdle. Unlike the JLTRG-R5 T-cell line I used in this report, primary T-cells do not expand indefinitely in

culture and therefore I had to often collect, isolate and activate new batches of T-cells. Nucleofection of T-cells is reported in the literature (Urnov et al., 2005), and I was able to get between 30-50% transfection of a GFP expressing plasmid. Similarly, I was able to transfect CCR5 ZFNs into the primary T-cells as detected by Western blot, but I was only able to detect very low levels of gene disruption and no evidence of gene targeting in primary CD4+ T-cells (Figure 4.1). Urnov et al (2005) demonstrated overall gene targeting frequencies of 5% in primary human T-cells by nucleofection of the IL2Rγ ZFNs and homologous donor template. In the toxicity assay described in Chapter 3, the IL2Rγ ZFNs and more so the CCR5 ZFNs are considerably cytotoxic in cell lines. One possibility for my inability to show targeting in T-cells could be related to the toxicity of the nucleases. It has been shown that, not surprisingly, in a population the most

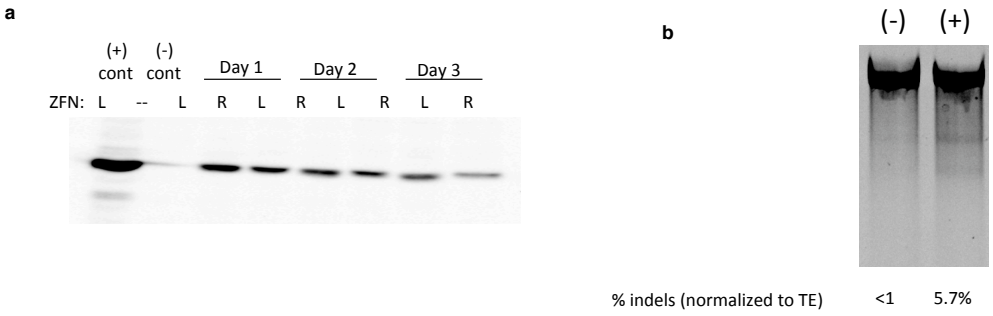


Figure 4.1: Expression and activity of CCR5 ZFNs in primary human T-cells. (a) Expression of left (L) and right (R) ZFNs in CD4+ T-cells. **(b)** Activity of CCR5 ZFNs in T-cells.

highly transfected cells undergo the highest rates of genome modification because they express higher levels of the nucleases (Kim et al., 2011). Higher levels of prolonged nuclease exposure also leads to higher toxicity (Pruett-Miller et al., 2008). In my experiments, perhaps I did achieve some low levels of targeting, but the targeted cells did not survive.

The work that first described the CCR5 ZFNs showed efficient cutting and survival of primary human T-cells using nuclease delivery by a non-integrating adenoviral vector (Perez et al., 2008). Similarly, the CD4+ T-cells that are CCR5 disrupted and reinfused into patients in the ZFN clinical trials used the same adenovirus to deliver the ZFNs (Ando et al., 2011; Mitsuyasu et al., 2011). In the current study, I attempted to stimulate ZFN cutting and gene addition by homologous recombination by nucleofection into T-cells. Through optimizations in the literature (Urnov et al., 2005) and in our lab, we found that the optimal ratio of ZFN 1 to ZFN 2 to targeting vector is 1:1:10, so trying to achieve targeting requires the transfection of 6-fold more DNA. There is a positive correlation between the amount of DNA transfected and the toxicity of the procedure. Also, the process of nucleofection itself that I used in this work can cause significant cell death, especially in primary cells (data not shown). Perhaps the combination of the higher amount of transfected DNA, the toxic CCR5 nucleases and the process of nucleofection prevented successfully targeted cells from surviving.

Improving cell viability after targeting CCR5 in primary cells

I have begun a series of preliminary experiments to address some of these concerns. First to reduce the amount of DNA required for transfection and to eliminate prolonged expression of the ZFNs, I began a set of experiments using ZFN delivery as mRNA. Because the half-life of mRNA in cells is significantly shorter compared to DNA plasmids, I hypothesized that transfection of ZFN mRNA would result in high initial expression, followed by a rapid decrease in protein levels. Transfection of mRNA has recently been used to generate iPS cells at efficiencies two orders of magnitude higher than standard protocols. Additionally, *in vitro* synthesis of the mRNA using the modified nucleotides 5'-methyl-cytosine (5mC) and pseudouridine (Psi) was determined to improve the survival of transfected cells compared to unmodified mRNA by avoiding innate intracellular anti-viral immune responses (Warren et al., 2010). In initial experiments, I synthesized GFP mRNA using either unmodified nucleotides, 5mC, Psi, or 5mC + Psi and nucleofected them into primary human CD4+ T-cells. Within 4 hours following transfection, I detected high levels of GFP expression from the unmodified and the Psi mRNA samples compared to cells transfected with a GFP-expressing plasmid. Interestingly, the 5mC and 5mC + Psi mRNAs showed much lower expression of GFP (Figure 4.2). As hypothesized, GFP positivity of mRNA transfected cells peaked not long after transfection (at 24 hours) and declined relatively rapidly after that, such that by 96 hours, there was

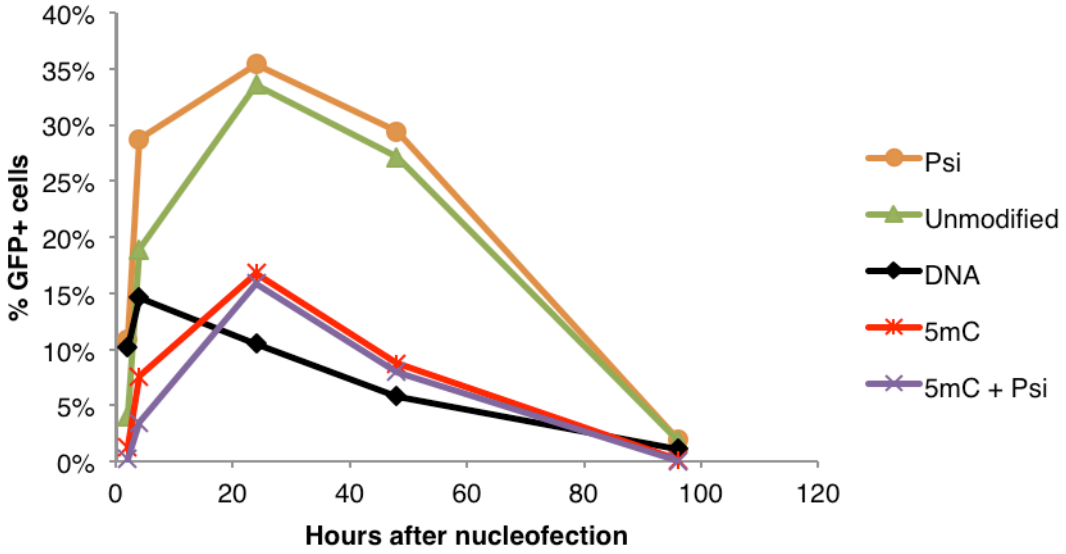


Figure 4.2: Transfection of primary human T-cells with GFP mRNA. Unmodified and Psi-modified mRNA are highly expressed at 24 hours after nucleofection.

almost no detectable GFP signal. The GFP profile of the DNA-transfected cells showed less of an increase and a slower decline of signal (Figure 4.2). These data suggest that delivery of ZFNs by mRNA would lead to an expression profile of the ZFNs that would be highly conducive to generate on-target cutting, without the prolonged expression that has been implicated as a source of toxicity (Pruett-Miller et al., 2009).

Next, I synthesized the CCR5 ZFNs as Psi mRNA and transfected them into K562 cells. The expression of the ZFNs peaked at 8 hours and by 72 hours was undetectable by Western blot (Figure 4.3). Compared to DNA, the mRNA

ZFNs stimulated significantly higher rates of cutting and repair by mutagenic NHEJ in the Surveyor nuclease assay (data not shown) and stimulated 25% more gene targeting in K562s (Figure 2.3c). Despite these improvements, I was still unable to target CCR5 in primary T-cells.

Besides synthesizing the ZFNs as mRNA, a second approach to reduce the nuclease-related toxicity is modifying the nuclease domain such that it can only heterodimerize. Since much of the toxicity is caused by off-target breaks, and since statistically two-thirds of off-target breaks result from the homodimerization of either two left ZFNs or two right ZFNs, obligate heterodimers have fewer potential off-target sites. Indeed, substituting charged amino acids at the interface of dimerization has been shown to reduce toxicity and maintain up to 70% of the activity of ZFNs with the wild type nuclease domain (Doyon et al., 2010; Miller

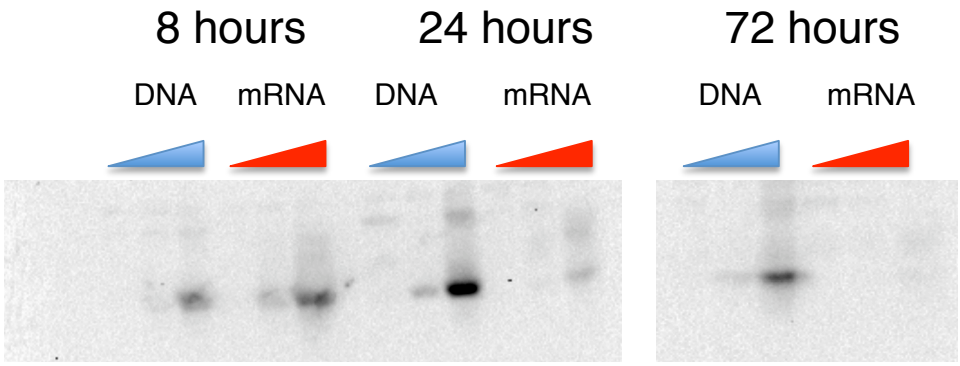


Figure 4.3 Expression of CCR5 ZFNs in K562s after delivery as DNA or mRNA. Increasing amount of FLAG-tagged CCR5 ZFN DNA (blue) or mRNA (red) were transfected into K562 cells and probed for expression by anti-FLAG antibody.

et al., 2007). I did not synthesize the obligate heterodimer CCR5 ZFNs, but in future applications of this targeting approach, using those versions may help further with the issue of nuclease-mediated toxicity.

The other option to address nuclease toxicity, either separate from or in combination with the obligate heterodimer nuclease domain is to use CCR5 TALENs instead of ZFNs. One of the first prominent papers describing TALEN architecture used as an example TALENs directed to CCR5 and showed gene disruption of greater than 20% and targeted integration of 16% in K562 cells (Miller et al., 2010). Combining the highly active, minimally toxic CCR5 TALENs with the gene targeting strategy outlined in this thesis is perhaps the best way to stack genetic resistance to both tropisms of HIV in primary human cells.

Therapeutic gene correction and endogenous locus tagging

In Chapter 3, I described the development of TALENs that target the human β- and γ−globin genes and the application of those TALENs in the context of sickle cell disease and β-thalassemia. This work described a significant advance in the field for five primary reasons. First, I made β-globin TALENs that were significantly more active and less toxic than published ZFNs that modify β-globin (Sebastiano et al., 2011; Zou et al., 2011a). For many of the same reasons discussed above concerning the toxicity of the CCR5 ZFNs, minimizing toxicity of nucleases in cell lines proves vital when attempting genome modification in

primary human cells. The β-globin TALENs induced modifications of greater than 50% of alleles, a frequency that is on par with the best nucleases that have been published (Reyon et al., 2012).

The second significant advance described in this work is the cDNA targeting strategy that would be curative for both sickle cell disease and β-thalassemia and would allow for direct selection of targeted cells *in vivo*. The two recently published β-globin ZFN papers (Sebastiano et al., 2011; Zou et al., 2011a) employed a gene targeting strategy that corrected the sickle mutation and integrated a drug resistance cassette 100 base pairs downstream in the first intron. Work from our lab has shown that integration of a gene 400 base pairs away from the site of a nuclease-induced break occurs approximately 7% as often as integration directly at the break site in HEK293T cells (Porteus, 2006). It is unknown what that frequency is when the break site and integration site are 100 base pairs apart as between the sickle mutation and the end of the first exon in β-globin. Nonetheless, integration of full-length cDNA directly at the site of the break is presumably much more efficient. Moreover, targeting full-length β-globin cDNA would also cure β-thalassemia, in which disease-causing mutations can occur throughout the length of the β-globin gene. Enrichment of targeted cells was performed using selection with O6BG and BCNU, FDA-approved drugs that can be administered *in vivo*.

Third, I introduce a new deep sequencing method to accurately quantify the frequencies of gene disruption and gene targeting. Single molecule, real time (SMRT) sequencing was first described in 2009 (Eid et al., 2009), and has since been used primarily to analyze DNA methylation patterns or damaged DNA bases (Clark et al., 2011a; Clark et al., 2011b; Flusberg et al., 2010; Song et al., 2011). In the genome engineering field, the standard way of confirming nuclease-induced modifications is by cloning PCR products and then picking and analyzing dozens of clones by Sanger sequencing. Not only is this process time and labor intensive, but it also requires screening of hundreds of clones individually to detect modifications at frequencies of 1%. Using SMRT sequencing, I was able to analyze almost 30,000 alleles for a fraction of the cost of runs on Illumina or 454 sequencing platforms. Although in K562s and HEK293T cells I showed very high rates of modification (>50%) and targeting (~20% of cells) using SMRT sequencing, the real strength of this method will come when analyzing targeting in cell types like HSCs in which modification events are much rarer.

Fourth, I successfully engineered a cell line that reports on the activity of endogenous promoters by tagging the chromosomal β- and γ-globin loci. A survey of peer-reviewed literature returns more than 30,000 papers using some sort of promoter reporter assay. In these studies, the authors primarily transfected cells with plasmids with a promoter of interest driving a reporter gene such as GFP. In this way, promoter activity is measured by extrachromosomal GFP

expression, but to my knowledge there have been no reports on the activity of endogenous promoters by nuclease-mediated locus tagging.

Finally, I used endogenous locus tagging of β-globin with GFP and γ-globin with tdTomato and screened for compounds that had differential effects on the two promoters. This is the first example of the combination of nuclease-mediated genomic engineering with a small molecule screen.

Autologous hematopoietic stem cell transplants to cure sickle cell disease

The ultimate goal of the β-globin cDNA targeting procedure developed in this thesis is to correct HSCs from sickle cell disease patients and retransfuse them to generate a functional cure for sickle cell disease. ZFN-mediated disruption has been reported in HSCs with the CCR5 ZFNs (Holt et al., 2010), but gene targeting with nucleases has not been shown in HSCs. In collaboration with Eric Kildebeck, I have demonstrated disruption of 9% of b-globin alleles in human HSCs (Figure 4.4). Targeting experiments are currently underway and can be analyzed with the SMRT sequencing method developed in this work to detect even low rates of targeting. After we show targeting in HSCs, our enrichment strategy will allow us to transplant the HSCs into an immunodeficient mouse and select for targeted cells *in vivo* as a proof-of-principle demonstration of the feasibility of this approach.

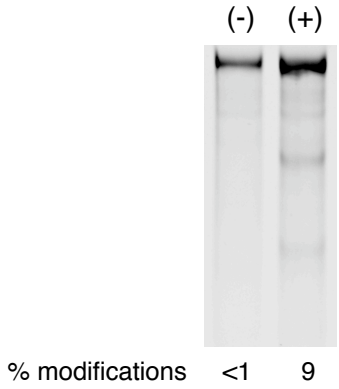


Figure 4.4 Gene disruption of β-globin in CD34+ HSCs. Nucleofection of TALENs into human HSCs resulted in modification of 9% of the β-globin alleles.

Overall conclusions

Targeted genome engineering holds great promise for treatment and cure of monogenic and other devastating human diseases. ZFNs in the treatment of HIV are showing early success in clinical trials, and the ease of synthesis and the significant success rate of TALENs make this genome-editing platform a powerful tool for research and for therapeutic gene editing. In this thesis, I have shown pre-clinical cures for HIV infection and the most common hemoglobinopathies, sickle cell disease and β-thalassemia, using targeted genome engineering strategies. This work provides the foundation for the eventual translation of nuclease-mediated gene editing into human HSCs to generate stem cells that are resistant to all HIV infection, and those that give rise to non-mutated

red blood cells in sickle cell disease patients. For all of the dissimilarities in the cause, pathogenesis and progression of HIV infection and sickle cell disease, they will soon have one very important aspect in common: a cure by targeted genome engineering.

BIBLIOGRAPHY

Abram, M.E., Ferris, A.L., Shao, W., Alvord, W.G., and Hughes, S.H. (2010). Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. J Virol *84*, 9864-9878.

Abuljadayel, I.S., Quereshi, H., Ahsan, T., Rizvi, S., Ahmed, T., Khan, S.M., Akhtar, J., and Dhoot, G. (2006). Infusion of autologous retrodifferentiated stem cells into patients with beta-thalassemia. ScientificWorldJournal *6*, 1278-1297.

Aiuti, A., Brigida, I., Ferrua, F., Cappelli, B., Chiesa, R., Marktel, S., and Roncarolo, M.G. (2009). Hematopoietic stem cell gene therapy for adenosine deaminase deficient-SCID. Immunol Res *44*, 150-159.

Albin, J.S., and Harris, R.S. (2010). Interactions of host APOBEC3 restriction factors with HIV-1 in vivo: implications for therapeutics. Expert Rev Mol Med *12*, e4.

Allers, K., Hutter, G., Hofmann, J., Loddenkemper, C., Rieger, K., Thiel, E., and Schneider, T. (2010). Evidence for the cure of HIV infection by CCR5Delta32/Delta32 stem cell transplantation. Blood *117*, 2791-2799.

Alwin, S., Gere, M.B., Guhl, E., Effertz, K., Barbas, C.F., 3rd, Segal, D.J., Weitzman, M.D., and Cathomen, T. (2005). Custom zinc-finger nucleases for use in human cells. Mol Ther *12*, 610-617.

Ando, D., Tebas, P., Stein, D., Wang, S., Lee, G., Holmes, M.C., Gregory, P.D., Giedlin, M., Tang, W., and June, C. (2011). HAART Treatment Interruption Following Adoptive Transfer of Zinc Finger Nuclease (ZFN) CCR5 Modified AutologousCD4 T-cells (SB-728-T) to HIV-infected Subjects Demonstrates Durable Engraftment and Suppression of Viral Load Paper presented at: 51st ICAAC (Chicago, IL).

Arnould, S., Chames, P., Perez, C., Lacroix, E., Duclert, A., Epinat, J.C., Stricher, F., Petit, A.S., Patin, A., Guillier, S., *et al.* (2006). Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. J Mol Biol *355*, 443-458.

Arnould, S., Delenda, C., Grizot, S., Desseaux, C., Paques, F., Silva, G.H., and Smith, J. (2010). The I-CreI meganuclease and its engineered derivatives: applications from cell modification to gene therapy. Protein Eng Des Sel *24*, 27-31

Arnould, S., Perez, C., Cabaniols, J.P., Smith, J., Gouble, A., Grizot, S., Epinat, J.C., Duclert, A., Duchateau, P., and Paques, F. (2007). Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. J Mol Biol *371*, 49-65.

Balazs, A.B., Chen, J., Hong, C.M., Rao, D.S., Yang, L., and Baltimore, D. (2011). Antibody-based protection against HIV infection by vectored immunoprophylaxis. Nature *481*, 81-84.

Banan, M., Esmaeilzadeh-Gharehdaghi, E., Nezami, M., Deilami, Z., Farashi, S., Philipsen, S., Esteghamat, F., Pourfarzad, F., Ali Imam, A.M., and Najmabadi, H. (2012). CREB1 is required for hydroxyurea-mediated induction of gamma-globin expression in K562 cells. Clin Exp Pharmacol Physiol.

Bank, A., Dorazio, R., and Leboulch, P. (2005). A phase I/II clinical trial of beta-globin gene therapy for beta-thalassemia. Ann N Y Acad Sci *1054*, 308-316.

Bauer, D.E., and Orkin, S.H. (2010). Update on fetal hemoglobin gene regulation in hemoglobinopathies. Curr Opin Pediatr *23*, 1-8.

Baumann, J.G. (2006). Intracellular restriction factors in mammalian cells--An ancient defense system finds a modern foe. Curr HIV Res *4*, 141-168.

Beerli, R.R., Dreier, B., and Barbas, C.F., 3rd (2000). Positive and negative regulation of endogenous genes by designed transcription factors. Proc Natl Acad Sci U S A *97*, 1495-1500.

Benabdallah, B.F., Allard, E., Yao, S., Friedman, G., Gregory, P.D., Eliopoulos, N., Fradette, J., Spees, J.L., Haddad, E., Holmes, M.C., *et al.* (2010). Targeted gene addition to human mesenchymal stromal cells as a cell-based plasma-soluble protein delivery platform. Cytotherapy *12*, 394-399.

Beumer, K., Bhattacharyya, G., Bibikova, M., Trautman, J.K., and Carroll, D. (2006). Efficient gene targeting in Drosophila with zinc-finger nucleases. Genetics *172*, 2391-2403.

Beumer, K.J., Trautman, J.K., Bozas, A., Liu, J.L., Rutter, J., Gall, J.G., and Carroll, D. (2008). Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. Proc Natl Acad Sci U S A *105*, 19821-19826.

Bianchi, N., Chiarabelli, C., Borgatti, M., Mischiati, C., Fibach, E., and Gambari, R. (2001). Accumulation of gamma-globin mRNA and induction of erythroid differentiation after treatment of human leukaemic K562 cells with tallimustine. Br J Haematol *113*, 951-961.

Bibikova, M., Carroll, D., Segal, D.J., Trautman, J.K., Smith, J., Kim, Y.G., and Chandrasegaran, S. (2001). Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. Mol Cell Biol *21*, 289-297.

Bibikova, M., Golic, M., Golic, K.G., and Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. Genetics *161*, 1169-1175.

Bitinaite, J., Wah, D.A., Aggarwal, A.K., and Schildkraut, I. (1998). FokI dimerization is required for DNA cleavage. Proc Natl Acad Sci U S A *95*, 10570-10575.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science *326*, 1509-1512.

Boutwell, C.L., Rolland, M.M., Herbeck, J.T., Mullins, J.I., and Allen, T.M. (2010). Viral evolution and escape during acute HIV-1 infection. J Infect Dis *202* S309-314.

Bozas, A., Beumer, K.J., Trautman, J.K., and Carroll, D. (2009). Genetic analysis of zinc-finger nuclease-induced gene targeting in Drosophila. Genetics *182*, 641-651.

Boztug, K., Schmidt, M., Schwarzer, A., Banerjee, P.P., Diez, I.A., Dewey, R.A., Bohm, M., Nowrouzi, A., Ball, C.R., Glimm, H., *et al.* (2010). Stem-cell gene therapy for the Wiskott-Aldrich syndrome. N Engl J Med *363*, 1918-1927.

Brunet, E., Simsek, D., Tomishima, M., DeKelver, R., Choi, V.M., Gregory, P., Urnov, F., Weinstock, D.M., and Jasin, M. (2009). Chromosomal translocations induced at specified loci in human stem cells. Proc Natl Acad Sci U S A *106*, 10620-10625.

Burma, S., Chen, B.P., and Chen, D.J. (2006). Role of non-homologous end joining (NHEJ) in maintaining genomic integrity. DNA Repair (Amst) *5*, 1042-1048.

Carbery, I.D., Ji, D., Harrington, A., Brown, V., Weinstein, E.J., Liaw, L., and Cui, X. (2010). Targeted genome modification in mice using zinc-finger nucleases. Genetics *186*, 451-459.

Cavazzana-Calvo, M., Hacein-Bey, S., de Saint Basile, G., Gross, F., Yvon, E., Nusbaum, P., Selz, F., Hue, C., Certain, S., Casanova, J.L., *et al.* (2000). Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. Science *288*, 669-672.

Cavazzana-Calvo, M., Lagresle, C., Hacein-Bey-Abina, S., and Fischer, A. (2005). Gene therapy for severe combined immunodeficiency. Annu Rev Med *56*, 585-602.

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 *467*, 318-322.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J., and Voytas, D.F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res *39*, e82.

Chan, K.S., Xu, J., Wardan, H., McColl, B., Orkin, S., and Vadolas, J. (2012). Generation of a genomic reporter assay system for analysis of gamma- and beta-globin gene regulation. FASEB J *26*, 1736-1744.

Charache, S., Terrin, M.L., Moore, R.D., Dover, G.J., Barton, F.B., Eckert, S.V., McMahon, R.P., and Bonds, D.R. (1995). Effect of hydroxyurea on the frequency of painful crises in sickle cell anemia. Investigators of the Multicenter Study of Hydroxyurea in Sickle Cell Anemia. N Engl J Med *332*, 1317-1322.

Chen, F., Pruett-Miller, S.M., Huang, Y., Gjoka, M., Duda, K., Taunton, J., Collingwood, T.N., Frodin, M., and Davis, G.D. (2011). High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. Nat Methods *8*, 753-755.

Chen, W., Zhan, P., De Clercq, E., and Liu, X. (2012). Recent progress in small molecule CCR5 antagonists as potential HIV-1 entry inhibitors. Curr Pharm Des *18*, 100-112.

Chiarle, R., Zhang, Y., Frock, R.L., Lewis, S.M., Molinie, B., Ho, Y.J., Myers, D.R., Choi, V.W., Compagno, M., Malkin, D.J., *et al.* (2011). Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. Cell *147*, 107-119.

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J., and Voytas, D.F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. Genetics *186*, 757-761.

Clark, T.A., Murray, I.A., Morgan, R.D., Kislyuk, A.O., Spittle, K.E., Boitano, M., Fomenkov, A., Roberts, R.J., and Korlach, J. (2011a). Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. Nucleic Acids Res *40*, e29.

Clark, T.A., Spittle, K.E., Turner, S.W., and Korlach, J. (2011b). Direct detection and sequencing of damaged DNA bases. Genome Integr *2*, 10.

Cohen, J. (2003). HIV. Escape artist par excellence. Science *299*, 1505-1508.

Connelly, J.P., Barker, J.C., Pruett-Miller, S., and Porteus, M.H. (2010). Gene correction by homologous recombination with zinc finger nucleases in primary cells from a mouse model of a generic recessive genetic disease. Mol Ther *18*, 1103-1110.

Cortesi, R., Gui, V., Gambari, R., and Nastruzzi, C. (1999). In vitro effect on human leukemic K562 cells of co-administration of liposome-associated retinoids and cytosine arabinoside (Ara-C). Am J Hematol *62*, 33-43.

Cui, X., Ji, D., Fisher, D.A., Wu, Y., Briner, D.M., and Weinstein, E.J. (2010). Targeted integration in rat and mouse embryos with zinc-finger nucleases. Nat Biotechnol *29*, 64-67.

Dar, A., Kollet, O., and Lapidot, T. (2006). Mutual, reciprocal SDF-1/CXCR4 interactions between hematopoietic and bone marrow stromal cells regulate human stem cell migration and development in NOD/SCID chimeric mice. Exp Hematol *34*, 967-975.

Davis, B.M., Roth, J.C., Liu, L., Xu-Welliver, M., Pegg, A.E., and Gerson, S.L. (1999). Characterization of the P140K, PVP(138-140)MLK, and G156A O6-methylguanine-DNA methyltransferase mutants: implications for drug resistance gene therapy. Hum Gene Ther *10*, 2769-2778.

DeKelver, R.C., Choi, V.M., Moehle, E.A., Paschon, D.E., Hockemeyer, D., Meijsing, S.H., Sancak, Y., Cui, X., Steine, E.J., Miller, J.C., *et al.* (2010). Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. Genome Res *20*, 1133-1142.

Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.K., Shi, Y., and Yan, N. (2012). Structural basis for sequence-specific recognition of DNA by TAL effectors. Science *335*, 720-723.

Ding, Q., Chang, C.J., Xie, X., Xia, W., Yang, J.Y., Wang, S.C., Wang, Y., Xia, J., Chen, L., Cai, C., *et al.* APOBEC3G promotes liver metastasis in an orthotopic mouse model of colorectal cancer and predicts human hepatic metastasis. J Clin Invest *121*, 4526-4536.

Doyon, J.B., Zeitler, B., Cheng, J., Cheng, A.T., Cherone, J.M., Santiago, Y., Lee, A.H., Vo, T.D., Doyon, Y., Miller, J.C., *et al.* (2011). Rapid and efficient clathrin-mediated endocytosis revealed in genome-edited mammalian cells. Nat Cell Biol *13*, 331-337.

Doyon, Y., McCammon, J.M., Miller, J.C., Faraji, F., Ngo, C., Katibah, G.E., Amora, R., Hocking, T.D., Zhang, L., Rebar, E.J., *et al.* (2008). Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. Nat Biotechnol *26*, 702-708.

Doyon, Y., Vo, T.D., Mendel, M.C., Greenberg, S.G., Wang, J., Xia, D.F., Miller, J.C., Urnov, F.D., Gregory, P.D., and Holmes, M.C. (2010). Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nat Methods *8*, 74-79.

Dreier, B., Beerli, R.R., Segal, D.J., Flippin, J.D., and Barbas, C.F., 3rd (2001). Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. J Biol Chem *276*, 29466-29478.

Dreier, B., Fuller, R.P., Segal, D.J., Lund, C.V., Blancafort, P., Huber, A., Koksch, B., and Barbas, C.F., 3rd (2005). Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. J Biol Chem *280*, 35588-35597.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., *et al.* (2009). Real-time DNA sequencing from single polymerase molecules. Science *323*, 133-138.

Erard, F., Dean, A., and Schechter, A.N. (1981). Inhibitors of cell division reversibly modify hemoglobin concentration in human erythroleukemia K562 cells. Blood *58*, 1236-1239.

Fibach, E., Kollia, P., Schechter, A.N., Noguchi, C.T., and Rodgers, G.P. (1995). Hemin-induced acceleration of hemoglobin production in immature cultured erythroid cells: preferential enhancement of fetal hemoglobin. Blood *85*, 2967-2974.

Flisikowska, T., Thorey, I.S., Offner, S., Ros, F., Lifke, V., Zeitler, B., Rottmann, O., Vincent, A., Zhang, L., Jenkins, S., *et al.* (2011). Efficient immunoglobulin gene disruption and targeted replacement in rabbit using zinc finger nucleases. PLoS One *6*, e21045.

Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., and Turner, S.W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods *7*, 461-465.

Foley, J.E., Yeh, J.R., Maeder, M.L., Reyon, D., Sander, J.D., Peterson, R.T., and Joung, J.K. (2009). Rapid mutation of endogenous zebrafish genes using zinc finger nucleases made by Oligomerized Pool ENgineering (OPEN). PLoS One *4*, e4348.

Fordis, C.M., Anagnou, N.P., Dean, A., Nienhuis, A.W., and Schechter, A.N. (1984). A beta-globin gene, inactive in the K562 leukemic cell, functions normally in a heterologous expression system. Proc Natl Acad Sci U S A *81*, 4485-4489.

Gabriel, R., Lombardo, A., Arens, A., Miller, J.C., Genovese, P., Kaeppel, C., Nowrouzi, A., Bartholomae, C.C., Wang, J., Friedman, G., *et al.* (2011). An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nat Biotechnol *29*, 816-823.

Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., Cummins, L.B., Arthur, L.O., Peeters, M., Shaw, G.M., *et al.* (1999). Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature *397*, 436-441.

Gaspar, H.B., Bjorkegren, E., Parsley, K., Gilmour, K.C., King, D., Sinclair, J., Zhang, F., Giannakopoulos, A., Adams, S., Fairbanks, L.D.*, et al.* (2006). Successful reconstitution of immunity in ADA-SCID by stem cell gene therapy following cessation of PEG-ADA and use of mild preconditioning. Mol Ther *14*, 505-513.

Geurts, A.M., Cost, G.J., Freyvert, Y., Zeitler, B., Miller, J.C., Choi, V.M., Jenkins, S.S., Wood, A., Cui, X., Meng, X.*, et al.* (2009). Knockout rats via embryo microinjection of zinc-finger nucleases. Science *325*, 433.

Girard, M.P., and Plotkin, S.A. (2011). HIV vaccine development at the turn of the 21st century. Curr Opin HIV AIDS *7*, 4-9.

Grizot, S., Epinat, J.C., Thomas, S., Duclert, A., Rolland, S., Paques, F., and Duchateau, P. (2009a). Generation of redesigned homing endonucleases comprising DNA-binding domains derived from two different scaffolds. Nucleic Acids Res *38*, 2006-2018.

Grizot, S., Smith, J., Daboussi, F., Prieto, J., Redondo, P., Merino, N., Villate, M., Thomas, S., Lemaire, L., Montoya, G.*, et al.* (2009b). Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. Nucleic Acids Res *37*, 5405-5419.

Guschin, D.Y., Waite, A.J., Katibah, G.E., Miller, J.C., Holmes, M.C., and Rebar, E.J. (2010). A rapid and general assay for monitoring endogenous gene modification. Methods Mol Biol *649*, 247-256.

Hacein-Bey-Abina, S., Garrigue, A., Wang, G.P., Soulier, J., Lim, A., Morillon, E., Clappier, E., Caccavelli, L., Delabesse, E., Beldjord, K.*, et al.* (2008). Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J Clin Invest *118*, 3132-3142.

Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., McCormack, M.P., Wulffraat, N., Leboulch, P., Lim, A., Osborne, C.S., Pawliuk, R., Morillon, E.*, et al.* (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science *302*, 415-419.

Harris, R.S., Petersen-Mahrt, S.K., and Neuberger, M.S. (2002). RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. Mol Cell *10*, 1247-1253.

Haynes, J., Jr., Baliga, B.S., Obiako, B., Ofori-Acquah, S., and Pace, B. (2004). Zileuton induces hemoglobin F synthesis in erythroid progenitors: role of the L-arginine-nitric oxide signaling pathway. Blood *103*, 3945-3950.

Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., DeKelver, R.C., Katibah, G.E., Amora, R., Boydston, E.A., Zeitler, B.*, et al.* (2009). Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat Biotechnol *27*, 851-857.

Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C.*, et al.* (2011). Genetic engineering of human pluripotent cells using TALE nucleases. Nat Biotechnol *29*, 731-734.

Holloman, W.K. (2011). Unraveling the mechanism of BRCA2 in homologous recombination. Nat Struct Mol Biol *18*, 748-754.

Holmes, C.B., Losina, E., Walensky, R.P., Yazdanpanah, Y., and Freedberg, K.A. (2003). Review of human immunodeficiency virus type 1-related opportunistic infections in sub-Saharan Africa. Clin Infect Dis *36*, 652-662.

Holt, N., Wang, J., Kim, K., Friedman, G., Wang, X., Taupin, V., Crooks, G.M., Kohn, D.B., Gregory, P.D., Holmes, M.C.*, et al.* (2010). Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. Nat Biotechnol *28*, 839-847.

Howden, S.E., Voullaire, L., Wardan, H., Williamson, R., and Vadolas, J. (2008). Site-specific, Rep-mediated integration of the intact beta-globin locus in the human erythroleukaemic cell line K562. Gene Ther *15*, 1372-1383.

Huang, P., Xiao, A., Zhou, M., Zhu, Z., Lin, S., and Zhang, B. (2011). Heritable gene targeting in zebrafish using customized TALENs. Nat Biotechnol *29*, 699-700.

Hutter, G., Nowak, D., Mossner, M., Ganepola, S., Mussig, A., Allers, K., Schneider, T., Hofmann, J., Kucherer, C., Blau, O.*, et al.* (2009). Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. N Engl J Med *360*, 692-698.

Kang, E.M., Choi, U., Theobald, N., Linton, G., Long Priel, D.A., Kuhns, D., and Malech, H.L. (2009). Retrovirus gene therapy for X-linked chronic granulomatous disease can achieve stable long-term correction of oxidase activity in peripheral blood neutrophils. Blood *115*, 783-791.

Keefer, J.R., Schneidereith, T.A., Mays, A., Purvis, S.H., Dover, G.J., and Smith, K.D. (2006). Role of cyclic nucleotides in fetal hemoglobin induction in cultured CD34+ cells. Exp Hematol *34*, 1151-1161.

Khan, I.F., Hirata, R.K., and Russell, D.W. (2011). AAV-mediated gene targeting methods for human cells. Nat Protoc *6*, 482-501.

Kim, H., Um, E., Cho, S.R., Jung, C., and Kim, J.S. (2011). Surrogate reporters for enrichment of cells with nuclease-induced mutations. Nat Methods *8*, 941-943.

Kim, H.J., Lee, H.J., Kim, H., Cho, S.W., and Kim, J.S. (2009). Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. Genome Res *19*, 1279-1288.

Kim, Y.G., and Chandrasegaran, S. (1994). Chimeric restriction endonuclease. Proc Natl Acad Sci U S A *91*, 883-887.

Lavelle, D.E. (2004). The molecular mechanism of fetal hemoglobin reactivation. Semin Hematol *41*, 3-10.

Lee, H.J., Kim, E., and Kim, J.S. (2009). Targeted chromosomal deletions in human cells using zinc finger nucleases. Genome Res *20*, 81-89.

Lei, Y., Lee, C.L., Joo, K.I., Zarzar, J., Liu, Y., Dai, B., Fox, V., and Wang, P. (2011). Gene Editing of Human Embryonic Stem Cells via an Engineered Baculoviral Vector Carrying Zinc-finger Nucleases. Mol Ther.

Li, T., Huang, S., Jiang, W.Z., Wright, D., Spalding, M.H., Weeks, D.P., and Yang, B. (2010). TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Res *39*, 359-372.

Li, T., Huang, S., Zhao, X., Wright, D.A., Carpenter, S., Spalding, M.H., Weeks, D.P., and Yang, B. (2011). Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. Nucleic Acids Res *39*, 6315-6325.

Liu, K., Xing, H., Zhang, S., Liu, S., and Fung, M. (2010). Cucurbitacin D induces fetal hemoglobin synthesis in K562 cells and human hematopoietic progenitors through activation of p38 pathway and stabilization of the gamma-globin mRNA. Blood Cells Mol Dis *45*, 269-275.

Liu, R., Paxton, W.A., Choe, S., Ceradini, D., Martin, S.R., Horuk, R., MacDonald, M.E., Stuhlmann, H., Koup, R.A., and Landau, N.R. (1996). Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. Cell *86*, 367-377.

Lloyd, A., Plaisier, C.L., Carroll, D., and Drews, G.N. (2005). Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. Proc Natl Acad Sci U S A *102*, 2232-2237.

Lombardo, A., Genovese, P., Beausejour, C.M., Colleoni, S., Lee, Y.L., Kim, K.A., Ando, D., Urnov, F.D., Galli, C., Gregory, P.D., *et al.* (2007). Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nat Biotechnol *25*, 1298-1306.

Maeder, M.L., Thibodeau-Beganny, S., Osiak, A., Wright, D.A., Anthony, R.M., Eichtinger, M., Jiang, T., Foley, J.E., Winfrey, R.J., Townsend, J.A.*, et al.* (2008). Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell *31*, 294-301.

Mak, A.N., Bradley, P., Cernadas, R.A., Bogdanove, A.J., and Stoddard, B.L. (2012). The crystal structure of TAL effector PthXo1 bound to its DNA target. Science *335*, 716-719.

Malech, H.L., Maples, P.B., Whiting-Theobald, N., Linton, G.F., Sekhsaria, S., Vowells, S.J., Li, F., Miller, J.A., DeCarlo, E., Holland, S.M.*, et al.* (1997). Prolonged production of NADPH oxidase-corrected granulocytes after gene therapy of chronic granulomatous disease. Proc Natl Acad Sci U S A *94*, 12133-12138.

Malim, M.H., Bohnlein, S., Hauber, J., and Cullen, B.R. (1989). Functional dissection of the HIV-1 Rev trans-activator--derivation of a trans-dominant repressor of Rev function. Cell *58*, 205-214.

Malim, M.H., McCarn, D.F., Tiley, L.S., and Cullen, B.R. (1991). Mutational definition of the human immunodeficiency virus type 1 Rev activation domain. J Virol *65*, 4248-4254.

Mashimo, T., Takizawa, A., Voigt, B., Yoshimi, K., Hiai, H., Kuramoto, T., and Serikawa, T. (2010). Generation of knockout rats with X-linked severe combined immunodeficiency (X-SCID) using zinc-finger nucleases. PLoS One *5*, e8870.

McMahon, M.A., Jilek, B.L., Brennan, T.P., Shen, L., Zhou, Y., Wind-Rotolo, M., Xing, S., Bhat, S., Hale, B., Hegarty, R.*, et al.* (2007). The HBV drug entecavir - effects on HIV-1 replication and resistance. N Engl J Med *356*, 2614-2621.

McMahon, M.A., Parsons, T.L., Shen, L., Siliciano, J.D., and Siliciano, R.F. (2011). Consistent inhibition of HIV-1 replication in CD4+ T cells by acyclovir without detection of human herpesviruses. J Virol *85*, 4618-4622.

Meiler, S.E., Wade, M., Kutlar, F., Yerigenahally, S.D., Xue, Y., Moutouh-de Parseval, L.A., Corral, L.G., Swerdlow, P.S., and Kutlar, A. (2011). Pomalidomide augments fetal hemoglobin production without the myelosuppressive effects of hydroxyurea in transgenic sickle cell mice. Blood *118*, 1109-1112.

Meng, X., Noyes, M.B., Zhu, L.J., Lawson, N.D., and Wolfe, S.A. (2008). Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. Nat Biotechnol *26*, 695-701.

Meyer, M., de Angelis, M.H., Wurst, W., and Kuhn, R. (2010). Gene targeting by homologous recombination in mouse zygotes mediated by zinc-finger nucleases. Proc Natl Acad Sci U S A *107*, 15022-15026.

Miller, J.C., Holmes, M.C., Wang, J., Guschin, D.Y., Lee, Y.L., Rupniewski, I., Beausejour, C.M., Waite, A.J., Wang, N.S., Kim, K.A., *et al.* (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol *25*, 778-785.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., *et al.* (2010). A TALE nuclease architecture for efficient genome editing. Nat Biotechnol *29*, 143-148.

Mischiati, C., Sereni, A., Lampronti, I., Bianchi, N., Borgatti, M., Prus, E., Fibach, E., and Gambari, R. (2004). Rapamycin-mediated induction of gamma-globin mRNA accumulation in human erythroid cells. Br J Haematol *126*, 612-621.

Mitsuyasu, R., Lalezari, J., Deeks, S., Wang, S., Lee, G., Holmes, M.C., Gregory, P.D., Giedlin, M., Tang, W., and Ando, D. (2011). Adoptive Transfer of Zinc Finger Nuclease CCR5 Modified Autologous CD4 T-cells (SB-728-T) to Aviremic HIV-infected Subjects with Suboptimal CD4 Counts (200-500 cells/mm3). Paper presented at: 51st ICAAC (Chicago, IL).

Mittelman, D., Moye, C., Morton, J., Sykoudis, K., Lin, Y., Carroll, D., and Wilson, J.H. (2009). Zinc-finger directed double-strand breaks within CAG repeat tracts promote repeat instability in human cells. Proc Natl Acad Sci U S A *106*, 9607-9612.

Moehle, E.A., Rock, J.M., Lee, Y.L., Jouvenot, Y., DeKelver, R.C., Gregory, P.D., Urnov, F.D., and Holmes, M.C. (2007). Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. Proc Natl Acad Sci U S A *104*, 3055-3060.

Morton, J., Davis, M.W., Jorgensen, E.M., and Carroll, D. (2006). Induction and repair of zinc-finger nuclease-targeted double-strand breaks in Caenorhabditis elegans somatic cells. Proc Natl Acad Sci U S A *103*, 16370-16375.

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

Moutouh-de Parseval, L.A., Verhelle, D., Glezer, E., Jensen-Pergakes, K., Ferguson, G.D., Corral, L.G., Morris, C.L., Muller, G., Brady, H., and Chan, K. (2008). Pomalidomide and lenalidomide regulate erythropoiesis and fetal hemoglobin production in human CD34+ cells. J Clin Invest *118*, 248-258.

Munoz, I.G., Prieto, J., Subramanian, S., Coloma, J., Redondo, P., Villate, M., Merino, N., Marenchino, M., D'Abramo, M., Gervasio, F.L., *et al.* (2010). Molecular basis of engineered meganuclease targeting of the endogenous human RAG1 locus. Nucleic Acids Res *39*, 729-743.

Mussolino, C., Morbitzer, R., Lutge, F., Dannemann, N., Lahaye, T., and Cathomen, T. (2011). A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res *39*, 9283-9293.

Nagasawa, T., Hirota, S., Tachibana, K., Takakura, N., Nishikawa, S., Kitamura, Y., Yoshida, N., Kikutani, H., and Kishimoto, T. (1996). Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1. Nature *382*, 635-638.

Neff, T., Horn, P.A., Peterson, L.J., Thomasson, B.M., Thompson, J., Williams, D.A., Schmidt, M., Georges, G.E., von Kalle, C., and Kiem, H.P. (2003). Methylguanine methyltransferase-mediated in vivo selection and chemoprotection of allogeneic stem cells in a large-animal model. J Clin Invest *112*, 1581-1588.

Ochiai, H., Fujita, K., Suzuki, K., Nishikawa, M., Shibata, T., Sakamoto, N., and Yamamoto, T. (2010). Targeted mutagenesis in the sea urchin embryo using zinc-finger nucleases. Genes Cells *15*, 875-885.

Ochsenbauer-Jambor, C., Jones, J., Heil, M., Zammit, K.P., and Kutsch, O. (2006). T-cell line for HIV drug screening using EGFP as a quantitative marker of HIV-1 replication. Biotechniques *40*, 91-100.

Odievre, M.H., Verger, E., Silva-Pinto, A.C., and Elion, J. (2011). Pathophysiological insights in sickle cell disease. Indian J Med Res *134*, 532-537.

Ogiwara, H., and Kohno, T. (2011). Essential factors for incompatible DNA end joining at chromosomal DNA double strand breaks in vivo. PLoS One *6*, e28756.

Osti, F., Corradini, F.G., Hanau, S., Matteuzzi, M., and Gambari, R. (1997). Human leukemia K562 cells: induction to erythroid differentiation by guanine, guanosine and guanine nucleotides. Haematologica *82*, 395-401.

Ott, M.G., Schmidt, M., Schwarzwaelder, K., Stein, S., Siler, U., Koehl, U., Glimm, H., Kuhlcke, K., Schilz, A., Kunkel, H., *et al.* (2006). Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. Nat Med *12*, 401-409.

Pacheco, B., Finzi, A., Stremlau, M., and Sodroski, J. (2010). Adaptation of HIV-1 to cells expressing rhesus monkey TRIM5alpha. Virology *408*, 204-212.

Papapetrou, E.P., Lee, G., Malani, N., Setty, M., Riviere, I., Tirunagari, L.M., Kadota, K., Roth, S.L., Giardina, P., Viale, A., *et al.* (2010). Genomic safe harbors permit high beta-globin transgene expression in thalassemia induced pluripotent stem cells. Nat Biotechnol *29*, 73-78.

Pattanayak, V., Ramirez, C.L., Joung, J.K., and Liu, D.R. (2011). Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. Nat Methods *8*, 765-770.

Perez, E.E., Wang, J., Miller, J.C., Jouvenot, Y., Kim, K.A., Liu, O., Wang, N., Lee, G., Bartsevich, V.V., Lee, Y.L., *et al.* (2008). Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat Biotechnol *26*, 808-816.

Pike-Overzet, K., van der Burg, M., Wagemaker, G., van Dongen, J.J., and Staal, F.J. (2007). New insights and unresolved issues regarding insertional mutagenesis in X-linked SCID gene therapy. Mol Ther *15*, 1910-1916.

Podsakoff, G.M., Engel, B.C., Carbonaro, D.A., Choi, C., Smogorzewska, E.M., Bauer, G., Selander, D., Csik, S., Wilson, K., Betts, M.R., *et al.* (2005). Selective survival of peripheral blood lymphocytes in children with HIV-1 following delivery of an anti-HIV gene to bone marrow CD34(+) cells. Mol Ther *12*, 77-86.

Porteus, M.H. (2006). Mammalian gene targeting with designed zinc finger nucleases. Mol Ther *13*, 438-446.

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

Pruett-Miller, S.M., Connelly, J.P., Maeder, M.L., Joung, J.K., and Porteus, M.H. (2008). Comparison of zinc finger nucleases for use in gene targeting in mammalian cells. Mol Ther *16*, 707-717.

Pruett-Miller, S.M., Reading, D.W., Porter, S.N., and Porteus, M.H. (2009). Attenuation of zinc finger nuclease toxicity by small-molecule regulation of protein levels. PLoS Genet *5*, e1000376.

Ranga, U., Woffendin, C., Verma, S., Xu, L., June, C.H., Bishop, D.K., and Nabel, G.J. (1998). Enhanced T cell engraftment after retroviral delivery of an antiviral gene in HIV-infected individuals. Proc Natl Acad Sci U S A *95*, 1201-1206.

Rerks-Ngarm, S., Pitisuttithum, P., Nitayaphan, S., Kaewkungwal, J., Chiu, J., Paris, R., Premsri, N., Namwat, C., de Souza, M., Adams, E., *et al.* (2009). Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. N Engl J Med *361*, 2209-2220.

Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D., and Joung, J.K. (2012). FLASH assembly of TALENs for high-throughput genome editing. Nat Biotechnol.

Richman, D.D., Little, S.J., Smith, D.M., Wrin, T., Petropoulos, C., and Wong, J.K. (2004). HIV evolution and escape. Trans Am Clin Climatol Assoc *115*, 289-303.

Rossi, J.J., June, C.H., and Kohn, D.B. (2007). Genetic therapies against HIV. Nat Biotechnol *25*, 1444-1454.

Rouet, P., Smih, F., and Jasin, M. (1994). Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol Cell Biol *14*, 8096-8106.

Sander, J.D., Cade, L., Khayter, C., Reyon, D., Peterson, R.T., Joung, J.K., and Yeh, J.R. (2011). Targeted gene disruption in somatic zebrafish cells using engineered TALENs. Nat Biotechnol *29*, 697-698.

Sander, J.D., Dahlborg, E.J., Goodwin, M.J., Cade, L., Zhang, F., Cifuentes, D., Curtin, S.J., Blackburn, J.S., Thibodeau-Beganny, S., Qi, Y.*, et al.* (2010). Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat Methods *8*, 67-69.

Sanjana, N.E., Cong, L., Zhou, Y., Cunniff, M.M., Feng, G., and Zhang, F. (2012). A transcription activator-like effector toolbox for genome engineering. Nat Protoc *7*, 171-192.

Santa-Marta, M., da Silva, F.A., Fonseca, A.M., and Goncalves, J. (2005). HIV-1 Vif can directly inhibit apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3G-mediated cytidine deamination by using a single amino acid interaction and without protein degradation. J Biol Chem *280*, 8765-8775.

Saracino, A., Monno, L., Cibelli, D.C., Punzi, G., Brindicci, G., Ladisa, N., Tartaglia, A., Lagioia, A., and Angarano, G. (2009). Co-receptor switch during HAART is independent of virological success. J Med Virol *81*, 2036-2044.

Sawyer, S.L., Wu, L.I., Emerman, M., and Malik, H.S. (2005). Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. Proc Natl Acad Sci U S A *102*, 2832-2837.

Schmit, J.C., and Weber, B. (1997). Recent advances in antiretroviral therapy and HIV infection monitoring. Intervirology *40*, 304-321.

Schuitemaker, H., Koot, M., Kootstra, N.A., Dercksen, M.W., de Goede, R.E., van Steenwijk, R.P., Lange, J.M., Schattenkerk, J.K., Miedema, F., and Tersmette, M. (1992). Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocytotropic to T-cell-tropic virus population. J Virol *66*, 1354-1360.

Schuitemaker, H., van 't Wout, A.B., and Lusso, P. Clinical significance of HIV-1 coreceptor usage. J Transl Med *9 Suppl 1*, S5.

Sebastiano, V., Maeder, M.L., Angstman, J.F., Haddad, B., Khayter, C., Yeo, D.T., Goodwin, M.J., Hawkins, J.S., Ramirez, C.L., Batista, L.F., *et al.* (2011). In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. Stem Cells *29*, 1717-1726.

Shenoy, S. (2011). Hematopoietic stem cell transplantation for sickle cell disease: current practice and emerging trends. Hematology Am Soc Hematol Educ Program *2011*, 273-279.

Smith, J., Grizot, S., Arnould, S., Duclert, A., Epinat, J.C., Chames, P., Prieto, J., Redondo, P., Blanco, F.J., Bravo, J., *et al.* (2006). A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. Nucleic Acids Res *34*, e149.

Sollu, C., Pars, K., Cornu, T.I., Thibodeau-Beganny, S., Maeder, M.L., Joung, J.K., Heilbronn, R., and Cathomen, T. (2010). Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. Nucleic Acids Res *38*, 8269-8276.

Song, C.X., Clark, T.A., Lu, X.Y., Kislyuk, A., Dai, Q., Turner, S.W., He, C., and Korlach, J. (2011). Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. Nat Methods *9*, 75-77.

Steinberg, M.H., and Rodgers, G.P. (2001). Pharmacologic modulation of fetal hemoglobin. Medicine (Baltimore) *80*, 328-344.

Stracker, T.H., and Petrini, J.H. (2011). The MRE11 complex: starting from the ends. Nat Rev Mol Cell Biol *12*, 90-103.

Stremlau, M., Owens, C.M., Perron, M.J., Kiessling, M., Autissier, P., and Sodroski, J. (2004). The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. Nature *427*, 848-853.

Sun, N., Liang, J., Abil, Z., and Zhao, H. (2012). Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. Mol Biosyst *8*, 1255-1263.

Szczepek, M., Brondani, V., Buchel, J., Serrano, L., Segal, D.J., and Cathomen, T. (2007). Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. Nat Biotechnol *25*, 786-793.

Takasu, Y., Kobayashi, I., Beumer, K., Uchino, K., Sezutsu, H., Sajwan, S., Carroll, D., Tamura, T., and Zurovec, M. (2010). Targeted mutagenesis in the silkworm Bombyx mori using zinc finger nuclease mRNA injection. Insect Biochem Mol Biol *40*, 759-765.

Tesson, L., Usal, C., Menoret, S., Leung, E., Niles, B.J., Remy, S., Santiago, Y., Vincent, A.I., Meng, X., Zhang, L.*, et al.* (2011). Knockout rats generated by embryo microinjection of TALENs. Nat Biotechnol *29*, 695-696.

Urnov, F.D., Miller, J.C., Lee, Y.L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D., and Holmes, M.C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature *435*, 646-651.

Vodros, D., and Fenyo, E.M. (2005). Quantitative evaluation of HIV and SIV co-receptor use with GHOST(3) cell assay. Methods Mol Biol *304*, 333-342.

Warren, L., Manos, P.D., Ahfeldt, T., Loh, Y.H., Li, H., Lau, F., Ebina, W., Mandal, P.K., Smith, Z.D., Meissner, A.*, et al.* (2010). Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. Cell Stem Cell *7*, 618-630.

Wilen, C.B., Wang, J., Tilton, J.C., Miller, J.C., Kim, K.A., Rebar, E.J., Sherrill-Mix, S.A., Patro, S.C., Secreto, A.J., Jordan, A.P.*, et al.* (2011). Engineering HIV-resistant human CD4+ T cells with CXCR4-specific zinc-finger nucleases. PLoS Pathog *7*, e1002020.

Witt, O., Monkemeyer, S., Ronndahl, G., Erdlenbruch, B., Reinhardt, D., Kanbach, K., and Pekrun, A. (2003). Induction of fetal hemoglobin expression by the histone deacetylase inhibitor apicidin. Blood *101*, 2001-2007.

Wongsrikeao, P., Saenz, D., Rinkoski, T., Otoi, T., and Poeschla, E. (2011). Antiviral restriction factor transgenesis in the domestic cat. Nat Methods *8*, 853-859.

Wood, A.J., Lo, T.W., Zeitler, B., Pickle, C.S., Ralston, E.J., Lee, A.H., Amora, R., Miller, J.C., Leung, E., Meng, X.*, et al.* (2011). Targeted genome editing across species using ZFNs and TALENs. Science *333*, 307.

Wright, D.A., Townsend, J.A., Winfrey, R.J., Jr., Irwin, P.A., Rajagopal, J., Lonosky, P.M., Hall, B.D., Jondle, M.D., and Voytas, D.F. (2005). High-frequency homologous recombination in plants mediated by zinc-finger nucleases. Plant J *44*, 693-705.

Xu, H., Svarovskaia, E.S., Barr, R., Zhang, Y., Khan, M.A., Strebel, K., and Pathak, V.K. (2004). A single amino acid substitution in human APOBEC3G antiretroviral enzyme confers resistance to HIV-1 virion infectivity factor-induced depletion. Proc Natl Acad Sci U S A *101*, 5652-5657.

Xu, J., Peng, C., Sankaran, V.G., Shao, Z., Esrick, E.B., Chong, B.G., Ippolito, G.C., Fujiwara, Y., Ebert, B.L., Tucker, P.W.*, et al.* (2011). Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. Science *334*, 993-996.

Yamanaka, S., Balestra, M.E., Ferrell, L.D., Fan, J., Arnold, K.S., Taylor, S., Taylor, J.M., and Innerarity, T.L. (1995). Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals. Proc Natl Acad Sci U S A *92*, 8483-8487.

Young, J.J., Cherone, J.M., Doyon, Y., Ankoudinova, I., Faraji, F.M., Lee, A.H., Ngo, C., Guschin, D.Y., Paschon, D.E., Miller, J.C., *et al.* (2011). Efficient targeted gene disruption in the soma and germ line of the frog Xenopus tropicalis using engineered zinc-finger nucleases. Proc Natl Acad Sci U S A *108*, 7052-7057.

Yuan, J., Wang, J., Crain, K., Fearns, C., Kim, K.A., Hua, K.L., Gregory, P.D., Holmes, M.C., and Torbett, B.E. (2012). Zinc-finger Nuclease Editing of Human cxcr4 Promotes HIV-1 CD4(+) T Cell Resistance and Enrichment. Mol Ther, doi: 10.1038/mt.2011.1310.

Yusa, K., Rashid, S.T., Strick-Marchand, H., Varela, I., Liu, P.Q., Paschon, D.E., Miranda, E., Ordonez, A., Hannan, N.R., Rouhani, F.J., *et al.* (2011). Targeted gene correction of alpha1-antitrypsin deficiency in induced pluripotent stem cells. Nature *478*, 391-394.

Zein, S., Li, W., Ramakrishnan, V., Lou, T.F., Sivanand, S., Mackie, A., and Pace, B. (2010). Identification of fetal hemoglobin-inducing agents using the human leukemia KU812 cell line. Exp Biol Med (Maywood) *235*, 1385-1394.

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G.M., and Arlotta, P. (2011). Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol *29*, 149-153.

Zou, J., Maeder, M.L., Mali, P., Pruett-Miller, S.M., Thibodeau-Beganny, S., Chou, B.K., Chen, G., Ye, Z., Park, I.H., Daley, G.Q., *et al.* (2009). Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. Cell Stem Cell *5*, 97-110.

Zou, J., Mali, P., Huang, X., Dowey, S.N., and Cheng, L. (2011a). Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease. Blood *118*, 4599-4608.

Zou, J., Sweeney, C.L., Chou, B.K., Choi, U., Pan, J., Wang, H., Dowey, S.N., Cheng, L., and Malech, H.L. (2011b). Oxidase-deficient neutrophils from X-linked chronic granulomatous disease iPS cells: functional correction by zinc finger nuclease-mediated safe harbor targeting. Blood *117*, 5561-5572.