GENE TARGETING IN A NOVEL MOUSE MODEL

AND THE CHICKEN DT40 CELL LINE

APPROVED BY SUPERVISORY COMMITTEE

Matthew Porteus, M.D., Ph.D.

Hongtao Yu, Ph.D.

David Chen, Ph.D.

Alec Zhang, Ph.D.

DEDICATION

Dedicated to my parents, Jon and Jean Connelly and Mike and Pam Holder for their continuous support and encouragement throughout my graduate career. I further wish to thank my wife, Helen, for her unconditional love and patience as I promised her I would be graduating “soon, in the next two months”, for almost a year now.

GENE TARGETING IN A NOVEL MOUSE MODEL AND THE CHICKEN DT40 CELL LINE

by

JON PATRICK CONNELLY

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, 2010

GENE TARGETING IN A NOVEL MOUSE MODEL AND THE CHICKEN DT40 CELL LINE

Jon Patrick Connelly

The University of Texas Southwestern Medical Center at Dallas, 2010

Matthew H. Porteus, M.D., Ph.D.

Gene targeting has the power to create precise changes at specific sites within the genome. In the context of gene therapy, this technology may be used to treat patients with monogenic diseases by fixing mutations in disease causing genes, followed by transplanting the corrected cells back into the patient. However, the natural rate of gene targeting is too low to be of practical use in most cells; an exception to this is the chicken DT40 cell line which has a high relative rate of gene targeting (gene targeting rate/ random integration rate). We therefore sought to determine the basis of this high rate of gene targeting using assays which quantitate the rates of repairing DNA double-strand breaks through different repair pathways. We show that compared to other cell types,

DT40 cells are deficient in random integration. Furthermore, we show this deficiency is due to a reduced ability to repair DNA breaks lacking homology at the ends. In other cell types, the naturally low rate of gene targeting can be stimulated 30-40,000 fold by inducing a double-strand break at the target site. These breaks can be created by proteins called zinc finger nucleases (ZFNs). ZFN mediated gene targeting is a powerful technology, but has not yet been fully characterized in primary cells. Furthermore, before clinical use in the treatment of monogenic diseases, it is necessary to first test this technology in animal models. In the second portion of this dissertation, we developed a mouse model of a generic recessive genetic disease. This model allows the study of gene targeting in any cell population isolated from the mouse. Using this model, we demonstrate ZFN mediated gene targeting in variety of primary cells isolated from the mouse, including ES cells, fibroblasts, and astrocytes. We further demonstrate that targeted stem cells retain their pluripotency, and show that targeted fibroblasts can be transplanted back into a recipient and continue to express protein from the corrected gene. This body of work contributes to bringing the technology of gene targeting closer to clinical application by detailing methods which can be used to further increase gene targeting rates, as well as providing a paradigm in which to study gene targeting followed by transplantation.

TABLE OF CONTENTS

PRIOR PUBLICATIONS

Connelly JP, Barker JC, Pruett-Miller SM, Porteus MH. Gene correction by homologous recombination with zinc finger nucleases in primary cells from a mouse model of a generic recessive genetic disease. (Molecular Therapy, In Press)

Connelly JP, Brinker SK, Takeda S, Porteus MH. Low rates of random integration and joining of non-compatible ends in chicken DT40 cells. (Manuscript in Preparation)

Pruett-Miller SM, Connelly JP, Maeder ML, Joung JK, Porteus MH. (2008) Comparison of zinc finger nucleases for use in gene targeting in mammalian cells. Mol Ther.; 16(4); 707-17.

Porteus MH, Connelly JP, Pruett SM. (2006) A look to future directions in gene therapy research for monogenic diseases. PLoS Genet. 2(9):e133. Review.

LIST OF FIGURES

LIST OF TABLES

LIST OF DEFINITIONS

ADA-SCID – Adenosine Deaminase-Deficient Severe Combined Immunodeficiency

ATM – Ataxia Telangiectasia Mutated

ATR – Ataxia Telangiectasia Mutated- and Rad3-related

BP – Base pair

CCR5 – C-C Chemokine Receptor type 5

CMV/CBA – Cytomegalovirus/ Chicken B-actin promoter

DSB – Double-strand break

DSB-GT – Double-strand Break Gene Targeting

DNA – Deoxyribonucleic Acid

DNA-PKcs – DNA-dependent Protein Kinase Catalytic Subunit

DRGFP – Direct Repeat GFP

DT40 – Chicken B Lymphoma DT40 Cell Line

ES – Embryonic Stem (cell)

FokI – Restriction Enzyme from Flavobacterium okeanokoites

GFP – Green Fluorescent Protein

HR – Homologous Recombination

I-SceI – Homing Endonuclease Isolated from *Saccharomyces cerevisiae*

LCA – Leber's Congenital Amaurosis

LigIV – Ligase IV

NHEJ – Non-homologous end-joining

RI – Random Integration

RFP – Red Fluorescent Protein

ROSA26 –Reverse-Orientation Splice Acceptor 26

S-GT – Spontaneous Gene Targeting

Ubx – Ultrabithorax

XLF – XRCC4-like Factor (also called Cernunnos)

X-SCID – X-linked Severe Combined Immunodeficiency

XRCC4 –X-ray Repair Complementing Defective Repair in Chinese Hamster Cells 4

ZFN – Zinc Finger Nuclease

CHAPTER I:
INTRODUCTION TO MONOGENIC DISEASES AND GENE THERAPY

The Importance of Gene Therapy

Monogenic diseases afflict millions of people worldwide, and as their name implies, these diseases are caused by a mutation in a single gene. Current estimates indicate that over 10,000 diseases are monogenic in nature, including cystic fibrosis, β-thalassemia, X-SCID, hemophilia, sickle cell anemia, and numerous others, and their prevalence in society is approximately 10/1000 (W.H.O. 2010). These diseases are both costly to treat and cause hardship on the affected individuals as well as their families. The goal of gene therapy focuses on correcting the mutant gene or compensating for the deficiency of its protein product. If performed correctly, it may be possible to cure or at least alleviate the symptoms of the disease and allow the affected individual to lead a normal life.

Gene therapy is performed using one of two methods. The first, known as "gene addition" is based on inserting a functional copy of the desired gene randomly into the cell genome, with the idea that the gene will then express a functional version of the protein that the patient lacks. The second method is known as "gene correction". In this method, a functional copy of the gene is used to "patch" the endogenous mutation present in the patient's genome. Each method has its own advantages and disadvantages as discussed below.

Gene Therapy by Gene Addition

Clinical trials of gene therapy have to date been based on the gene addition method. In these trials, a viral vector is developed that contains a functional copy of the gene needed to 'cure' the patient. Cells are then either taken from the patient, infected with the virus, and re-administered back to the patient, or alternatively, the virus may be directly administered into patient. In some instances, the therapy was successful and the patient showed long-term improvement (Hacein-Bey-Abina, Le Deist et al. 2002; Bainbridge, Smith et al. 2008; Aiuti, Cattaneo et al. 2009; Cartier, Hacein-Bey-Abina et al. 2009; Maguire, High et al. 2009), more often however, clinical improvement was transient or absent (Aiuti, Slavin et al. 2002; Raper, Yudkoff et al. 2002; Manno, Chew et al. 2003; Raper, Chirmule et al. 2003; Gaspar, Parsley et al. 2004; Manno, Pierce et al. 2006; Ott, Schmidt et al. 2006). More importantly, some patients suffered severe complications including myeloproliferations and leukemias (Hacein-Bey-Abina, von Kalle et al. 2003; Hacein-Bey-Abina, Von Kalle et al. 2003; Cavazzana-Calvo, Lagresle et al. 2005; Ott, Schmidt et al. 2006; Bainbridge, Smith et al. 2008), and, in one instance, a severe immunological response was induced by the direct administration of the viral vector into the patient (Raper, Chirmule et al. 2003). These trials indicate that although promising, gene therapy in the clinic still must overcome some serious obstacles.

Gene Therapy by Gene Correction

Instead of adding a desired gene randomly into the genome, a more ideal method would be to fix the disease causing mutation that is present within the patient's genome. This method is known as gene correction or gene targeting. Gene correction uses the body's natural DNA repair mechanisms to fix the mutant gene. This can be performed by

introducing into the cell a piece of DNA that contains a correct copy of the desired gene. At a low frequency (10^{-6}), the cell will use the introduced DNA to replace the mutant nucleotides with the correct ones found in the introduced DNA. This rate is too low to be of therapeutic use; however, by using this method, the 'patched' gene will remain under control of the natural promoter and regulatory elements present at the locus. This method also reduces the problems of random viral insertion, vector silencing, and immune responses associated with the gene addition method. To achieve higher rates of gene targeting it is necessary to create a DNA double-strand break at the target site (discussed below). Creation of a double-strand break can increase gene targeting rates several thousand fold, potentially making this strategy therapeutically relevant.

Research Goals

My graduate school work focused on studying two aspects relevant to gene targeting and gene therapy, (i) why a unique chicken B lymphoma cell line has an unusually high rate of relative gene targeting, and, (ii) the creation of a mouse model that allows the study of DNA double-strand break mediated gene targeting and transplantation of primary cells isolated from the mouse prior to performing clinical trials in patients.

Study DNA Repair in the Chicken DT40 Cell Line

When a DNA double-strand break is induced in a cell, there are two main pathways by which it is repaired, non-homologous end joining (NHEJ), and homologous recombination (HR). Non-homologous end joining re-joins the broken ends, sometimes in a mutagenic fashion, while homologous recombination uses a template to "patch" the break. When a double-strand break is made in a cell, one of these two pathways is used,

but in order to precisely alter the DNA sequence at the break, repair must proceed through the HR pathway using a defined template. Most cells tend to repair the break using NHEJ. An exception to this is the chicken DT40 cell line which has been shown to have a high relative rate of gene targeting. We therefore desired to learn why this line has a higher than average rate of gene targeting with the hope that the information learned could be applied to other cell types to increase gene targeting rates. I modified a previously described assay that allows the rates of repair of compatible and non-compatible DNA ends (by the NHEJ pathway) to be easily quantified and studied. We used this assay along with assays measuring gene targeting (both spontaneous and double-strand break induced) and random integration assays to determine why chicken DT40 cells have a high rate of gene targeting compared to other cells. Furthermore, we studied the effects of inhibiting different aspects of DNA repair with chemicals to explore whether inhibition of NHEJ could be used to increase gene targeting.

Development of a Mouse Model to Study Gene Targeting in Primary Cells

My second goal was to develop a relevant mouse model in which gene targeting stimulated by double-strand breaks could be studied and characterized in primary cells. Gene targeting may have unforeseen consequences, as was demonstrated in prior gene addition clinical trials and because of this, it is important that studies should first be tested and carried out in a relevant animal model before being used in the clinic. At the beginning of my graduate school career, most studies of gene correction had been performed in transformed cell lines. These lines provided insight into the process of gene correction, however because they are transformed, they do not accurately reflect how the process may work in primary cells. With this mouse model I wanted to study what cell

types could be corrected using double-stranded break mediated gene targeting, and at what rates. I also desired to learn whether stem cells undergoing gene correction retain their pluripotency. Finally, my ultimate goal was to show that gene correction could be accomplished in therapeutically relevant cell types *ex vivo*, and demonstrate that these cells could be transplanted back into recipient mice and maintain their corrected phenotype.

CHAPTER II:
REVIEW OF THE LITERATURE

DNA Double-Strand Break Repair in Mammalian Cells

DNA double-strand breaks (DSBs) are potentially dangerous cellular lesions that can be induced through a number of processes including ionizing radiation, reactive oxygen species, and topoisomerase inhibitors (Ross, Rowe et al. 1984; Kolachana, Subrahmanyam et al. 1993; Olive and Banath 1993), and reviewed in (Houtgraaf, Versmissen et al. 2006). If not repaired correctly, DSBs may lead to cell death or chromosomal rearrangements resulting in cancer (Burma, Chen et al. 2006; O'Driscoll and Jeggo 2006). Two main mechanisms exist to repair DSBs, the non-homologous end-joining pathway (NHEJ) and homologous recombination (HR), (reviewed in (Karran 2000; van Gent, Hoeijmakers et al. 2001)). NHEJ is performed by re-joining the DNA ends, sometimes in a mutagenic fashion. In this pathway, when a double-strand break occurs, Ku70 and Ku80 bind to each free end of the DNA at the break site as a heterodimer and recruit DNA-dependent protein kinase catalytic subunit (DNA-PK_{cs}). This complex, along with Artemis is able to process the ends if needed and recruit LigIV, XRCC4, and XLF which ligates the ends back together. Due to the processing activity by Artemis, nucleotides may be deleted or inserted at the break site causing mutations (Roth and Wilson 1986; Guirouilh-Barbat, Huck et al. 2004). In contrast to NHEJ, HR uses a homologous sequence of DNA to repair the break. Because a homologous template is needed, it is thought that this repair pathways functions primarily during the

S/G2 phase of the cell cycle when a sister chromatid is present (Saleh-Gohari and Helleday 2004). During HR, the DNA ends are first resected forming 3' overhangs. Next Rad51 mediates strand invasion of the 3' tail into the homologous DNA sequence being used for repair and new DNA is synthesized using the homologous strand as a template (Figure 2.1). At this point HR may diverge into two pathways to complete repair of the break, (i) double-strand break repair, and (ii) synthesis dependent strand annealing (reviewed in (Sung and Klein 2006)). If the double-strand break repair pathway is used, the second 3' tail of the double-strand break will anneal to the homologous strand being used as a template forming a Holliday junction and synthesize new DNA using the homologous strand as a template. The Holliday junctions can then be resolved in a crossover or non-crossover manner (2.1 A1-3). If synthesis dependent strand annealing is used, the 3' tail will anneal to its homologous DNA strand being used as a repair template and synthesize new DNA (2.1 B1). Next, the strand is released, and anneals back to the 3' overhang on the opposite side of the double-strand break. The remaining 3' tail then uses the strand as a template to synthesize new DNA.

In conclusion, when a double-strand break is created in a cell, there are two main pathways with which to repair the break- NHEJ and HR. NHEJ may fix the break by perfectly ligating the ends back together, or may first process the ends through deleting/inserting nucleotides, followed by rejoining. On the other hand, HR fixes the break by copying in information found on a homologous template strand. An immediate question is how one method is chosen over the other. This remains unknown, however studies have demonstrated that mammalian cells preferentially fix the break using NHEJ (Guirouilh-Barbat, Huck et al. 2004; Mao, Bozzella et al. 2008).

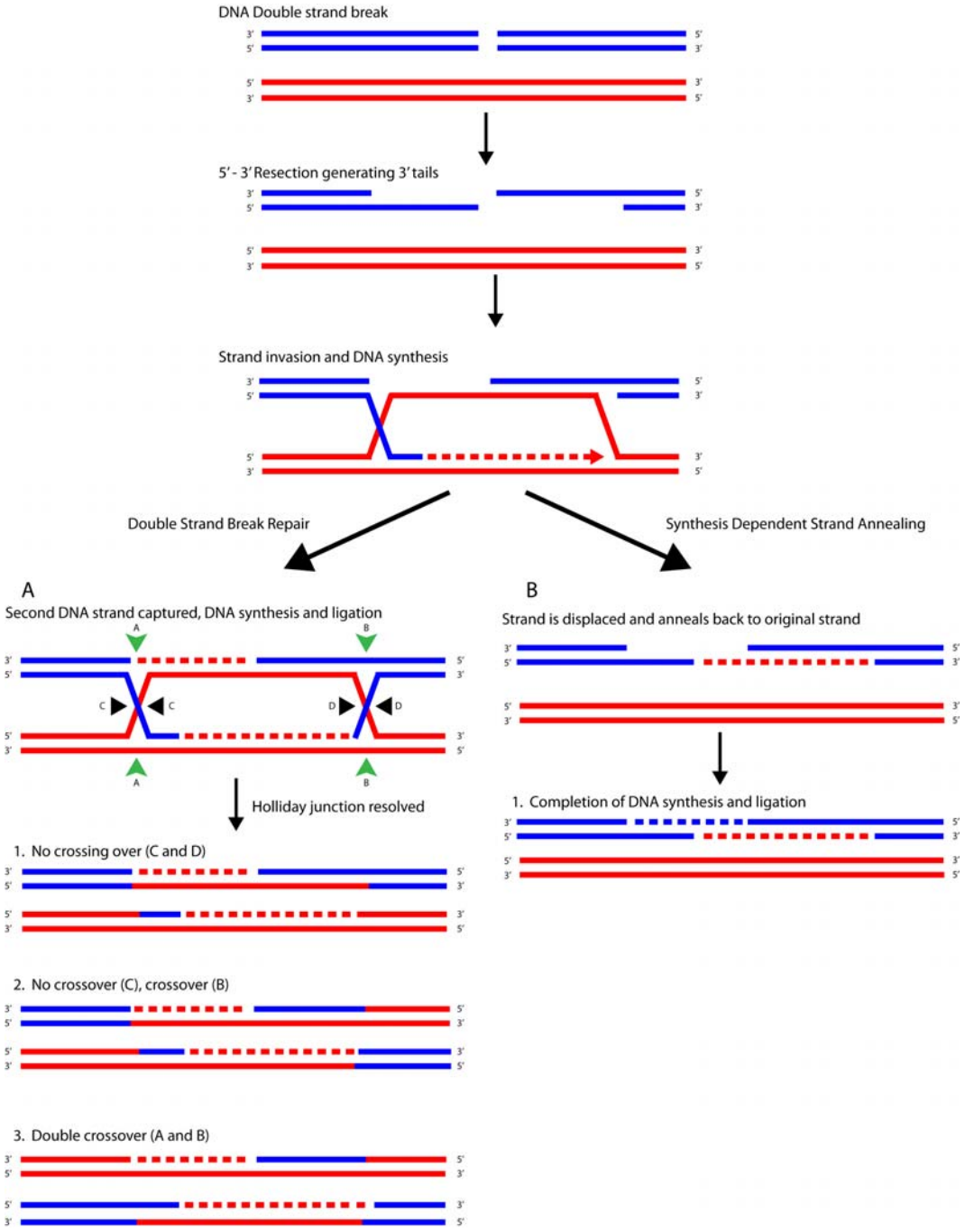

Figure 2.1: Double-Strand Break Repair by Homologous Recombination.

Repair of a DNA double-strand break by homologous recombination. If repair occurs using "double-strand break repair" (A), one of four types of repair products may be generated depending on how the Holliday junctions are resolved (green triangles result in crossover, black arrows do not lead to crossover). (A1) No crossover (C and D), (A2) No crossover at C, with crossover at B, or, alternatively, crossover at A and no crossover at D (not shown), (A3) Double crossover resulting from crossover at A and B. (B1) If repair occurs using synthesis dependent strand annealing, no crossover occurs.

Gene Targeting

When a DSB is naturally repaired by a cell through HR, the template for repair is most often the sister chromatid. The field of gene targeting began over 30 years ago when Hinnen et al. (1978) demonstrated that exogenously supplied DNA could also be used as a template substrate for HR in the budding yeast *S. cerevisiae* (Hinnen, Hicks et al. 1978). In this work, they showed that a leu2- strain of yeast could undergo homologous recombination with an exogenously supplied plasmid containing a functional LEU2+ gene. Analysis of the colonies that were able to grow on media lacking leucine indicated that the majority of colonies had randomly integrated the plasmid into their genome. However, approximately 15% of the transformants contained a functional LEU2 gene at the endogenous leu2 locus, demonstrating that the mutant locus was able to recombine with the plasmid and replace the mutant nucleotides with correct ones. Later studies by Smith and Berg, and Smithies in 1984 demonstrated gene targeting by homologous recombination could also be performed in mouse and human cell lines between an introduced plasmid and chromosomally integrated reporter system

(Smith and Berg 1984; Smithies, Koralewski et al. 1984). A year later, Smithies et al. also demonstrated homologous recombination at an endogenous locus (the human β-globin locus) (Smithies, Gregg et al. 1985). Importantly, it was shown that although gene targeting occurred, its frequency was extremely low. Furthermore, Smithies et al. (1984) noted that when the incoming plasmid DNA was linearized within a region of homology between the plasmid and chromosomal target, recombinants occurred at a higher frequency. However, the frequency of correctly targeted clones (1 per 1000 selected clones), was too low to be of practical use for gene therapy. With the further development of selection strategies (both positive and negative) and the finding that large homologous sequences that flank the transgene increase targeting, cells that have undergone gene targeting can be screened for with minimal difficulty. The power of this technology is great, the genomic DNA of a cell can be precisely altered- nucleotides can be changed, deleted, or added with precision.

Today, gene targeting is widely applied in the field of mouse genetics. Gene targeting is performed in mouse ES cells to delete or modify a locus, and these modified ES cells are used to derive mice containing the modifications. Today, thousands of lines of gene targeted mice have been developed and are extensively used in research studies. Gene targeting also holds great promise in the field of gene therapy. The major goal of gene therapy by gene targeting is to isolate therapeutically relevant cell types from a diseased individual, correct the causative mutations through gene targeting, and then transplant the corrected cells back into the patient thus curing or alleviating the symptoms of the disease. However, the overall gene targeting rate is still too low in mammalian cells and as mentioned above selection strategies must be utilized in order to find cells

containing the desired modification. In order for gene targeting to be used therapeutically, higher rates of targeting needed to be achieved.

DNA Double-strand Breaks Increase Gene Targeting Rates

Previous studies had demonstrated that double-strand breaks introduced into the donor plasmid increased homologous recombination rates 4-40 fold (Smith and Berg 1984; Jasin, de Villiers et al. 1985). Rouet et al. showed that double-strand breaks introduced into the actual chromosomal target site also increase homologous recombination 2-3 orders of magnitude in a transformed cell line (Rouet, Smih et al. 1994). This was performed by introducing an I-Sce site into a neomycin resistance gene (rendering it non-functional) and integrating it into the genome. Next, Rouet et al. transfected into the cells a plasmid expressing the homing endonuclease I-SceI along with a plasmid containing a truncated neo fragment. I-SceI was expressed, and cut the integrated neo gene, forming a double-strand break. The cleaved neo transgene then underwent homologous recombination with the truncated donor plasmid, and in the process corrected the I-SceI mutation. Neomycin phosphotransferase was expressed from the corrected gene, and resistant clones could be screened for. Their work demonstrated that homologous recombination between an integrated transgene and extrachromosmal donor plasmid could be stimulated dramatically by the induction of a double-strand break at the chromosomal target site, with information being copied from the plasmid to the chromosomal site. In further studies, homologous recombination stimulated by an I-SceI induced double-strand break was demonstrated in mouse ES cells (Smih, Rouet et al. 1995) and Xenopus oocytes (Segal and Carroll 1995). These

experiments all relied on creating a double-strand break with the I-SceI endonuclease, and targeting rates were measured using an integrated transgene. When put into the context of gene therapy however, disease genes are not transgenes, and selection for corrected cells is rarely possible. Furthermore, the target gene will not contain an I-SceI site at which a double-strand break can be readily induced. A method to create high levels of site specific DSBs needed to be created.

Development of Zinc Finger Nucleases

Srinivasan Chandrasegaran's lab became interested in developing restriction enzymes that could recognize and cleave DNA sequences containing larger recognition sequences than the 6-8 bp sites that are most commonly found in nature. It had already been shown that the type IIS restriction enzymes such as FokI, contained two separable domains- an N-terminal domain that conferred DNA binding activity, and a C-terminal domain that had non-specific nuclease activity (Li, Wu et al. 1992; Wah, Hirsch et al. 1997). They reasoned that it should therefore be possible to fuse the C-terminal nuclease domain to a new DNA binding domain and thereby create a nuclease that cleaves DNA at the new DNA binding recognition sequence. This was first demonstrated by fusing the FokI nuclease domain to the Ultrabithorax (Ubx) homeodomain of *Drosophila* (Kim and Chandrasegaran 1994). Ubx is a transcription factor important for embryonic development, and contains a helix-turn-helix DNA binding motif that recognizes a 9bp degenerate DNA binding site. They successfully demonstrated that this novel chimeric restriction endonuclease could bind and cleave the Ubx site in vitro (Kim and Chandrasegaran 1994). Expanding on this technique, they next demonstrated that the

nuclease domain could be fused to arrays of zinc finger motifs termed a zinc finger nuclease (ZFN) (Kim, Cha et al. 1996). To date, chimeric nucleases have been created from DNA binding motifs such as the zinc-finger motif, helix-turn-helix motif, and the basic helix-loop-helix motif. (reviewed in (Chandrasegaran and Smith 1999)). Of these DNA binding proteins, the Cys_2-His_2 class of zinc finger proteins is most commonly used due to its semi-modular nature (Pavletich and Pabo 1991) compared to other DNA binding motifs. In their work, Pavletich and Pabo determined the crystal structure of the murine transcription factor Zif268. This zinc finger is composed of an array of three zinc finger motifs that recognize a 9bp sequence. In their work, they found that each motif formed three major contacts on the same DNA strand. The contacts of each zinc finger motif with the DNA were mostly independent, suggesting that different zinc finger motifs could be mixed and matched to recognize novel sequences. Each motif formed a $\beta\beta\alpha$ structure stabilized by a zinc ion. The motif binds DNA by inserting its α helix (the finger) into the major groove of the DNA strand (binding to a three nucleotide sequence). Because of the modular nature of the zinc finger motif, they can be fused together in an array forming a zinc finger protein containing 3-6 zinc finger motifs with a novel recognition sequence of 9-18 bp. Because two FokI nuclease domains must dimerize together for efficient cleavage, two chimeric nucleases must bind to opposite strands of a DNA sequence in an inverted orientation to cleave and stimulate gene targeting (Smith, Bibikova et al. 2000). In 2001, Bibikova et al. (2001) demonstrated these chimeric restriction endonucleases could bind, cleave, and stimulate homologous recombination on an extra-chromosomal target in Xenopus oocytes (Bibikova, Carroll et al. 2001). A year later, they also demonstrated that zinc finger nucleases could bind and cleave

chromosomal targets in Drosophila (Bibikova, Golic et al. 2002). In order to increase the potential number of ZFN DNA binding sites, phage display strategies have been used to generate large numbers of zinc finger motifs with known binding sequences that can be fused together to recognize novel sites and these efforts have increased the range of potential cleavage target sites (Choo and Klug 1994; Choo and Klug 1994; Jamieson, Kim et al. 1994; Rebar and Pabo 1994; Wu, Yang et al. 1995). Furthermore, because an individual zinc finger motif in an array does have a modest influence on the binding of an adjoining finger (Isalan, Choo et al. 1997), selection and shuffling strategies are being employed to take this context-dependent interaction into account (Greisman and Pabo 1997; Isalan, Klug et al. 2001; Hurt, Thibodeau et al. 2003), and reviewed in (Cathomen and Joung 2008). To date, ZFNs have been shown to cleave and stimulate homologous recombination in a variety of sequences from plants, Drosophila, Xenopus, mice, and humans.

Summary

Gene targeting continues to hold great promise to the field of gene therapy. Through the use of ZFNs, therapeutically relevant targeting frequencies have been achieved in a variety of cell types at a number of different loci. Still important questions remain. For instance, studies on the chicken B lymphocyte DT40 cell line (a cell line derived from an avian leukosis virus induced bursal lymphoma) have demonstrated that this line has a high relative rate of gene targeting compared to other cells. In Chapter III, I studied the basis of this high gene targeting rate, with the prospect that the resulting data

could be used to increase targeting rates in other cell types. Other questions I had were more focused on the application of gene targeting to the field of gene therapy. What is the range of primary cells that are amenable to gene targeting? When ES cells are subjected to gene targeting, do they retain their pluripotency? And furthermore, can targeted cells be transplanted back into a recipient and remain viable? To address these questions, I first created a mouse model of a generic genetic recessive disease as a paradigm in which to study gene therapy by gene correction prior to being performed in patients (Figure 2.2). This mouse model allowed us to isolate a variety of primary cell types from the mouse and study gene targeting within them. Once correction of the mutant gene was demonstrated, we next showed that the corrected cells could be transplanted back into a host recipient as would be necessary in patients to correct their disease phenotype. These studies resulted in a publication, reprinted in Chapter IV.

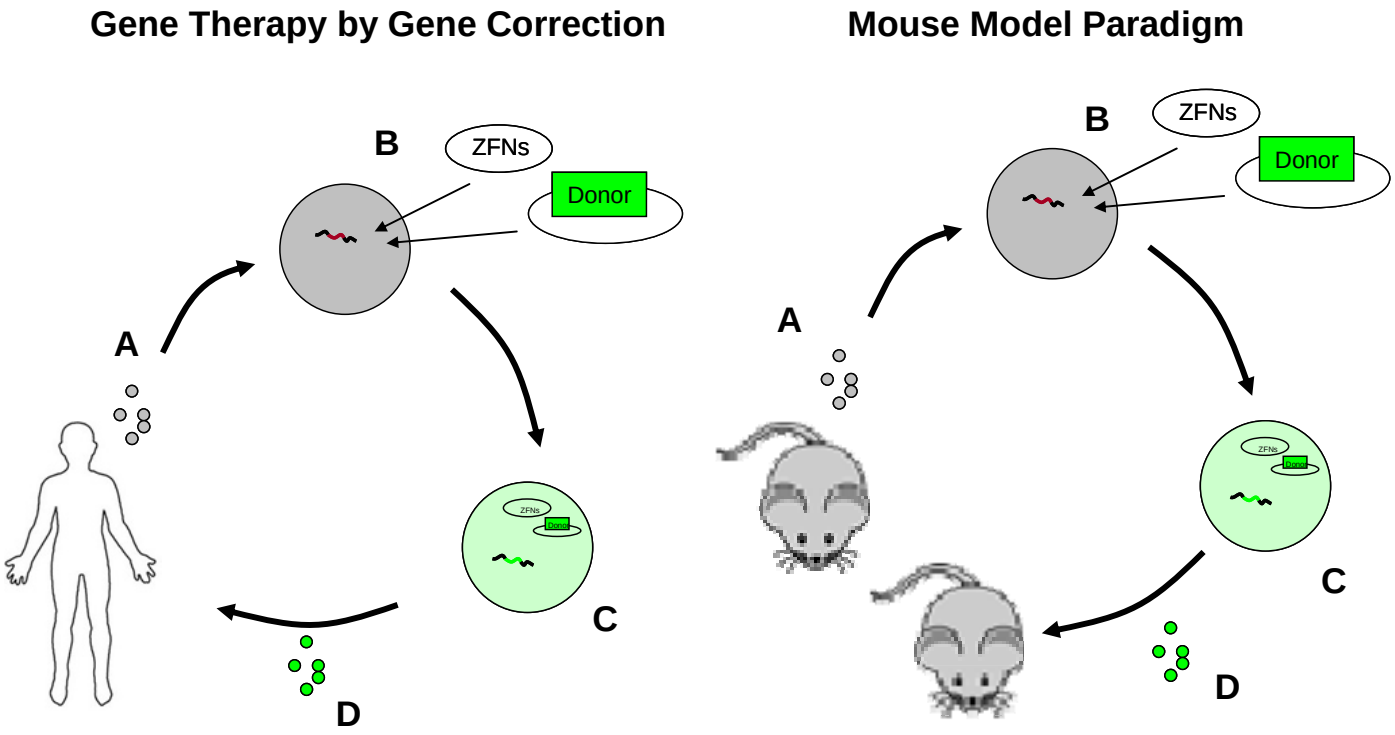

Figure 2.2: Gene Therapy using Gene Correction. Gene therapy by gene correction is performed by (A) first isolating the desired cell type from the patient/ mouse. (B) Next ZFNs designed to the target site and a donor plasmid are transfected into the isolated cells. (C) The ZFNs are expressed, cleave the target site, and the cell uses the donor plasmid as a template for homologous recombination. (D) Finally, the targeted cells are transplanted back into the patient/ recipient mouse

CHAPTER III:
HIGH RELATIVE RATES OF GENE TARGETING RESULT FROM LOW RATES OF END-JOINING IN CHICKEN DT40 CELLS

Abstract

Gene targeting is a method to introduce precise changes into the genome of a cell. The rate of gene targeting is determined by the balance between targeted events and random integration events. DT40 cells have a high relative rate of gene targeting however the reason for this high rate is unknown. We find that this high gene targeting rate is the result of a low random integration rate. Furthermore, we show that compared to other cell lines, DT40 cells are deficient in end-joining of compatible ends but markedly deficient in joining of non-compatible ends. This deficiency in non-compatible end-joining explains the decrease in random integration and the consequent increase in the relative rate of gene targeting. These results also suggest that determining methods to lower the frequency of end-joining may be helpful in achieving high rates of gene targeting in other cell types.

Introduction

Gene targeting is the replacement of an endogenous segment of genomic DNA with a different but homologous segment of exogenous DNA by homologous recombination. It is the most precise way to manipulate the genome of a cell and has the flexibility to create both subtle and dramatic genomic change. Because of these reasons, gene targeting could be an ideal form of gene correction therapy for human monogenic diseases. While widely used experimentally in certain cell types such as yeast, murine embryonic stem cells, and the chicken DT40 cell line, the rate of gene targeting in most mammalian somatic cells is too low to be of experimental use (Sedivy and Sharp 1989; Porteus and Baltimore 2003). Therefore, finding ways to increase the rate of gene targeting could have important experimental and therapeutic applications.

The relative rate of gene targeting is a ratio between the rate that the introduced fragment integrates homologously and the rate that the fragment integrates randomly (a form of "illegitimate recombination"). That is, the relative rate of gene targeting is the ratio of the absolute rate of gene targeting compared to the absolute rate of random integration. One way of increasing the relative rate of gene targeting would be to increase the absolute rate of gene targeting. Work in the last decade has shown that the creation of a DNA double-strand break can increase the absolute rate of gene targeting by 4-5 orders of magnitude (Rouet, Smih et al. 1994; Choulika, Perrin et al. 1995; Brenneman, Gimble et al. 1996; Donoho, Jasin et al. 1998; Porteus and Baltimore 2003). Zinc finger nucleases have become a promising approach to create site specific DNA double-strand breaks (Bibikova, Golic et al. 2002; Bibikova, Beumer et al. 2003; Porteus

and Baltimore 2003; Urnov, Miller et al. 2005; Beumer, Bhattacharyya et al. 2006; Morton, Davis et al. 2006; Carroll 2008; Doyon, McCammon et al. 2008; Meng, Noyes et al. 2008; Geurts, Cost et al. 2009; Shukla, Doyon et al. 2009; Townsend, Wright et al. 2009). However the generation of a double-strand break (DSB) does not specifically stimulate gene targeting. Eukaryotic cells have redundant mechanisms to repair DNA double-strand breaks (Paques and Haber 1999; West, Chappell et al. 2000; van Gent, Hoeijmakers et al. 2001). The two major pathways of DNA double-strand break repair are non-homologous end-joining (NHEJ) and homologous recombination. In classical NHEJ, the two free ends of the DSB are processed and ligated back together using a complex of genes that includes Ku70, Ku80, DNA-PK_{cs}, XRCC4 and Lig4 among others. In homologous recombination, the DSB is processed to generate free 3' ends, the 3' ends then invade a homologous DNA substrate and branch migration occurs through the simultaneous action of DNA polymerases and helicases to generate a copy of the undamaged template. Finally, the homologous recombination complex and accompanying Holliday junction are resolved, almost always in a non-crossover fashion, to finish repair of the DSB. Gene targeting occurs when homologous recombination repairs the DSB using the extra-chromosomal fragment of DNA as the repair substrate. In this process genetic differences present in the substrate can be introduced into the genomic target thereby creating a targeting event.

In 1991, Buerstedde and Takeda noted that the avian leukosis transformed chicken B-cell DT40 cell line had a thousand-fold increase in the relative rate of gene targeting (Buerstedde and Takeda 1991). It remains unclear whether the high rate of gene targeting in DT40 cells is the result of an increase in the absolute rate of gene targeting, a

decrease in the absolute rate of random integration or some combination of both. Sonoda et al. (1999) showed that the rate of stable integrants in wild-type DT40 cells was $6x10^{-5}$ (Sonoda, Sasaki et al. 1999). But they did not report the initial transfection efficiency and so the random integration rate could not be determined. For instance, if the transfection efficiency was 10^{-2} (or 1%), an efficiency not uncommon in difficult to transfect cell lines, then the random integration rate in DT40 cells would be 0.6% a rate not significantly different than many other cell lines. Consequently, they made no conclusion about the rate of random integration in DT40 cells (Sonoda, Sasaki et al. 1999).

In this work we studied the mechanism for the high relative rate of gene targeting in DT40 cells using quantitative assays for homologous recombination, NHEJ, and random integration. We find that DT40 cells do not have a markedly increased absolute rate of homologous recombination or gene targeting, but instead are intrinsically deficient in joining non-compatible overhangs which leads to a markedly decreased rate of random integration and a consequent increase in the relative rate of gene targeting.

Results

DT40 Cells Have a Higher Absolute Rate of Gene Targeting and Homology Directed Repair than 293 Cells

We used two previously described assays that both measure the absolute rate of homologous recombination (schematized in Figure 3.1A) (Pierce, Johnson et al. 1999; Porteus and Baltimore 2003). In the GFP gene targeting system, cell lines are created

with a single copy of a mutated GFP gene that contains a recognition site for the I-SceI (Sce) endonuclease. The rate of spontaneous gene targeting (S-GT) is measured by the transfection of a plasmid ("repair plasmid") that can repair the GFP mutation, counting the number of GFP positive cells by flow cytometry and normalizing the number of GFP positive cells to the transfection efficiency. In 293 and DT40 cells, the rate of double-strand break mediated gene targeting (DSB-GT) is measured by transfecting a single plasmid ("Sce/repair plasmid") that contains both an expression cassette for Sce and the GFP repair fragment. Next, as described above, GFP positive cells are quantified by flow cytometry and normalized to transfection efficiency to determine the gene targeting rate. In 293 cells, the rate of spontaneous gene targeting was 7.1×10^{-7} (Table 3.1), and the rate of double-strand break mediated gene targeting was 3.5×10^{-4} (Figure 3.1B and Table 3.1). In contrast, the rates of spontaneous and DSB-mediated gene targeting in DT40 cells were 3×10^{-6} and 1.2×10^{-3} respectively; rates that are 3-5 fold higher than in 293 cells (Figure 3.1B and Table 3.1). In comparison, the rate of spontaneous gene targeting in ES cells has been measured at $1.1 - 1.4 \times 10^{-6}$ (Doetschman, Gregg et al. 1987; Hasty, Rivera-Perez et al. 1991). To measure DSB-mediated gene targeting in ES cells, we used zinc finger nucleases that cleave an endogenous site within GFP (Pruett-Miller, Connelly et al. 2008). The rate of DSB-mediated gene targeting using these zinc finger nucleases was 2×10^{-3}. These rates are ~1.4 - 6 fold higher than those found in 293 cells (Table 3.1).

The second assay we used to measure homologous recombination was the DRGFP assay developed and described by Pierce et al. (1999). In this assay, a DSB is created in a mutated GFP chromosomal gene target by Sce just as in the DSB-mediated

gene targeting assay. In the DRGFP assay the repair GFP fragment is provided as a linked allele rather than as an extra-chromosomal piece of DNA (Figure 3.2A). The rate of homology directed repair (HDR) is determined by normalizing the percentage of GFP positive cells to the transfection efficiency. We found that DT40 cells had a 1.6-fold higher rate of homology directed repair than 293 cells (Figure 3.2B). This increase in HDR is consistent with the increase in homologous recombination found using the gene targeting assay for DT40 cells. These experiments demonstrate that DT40 cells have a slightly increased absolute rate of homologous recombination but not nearly high enough to account for the dramatically increased relative rate of gene targeting.

The Rate of Random Integration is Significantly Decreased in DT40 cells Compared to 293 Cells

We developed a GFP based assay to measure the rate of random integration (Figure 3.3A). In this assay, we transfect cells with a GFP expression cassette and serially measure the number of GFP positive cells by flow cytometry after transfection. We find that at day 1-2 after electroporation, the percentage of GFP positive cells is maximal and then declines thereafter (Figure 3.3B). At day 10 the percentage of GFP positive cells reaches a plateau and remains essentially stable for at least two weeks (Figure 3.3B and data not shown). We found that the initial transfection efficiencies in 293 cells (10-20%), DT40 cells (10-15%), and ES cells (20%) were similar suggesting that transfected DNA could reach the nucleus and be expressed at similar rates in both cell types (data not shown). In addition, we found similar kinetics in the decline of GFP positive cells in 293, DT40 cells and ES cells (Figure 3.3B). The rate of random integration was calculated by dividing the percentage of GFP positive cells at plateau by

the maximal percentage of GFP positive cells (Figure 3.3B). In 293 cells we found that the rate of random integration for a supercoiled plasmid was approximately 2 x 10^{-2} (2%) (Table 3.1) and approximately 1.5-2 fold higher for a linearized plasmid (data not shown). We found that the rate of random integration in 293 cells did not significantly vary based on the method of transfection and that the rate of random integration was the same in 293 cells and murine 3T3 cells (data not shown). In contrast to 293 cells, we found that the rate of random integration in DT40 cells was 2.1 x 10^{-5} (0.002%), or 1000-fold lower than the rate in 293 cells (Figure 3.3B and Table 3.1). In comparison, the rate of random integration in ES cells is 1.1 x 10^{-3} (0.11%) or ~ 20 fold lower than that of 293 and 55 fold higher than DT40 cells (Figure 3.3B and Table 3.I). These data suggest that the high relative rate of gene targeting in DT40 cells can be explained by a decrease in the rate of random integration.

Role of Ku70 and Rad54 in Gene Targeting and Random Integration in DT40 Cells

We next examined the rates of gene targeting, homology directed repair, and random integration in DT40 cell lines in which Ku70, Rad52, or Rad54 had been knocked out. Ku70 is part of the end-binding complex central to repair of DSBs by NHEJ while Rad52 and Rad54 are genes that are involved in the homologous recombinational repair of DSBs. The phenotypes of these cell lines have been previously described (Bezzubova, Silbergleit et al. 1997; Takata, Sasaki et al. 1998).

Rad54$^{-/-}$ DT40 cells have increased sensitivity to ionizing radiation and are 30-fold decreased in the relative rate of spontaneous gene targeting (Bezzubova, Silbergleit et al. 1997). In contrast, Rad52$^{-/-}$ DT40 cells do not have increased sensitivity to ionizing radiation and are only 2-4 fold decreased in the relative rate of spontaneous gene

targeting (Yamaguchi-Iwai, Sonoda et al. 1998). We found that the rate of DSB-mediated gene targeting and homologous recombination as measured in the DRGFP assay were decreased by 3-5-fold in the Rad54 knockout line (Figure 3.4). Interestingly, this decrease is not as significant as the decrease found in the relative rate of spontaneous gene targeting observed by Bezzubova et al. (1997) when they initially characterized these cells. In these experiments we created a DSB to initiate the gene targeting process while their work measured the spontaneous rate. It may be that by creating a break with an artificial endonuclease, some important role of Rad54 in spontaneous gene targeting is bypassed. In the Rad52$^{-/-}$ cell lines we found no decrease in DSB-mediated gene targeting and less than a two-fold decrease in homology directed repair (Figure 3.4). The minor effect of Rad52 mutation on the rate of homologous recombination is consistent with the reduced role of Rad52 in homologous recombination in vertebrate cells as compared to its central importance in homologous recombination in yeast (Rijkers, Van Den Ouweland et al. 1998; Yamaguchi-Iwai, Sonoda et al. 1998; Paques and Haber 1999). We found no significant difference in the rate of random integration in Rad52$^{-/-}$ cells compared to wild-type cells (Figure 3.4C).

Ku70$^{-/-}$ DT40 cells showed normal cell cycling properties, showed no increase in spontaneous chromosomal rearrangements, but had increased sensitivity to ionizing radiation, particularly during the G1-early S phase of the cell cycle (Takata, Sasaki et al. 1998). Ku70$^{-/-}$ DT40 cells had a five-fold increase in DSB-mediated gene targeting and a two-fold increase in homology directed repair (Figure 3.4A, B). This increase is consistent with the findings of Pierce et al. (2001) who also found an increase in homologous recombination in Ku70 knockout hamster cells (Pierce, Hu et al. 2001).

Even though wild-type DT40 cells have an already low rate of random integration, the elimination of Ku70 further decreases the rate by 5-fold but does not eliminate random integration entirely (Figure 3.4C). Thus, there seems to be Ku70-dependent and Ku70-indpendent pathways for random integration in DT40 cells.

Description of End-Joining Assay

We developed a GFP based assay for end-joining to explore the potential mechanistic similarity between random integration and NHEJ (Figure 3.5A). In this assay we made constructs in which a strong promoter drives a GFP gene flanked by Sce sites. Downstream from the GFP gene is a CD8α gene (CD8) or a red fluorescent protein gene (RFP). For clarity, this portion of the end-joining cassette will be referred to as CD8/RFP. In cell lines that contain a single copy of these constructs the cells are GFP positive but CD8/RFP negative because there is no internal ribosomal entry site (IRES) to initiate cap-independent translation of the CD8/RFP gene. If Sce cuts both sites simultaneously, however, the GFP gene will drop out and if end-joining occurs then the CD8/RFP gene will be joined to the strong promoter and the cells will become GFP-CD8/RFP+ (Figure 3.5A, B). We call this assay an end-joining assay because it measures a product that is created by a small intra-chromosomal deletion followed by joining of the 5' and 3' ends. In the compatible ("C") orientation, the Sce sites are oriented as direct repeats such that if both are cut simultaneously the remaining two 3' four basepair overhang ends created by Sce can overlap and anneal (Figure 3.5A). In the non-compatible ("NC") orientation the Sce sites are oriented as inverted repeats and if

both sites are cut simultaneously two 3'overhang non-compatible ends will have to be joined (Figure 3.5A).

End-Joining in 293 Cells

293 cells are able to join both compatible and non-compatible ends efficiently and approximately equally (Figure 3.5C, D). To determine the precision and type of end joining, we sequenced the junctions. The junction sequences with the compatible ends showed, as expected, precise joining of the compatible 3' overhangs without insertion or deletion of nucleotides most often (Figure 3.6, Figure 3.7). In non-compatible joining, 293 cells used a two nucleotide microhomology based repair to join the ends about half of the time (8/14) of the time (Figure 3.6). In the rest, the junctions showed evidence of other repair mechanisms involving small insertions or deletions without obvious evidence of the use of microhomology (Figure 3.7).

End-Joining in ES cells

ES cells were less efficient at end-joining than 293 cells. For compatible ends, ES cells were 3-fold deficient and showed a higher rate of incorrect end-joining (~40%) (Figure 3.5C, D, Figure 3.6, Figure 3.7). For non-compatible ends, ES cells demonstrated a ~3 fold decrease compared to 293 cells. These cells also demonstrated the use of small two nucleotide microhomologies in the formation of junctions >50% of the time (Figure 3.7).

End-Joining in DT40 Cells

DT40 cells were less efficient at end-joining than 293 cells. In the case of compatible ends, DT40 cells were approximately 10-fold less efficient than 293 cells in joining the ends (Figure 3.5C, D). Like 293 cells, the compatible junctions in wild-type

DT40 cells were precise with no evidence of insertions or deletions (Figure 3.6, Figure 3.7). With non-compatible ends, DT40 cells were over 100-fold less efficient than 293 cells in joining ends (Figure 3.5B, D). Wild-type DT40 cells, like 293 cells, used microhomology based end-joining about 50% of the time (Figure 3.6, Figure 3.7). There was a subtle difference with 293 cells, however, as DT40 cells made a higher percentage of junctions using microhomology joint #1 rather than microhomology joints #2 or #3. To make joint #2, for example, internal homology must be found and joining requires the creation and subsequent trimming of single nucleotide tails. Microhomology joint #1, however, can be created by aligning the two most distal nucleotides of the overhang followed by filling in of the two nucleotide gap one each strand (Figure 3.6). Thus, microhomology joint #2 seems to require a more complex end-joining mechanism than microhomology joint #1.

As expected $Ku70^{-/-}$ DT40 cells were even less proficient at joining ends than wild-type DT40 cells (Figure 3.5C, D). The joining of compatible ends seemed to have a greater Ku70 dependence as the rate of joining compatible ends fell 10-fold with the loss of Ku70 while the rate only fell approximately 1.5-fold for non-compatible ends. In the absence of Ku70 the rate of joining non-compatible and compatible ends was essentially equal suggesting that without Ku70, compatible and non-compatible ends in DT40 cells were joined using the same alternative end-joining pathway. While the joining of non-compatible ends in $Ku70^{-/-}$ cells was quantitatively decreased by 1.5-fold, the junction sequences for the non-compatible ends were not discernibly different than in wild-type cells (Figure 3.7). In contrast, not only were $Ku70^{-/-}$ cells quantitatively decreased in joining compatible ends but they also showed a striking difference in the accuracy of

joining as over half (4/7) of the junctions contained small deletions or insertions instead of being precise (Figure 3.6, Figure 3.7). This data clearly demonstrates the requirement of Ku70 for precise end-joining repair of compatible ends from DNA double-strand breaks.

Inhibition of DNA Repair Changes the Rates of Random Integration, Gene Targeting and End-Joining

We next tested whether inhibiting the ATM/ATR and DNA-PK_{cs} repair protein kinases would alter the rates of gene targeting, random integration, and end joining. NU7026 (Sigma Aldrich, St. Louis MO) has been shown to inhibit DNA-PK_{cs}, a protein involved in NHEJ (Willmore, de Caux et al. 2004). When treated with NU7026, 293 cells did not display a change in gene targeting (Figure 3.1C), however a 1.4-fold increase in random integration was observed (from 2% without drug to ~3% with drug) (Figure 3C), and a ~2 fold decrease in end-joining of non-compatible and compatible ends compared to untreated control cells (Figure 3.5E). Next cells were treated with the ATM inhibitor KU-55933 (KuDOS Pharmaceuticals Ltd, Cambridge, UK) at 10 μM, a concentration that inhibits ATM (Hickson, Zhao et al. 2004). Under these conditions, we did not observe a significant change in gene targeting (Figure 3.1C), however, the rate of random integration increased to 5% (2.4 fold) (Figure 3.3C), and a there was a small decrease (1.6 fold) in end-joining of non-compatible ends (Figure 3.5E). Finally, cells were treated with caffeine, a drug shown to inhibit ATM and ATR (Sarkaria, Busby et al. 1999). Cells treated with 4 mM caffeine exhibited a 3-fold decrease in gene targeting (Figure 3.1C), the random integration rate increased to ~8% (3.5-fold higher than 293 cells) (Figure 3.3C), and there was a 4 and 22 fold decrease in joining of non-compatible

and compatible ends, respectively (Figure 3.5E). These data suggest that inhibiting DNA-PK_{cs}, ATM, and ATM/ATR results in increased rates of random integration and decreased rates of end-joining. Sequencing of compatible-end junctions formed under drug treatment revealed a modest increase in the amount of imprecise joining (Figure 3.8). Inhibiting DNA-PK_{cs} or ATM specifically either did not change or slightly increased the rate of homologous recombination, while inhibiting ATM/ATR with caffeine substantially increased the rate of random integration and decreased the rate of DSB gene targeting and end-joining of compatible and non-compatible ends. These results suggest that random integration and end-joining are not necessarily dependent on one another.

Figures

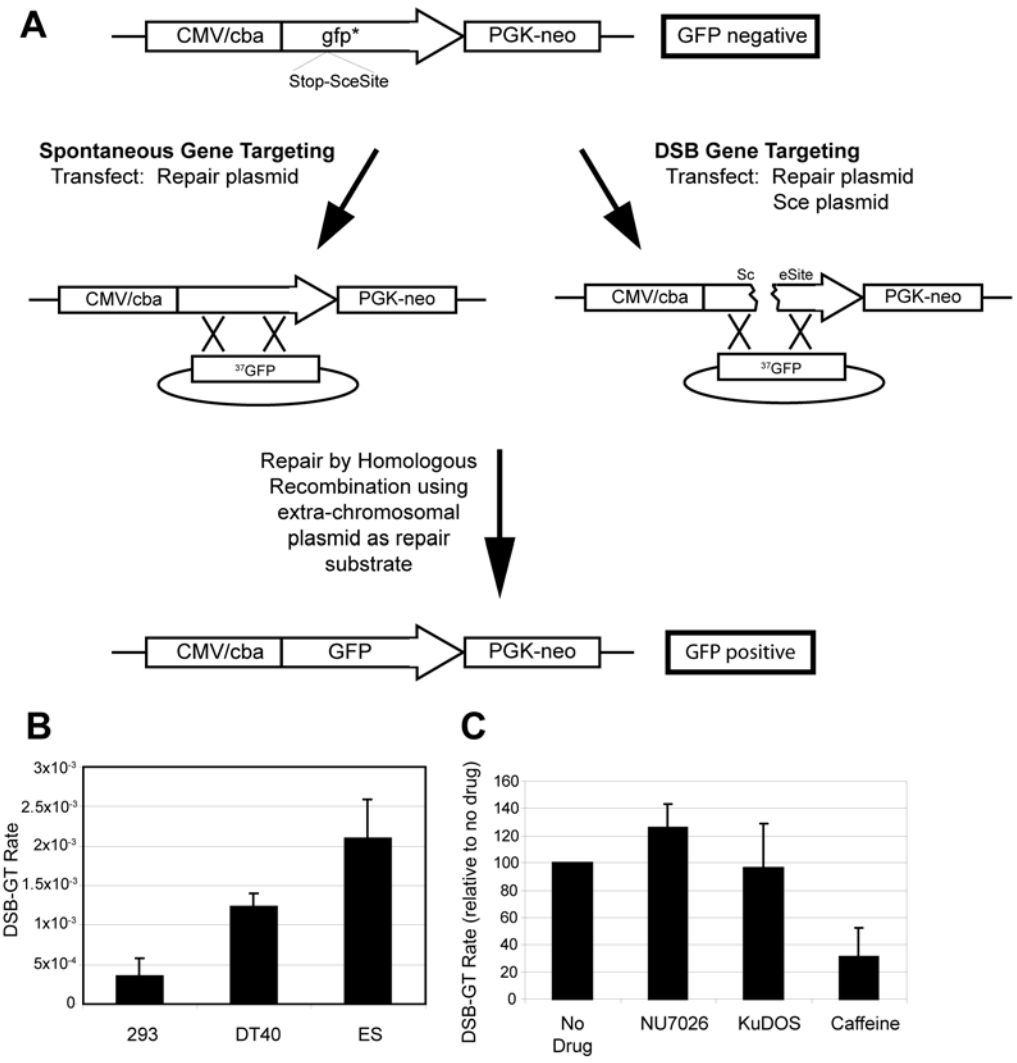

Figure 3.1: Schematic of Assay for Gene Targeting

A) Schematic for GFP Gene Targeting Assay. In this assay homologous recombination is measured by the conversion of a mutated single-copy chromosomally integrated GFP target gene by a homologous transfected plasmid. The details of this assay have been

previously described (Porteus and Baltimore 2003). In spontaneous gene targeting (S-GT), the rate of conversion is measured when a DSB is not induced by the I-SceI (Sce) endonuclease. In double-strand break induced gene targeting (DSB-GT), the rate of conversion is measured after the induction of a DSB by Sce or zinc finger nucleases designed to cleave at an endogenous site present in GFP. B) Rates of DSB-GT in 293, DT40 cells, and ES cells. C) Inhibition of DNA repair using the drugs NU7026 [10 μM], KU-55933 [10 μM], and Caffeine [4 mM].

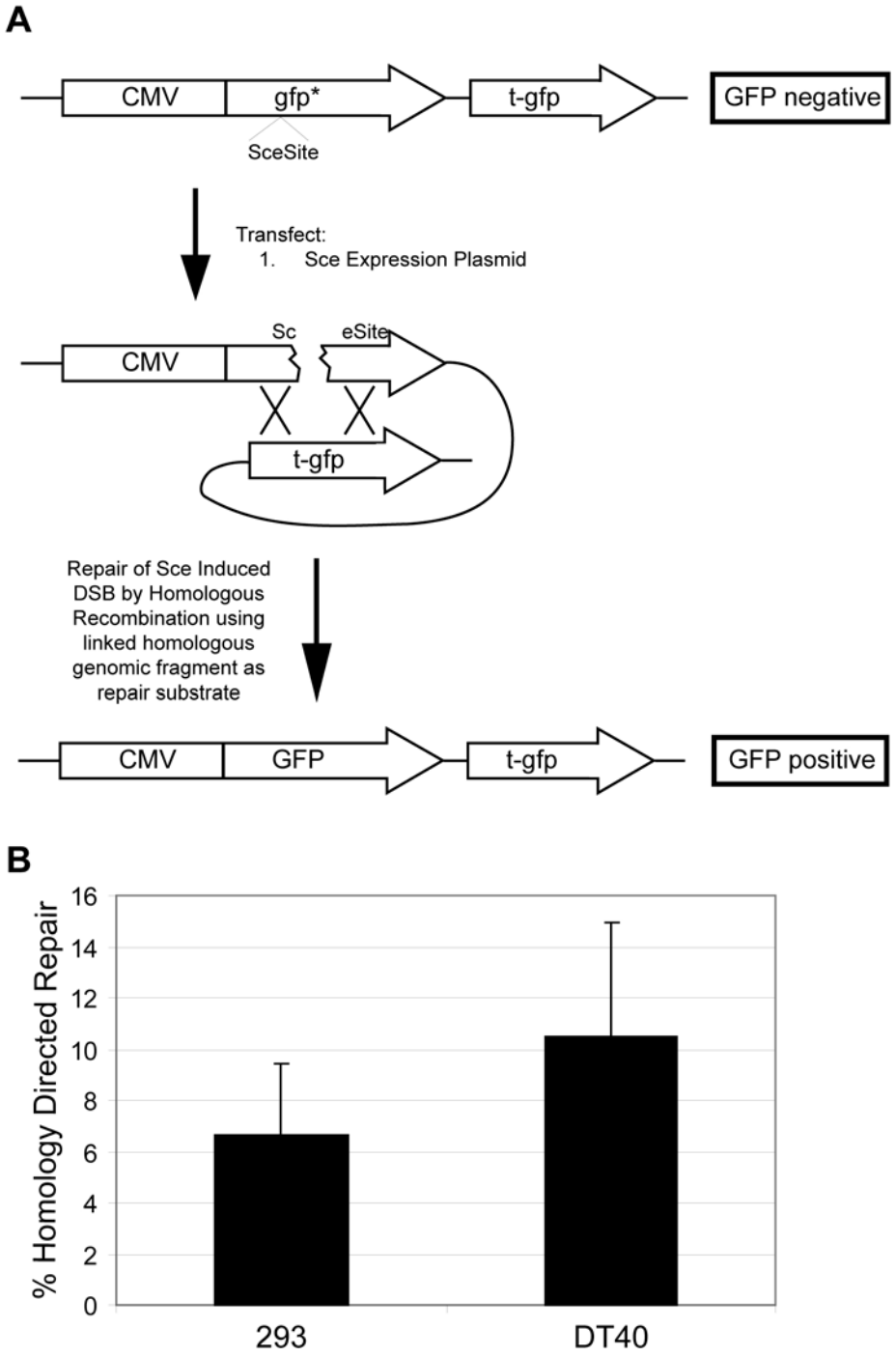

Figure 3.2: Schematic of Assay for DRGFP

A) Schematic of DRGFP Assay. In this assay homologous recombination is measure by the conversion of a mutated single-copy chromosomally integrated GFP target gene from a linked truncated GFP fragment. The details of this assay have been previously described (Pierce, Johnson et al. 1999). Abbreviations: GFP: green fluorescent protein gene; Stop: in-frame stop codon; Sce Site: Recognition site for the I-SceI endonuclease; gfp*: mutated green fluorescent protein gene; t-GFP: plasmid containing fragment of GFP that begins at nucleotide 37 of the coding region, so truncated at 5' end; CMV: Cytomegalovirus promoter/enhancer.

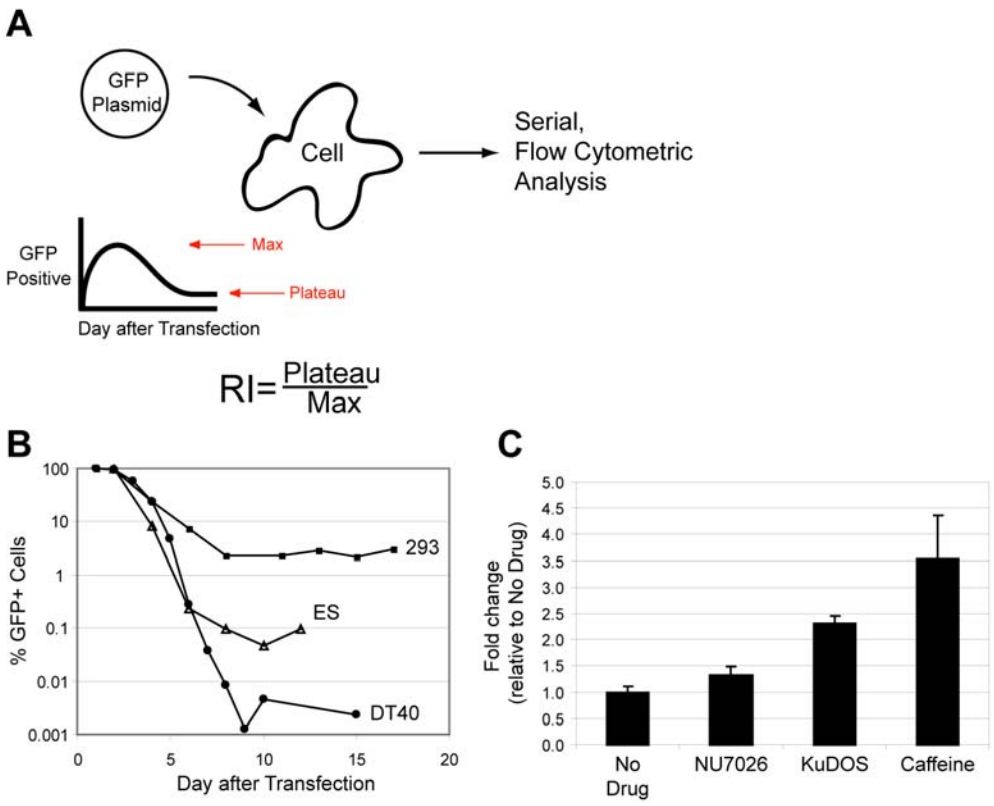

Figure 3.3: Random Integration Assay

A) Schematic of Random Integration Assay. A GFP expression plasmid is transfected into cells and the cells are analyzed by flow cytometry to determine the fraction that is GFP positive. The Random Integration Rate (RI) is determined by dividing the fraction of GFP positive cells at "Plateau" by the fraction positive at "Max." B) Random Integration in 293, DT40 cells and ES cells. Along the Y-axis is the percent GFP positive cells on a logarithmic scale normalized to the day two transfection efficiency. On the X-axis is the day after transfection. The solid square line is the data for 293 cells after transfection of a supercoiled plasmid, the open triangle line represent random integration in ES cells, and the sold circle line is the data obtained for DT40 cells after transfection

of a supercoiled plasmid. The shape and time course of the percentage of GFP positive cells is similar between 293, DT40 and ES cells.

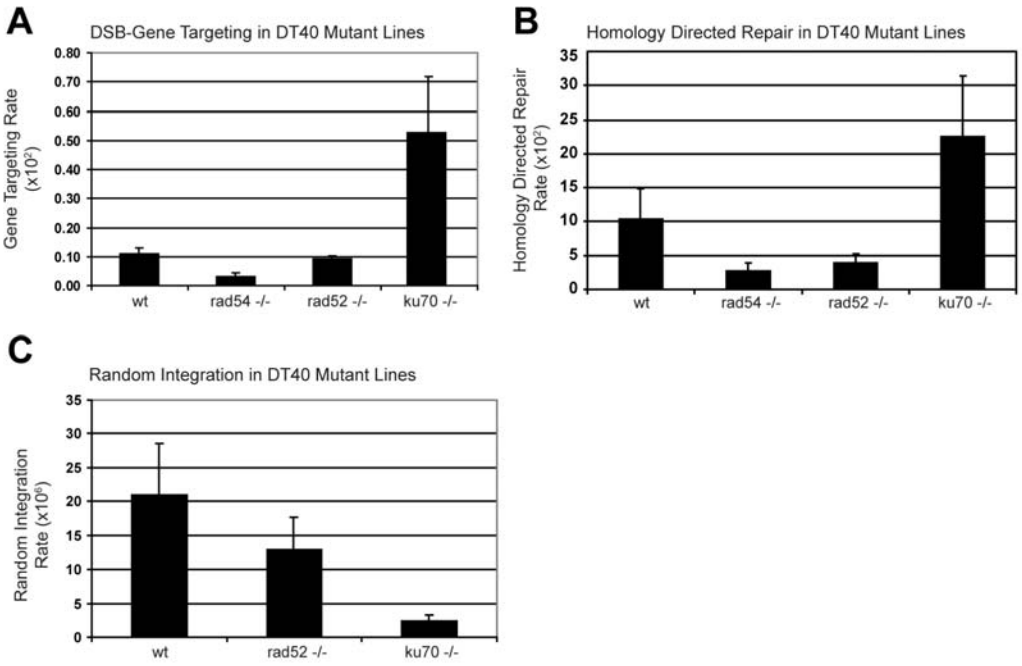

Figure 3.4: Effect of mutations in Rad52, Rad54, and Ku70 on Homologous Recombination and Random Integration in DT40 Cells.

A) Rate of gene targeting in DT40 mutant lines after induction of a double-strand break. B) Rate of Homology Directed Repair in DT40 mutant lines. C) Random Integration rate in DT40 mutant lines.

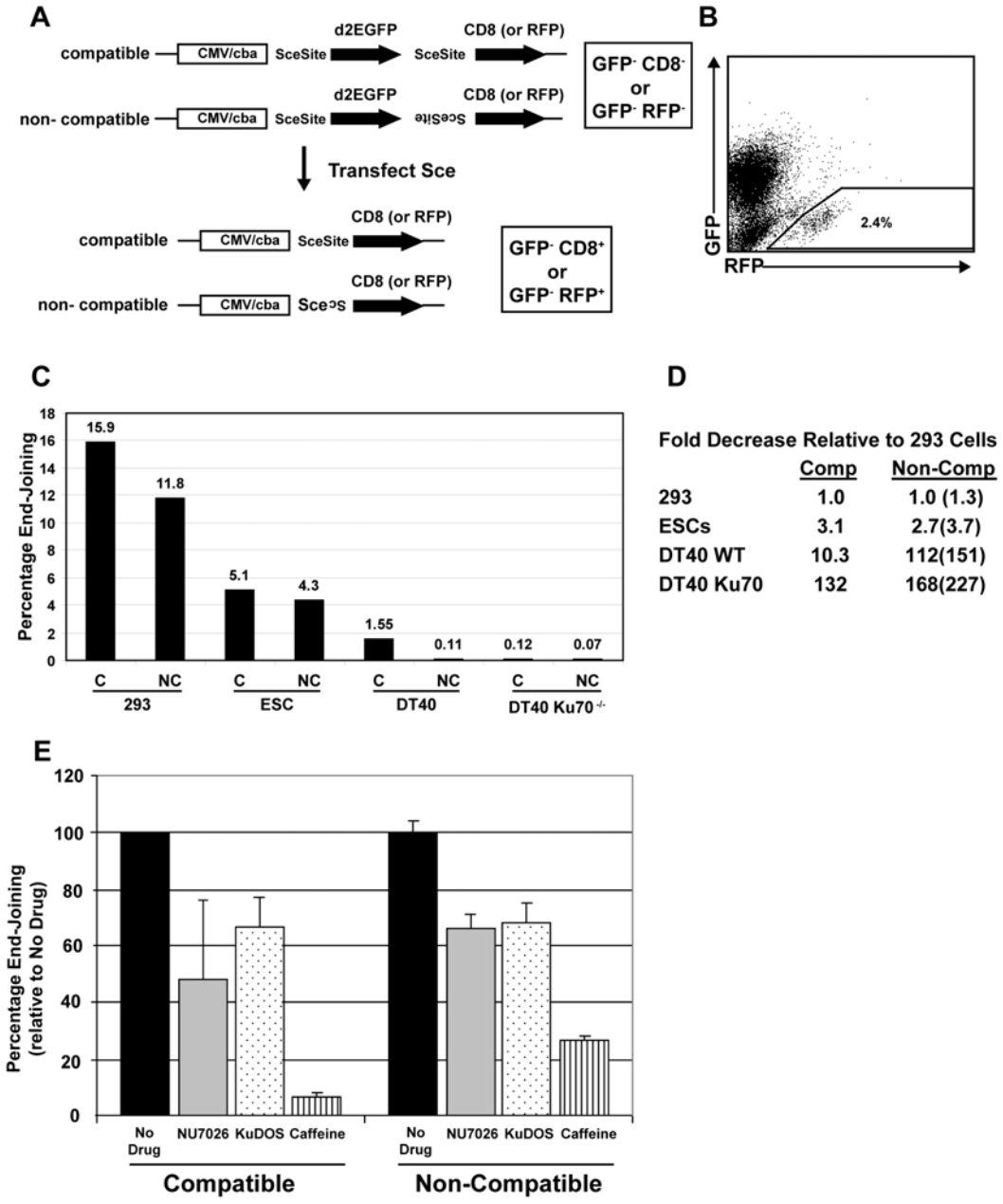

Figure 3.5: Measurement of End-Joining Using the End-Joining Assay.

A) Schema of end-joining assay. Cell lines with the substrates prior to transfection are GFP+ CD8/RFP-. If end-joining occurs, the cells will become GFP- CD8/RFP+. The

rate of end-joining is determined by normalizing the number of GFP- CD8/RFP+ cells to the transfection efficiency. In the non-compatible substrate, the recognition sites for Sce ("Sce site") are inversely oriented (as depicted by one site being upside down and backwards). B) Example of end-joining measured by flow cytometry. Along the Y-axis is GFP fluorescence, and the along the X-axis is RFP fluorescence. Cells that have undergone end-joining express either CD8 which can be stained with a CD8 antibody conjugated to phycoerytherin or RFP which can be measured using flow cytometry. C) Rate of End-Joining in 293, DT40, and ES cells. The Y-axis is the percentage of end-joining normalized to the transfection efficiency. Along the X-axis are the different cell types and end-joining substrates examined. C=compatible end substrate. NC=non-compatible end substrate. D) The table shows the relative rate of end-joining compared to 293 cells derived from the graphical data shown in C). C = compatible end substrate. NC = non-compatible end substrate. The number in parentheses in the non-comp column is the fold decrease relative to the rate of end-joining in 293 cells using the compatible end substrate. E) Effects of drugs on end-joining in 293 cells. The Y-axis is the percentage of end-joining relative to the no drug control. Along the X-axis is the drug used in the experiment. Abbreviations: As in Figure 3.1. d2EGFP: destabilized Green Fluorescent Protein gene. CD8: CD8α gene. RFP: RFP gene. NU7026 [10 μM]. KuDOS: KU-55933 [10 μM]. Caffeine [4 mM].

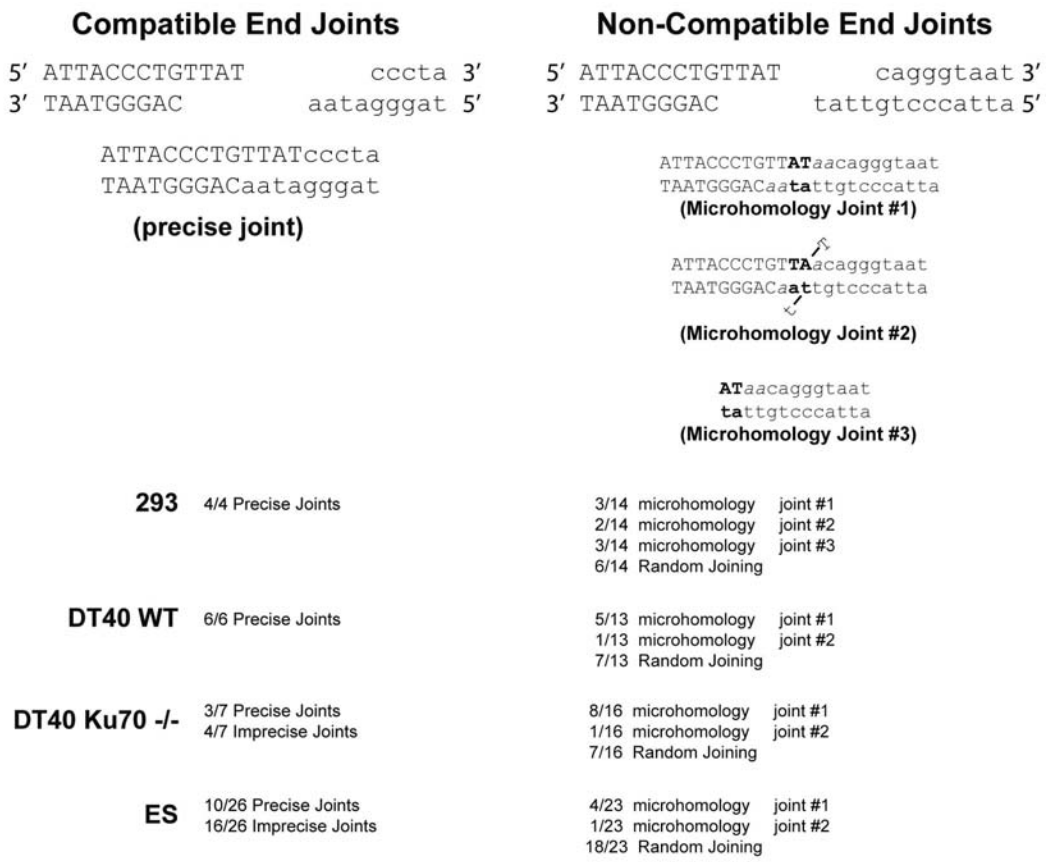

Figure 3.6: Summary of Sequences Found at End-Joining Junctions.

In the left column is a depiction of the ends that would be joined after simultaneous cleavage of both Sce sites in the compatible end construct and the resulting precise joint that would be formed if there were non-mutagenic end-joining. In the right column is a depiction of the ends that would be joined after simultaneous cleavage of both Sce sites in the non-compatible end construct. The sequence coming from the 5' site is in uppercase and the sequence coming from the 3' site is in lowercase. Below that are shown the three sequence joints that were found using identifiable microhomology (labeled microhomology joints #1, #2, and #3). The nucleotides in bold highlight the two

basepair microhomology that was used in end-joining. The italicized nucleotides highlight the templated nucleotides created by a DNA polymerase filling small gaps in the process of joint formation. Below the three joints are the types of joints found in each of the cell types tested.

A

Compatible Junctions

293 cells

	Event	Total
GGG CCC GGG ATC CAT TAC CCT GTT AT C CCT A GC GGC CGC TCT AGA A	Perfect	36 (~67%)
GGG CCC GGG ATC CAT TAC CCT GTT AT A GC GGC CGC TCT AGA A	Deletion	4 (~7%)
GGG CCC GGG ATC CAT TAC CCT GTT AT GA A	Deletion	1 (~2%)
GGG CCC GGG ATC CAT TAC CCT GTT AT AGTGGATCCCC	Del 9bp 3' NotI	1 (~2%)
GGG CCC GGG ATC CAT TAC CCT GTT A CC CCT A GC GGC CGC TCT AGA A	Insert/Deletion	1 (~2%)
GGG CCC GGG ATC CAT TAC CCT GTT A GC GGC CGC TCT AGA A	Deletion	1 (~2%)
GGG CCC GGG ATC CAT TAC CCT GTT TATC CCT A GC GGC CGC TCT AGA A	Insertion/Deletion	1 (~2%)
GGG CCC GGG ATC CAT TAC CCT A GC GGC CGC TCT AGA A	Deletion	1 (~2%)
GGG CCC GGG ATC CAT TA TC CCT A GC GGC CGC TCT AGA A	Insert/Deletion	3 (~6%)
GGG CCC GGG ATC CAT TA CCCCCGGGCTG	Del 16bp 3' Not1	1 (~2%)
GGG CCC GGG ATC CAT C CCT A GC GGC CGC TCT AGA A	Deletion	1 (~2%)
GGG CCC GGG ATC CAT TAC CCTACCCTGTTATATCCCT A GC GGC CGC TCT AGA A	Insert/copy	1 (~2%)
GGG CCC GGG ATC C GA A	Deletion	1 (~2%)
AGA TGC GCG GCC GAT TAC CCT GTT AT C CCT A GC GGC CGC TCT AGA A	5' Rearrangement	1 (~2%)
		Total [54]

DT40 Wt

GGG CCC GGG ATC CAT TAC CCT GTT AT C CCT A GC GGC CGC GC ATG GCC TTA	Perfect	15 (100%)
		Total [15]

DT40-77

GGG CCC GGG ATC CAT TAC CCT GTT AT C CCT A GC GGC CGC GC ATG GCC TTA	Perfect	3 (~25%)
GGG CCC GGG ATC CAT TAC CCT GTT AT TA C CCT A GC GGC CGC GC ATG GCC TTA	Insert	1 (~8%)
GGG CCC GGG ATC CAT TAC CCT GTT AT AGGGTATA T A GC GGC CGC GC ATG GCC TTA	Insert/Deletion	1 (~8%)
GGG CCC GGG ATC CAT TAC CCT GTT AT TTTTGGCCAAAGTGAA ATG GCC TTA	Insert/Deletion	1 (~8%)
GGG CCC GGG ATC CAT TAC CC CT A GC GGC CGC GC ATG GCC TTA	Deletion	1 (~8%)
GGG CCC GGG ATC CAT TAC C CCT A GC GGC CGC GC ATG GCC TTA	Deletion	1 (~8%)
GGG CCC GGG ATC CAT TA T C CCT A GC GGC CGC GC ATG GCC TTA	Insert/Deletion	1 (~8%)
GGG CCC GGG ATC CAT T GC GGC CGC GC ATG GCC TTA	Deletion	1 (~8%)
GGG CCC GGG ATC CCCCTGTCATCCCTGTTAT C CCT A GC GGC CGC GC ATG GCC TTA	Rearrangement	1 (~8%)
GGG CCC GGG ATC CAT TAC CCT GTT AT CCCTGTTAT C CCT A GC GGC CGC GC ATG GCC TTA	Rearrangement	1 (~8%)
		Total [12]

ES CELLS

GGG CCC GGG ATC CAT TAC CCT GTT AT C CCT A GC GGC CGC TCT AGA A	Perfect	10 (~38%)
GGG CCC GGG ATC CAT TAC CCT GTT AT A GC GGC CGC TCT AGA A	Deletion	3 (~12%)
GGG CCC GGG ATC CAT TAC CCT GTT AT AAC AGG GC GGC CGC TCT AGA A	Insert/deletion	1 (~4%)
GGG CCC GGG ATC CAT TAC CCT GTT TATC CCT A GC GGC CGC TCT AGA A	Insert/deletion	1 (~4%)
GGG CCC GGG ATC CAT TAC CCT GT CCT A GC GGC CGC TCT AGA A	Deletion	1 (~4%)
GGG CCC GGG ATC CAT TAC CCT GT A GC GGC CGC TCT AGA A	Deletion	1 (~4%)
GGG CCC GGG ATC CAT TAT CC CT A GC GGC CGC TCT AGA A	Deletion	1 (~4%)
GGG CCC GGG ATC CAT TA T C CCT A GC GGC CGC TCT AGA A	Insert/deletion	3 (~12%)
GGG CCC GGG ATC CA C TCT AGA A	Deletion	1 (~4%)
GGG CCC GGG AT T A GC GGC CGC TCT AGA A	Deletion	1 (~4%)
GGG CC A A	Deletion	1 (~4%)
GGG CCC GGG ATC CAT TAC CCT GTT AT C CCT AGA TCC ACC GGT CGC TCT AGA A	Insert	1 (~4%)
GGG CCC GGG ATC CAT TAC CCT GTT GCC AAA ATG ATG TAA CCA GCA C	Rearrangement	1 (~4%)
		Total [26]

B

Non-Compatible Junctions

293 Cells

```
GGG CCC GGG ATC CAT TAC CCT GTT AT AA CAG GGT AAT GC GGC CGC GC ATG GCC TTA      Insert                  3 (~20%)
GGG CCC GGG ATC CAT TAC CCT GTT A  A CAG GGT AAT GC GGC CGC GC ATG GCC TTA      Insert                  2 (~13%)
GGG CCC GGG ATC CAT TAC CCT GTT A    CAG GGT AAT GC GGC CGC GC ATG GCC TTA      Deletion                1 (~7%)
GGG CCC GGG ATC CAT TAC CCT GT     AA CAG GGT AAT GC GGC CGC GC ATG GCC TTA      Insert/ Deletion        2 (13%)
GGG CCC GGG ATC CAT TAC CCT G        GGT AAT GC GGC CGC GC ATG GCC TTA      Deletion                1 (~7%)
GGG CCC GGG ATC CAT                AA CAG GGT AAT GC GGC CGC GC ATG GCC TTA      Insert/ Deletion        3 (~20%)
GGG CCC GGG ATC CA           CTAGTCCA CAG GGT AAT GC GGC CGC GC ATG GCC TTA      Insert/ Deletion        2 (~13%)
ACC              CTGCCGGCAG          CAG GGT AAT GC GGC CGC GC ATG GCC TTA      Insert/Del 15bp 5' Sce  1 (~7%)
                                                                                                        Total [15]
```

DT40 Wt Cells

```
GGG CCC GGG ATC CAT TAC CCT GTT AT AA   CAG GGT AAT GC GGC CGC GCA TGG C        Insert                  5 (~38%)
GGG CCC GGG ATC CAT TAC CCT GTT AT AG   CAG GGT AAT GC GGC CGC GCA TGG C        Insert                  1 (~8%)
GGG CCC GGG ATC CAT TAC CCT GTT AT GGGTATATAA CAG GGT AAT GC GGC CGC GCA TGG    Insert                  1 (~8%)
GGG CCC GGG ATC CAT TAC CCT GTT AT ATATAA CAG GGT AAT GC GGC CGC GCA TGG C      Insert                  1 (~8%)
GGG CCC GGG ATC CAT TAC CCT GTT A  A     CAG GGT AAT GC GGC CGC GCA TGG C      Insert/Deletion         1 (~8%)
GGG CCC GGG ATC CAT TAC CCT GT     AA    CAG GGT AAT GC GGC CGC GCA TGG C      Insert/Deletion         1 (~8%)
GGG CCC GGG ATC CAT TAC CCT GT     ATAA  CAG GGT AAT GC GGC CGC GCA TGG C      Insert/Deletion         1 (~8%)
GGG CCC GGG ATC CAT TAC CC                                GC GCA TGG C        Deletion                1 (~8%)
GGG CCC GGG AT                     AA    CAG GGT AAT GC GGC CGC GCA TGG C      Insert/Deletion         1 (~8%)
                                                                                                        Total [13]
```

DT40-77

```
GGG CCC GGG ATC CAT TAC CCT GTT AT AA   CAG GGT AAT GC GGC CGC GCA TGG C        Insert                  8  (50%)
GGG CCC GGG ATC CAT TAC CCT GTT AT      CAG GGT AAT GC GGC CGC GCA TGG C        Blunt ended             1  (~6%)
GGG CCC GGG ATC CAT TAC CCT GTT AT       G GGT AAT GC GGC CGC GCA TGG C        Deletion                1  (~6%)
GGG CCC GGG ATC CAT TAC CCT GTT AT GAGACC          AT GC GGC CGC GCA TGG C      Insert/Deletion         1  (~6%)
GGG CCC GGG ATC CAT TAC CCT GTT A  A    CAG GGT AAT GC GGC CGC GCA TGG C      Insert/Deletion         1  (~6%)
GGG CCC GGG ATC CAT TAC CCT GT     ATA  CAG GGT AAT GC GGC CGC GCA TGG C      Insert/Deletion         1  (~6%)
GGG CCC GGG ATC CAT TAC CCT GT     AAT            GC GGC CGC GCA TGG C      Insert/Deletion         1  (~6%)
GGG CCC GGG ATC CAT                 G GGT AAT GC GGC CGC GCA TGG C          Deletion                1  (~6%)
GGG CCC GGG ATC CAT                CAG GGT AAT GC GGC CGC GCA TGG C         Deletion                1  (~6%)
                                                                                                        Total [16]
```

ES Cells

```
GGG CCC GGG ATC CAT TAC CCT GTT AT A  CAG GGT AAT GC GGC CGC TCT AGA A          Insert                  5  (~22%)
GGG CCC GGG ATC CAT TAC CCT GTT AT AA CAG GGT AAT GC GGC CGC TCT AGA A          Insert                  4  (~17%)
GGG CCC GGG ATC CAT TAC CCT GTT AT CCC TA        GC GGC CGC TCT AGA A           Insert/deletion         1  (~4%)
GGG CCC GGG ATC CAT TAC CCT GTT AT GTA ATG CGT AT G GGT AAT GC GGC CGC TCT AGA A  Insert/deletion       1  (~4%)
GGG CCC GGG ATC CAT TAC CCT GTT A A   CAG GGT AAT GC GGC CGC TCT AGA A          Insert/deletion         2  (~9%)
GGG CCC GGG ATC CAT TAC CCT GTT A              AT GC GGC CGC TCT AGA A          Deletion                1  (~4%)
GGG CCC GGG ATC CAT TAC CCT GTT (27bp insert*) T AAT GC GGC CGC TCT AGA A       Insert/deletion         1  (~4%)
GGG CCC GGG ATC CAT TAC CCT GT AAT G             T GGC CGC TCT AGA A            Insert/deletion         1  (~4%)
GGG CCC GGG ATC CAT TAC CCT GT                AAT GC GGC CGC TCT AGA A          Deletion                1  (~4%)
GGG CCC GGG ATC CAT TAC CCT G ATA A   CAG GGT AAT GC GGC CGC TCT AGA A          Insert/deletion         1  (~4%)
GGG CCC GGG ATC CAT TAC TAA           CAG GGT AAT GC GGC CGC TCT AGA A          Insert/deletion         1  (~4%)
GGG CCC GGG ATC CAT TAC                AG GGT AAT GC GGC CGC TCT AGA A          Deletion                1  (~4%)
GGG CCC GGG ATC CAT TAC GGC C                 AAT GC GGC CGC TCT AGA A          Insert/deletion         1  (~4%)
GGG CCC GGG ATC CAT TAC CCT GTT AT AAC A TTA CCC TGT TAT AAC AGG GTA ATGC GGC CGC TCT AGA A   Insert    1  (~4%)
AAAGAA                             CAG GGT AAT GC GGC CGC TCT AGA A            Del 28bp 5' Sce         1  (~4%)
                                                                                                        Total [23]
```

*CCA TTG ACG TCA ATG GAA AGT CCC TAT = 27bp insert from CMV promoter

Figure 3.7: Sequences Found by End-Joining in Different Cell Lines.

Sequences of all junctions found using the end-joining assay with: A) Compatible end-joining, and B) Non-compatible end-joining. At the top are the sequences from junctions formed by compatible end-joining. In the right column are total counts of each sequence observed. At the bottom of the figure are sequences formed by non-compatible end-joining. Yellow highlighting: I-SceI sequence from 5' Sce site; Green highlighting: I-SceI sequence from 3' Sce site; Red lettering: NotI site, used as a reference.

Junctions with Drugs

293 Cells with compatible end joining assay

(No Drug)

```
GGG CCC GGG ATC CAT TAC CCT GTT AT  C CCT A GC GGC CGC TCT AGA A
GGG CCC GGG ATC CAT TAC CCT GTT AT         A GC GGC CGC TCT AGA A
GGG CCC GGG ATC CAT TAC CCT GTT AT                        GA A
GGG CCC GGG ATC CAT TAC CCT GTT AT                  AGTGGATCCCC
GGG CCC GGG ATC CAT TAC CCT GTT A C C CCT A GC GGC CGC TCT AGA A
GGG CCC GGG ATC CAT TAC CCT GTT A           GC GGC CGC TCT AGA A
GGG CCC GGG ATC CAT TAC CCT GTT TAT C CCT A GC GGC CGC TCT AGA A
GGG CCC GGG ATC CAT TAC CCT              A GC GGC CGC TCT AGA A
GGG CCC GGG ATC CAT TA             T C CCT A GC GGC CGC TCT AGA A
GGG CCC GGG ATC CAT TA                      …CCCCCGGGCTG
GGG CCC GGG ATC CAT                  C CCT A GC GGC CGC TCT AGA A
GGG CCC GGG ATC CAT TAC CCTACCCTGTTATATC CCT A GC GGC CGC TCT AGA A
GGG CCC GGG ATC C                                     GA A
AGA TGC GCG GCC GAT TAC CCT GTT AT  C CCT A GC GGC CGC TCT AGA A
```

Event	Total
Perfect	36 (~67%)
Deletion	4 (~7%)
Deletion	1 (~2%)
Del 9bp 3′ NotI	1 (~2%)
Insert/Deletion	1 (~2%)
Deletion	1 (~2%)
Insertion/Deletion	1 (~2%)
Deletion	1 (~2%)
Insert/Deletion	3 (~6%)
Del 16bp 3′ Not1	1 (~2%)
Deletion	1 (~2%)
Insert/copy	1 (~2%)
Deletion	1 (~2%)
5′ Rearrangement	1 (~2%)
	Total [54]

NU7026 (10uM)

```
GGG CCC GGG ATC CAT TAC CCT GTT AT  C CCT A GC GGC CGC TCT AGA A
          CGGCCG AT TAC CCT GTT AT  C CCT A GC GGC CGC TCT AGA A
GGG CCC GGG ATC CAT TAC CCT GTT AT A  CCT A GC GGC CGC TCT AGA A
GGG CCC GGG ATC CAT TAC CCT GTT A           GC GGC CGC TCT AGA A
GGG CCC GGG ATC CAT TAC CCT GT   AT C CCT A GC GGC CGC TCT AGA A
GGG CCC GGG ATC CAT TAT CCC T              A GC GGC CGC TCT AGA A
GGG CCC GGG ATC CAT TAC CC       CA                  CT AGA A
```

Event	Total
Perfect	6 (~50%)
Complex	1 (~8%)
Insert/Deletion	1 (~8%)
Deletion	1 (~8%)
Insert/Deletion	1 (~8%)
Deletion	1 (~8%)
Insert/Deletion	1 (~8%)
	Total [12]

KU55933 (10uM)

```
GGG CCC GGG ATC CAT TAC CCT GTT AT  C CCT A GC GGC CGC TCT AGA A
GGG CCC GGG ATC CAT TAC CCT GTT A G C CCT A GC GGC CGC TCT AGA A
GGG CCC GGG ATC CAT TA             T C CCT A GC GGC CGC TCT AGA A
AAAGAATTGATTTGATA                    C CGC TCT AGA A
GGG CCC                                    GC TCT AGA A
GGG CCC GGG ATC CAT TAC CCT GT AT C CCT GTATCCATTATCCCTAGC GGC CGC TCT AGA A
```

Event	Total
Perfect	6 (~55%)
Insert/Deletion	1 (~9%)
Insert/ Deletion	1 (~9%)
Del 17bp 5′ Sce	1 (~9%)
Deletion	1 (~9%)
Rearrangement	1 (~9%)
	Total [11]

Caffeine (4mM)

```
GGG CCC GGG ATC CAT TAC CCT GTT AT  C CCT A GC GGC CGC TCT AGA A
GGG CCC GGG ATC CAT TAC CCT GTT A           GC GGC CGC TCC AGA A
GGG CCC GGG ATC CAT                           AGTGGATCCCCCGGG
GGG CCC GGG ACA C CAT TA           T C CCT A GC GGC CGC TCT AGA A
AAAGAA                           CAGGGTAAT GC GGC CGC TCT AGA A
GGG CCC GGG ATC     TATTACCCTGTTA T C CCT A GC GGC CGC TCT AGA A
```

Event	Total
Perfect	2 (~11%)
Deletion	1 (~6%)
Deletion 17bp 3′ Sce	1 (~6%)
Rearrangement	12 (~67%)
Insert/Deletion	1 (~6%)
Deletion/ Rearrangement	1 (~6%)
	Total [18]

Figure 3.8: Sequences in End-Joining Assay with Drug Treatment.

Sequences from 293 cells that have undergone end-joining while being treated with the drugs NU7026, KU-55933, and Caffeine are recorded in the left column, and the total number of each sequence observed is given in the right column. Color formatting is the same as in Figure 3.7. Yellow highlighting: I-SceI sequence from 5' Sce site; Green highlighting: I-SceI sequence from 3' Sce site; Red lettering: NotI site, used as a reference.

Table 1: Comparison of 293, DT40 and ES Cells

Absolute Rate	S-GT	DSB-GT	HDR-DRGFP	RI
293	7.1×10^{-7} (1 in 1,400,000)	3.5×10^{-4} (1 in 3,000)	6.7×10^{-2} (1 in 15)	2×10^{-2} (1 in 50)
ES	1×10^{-6} (↓1.4) (1 in 1,000,000)	2×10^{-3} (↓6) (1 in 500)	N.D	1.1×10^{-3} (↑20) (1 in 900)
DT40	3×10^{-6} (↓5) (1 in 300,000)	1.2×10^{-3} (↓3) (1 in 1,000)	1.1×10^{-1} (↓1.7) (1 in 10)	2.1×10^{-5} (↑1000) (1 in 50,000)

Table 3.1: Comparison of Homologous Recombination and Random Integration in 293, DT40, and ES cells.

Fold change (in superscript) is the difference found in DT40 cells compared to 293 cells. The rate of spontaneous gene targeting in 293 cells has been previously reported (Porteus and Baltimore 2003). Abbreviations: S-GT: Spontaneous Gene Targeting; DSB-GT: Double-Strand Break Induced Gene Targeting; HDR-DRGFP: Rate of GFP positive cells using the DRGFP assay to detect homology directed repair (HDR). RI: Random Integration.

Discussion

Increasing the rate of gene targeting in mammalian somatic cells has both experimental and therapeutic importance. One approach to increasing the rate of gene targeting is to increase the absolute rate of gene targeting by the creation of a DNA double-strand break in the target gene. A second potential method of increasing the relative rate of gene targeting is to decrease the background rate of random integration. This work demonstrates that DT40 cells have a naturally decreased rate of random integration and that this decreased rate of random integration may account for the increase in the relative rate of gene targeting previously described (Buerstedde and Takeda 1991).

Random integration and translocations share certain similarities in that both are forms of recombination in which two segments of DNA that are normally not juxtaposed become joined. We developed the end-joining assay as a way to quantitatively measure the joining of two normally disparate ends. We characterized the assay in 293 cells and found that 293 cells efficiently joined both compatible and non-compatible ends, they joined compatible ends in a precise fashion, and they joined non-compatible ends primarily using a microhomology based mechanism. Our data regarding the use of microhomology of joining non-compatible ends is consistent with the finding that half of random integration events in hamster cells occur with evidence of a microhomology based joining mechanism (Merrihew, Marburger et al. 1996). When we studied DT40 cells using the end-joining assay, we found that DT40 cells were able to join compatible ends accurately but with decreased efficiency compared to 293 cells. More strikingly,

DT40 cells were over 100-fold less efficient at joining non-compatible ends. Thus, the 1000-fold decrease in random integration can largely be accounted for by the over 100-fold decrease in joining of non-compatible ends. The mechanism of random integration remains a relatively unstudied field. Merrihew et al (1996) showed that sites of random integration were associated with complex genomic rearrangements (Merrihew, Marburger et al. 1996). Lin and Waldman (2001) and Miller et al. (2004) have both shown that extra-chromosomal DNA, either as a plasmid or as an adeno-associated viral vector, can integrate into chromosomal sites of induced double-strand breaks (Lin and Waldman 2001; Miller, Petek et al. 2004). Like random integrations of plasmid DNA, Miller et al. (2002) also showed that adeno-associated viral vector integration is associated with chromosomal rearrangements (Miller, Rutledge et al. 2002). Moreover, the finding that plasmids integrate into fragile sites after aphidicolin and bleomycin treatment is also consistent with the idea that random integration occurs at the sites of double-strand breaks (Rassool, McKeithan et al. 1991; Nakayama, Adachi et al. 1998). Finally, small oligonucleotides can be captured in the repair of an extra-chromosomal DSB (Roth, Proctor et al. 1991). All of these studies are consistent with the idea that an alternative pathway of DSB repair is to patch unrepaired DSBs with fragments of extra-chromosomal DNA. Moreover, they are also consistent with the idea that a major mechanism of random integration is through the patching of unrepaired, spontaneous DSBs. If random integration occurs through this mechanism, then the ends of the DSB and the ends of the extra-chromosomal DNA are unlikely to share sequence homology and efficient joining would require the joining of non-compatible ends. Thus, cells deficient in joining non-compatible ends would be expected to show decreased rates of

random integration. Our data comparing end-joining and random integration rates between 293, DT40 cells and ES cells without the use of chemical inhibitors confirm this prediction and thus provide support for such a model of random integration. However, our finding that using compounds to inhibit DNA repair enzymes can cause both an increased rate of random integration as well as decreases in end-joining seem to contradict this prediction. We found that NU7026 had no effect on random integration but decreased end-joining, while KU-55933 and caffeine both increased random integration and decreased end-joining. Furthermore none of these compounds substantially increased gene targeting frequencies. We can not readily explain these observations, but they seem to imply that the process of random integration does not require DNA-PK_{cs}, ATM, or ATR, but the process of end-joining does require these proteins. By inhibiting these DNA repair proteins, an alternative mechanism may be used for random integration that is independent of end-joining. Clearly further details on the process of random integration need to be determined to explain these results.

We found that the end-joining gene Ku70 is important in random integration as DT40 cells deficient in end-joining are also decreased in random integration. Furthermore, our data supports the findings of Pierce et al. (2001) who found that mutations in Ku70 cause an increase in the homologous recombinational repair of DSBs probably by the shunting of DSBs that would normally be repaired by end-joining into a recombinational repair pathway (Pierce and Jasin 2001). One intriguing hypothesis to increase the relative rate of gene targeting in mammalian somatic cells is to inhibit the end-joining process. By inhibiting end-joining one might get both an increase in the absolute rate of gene targeting and a decrease in the random integration rate thereby

getting a significant increase in the relative rate of gene targeting. Indeed this hypothesis is supported by the fact that when cells were treated with NU7026 we observed a modest increase in gene targeting as well as a decrease in end-joining. Our work with compounds that inhibit DNA repair as well as experiments with the Ku70$^{-/-}$ DT40 cells provides a cautionary note to the approach of inhibiting end-joining to increase the relative gene targeting rate. By inhibiting end-joining, one may turn spontaneously occurring lesions such as double-strand breaks with compatible ends that normally are non-mutagenic into mutagenic lesions. If one were to trying to create knock-outs in mammalian somatic cell lines for experimental purposes, the potential increase in the mutation load to the cell may or may not be problematic. If one were trying to use gene targeting to correct mutations therapeutically, however, one would want to carefully evaluate whether the benefit of increasing the relative rate of gene targeting by inhibiting end-joining would be worth the potential risk of creating additional, unwanted mutations or possible random integration events.

The answer that the increase in the relative rate of gene targeting in DT40 cells is the result of a deficiency in joining non-compatible ends inevitably leads to the question of what causes the decreased rate of joining of non-compatible ends. The simple answer that DT40 cells are completely deficient in the classic Ku70 end-joining pathway is not supported by the data. First, our results demonstrate that DT40 cells have Ku70 dependent pathways of re-joining double-strand breaks. Second, Ku70$^{-/-}$ DT40 cells have increased chromosomal instability in response to DNA damage compared to wild-type DT40 cells suggesting that the Ku70 pathway of double-strand break repair is intact in DT40 cells (Takata, Sasaki et al. 1998). These data do not rule out the possibility that

DT40 cells are just relatively deficient in Ku70 end-joining which leads to a preferential defect in joining non-compatible ends. They also do not rule out the more interesting possibility that there is a regulator of DNA double-strand break repair in either 293 or DT40 cells that changes the way in which ends of a double-strand break, particularly those that are non-compatible such as would be created by RAG mediated V(D)J recombination, are processed and repaired.

Materials and Methods

DNA Manipulations and Cloning

Standard molecular biology procedures were used to make all plasmids (Ausubel, Brent et al. 1996). The Sce expression plasmids, GFP gene targeting plasmids, and DRGFP plasmids have been previously described (Pierce, Johnson et al. 1999; Porteus and Baltimore 2003) as well as the GFP zinc finger nucleases used to perform gene targeting in ES cells (Pruett-Miller, Connelly et al. 2008). To determine that the random integration rates were not promoter dependent, we used different GFP expression plasmids that contained either the phosphoglycerate kinase (PGK) promoter, cytomegalovirus (CMV) promoter, EF1α promoter, or CMV/Chicken β-actin (CMV/CBA) hybrid promoter to drive GFP expression. The end-joining compatible plasmid was made by using PCR with the following primers to amplify the destabilized GFP gene (d2EGFP) from Clontech (Palo Alto, CA). This fragment was cloned into pHFBW (a plasmid with a lentiviral backbone) using standard techniques.

Cell Culture Conditions

293 cells were grown in DMEM supplemented with 10% Bovine Growth Serum (Hyclone, Logan, UT), 2 mM L-Glutamine, 100 IU/ml penicillin, 100 μg/ml streptomycin ("Pen-Strep") and grown at 37^0C/5% CO_2 in a humidified incubator. DT40 Cells were grown in RPMI supplemented with 10% Bovine Growth Serum/ 1% Chicken Serum/2 mM L-Glutamine/ Pen-Strep/1X Non-Essential Amino Acids (Invitrogen/Gibco; Carlsbad, CA)/10 mM Hepes pH 7.4 and grown at 37^0C/5% CO_2 in a humidified incubator. ES cells were cultured in ESLX media consisting of ES-DMEM, 20% ES Qualified FBS, 1X non-essential amino acids, 1X nucleosides, 1000 U/ml ESGRO LIF all purchased from Chemicon (Chemicon, Billerica, MA), 2 mM L-Glutamine, 1X Pen/ Strep (Invitrogen, Carlsbad, CA), and 0.12 mg/ml sodium pyruvate and 0.1 mM BME (Sigma-Aldrich, St. Louis, MO).

Method of Transfection

Several different transfection techniques were examined for DT40 cells including Fugene 6 (Roche Applied Science; Indianapolis, IN) and Effectene (Qiagen: Valencia, CA) but we found the only effective method of transfecting DT40 cells was by electroporation. To transfect by electroporation we grew 1 x 10^7 DT40 cells to mid-log phase, washed the cells twice in serum-free RPMI, resuspended in 180 μl serum-free RPMI, added 10 μg of plasmid DNA, incubated on ice for 5 minutes, then electroporated in a 2 mm cuvette at 150 Volts, 1200 μF, R3 resistance using a BTX 600 machine (Genetronics; San Diego, CA). Following electroporation we incubated the cells on ice for 5 minutes to allow cell recovery before adding the cells to full media. All results for DT40 cells were obtained using electroporation as described above. We routinely obtained transfection efficiencies of 10-15% using this protocol.

Except for the spontaneous gene targeting rate and gene targeting with chemical inhibitors, all results for 293 cells were obtained by electroporation to allow direct comparison to the results obtained by electroporating DT40 cells. The spontaneous gene targeting rate in 293 cells was obtained using the standard calcium phosphate transfection technique and has been previously reported (Ausubel, Brent et al. 1996; Porteus and Baltimore 2003). We used the following procedure to electroporate 293 cells. 3×10^6 293 cells were grown to mid-log phase on a 10 cm plate, trypsinized to remove from the plate, washed twice in serum-free DMEM and resuspended in 360 μl of serum-free DMEM. The cell suspension was mixed with 10 μg of plasmid DNA in a 4 mm cuvette, incubated on ice for 5 minutes, electroporated at 150 V, 1000 μF, R3 resistance using a BTX 600 machine, incubated on ice for 5 minutes to allow cell recovery and then added to full media. We routinely obtained transfection efficiencies of 5-20% using this protocol. Using this protocol we found that transfection of a GFP expression plasmid and a RFP expression plasmid resulted in %10 single-positive cells, but the rate of double positive cells was only 1% (data not shown). Thus, the rate of co-transfection of two plasmids using this protocol was low. In our gene targeting experiments we used a single plasmid system because of this low rate of co-transfection. For random integration measurements in ES cells, cells were harvested from sub-confluent plates using trypsin, washed with PBS, and 2×10^6 cells per sample were resuspended in 700 ul PBS containing 5 μg of GFP expression plasmid. Cells were electroporated with 250 V and 960 μF using a Bio-Rad Gene Pulser with a capacitance extender (Bio-Rad, Hercules, CA), followed by a 5 minute incubation on ice, and replated in warm ESLX. This

procedure gave electroporation efficiencies of ~20%. For gene targeting rates in ES cells, cells were seeded in a 24 well plate at 100,000 cells per well and transfected using Lipofectamine 2000 (Invitrogen Corp., Carlsbad, CA). Each transfection sample contained 100 ng of each GFP zinc finger nuclease and 600 ng of a repair plasmid containing a fragment of the GFP gene.

Production of Lentivirus

The low rate of random integration in DT40 cells made production of stable DT40 cell lines difficult. Therefore, to create DT40 lines with the various reporter constructs we used lentiviral infection. We used the third generation lentiviral production system previously described (http://tronolab.com/protocols_lentivectors.php). Briefly, a 10 cm plate of 293-T cells was co-transfected with pMD2G, pMDLg/pRRE, pRSV-Rev and the apporopriate lentiviral construct (either plv658, plvDRGFP, plv1039-NC, plv1040-C, plvPC221-NC, or plvPC222-C) by the calcium phosphate technique. The next morning the media was changed and the cells were incubated for a further 48 hours. At 48 hours the supernatant was harvested, the debris removed by centrifugation followed by filtration through a 0.4 micron filter and the virus was concentrated by ultracentrifugation. The viral pellet was resuspended in PBS without calcium or magnesium and stored at –80^{o}C.

Creation of GFP Gene Targeting and DRGFP Cell Lines

The 293 GFP gene targeting cell lines have been previously described (Porteus and Baltimore 2003). The 293 DRGFP cell line was made by electroporating 10 μg of pDRGFP as described above into 293 cells to generate cells with single copy integration events. The cells were then selected in puromycin at 1 μg/ml for two weeks. Puromycin

resistant colonies were pooled and FACS sorting used to eliminate those cells that had already undergone rearrangement of the mutated GFP gene. DT40 cell lines were made by infecting mid log-phase cells with either lv658 or lvDRGFP at a multiplicity of infection of 0.1 to create cells with single copy integrations. To create DT40/658 cell lines, the infected cells were sorted three times for CD8+ cells using a Miltenyi Biotech (Auburn, CA) magnetic bead sorter and anti-CD8 beads until the percentage of CD8+ cells was greater than 90%. In the GFP gene targeting reporter construct, pA658, CD8 expression is through an internal ribosomal entry site driven by the same promoter that drives expression of the mutated GFP gene (see (Porteus and Baltimore 2003) for a schematic of the pA658). To create DT40/DRGFP cells, the infected cells were selected in puromycin for two weeks until a polyclonal population expanded. FACS sorting was used to eliminate the GFP+ cells that spontaneously arose during the production of the DT40/DRGFP cell line. The ES cell line was created by the UT Southwestern Transgenic Core facility by electroporating 129/SvEvTac (SM-1) ES cells with a version of the 658D gene targeting construct. These cells contain a single copy of the GFP gene targeting reporter integrated into the ROSA26 locus and expressed by the ROSA26 promoter.

Creation of End-Joining Cell Lines

The end-joining cell lines were created by infecting 293, DT40 cells, or ES cells with lv1039-NC, lv1040-C, plvPC221-NC, or plvPC222-C. lv1039-NC and lvPC221-NC are lentivirus that contain the end-joining reporter with non-compatible ends and the CD8 or RFP gene respectively. lv1040-C and lvPC222-C are lentivirus that contain the end-joining reporter with compatible ends and the CD8 or RFP gene respectively (Figure

3.5A). The infected cells were then expanded for several days and polyclonal cell lines created by fluorescent activated cell sorting for GFP+ CD8/RFP- cells. After fluorescent activated cell sorting we obtained a population of cells in which the majority of cells were GFP+ CD8/RFP-. We would periodically analyze the cell lines and found that the cell lines were phenotypically stable for their GFP+ CD8/RFP- phenotype.

Measurement of Gene Targeting

Spontaneous gene targeting (S-GT) was measured by electroporating 10 μg of the repair plasmid (RS2700) (Porteus and Baltimore 2003) into the 293 and DT40 GFP gene targeting cell lines. The targeting rate was determined by counting the number of GFP positive cells by flow cytometry three days after transfection and normalizing to the transfection efficiency.

Double-strand break (DSB) mediated gene targeting (DSB-GT) in 293 and DT40 cells was measured by electroporating 10 μg of a single plasmid (pA979 or Sce/Repair Plasmid) that contains two elements (Porteus and Baltimore 2003). The first element contains an expression cassette for Sce expression. The second element contains a fragment that can correct the GFP mutation in the reporter gene and has 2100 bp of homology to the integrated target. Three days after transfection the number of GFP positive cells was determined by flow cytometry and the targeting rate determined by normalizing the number of GFP+ cells to the transfection efficiency for that experiment. Gene targeting in ES cells was performed as described above in transfection methods, and gene targeting rates were measured three days post transfection using flow cytometry.

Measurement of Homology Directed Repair by DRGFP Assay

Homology directed repair (HDR) was measured by electroporating 10 μg of a Sce expression plasmid (pA961), in which Sce is driven by the CMV/CBA hybrid enhancer/promoter (Porteus and Baltimore 2003), into either DT40/DRGFP cells or 293/DRGFP cells. The HDR rate was determined three days after transfection by counting the number of GFP+ cells by flow cytometry and normalizing to the transfection efficiency.

Measurement of Random Integration

Random Integration (RI) was measured by electroporation of 10 μg of a GFP expression plasmid into either 293 or DT40 cells, for ES cells, 5 ug of plasmid was used. We then analyzed a portion of the transfected cells by flow cytometry for GFP positive cells at day 1 and day 2 after transfection to determine the maximal transient transfection efficiency. We then continued to passage with serial monitoring by flow cytometry for the percentage of GFP positive cells until the percentage of positive cells was stable over six days.

Measurement of End-Joining using the End-Joining Assay

To measure end-joining we electroporated 10 μg of a Sce expression plasmid (A961) into the end-joining reporter cell lines. If Sce cuts at both of its target sites in the reporter construct, the GFP gene should drop out and the cells should become GFP-. If the two ends are then joined, the CD8α/RFP should become linked to the CMV/CBA promoter and the cells should become positive for cell surface CD8 or internal RFP. To assure depletion of GFP and expression CD8/RFP after the joining reaction, we expanded the cells for 3-5 days following transfection. For CD8 expression, we analyzed the cells by staining with phycoerytherin-conjugated anti-CD8 monoclonal antibody (Ditech; Oslo,

Norway). The end-joining rate was determined by counting the number of GFP- CD8+ cells by flow cytometry and normalizing to the transfection efficiency. For cells containing the GFP/RFP end-joining construct, cells were expanded and analyzed using flow cytometry.

Determining the Sequence of the End-Joining Junctions

To determine the sequence of the junctions after end-joining we induced end-joining as described above. We then purified the CD8+ cells using a Miltenyi column as described above until the percentage of CD8+ cells was greater than 80%. After purification of the CD8+ population we harvested genomic DNA using the DNeasy Kit (Qiagen; Valencia, CA). We amplified the junction by PCR of the bulk genomic DNA using primers A1020B (5' GGCTCTAGAGCGTCTGCTAACC 3' and A1070D (5' ACGTCGAGCTCGACTGTCTCGCCGAGGTTCCAG 3') that flank the junction, digesting the PCR fragment with XhoI/XbaI and cloning into the XhoI/XbaI sites of pBluescript (Stratagene, La Jolla, CA). In some experiments, PCR products were subsequently ligated into pGEM-T Easy (Promega, San Luis Obispo, CA). We then sequenced the junction using the T7 primer.

Inhibition of DNA Repair Proteins Using Drug Inhibitors

293 Cells were plated in 24 well plates the day of transfection. At the time of transfection, media was changed to DMEM supplemented with 10% Bovine Growth Serum (Hyclone, Logan, UT), 2 mM L-Glutamine, 100 IU/ml penicillin, 100 μg/ml streptomycin ("Pen-Strep") with or without drug. NU7026 was added to a final concentration of 10 μM, KU-55933 at 10 μM, and caffeine was added to a final concentration of 4 mM. Cells were transfected using calcium phosphate transfection.

The next morning (day 1), media was replaced with supplemented DMEM (as above) with the addition of inhibitors at the above concentrations. On day 3 media was changed again with supplemented DMEM and no drug. Gene targeting analysis was performed on day 3, while end-joining and random integration analyses were performed on day 8 and day 15 respectively.

Acknowledgements

We thank David Baltimore for his support during the early part of this work. We thank Eric Brown for his generous gift of the lentiviral vectors for the GFP gene targeting system, DRGFP system and pHFBW. This work was supported by career development awards from the NIH (K08 Hl70268) and the Burroughs-Wellcome Fund.

CHAPTER IV:
GENE CORRECTION BY HOMOLOGOUS RECOMBINATION WITH ZINC FINGER NUCLEASES IN PRIMARY CELLS FROM A MOUSE MODEL OF A GENERIC RECESSIVE GENETIC DISEASE

Jon P Connelly[1,2*], Jenny C Barker[1,2*], Shondra Pruett-Miller[1,2], Matthew H Porteus[1,2]

[1]Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX, USA

[2]Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX, USA

*These authors contributed equally to the work and should be considered co-first authors.

Correspondence: Matthew Porteus, Department of Pediatrics and Department of Biochemistry, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas, USA. E-mail: matthew.porteus@utsouthwestern.edu

Short title: Gene Correction using Zinc Finger Nucleases

Abstract

Zinc finger nucleases have been used to create precise genome modifications at frequencies that might be therapeutically useful in gene therapy. We created a mouse model of a generic recessive genetic disease to establish a pre-clinical system to develop the use of zinc finger nuclease mediated gene correction for gene therapy. We knocked a mutated GFP gene into the ROSA26 locus in murine embryonic stem cells and used these cells to create a transgenic mouse. We used zinc finger nucleases to determine the frequency of gene correction by gene targeting in different primary cells from this model. We achieved targeting frequencies from 0.17- 6% in different cell types, including primary fibroblasts and astrocytes. We demonstrate that *ex vivo* gene corrected fibroblasts can be transplanted back into a mouse where they retained the corrected phenotype. In addition, we achieved targeting frequencies of over 1% in embryonic stem cells and the targeted embryonic stem cells retained the ability to differentiate into cell types from all three germ line lineages. In summary, potentially therapeutically relevant frequencies of zinc finger nuclease mediated gene targeting can be achieved in a variety of primary cells and these cells can then be transplanted back into a recipient.

Introduction

Conceptually, the simplest application of gene therapy is for diseases caused by mutations in a single gene, the so-called monogenic diseases. While millions of people suffer from monogenic diseases, a cure is only possible for a small fraction for whom either hematopoietic stem cell transplantation or organ transplantation is available. In contrast, gene therapy uses the patient's own cells and has the potential to cure many of these diseases. In the last decade, several clinical trials have been carried out which have highlighted both the promise of gene therapy (the benefit of tens of patients with severe combined immunodeficiency (Hacein-Bey-Abina, Le Deist et al. 2002; Aiuti, Cattaneo et al. 2009) a handful of patients with Leber's congenital amaurosis (Bainbridge, Smith et al. 2008; Maguire, High et al. 2009), and two patients with X-linked adrenoleukodystrohy (Cartier, Hacein-Bey-Abina et al. 2009) from gene therapy based on viral delivery), and the potential harm from the uncontrolled integrations of the viral vectors used to deliver the therapeutic transgene (Hacein-Bey-Abina, Von Kalle et al. 2003). An alternative to using integrating viruses is to use gene targeting by homologous recombination to precisely control the genomic modification either through directly correcting a mutation or through controlling the site of transgene integration (Porteus and Carroll 2005; Lombardo, Genovese et al. 2007).

The natural rate of gene targeting by homologous recombination, hereafter referred to as "gene targeting," is $1x10^{-5}$ -$1x10^{-8}$ (Doetschman, Gregg et al. 1987; Porteus and Baltimore 2003) and is too low to be therapeutically useful. This barrier has been overcome by the recognition that the creation of a gene specific DNA double-strand

break can stimulate gene targeting several thousand fold (Rouet, Smih et al. 1994; Choulika, Perrin et al. 1995; Brenneman, Gimble et al. 1996; Donoho, Jasin et al. 1998; Porteus and Baltimore 2003), potentially to therapeutically relevant levels. To translate this finding to the field of gene therapy, it was necessary to devise a method to generate site specific DNA double-strand breaks. There have been two major approaches to this problem. The first is to re-design homing endonucleases to recognize target sites in endogenous genes (Arnould, Chames et al. 2006; Ashworth, Havranek et al. 2006; Paques and Duchateau 2007). The second is to design zinc finger nucleases to recognize target sites in endogenous genes. Zinc finger nucleases (ZFNs) are artificial proteins in which the non-specific nuclease domain from the FokI restriction endonuclease is fused to a zinc finger DNA binding domain (reviewed in (Durai, Mani et al. 2005; Porteus and Carroll 2005)). A ZFN can have 3-6 individual zinc finger domains arranged in tandem recognizing a target site 9-18 basepairs long. Additionally, the FokI nuclease domain functions as a dimer (Bitinaite, Wah et al. 1998; Smith, Bibikova et al. 2000). Therefore, a pair of ZFNs must be engineered to bind the target site in a way that permits the nuclease domain to dimerize and create the double-strand break. Thus, even with a pair of 3-finger ZFNs, the full target site is 18 basepairs long. An 18 basepair sequence should only occur once in the mammalian genome based on probability and can be empirically determined for any given sequence by BLAST searches. There are a number of different approaches to engineer ZFNs, each of which has their advantages and disadvantages (reviewed in Cathomen and Joung (2008) (Cathomen and Joung 2008)). Nonetheless, ZFNs have been successfully engineered to a wide variety of different gene targets in a range of different species (Bibikova, Beumer et al. 2003; Porteus and

Baltimore 2003; Urnov, Miller et al. 2005; Morton, Davis et al. 2006; Carroll 2008; Doyon, McCammon et al. 2008; Meng, Noyes et al. 2008; Geurts, Cost et al. 2009; Shukla, Doyon et al. 2009; Townsend, Wright et al. 2009). These ZFNs have been used to generate high rates of precise genome modifications either by the use of mutagenic non-homologous end-joining (in which short insertions or deletions are created at the site of the ZFN induced double-strand break) or by the use of gene targeting, including creating genetically modified zebrafish and rats (Doyon, McCammon et al. 2008; Meng, Noyes et al. 2008; Geurts, Cost et al. 2009).

In human cells, ZFNs have been used to stimulate gene targeting in a variety of different cell lines. The most recent advances demonstrated that ZFNs can stimulate gene targeting in human embryonic stem (ES) cells and induced pluripotent stem (iPS) cells (Hockemeyer, Soldner et al. 2009; Zou, Maeder et al. 2009). Moreover, Perez et al. demonstrated that human T-cells modified at the CCR5 gene by mutagenic repair of a ZFN induced double-strand break could survive when transplanted back into an immunodeficient mouse (Perez, Wang et al. 2008). Nonetheless, to date there has been no easy way to model a therapeutic paradigm in which host derived cells are precisely modified by ZFN mediated gene targeting *ex vivo* and then transplanted back into a recipient, as might be done when trying to treat a patient with a genetic disease. To this end, we have created a mouse model of a generic recessive genetic disease in which a mutated GFP gene has been knocked-in to the ubiquitously expressed ROSA26 locus (Zambrowicz, Imamoto et al. 1997). Using this model, we show that gene correction of 0.17-6% of murine ES cells, novel ROSA-3T3 cell lines, primary embryonic fibroblasts, primary adult fibroblasts, and primary astrocytes can be achieved. We also demonstrate

that gene corrected cells can be transplanted back into a recipient mouse and the transplanted cells retain their gene corrected phenotype.

Results

Generation of a Mouse Model of a Generic Recessive Genetic Disease

A critical aspect of translating the use of ZFN mediated gene targeting to clinical use is to develop appropriate animal models to evaluate the feasibility of the strategy. In this work we generated a mouse model that allows us to study the efficiency of ZFN mediated gene targeting by homologous recombination in any primary cell type using a previously well defined pair of ZFNs. This model provides a platform to directly determine the efficiency of gene targeting in a specific cell type and whether targeted cells can be successfully transplanted back into a host. We created a mouse model of a generic recessive genetic disease by knocking-in a mutated GFP gene into the ROSA26 locus using standard homologous recombination technology in murine ES cells (Figure 4.1A). From two of the targeted ES cell clones we generated transgenic mouse lines in which either one ($ROSA26^{GFP*/+}$) or both alleles ($ROSA26^{GFP*/GFP*}$) of the ROSA26 locus contain a knock-in of the mutated GFP gene. The mutation in the GFP gene consists of an 85 nucleotide sequence which includes both an in-frame stop codon and the recognition site for the I-SceI homing endonuclease. This insertion is 12bp downstream from the ZFN target site in the GFP gene that we have previously designed and validated several different pairs of “GFP”- ZFNs to recognize (Pruett-Miller, Connelly et al. 2008). This animal model mimics the GFP gene-targeting system that we have used to better

understand ZFN mediated gene targeting in tissue culture cells and human ES and iPS cells (Porteus and Baltimore 2003; Porteus 2006; Pruett-Miller, Connelly et al. 2008; Zou, Maeder et al. 2009). In this system, cells containing the integrated GFP* gene are transfected with two plasmids that each express a ZFN, and a third "donor" plasmid that carries sequence information needed to correct the mutation (by serving as a donor template during homologous recombination) (Porteus and Baltimore 2003; Durai, Mani et al. 2005). Because the mutated GFP gene is knocked-in to the ubiquitously expressed ROSA26 locus, this model can be used to study the efficiency of gene correction by gene targeting in potentially every cell type in the mouse. There are two possible strategies to use gene targeting for gene therapy. The first is to attempt gene targeting *in vivo* and directly correct disease-causing mutations in cells without removing them from the animal beforehand. The second is to purify cells from the animal first, and correct the mutation *ex vivo* prior to transplanting back into the animal. Here we report on our results at purifying primary cells from this generic genetic disease model and using ZFN mediated gene targeting to directly correct the mutation in the GFP gene and then transplanting those cells back into a mouse.

ZFN Mediated Gene Targeting in Murine Embryonic Stem Cells

While we were generating the mouse line from the ROSA26$^{GFP*/+}$ targeted ES cell clones, we performed a series of targeting experiments in the ES cell clones (Figure 4.2). To perform these experiments we transfected ES cells with three plasmids—an EGFP donor vector that contains sequence information necessary to correct the mutation in the integrated target but is non-fluorescent because it is missing the first 37 nucleotides of the coding region, and two ZFN expression plasmids where the ZFN is driven by the

ubiquitin C promoter. In prior work we have shown that titrating the amount of the donor plasmid and nuclease expression plasmid is important in maximizing the rate of gene targeting (Porteus 2006; Pruett-Miller, Connelly et al. 2008). In the gene targeting titration experiment we varied the amount of donor plasmid and ZFN expression plasmids while keeping the total amount of DNA transfected the same. We achieved the maximal rate of targeting when the transfection mix consisted of ~90% donor plasmid and ~10% of ZFN expression plasmids (Figure 4.2A). ROSA26$^{GFP*/+}$ ES cells that were transfected with only the donor plasmid without ZFNs had an absolute targeting rate of less than 0.001% (data not shown). The targeting rate with ZFNs was 0.21%. An important concern involving the use of ZFNs to create DNA double-strand breaks is the potential for off-target cleavage. We have previously reported that the ZFNs used in this study demonstrate low levels of off-target toxicity in the murine embryonic stem cells used in this paper (Pruett-Miller, Connelly et al. 2008), and no gross toxicity was observable in these experiments. We also tested whether exposing the cells to vinblastine would increase the rate of targeting in murine ES cells as has been demonstrated in other cell types arrested at the S/G2 phase of the cell cycle (Urnov, Miller et al. 2005; Potts, Porteus et al. 2006; Maeder, Thibodeau-Beganny et al. 2008), and found that vinblastine exposure for the first 15 hours after transfection increased the rate of gene targeting in murine ES cells by approximately 4-5 fold to an overall rate of 1.6% (Figure 4.2B). At this concentration of vinblastine (100nM for 15 hours) no gross cytotoxic effects were readily apparent in this cell type. We also compared two different pairs of ZFNs that target the same sequence in the GFP gene (Table 4.1) and found that ZFN pair 3/4 was approximately 50% better than pair 1/2 (the pair we have previously published and

characterized most extensively (Pruett-Miller, Connelly et al. 2008)). ZFN 3 only differs from ZFN 1 by a single amino acid in finger 1 and ZFN 4 has a different finger 3 than ZFN 2.

To determine if the targeted ES cells retained pluripotency, we evaluated whether the targeted cells could differentiate into all three germ line lineages using a teratoma formation assay. GFP^+ targeted ES cells were purified using fluorescence-activated cells sorting (FACS), and these ES cells were injected subcutaneously into nude mice to create fluorescent teratomas (Figure 4.3). Histological examination of the teratomas showed that targeted ES cells formed tissues representative of all three germ line lineages—endoderm, mesoderm, and ectoderm. In summary, ZFNs can stimulate gene targeting in murine ES cells to an absolute rate of 1.6% (a greater than 1000-fold stimulation over the targeting rate without ZFNs) and these targeted cells retain the potential to differentiate into all of the major cell lineages.

Gene Targeting in ROSA-3T3 Cells

We generated immortalized fibroblast cell lines "ROSA-3T3s" from both $ROSA26^{GFP*/+}$ and $ROSA26^{GFP*/GFP*}$ mice using a standard fibroblast immortalization protocol (Todaro and Green 1963), and performed a gene targeting titration experiment with varying amounts of ZFN expression plasmids and donor plasmid (Figure 4.4A). As in the murine ES cells, we observed the maximal rate of targeting when the transfection mix contained ~90% donor plasmid and ~10% ZFN expression plasmids. Gene targeting rates reached a maximum of 1.8% and 6.7% in heterozygous and homozygous lines respectively. Vinblastine exposure did not affect targeting rates in these lines (data not shown).

Gene Targeting in Primary Embryonic and Adult Fibroblasts

We also studied gene targeting rates in fibroblasts derived from both embryonic (E13.5) and adult (3-6 months old) mice. In these experiments we used nucleofection rather than Lipofectamine 2000 as the method of transfection and found that increasing the amount of donor plasmid while keeping the amount of ZFN expression plasmid constant increased the frequency of gene targeting (Figures 4.4B and 4.4C). The maximal gene targeting rate in primary adult fibroblasts from ROSA26$^{GFP*/GFP*}$ mice was over 2% (Figure 4.4B). In murine embryonic fibroblasts the maximal rate of gene targeting was ~1.8%. Interestingly, the targeting rate in murine embryonic fibroblasts from heterozygous mice was significantly different from fibroblasts derived from homozygous mice when low amounts of donor was transfected but this difference disappeared when the highest amount of donor plasmid was transfected (Figure 4.4C). In summary, ZFN mediated gene targeting rates of ~2% can be achieved in primary fibroblasts of either adult or embryonic origin.

Gene Targeting in Primary Astrocytes

The final primary somatic cell type that we examined for ZFN mediated gene targeting was astrocytes. Using nucleofection with a constant amount of ZFN expression plasmid and increasing amounts of donor plasmid, we found that 0.03-0.17% of primary astrocytes could be targeted (Figure 4.4D). As in all of the other cell types examined, the rate of targeting increased as the amount of donor plasmid transfected increased. Overall this rate of targeting is lower than what we found in primary fibroblasts. This difference may reflect an intrinsic difference between astrocytes and fibroblasts but may also reflect an underestimation of the rate of targeting in astrocytes because of a decreased ability to

detect targeted cells secondary to the low level of GFP expression from the ROSA26 promoter in astrocytes compared to ROSA-3T3s and primary fibroblasts (Figure 4.4A, C, D). As in the immortalized cells and primary fibroblasts, there was no effect of vinblastine exposure on the rate of targeting in primary astrocytes (data not shown).

Transplantation of Gene Corrected Primary Mouse Fibroblasts into an Immunocompetent Recipient

To perform cell based gene correction therapy clinically, gene corrected cells must be able to survive transplantation back into a recipient. A potential treatment for hemophilia is to modify fibroblasts *ex vivo* to secrete Factor VIII or Factor IX, followed by transplanting the cells back into recipients. This has already been performed with transient but significant clinical benefits observed in mice, rats, and human patients (Palmer, Thompson et al. 1989; Qiu, Lu et al. 1996; Roth, Tawa et al. 2001). As mentioned above, ZFNs may be toxic to cells and affect their transplantability. Although no toxic effects were readily apparent upon microscopic examination of treated fibroblasts, we further looked for toxic effects using a previously described toxicity assay that measures how well cells survive after expression of the ZFNs. When fibroblasts were subjected to this assay, a small degree of toxicity was measured (Figure 4.5A). This was minor in comparison with a previously published pair of ZFNs designed to target the CCR5 locus (Perez, Wang et al. 2008), and which exhibit noticeable toxic effects on cells. Both pairs of ZFNs contained the wild-type nuclease domain and were not modified to prevent homodimerization (Miller, Holmes et al. 2007; Szczepek, Brondani et al. 2007; Pruett-Miller, Connelly et al. 2008). To test whether our gene corrected fibroblasts were capable of transplantation, we isolated primary adult fibroblasts and

performed another round of gene targeting. On day 6 post-nucleofection, we analyzed a portion of the total cell population and found that 0.21% of the total cells had undergone gene correction. We next took the remaining population, (without selecting for GFP positive cells), embedded the cells in Matrigel and transplanted the cells subcutaneously into an immunocompetent, isogenic mouse. Two weeks after transplantation we excised the Matrigel plug, cultured the isolated cells for 6 days, analyzed the transplanted cells by flow cytometry, and found that 0.10% of the isolated cells were GFP positive. The lower frequency of targeted cells may be due to contaminating host cells that were excised with the plug (Figure 4.5B). This study demonstrated that *ex vivo* zinc finger nuclease gene corrected fibroblasts could be successfully transplanted.

Figures

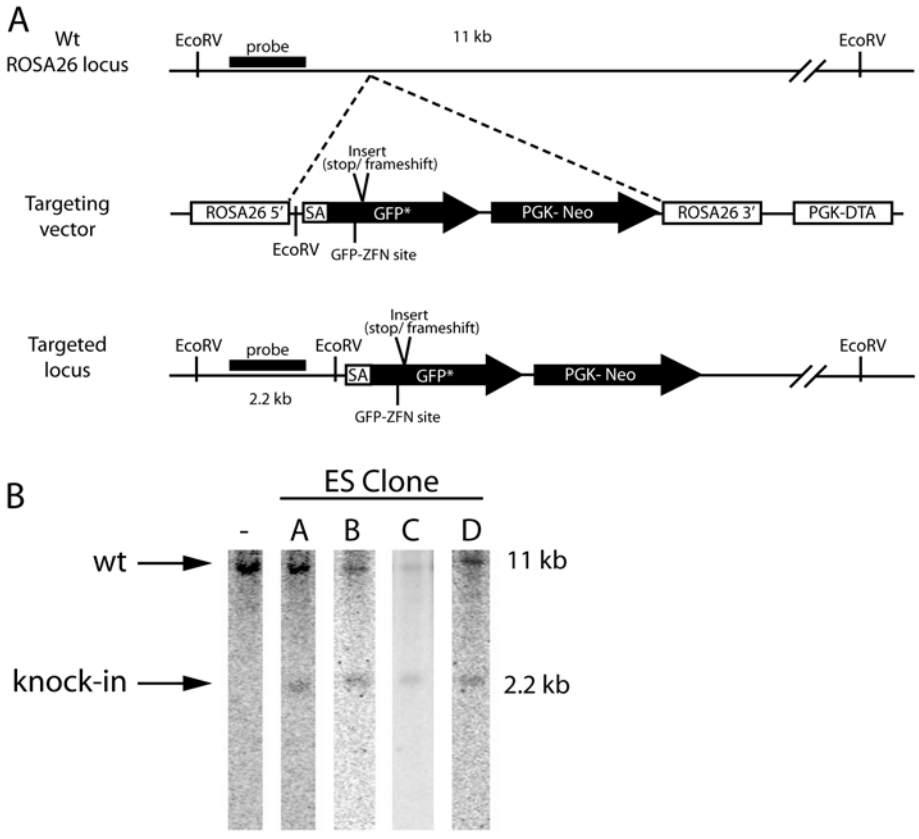

Figure 4.1: Construction of targeting vector and screening of ES clones. (A) Schematic of targeting vector and screening strategy. Correct knock-in of the GFP* reporter cassette to the ROSA26 locus causes the addition of an upstream EcoRV site resulting in a 2.2kb fragment upon digestion. (B) Southern blot analysis of correct knock-in ES clones. Genomic DNA was purified from ES clones, digested with EcoRV and used for Southern analysis. SA, splice acceptor; PGK-Neo, Neomycin resistance cassette; PGK-DTA, diphtheria toxin cassette.

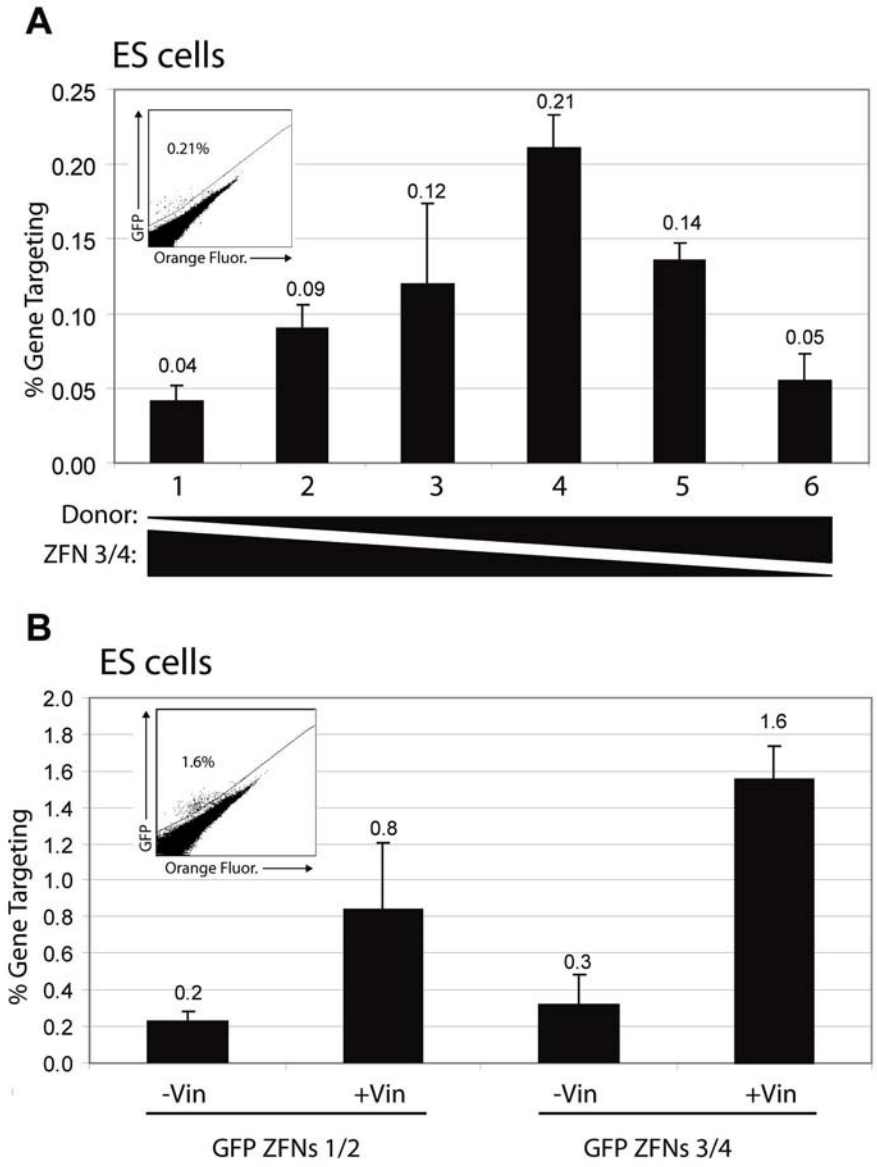

Figure 4.2: Gene targeting in ES cells. (A) Titration of donor plasmid and ZFNs in ES cells. Cells were transfected using Lipofectamine 2000 with different amounts of donor and ZFN plasmids. From left to right, amounts of donor plasmid increased while ZFN amounts decreased. Transfection was performed with Lipofectamine 2000 and the indicated amounts of donor plasmid and ZFNs in each lane, indicated as Donor (ng),

ZFN1 (ng)/ ZFN2 (ng). (1) 100, 350/350; (2) 400, 200/200; (3) 600, 100/100; (4) 700, 50/50; (5) 750, 25/25; (6) 775, 13/13. Fifteen hours post transfection, media was changed to normal ESLX. Gene targeting events were analyzed 4 days post transfection using flow cytometry. (B) Gene targeting in ES cells using two sets of GFP-ZFNs, with and without vinblastine treatment. Cells were plated in ESLX with and without vinblastine (100 nm). Transfection mix was added, and 15 hours later removed and replated with ESLX. Gene targeting events were analyzed as in (A). In the upper left of each graph is a representative flow cytometry plot after targeting in which GFP fluorescence is measured in the y-axis and background orange fluorescence along the x-axis. The number in the left corner of the flow plot is the percentage of $GFP^{(+)}$ cells.

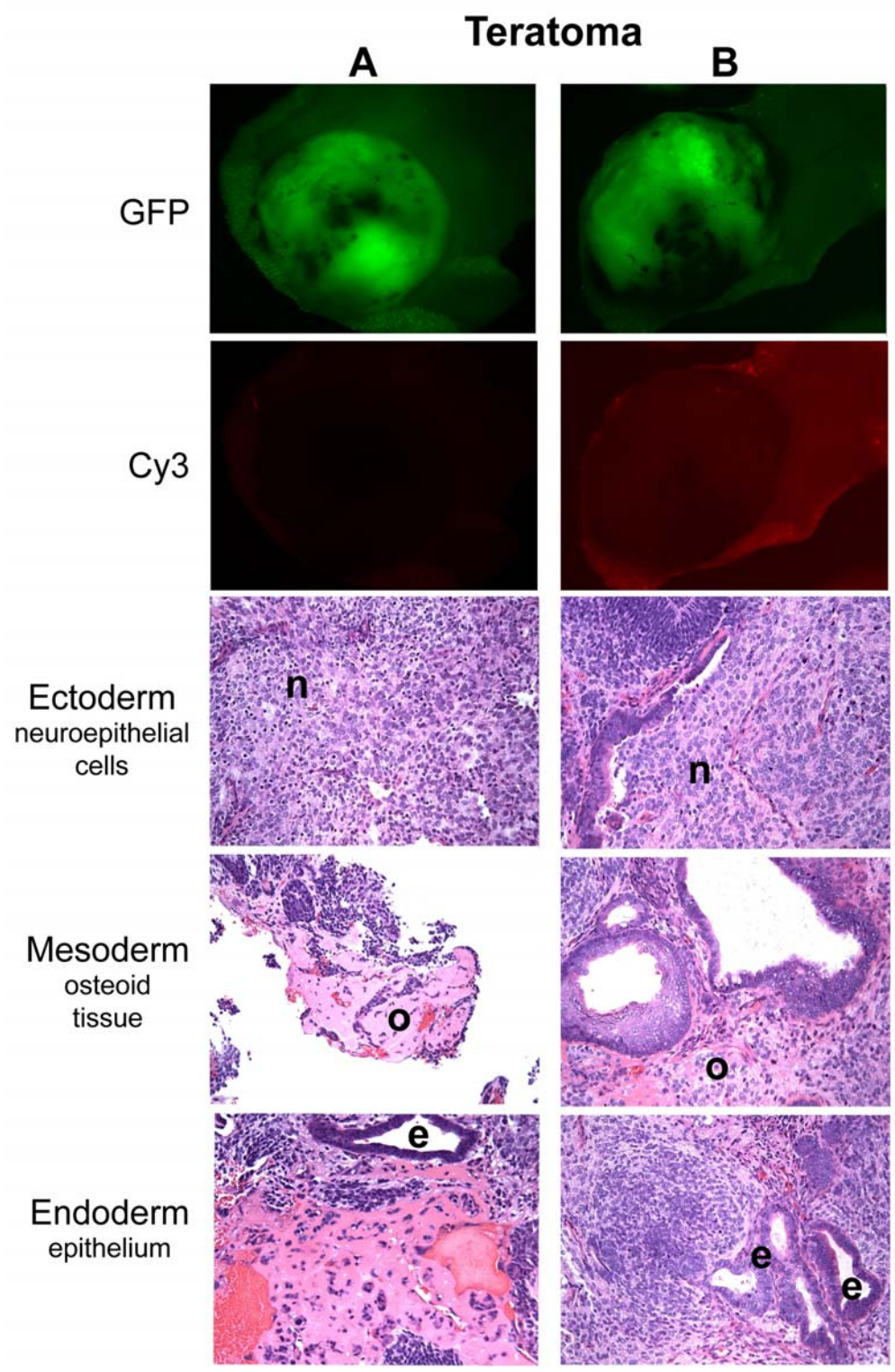
Teratoma
A
B
GFP
Cy3
Ectoderm
neuroepithelial
cells
Mesoderm
osteoid
tissue
Endoderm
epithelium
n
n
o
o
e
e
e

Figure 4.3: Teratoma formation assay. Two teratomas resulting from subcutaneous injection of targeted cells were harvested from nude mice. (A) Teratomas were initiated by gene targeted ES cells as demonstrated by GFP fluorescence. Teratomas were also photographed under a Cy3 filter to show background fluorescence. (B) Sections were cut and stained with hematoxylin and eosin to identify tissue structures indicated in figure. n, neuroepithelial cells; o, osteoid tissue; e, epithelium.

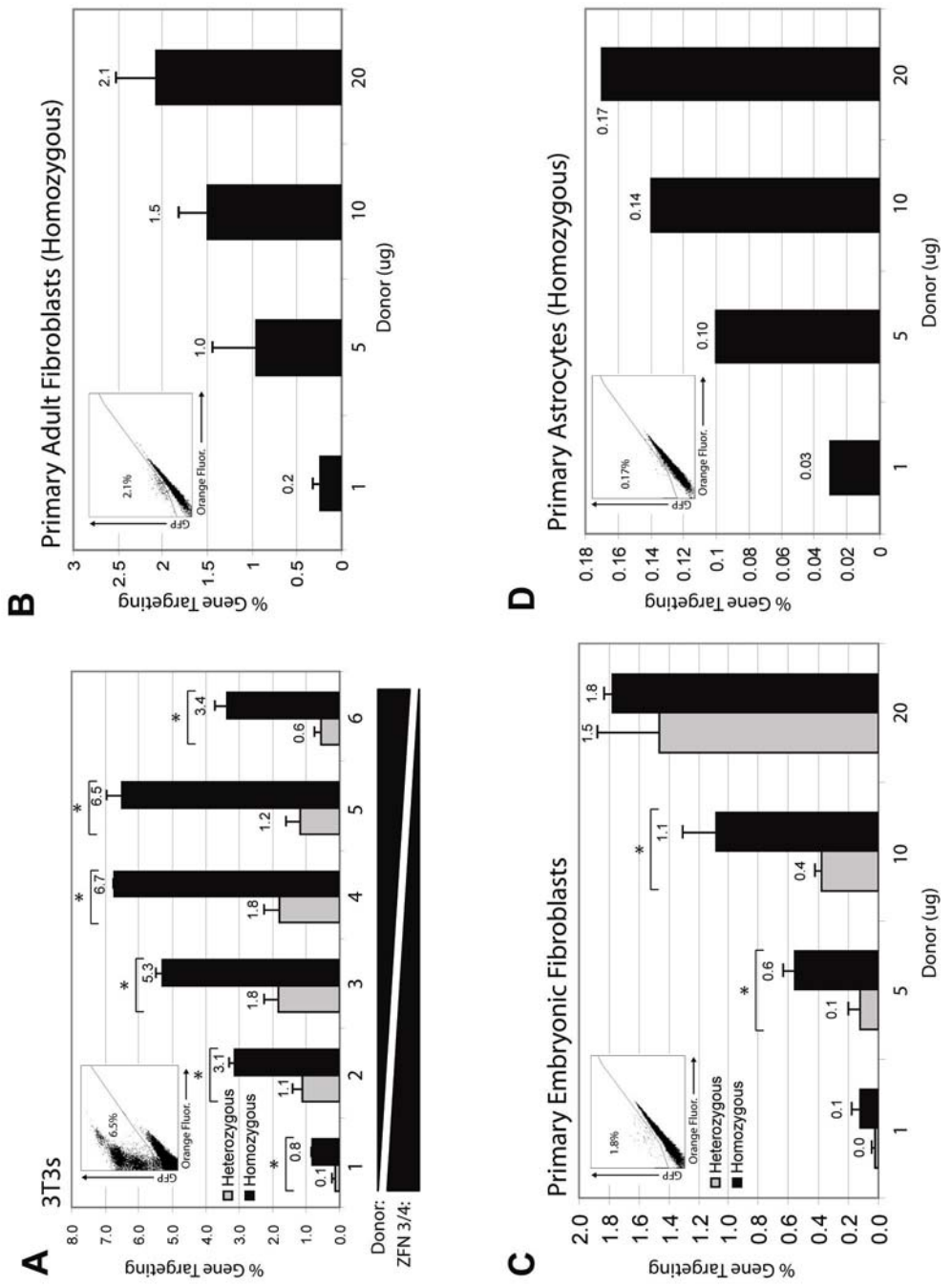

Figure 4.4: ZFN-mediated gene targeting in primary cells. (A) Gene targeting in homozygous and heterozygous ROSA-3T3s: transfections were performed using

Lipofectamine 2000 with the indicated amounts of donor plasmid and ZFNs. The next day, media was changed and on day 4 gene targeting events were analyzed. (B,C) Gene targeting in MAF/ MEFs: Transfection of plasmids was performed by nucleofection using 2 ug of each ZFN and the indicated amounts of donor plasmid. Gene targeting events were analyzed 4 days post transfection. (D) Gene targeting in astrocytes: targeting was performed in the same manner as MAFs/ MEFs. In the upper left of each graph is a representative flow cytometry plot after targeting in which GFP fluorescence is measured in the y-axis and background orange fluorescence along the x-axis. The number in the left corner of the flow plot is the percentage of $GFP^{(+)}$ cells. The corrected ROSA-3T3 cells show much higher GFP fluorescence than the primary cells demonstrating that while the ROSA26 locus is ubiquitously expressed, expression levels from the locus vary significantly depending on the cell type. Data are presented as mean $\pm$ SEM (*, $P<0.05$).

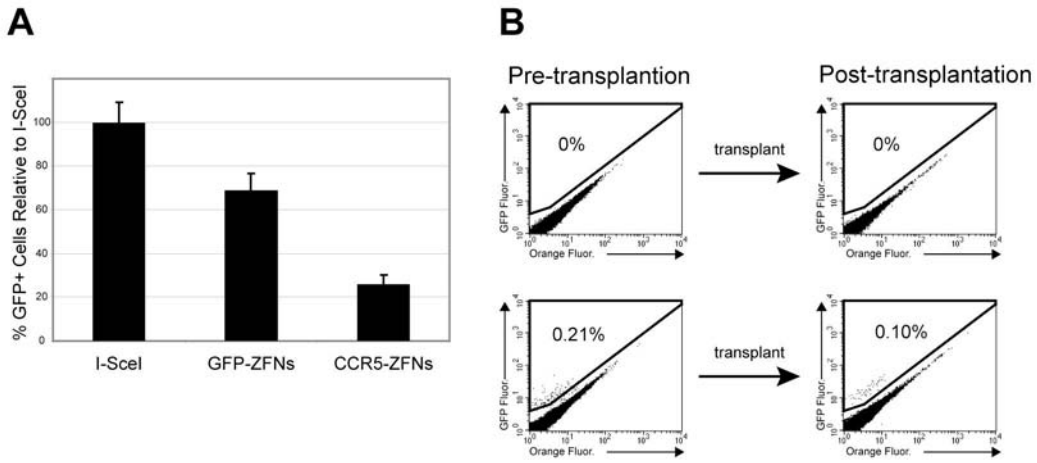

Figure 4.5: Toxicity assay and transplantation of gene targeted adult fibroblasts. (A) Toxicity of GFP-ZFNs in adult fibroblasts measured using a fluorescence reporter assay. Cells were transfected with a tdTomato reporter along with an I-SceI, GFP-ZFN, or CCR5-ZFN expression plasmids. ZFN toxicity in cells is reported as fluorescence lost (due to cytotoxic effect) compared to cells transfected with the I-SceI expression plasmid. (B) For transplantation, adult fibroblasts underwent gene correction by nucleofection of 2 ug of each ZFN expression plasmid and 10 ug of donor plasmid. Gene targeting was measured by flow cytometry immediately before transplantation 6 days after nucleofectionon. Fibroblasts were then injected subcutaneously in a Matrigel matrix. Two weeks after transplantion, the Matrigel plug and surrounding skin was excised, cells dissociated, and plated to allow for fibroblast enrichment from other host-derived infiltrating cell types. On day 6, cells were harvest and analyzed using flow cytometry.

ZFN abbreviation	Finger-1	Finger-2	Finger-3
Target site	GGT	GAT	GAA
GFP1	TRQKLGV	VAHNLTR	QHPNLTR
GFP3	TKQKLDV	VAHNLTR	QHPNLTR
Target site	GGC	GAC	GAC
GFP2	APSKLDR	DRSNLTR	EGGNLMR
GFP4	APSKLDR	DRSNLTR	DQGNLIR

Table 4.1: Amino acid sequences to ZFNs. Table indicates ZFN name along with the 7 amino acid stretch that mediates DNA binding and the 3bp DNA sequence to which it binds. All ZFNs used in this work contained the wild-type nuclease domain and were not modified to prevent homodimerization.

Discussion

ZFNs have now been used in a wide variety of situations to create precise genome modifications (Carroll 2008). The precise genome modifications caused by ZFNs fall into two general classes. The first is to create small insertions/ deletions at a specific locus by the mutagenic repair of a ZFN induced double-strand break. While the exact mutation cannot be controlled using this strategy, the precise location of the mutation is controlled by the specificity of the ZFN induced double-strand break. This strategy is being increasingly used to create knockout cell lines and organisms in which the ability to efficiently create such knockouts was not previously available (Bibikova, Golic et al. 2002; Lloyd, Plaisier et al. 2005; Morton, Davis et al. 2006; Doyon, McCammon et al. 2008; Meng, Noyes et al. 2008; Santiago, Chan et al. 2008; Geurts,

Cost et al. 2009; Shukla, Doyon et al. 2009; Townsend, Wright et al. 2009). In addition, this strategy has been used to create targeted mutations in the CCR5 gene in human T-cells, thereby creating a population of T-cells that are resistant to HIV infection (Perez, Wang et al. 2008)—a strategy that has now entered a Phase I clinical trial (Clinicaltrials.gov identifier NCT00842634). Theoretically this approach could also be used to treat dominant genetic diseases by selectively mutating the dominantly acting allele. The second way to create precise genome modifications (gene targeting) is to use the ZFNs to create a gene specific double-strand break and then have the cell repair that break by homologous recombination using an introduced donor sequence as the template for repair. Using this strategy both the specific site of the genome change and the specific sequence change can be controlled. In addition to the double-precision of gene targeting, it also has the advantage that one can precisely create both small (single nucleotide) or large (the insertion of full a transgene cassette) changes in the genome (Porteus 2006; Moehle, Rock et al. 2007) and reviewed in (Porteus and Carroll 2005). In addition to being able to create inactivating mutations as the first strategy does, this strategy also allows the controlled integration of a transgene or the direct correction of a disease causing mutation and could theoretically be used to treat recessive genetic diseases. ZFN mediated gene targeting has now been used in a wide variety of different cell types at a large number of different loci. These include the modification of the IL2RG gene in various cell lines (Urnov, Miller et al. 2005), the modification of the CCR5 gene in a variety of cell lines including primary human embryonic stem cells and hematopoietic stem cells (Perez, Wang et al. 2008), the correction of GFP reporter genes and the modification of the PIG-A gene in human embryonic stem (hES) and human

induced pluripotent stem (iPS) cells (Zou, Maeder et al. 2009), and the modification of the OCT4 (POU5F1) and AAVS1 locus in hES and iPS cells (Hockemeyer, Soldner et al. 2009).

As part of the effort to translate the use of ZFN mediated gene targeting to clinical use, we have generated a mouse model of a generic recessive disease by knocking-in a mutated GFP gene into the murine ROSA26 locus. Our goal is to generate a model in which one can either test the ability to use ZFN mediated gene targeting to correct mutations directly *in vivo* (similar to work performed by Miller et al. (2006) to assess AAV gene targeting (Miller, Wang et al. 2006)), or to mimic a paradigm in which patient cells are purified and then precisely modified *ex vivo* before transplanting the modified cells back into the patient. Here we report a first step in establishing the *ex vivo* cell modification paradigm by demonstrating that primary embryonic fibroblasts, primary adult fibroblasts, and primary astrocytes can all be isolated from the transgenic mouse line and the mutation in the GFP gene efficiently corrected using ZFN mediated gene targeting *ex vivo*. The rates of correction in primary fibroblasts (~2%) are of a sufficient magnitude to suggest clinical utility in the appropriate situation. Moreover, the gene corrected adult fibroblasts could be successfully transplanted back into an immunocompetent mouse where they both survived and retained their corrected phenotype. Unlike previous studies where virally modified fibroblasts were transplanted back into a recipient mouse, gene correction allows for the corrected gene to be driven by its own promoter, thus preventing the silencing that can occur with transgenes that are driven by viral elements.

The use of gene targeting by homologous recombination without the use of nucleases in murine ES cells is a well-established procedure. Here we demonstrate that ZFNs can stimulate gene targeting in >1% of cells without selection which is an order of magnitude or more than the rate of targeting in murine ES cells using I-SceI (Donoho, Jasin et al. 1998) or in human ES or iPS cells (Hockemeyer, Soldner et al. 2009; Zou, Maeder et al. 2009). Thus this efficiency is high enough that one might be able to use ZFNs to create ES cells with extremely precise genetic modifications without using selectable markers.

This work also highlights an alternative strategy to creating gene corrected iPS cells. In prior work, investigators have first converted fibroblasts into iPS cells and then used gene targeting by homologous recombination in the iPS cells to correct a mutation (Hanna, Wernig et al. 2007). The relatively high rates of gene correction stimulated by zinc finger nucleases in primary fibroblasts that we demonstrate here suggests that one could first correct disease causing mutations in patient fibroblasts, and subsequently convert those gene corrected fibroblasts into iPS cells. The advantages and disadvantages of whether to correct disease causing mutations at the primary cell stage or after conversion to iPS cells will likely be assessed in future studies.

This work shows that in primary cells, just as in cell lines, that the best gene targeting rates are obtained when the optimal mixture of ZFN expression plasmids and donor plasmid are introduced. Our finding that optimal targeting rates are achieved when the donor plasmid is at least 10-fold more abundant than the ZFN expression plasmid suggests that keeping the two elements separate so that each can be introduced at their

optimal amounts will be important in achieving optimal targeting rates in different cell types.

In summary, we have developed a mouse model of a generic recessive disease in which ZFN mediated gene targeting can be studied in any cell of the mouse. A particular advantage of this model is that correctly targeted cells can easily be quantified and isolated for subsequent experiments using flow cytometry. In this paper, we demonstrated gene targeting and quantified the rates within several cell types from this mouse model and that targeted adult fibroblasts, a potentially clinically relevant cell type, can be transplanted back into a mouse. Our current work focuses on isolating stem cells (such as hematopoietic stem cells, mesenchymal stem cells, muscle progenitor cells, and adipocyte precursor cells) from the mouse followed by performing gene targeting and subsequently transplanting the targeted stem cells back into a mouse, thus extending our current results of correcting and then transplanting a somatic cell. This study represents an important step in developing a paradigm for gene correction based gene therapy in an animal model prior to human clinical trials.

Materials and Methods

Generation of ROSA26-GFP* targeting construct

We constructed the ROSA26-GFP* targeting construct by first destroying the XhoI and XbaI sites present in the insert of the published GFP* reporter gene. This was done by digesting GFP* with XhoI followed by blunting of ends with Klenow. Next the XbaI site was destroyed in the same manner creating GFP*-B. We then used PCR to fuse a XhoI-

XbaI-ClaI linker to GFP* using the follower primers: GFP*-XhoI-XbaI-ClaI-Forward: 5'-GTCTCGAGTCTAGAATCGATATGGTGAGCAAGGGCGAGG- 3' and GFP*-XhoI-Reverse: 5'-GACTCGAGTTACTTGTACAGCTCGTCCATGCCG- 3'. The vector GFP*-B-Pgk-Neo was created by ligating GFP*-B into the XhoI site of Pgk-Neo-Pgk-PA, giving GFP*-B-Neo. Next a splice acceptor with ClaI ends was generated by PCR of pSAβGeo with primers SAClaI-Forward: 5'-GGATCGATATCTGTAGGGCGCAGTAGTCCAG-3' and SAClaI-Reverse: 5'-GTATCGATACCGTCGATCCCCACTG-3'. We digested the PCR product with ClaI and ligated it into the ClaI site of GFP*-B-Neo, generating SA-GFP*-B-Neo. Finally this vector was digested with XbaI, and ligated into the XbaI site of the ROSA26-1 targeting vector. All restriction enzymes were ordered from New England Biolabs Inc. The vectors pSAβGeo and pROSA26-1 were generous gifts from Dr. Philippe Soriano (Mount Sinai, New York, NY).

Generation of Reporter Mice

All experiments involving mice were approved by the IACUC at the University of Texas Southwestern Medical Center. Targeting of the GFP* reporter construct to the ROSA26 locus was performed by the UT Southwestern Transgenic core facility. Briefly, 50ug of targeting vector was linearized with KpnI and electroporated into $1x10^7$ 129/SvEvTac (SM-1) ES cells. Selection for targeted clones was performed using G418 at 250ug/ml. Resistant clones were picked and screened using southern analysis (Figure 4.1). Correctly targeted ES cells were injected into pseudo pregnant C57/Bl6 females. Chimeric offspring were bred to identify mice transmitting the ROSA26-GFP* allele to their progeny.

Gene targeting in ES cells

ES cells were cultured in ESLX media consisting of ES-DMEM, 20% ES Qualified FBS, 1X non-essential amino acids, 1X nucleosides, 1000 U/ml ESGRO LIF all purchased from Chemicon (Chemicon, Billerica, MA), 2mM L-Glutamine, 1X Pen/ Strep (Invitrogen, Carlsbad, CA), and 0.12mg/ ml sodium pyruvate and 0.1mM BME (Sigma-Aldrich, St. Louis, MO). Before reaching confluency, cells were harvested and plated 100,000 cells per well in a 24 well gelatinized plate with and without 100nM vinblastine. We then transfected ES cells using Lipofectamine 2000 (Invitrogen, Carlsbad, CA) with the indicated amount of donor and ZFN plasmids. Two sets of ZFNs designed to target GFP were compared. Both sets bind to the same recognition site; however, their amino acid sequences differ (Table 1). Fifteen hours post-transfection media was changed to ESLX without vinblastine. Gene targeting events were analyzed four days post transfection using a FACSCalibur (Becton Dickinson).

Gene targeting and teratoma formation in ES cells

Gene targeting experiments in ES cells were performed by nucleofection using the Mouse ES Cell kit (Lonza, Switzerland, Cat. VPH-1001). Briefly $1.5x10^6$ cells were nucleofected with 1.8 ug donor substrate and 1.6 ug of each GFP-ZFN using program A-30, giving an initial targeting rate of 0.07%. Cells underwent two rounds of sorting for targeted GFP^+ cells. For teratoma analysis, nude mice were sublethally irradiated with 350 rads using a cesium-137 source. The following day mice received subcutaneous injects in the hind flank with $1.3x10^6$ ES cells suspended in PBS. Teratomas were harvested 16-20 days after injections, photographed for GFP and Cy3 fluorescence using a Zeiss Stemi-11 Stereoscope equipped with an epifluorescence illuminator and

Optronics Macrofire CCD camera. Teratomas were subsequently fixed in 4% paraformaldehyde for 48hrs and paraffin processed for histopathologic analysis. Resulting H&Es stains were photographed on a Leica DM2000 upright microscope with standard bright-field optics and Optronics Microfire CCD camera.

Gene targeting in primary cells

MEFs: MEFs were isolated from E13.5 embryos using the WiCell protocol (WiCell Research Institute). Cells were cultured in DMEM, 10% FBS, Pen/Strep, and L-Glut. Before senescence occurred, cells were harvested and nucleofected in triplicate using the Basic Fibroblast kit (Lonza, Switzerland, Cat. VPI-1002) with program U-23. Each nucleofection consisted of $4x10^5$ cells plus 2 ug of each GFP-ZFN and 1, 5, 10 or 20 ug of donor plasmid. We analyzed gene targeting events on day 4 post-transfection.

MAFs: MAFs were isolated from the ears of 3-6 month old mice and cultured in DMEM, 20% FBS, Pen/Strep, L-Glut, Fungizone and 1X non-essential amino acids. $5x10^5$ cells per sample were nucleofected per sample using the Basic Fibroblast kit (Lonza, Switzerland, Cat. VPI-1002) with program U-23 and analysis performed on day 4 post-transfection.

Astrocytes: Astrocytes were isolated from newborn mice and cultured in the same media used for MAFs. Nucleofections were performed on $5x10^5$ cells per sample using the Mouse Astrocyte kit (Lonza, Switzerland, Cat. VPG-1006) with program T-20.

Transplantation of gene targeted primary adult fibroblasts and cell survival assay

Toxicity of GFP-ZFNs was measured as previously described (Pruett-Miller, Connelly et al. 2008). Briefly, $1x10^6$ primary fibroblasts were nucleofected with 4 ug of a tdTomato expression plasmid along with 4 ug of an I-SceI expression plasmid or 2ug of each GFP-

ZFN, and replated in a 12 well plate. On day 2, a portion of the cells was analyzed using flow cytometry and the remainder of cells re-plated in a 12 well plate. On day 6, cells were harvested and analyzed for tdTomato expression using flow cytometry. Toxicity compared to I-SceI was determined by first calculating the change in GFP expression from day 6 to day 2 of the I-SceI transfected cells and then the GFP-ZFN transfected cells ($\Delta GFP_{nuclease} = \%GFP^{+}_{Day6} / \%GFP^{+}_{Day2}$). Toxicity compared to I-SceI was then calculated as Toxicity = $\Delta GFP_{ZFNs} / \Delta GFP_{I\text{-}SceI}$. Analysis of toxicity for ZFNs targeting the CCR5 gene "CCR5 ZFNs" was performed in parallel using 2 ug of each CCR5 ZFN.

For transplantation experiments, fibroblasts underwent gene targeting by nucleofection as described above, using 2 ug of each ZFN expression plasmid and 10 ug of donor plasmid. Cells were analyzed by flow cytometry 6 days after nucleofection, immediately before transplantation. $9.3x10^5$ fibroblasts, of which 0.21% were GFP+, were injected subcutaneously in a Matrigel (BD Biosciences, San Jose, CA) matrix in the back of a wild type, immunocompetent mouse not containing our *ROSA26-GFP** transgene. The Matrigel plug and surrounding tissue were excised 2 weeks later. Cells were dissociated by incubation in collagenase/dispase (25mg/ml) (Roche Diagnostics, Indianapolis, IN) for 1 hour at 37 degrees. One ml MAF media was then added and cells incubated overnight at 37 degrees. The next morning, cells were triturated, filtered with a 70uM cell strainer (BD Biosciences, San Jose, CA) and plated in 24 well plates. On day 6 cells were harvest and analyzed using flow cytometry.

Acknowledgements

We would like to thank (and so we will) Dr. Robert Hammer and Robin Nguyen at the UT Southwestern Transgenic Core for creation of the ROSA26 knock-in ES cells, derivation of the transgenic mouse and ES cell culture protocols. We also thank Dr. James Richardson and John Shelton at the JAR Molecular Pathology Core for help with photography and performing histological analysis of the teratomas. We thank Matthew Foglia for his careful reading and editing of the manuscript and Dr. Robert Bachoo for astrocyte isolation instruction. Finally we thank Dr. Philippe Soriano for the pROSA26-1 and pSAβGeo vectors. The work in the Porteus lab is supported by National Institutes of Health (NIH) grants PN2EY018244 (a Nanomedicine Development Center grant as part of the Director's initiative), R01 HL079295, and K08 H1070268, a career development award from the Burroughs Welcome Fund, and state of Texas funding through the University of Texas Southwestern Medical Center. JB was supported by the UTSW Medical Scientist Training Program. SPM was supported by a National Institutes of Health pharmacology training grant at the University of Texas Southwestern Medical Center.

CHAPTER V:
CONCLUSIONS AND FUTURE DIRECTIONS

High Relative Rate of Gene Targeting in DT40 Cells

Chicken DT40 cells are widely used experimentally due to the relative ease in which genes may be knocked-in/ knocked-out through gene targeting by homologous recombination. The gene targeting rate within this cell line is much higher than that found in most other primary cells, and for this reason DT40 cells are of valuable interest to the field of gene targeting. We undertook studies to determine the reason why DT40 cells have a high rate of gene targeting with the idea that the results may lead to new strategies to increase targeting rates in other primary cells. In studying DT40 cells, we learned that although their absolute gene targeting rate is not much higher than other cell lines, their rates of random integration are significantly lower. When DNA is introduced into a cell (as is necessary to perform gene targeting), there are generally two possible end fates for the DNA. The first is that the DNA remains unintegrated within the cell and, as time proceeds and the cell undergoes subsequent divisions, the DNA will be lost/ degraded. The second possibility is that the DNA will randomly integrate into the genome. This is undesirable because of the potential consequences associated with random integration. The DNA may integrate into tumor suppressor genes, resulting in their inactivation, or may integrate near a proto-oncogene, resulting in its subsequent activation. In either case, these random integrations could result in pushing the cell towards a cancerous state. To make gene targeting (and gene therapy) as safe as possible, it is necessary to increase the targeting rate while maintaining or reducing the random

integration rate. Eliminating cells which have randomly integrated the transfected DNA can in part be achieved by incorporating into the introduced DNA markers which may be selected against (such as the Herpes Simplex Virus thymidine kinase gene, or the Diphtheria Toxin A gene).

We further showed that the low random integration rate was due to an inability of the cells to repair DNA breaks containing non-compatible ends through the NHEJ pathway. This is supported by observations that DT40 cells deficient in NHEJ (such as Ku70-/- cells) have a lower random integration rate as well as an increased gene targeting rate compared to wild-type cells. From these results, we next tried inhibiting NHEJ using the compound NU7026 (which inhibits DNA-PK_{cs}), to determine whether random integration would be reduced and/ or the gene targeting rate increased. Interestingly, although this compound inhibited end-joining in 293 cells, there was no decrease in the random integration rate or increase in gene targeting. This suggests that inhibiting specific proteins (such as Ku70) may be necessary to reduce random integration and increase gene targeting frequencies.

Gene Targeting in a Mouse Model of a Generic Recessive Genetic Disease.

As mentioned above, ZFN mediated gene targeting holds great potential in the field of gene therapy. The technology is new though and may carry potential risks. For this reason, we established a mouse model of a generic recessive genetic disease which may be used as a paradigm for gene therapy prior to clinical trials. With this model, we demonstrated ZFN mediated gene targeting in a variety of primary cells, studied the effects of gene targeting on the pluripotency of ES cells, and determined whether targeted

fibroblasts remained competent for transplantation. We found that all cell types in which we could transfect the requisite plasmids were able to undergo gene targeting at frequencies of 0.17-6 %. Furthermore, gene targeted ES cells retained their pluripotency, and gene targeted fibroblasts could be transplanted into recipient mice where they remained viable. Our results continue to support the application of ZFN mediated gene targeting towards gene therapy.

In these studies, we targeted cell types that currently have a modest utility towards gene therapy. Gene therapy using genetically modified ES cells remains controversial, and currently no protocols exist for the differentiation of large numbers of ES cells into therapeutic cell types. Moreover, any cells which remain undifferentiated and are transplanted may form teratomas in the patient. However, our demonstration that gene targeting can be performed in primary fibroblasts, and that these cells can be transplanted back into a host recipient parallels a clinical study already performed with patients afflicted with hemophilia A. Selden et al. removed fibroblasts from Hemophilia A patients, and randomly integrated a cassette expressing Factor VIII into the cells. Next the modified cells were re-transplanted back into the patients. Over a 10 month period, the patients showed increased levels of Factor VIII. The patients also showed reduced episodes of spontaneous bleeding as well as a decreased dependence on exogenous Factor VIII. These results were temporary however, and the reason the effect was un-sustained remained undetermined. Nonetheless, this study suggests that performing gene targeting in cells at specific disease-associated sites within the genome can offer clinical benefits. Our future studies are aimed at performing gene targeting in therapeutically relevant cell types such as hematopoietic stem cells. Diseases such as X-SCID, ADA-SCID, and

sickle cell anemia are diseases of the blood. Demonstrating that hematopoietic stem cells can undergo gene targeting and be transplanted back into a recipient would offer great hope for the potential cure of these diseases. Thus far, gene targeting in hematopoietic cells has proven difficult due to the inefficiency of introducing donor and zinc finger nuclease expression plasmids into these cells through processes such as electroporation, nucleofection, and transfection techniques. The most promising delivery method is the use of non-integrating lentiviral vectors. Lentiviral vectors readily infect hematopoietic stem cells, but so far (in our hands) have failed to demonstrate the ability to stimulate gene targeting. Ongoing research in our lab is focused on performing zinc finger nuclease mediated gene targeting in these cells. Another potential use for gene targeting is to target transgene cassettes expressing therapeutic proteins to specific sites within the genome. Not all mutations resulting in disease are simple point mutations or small deletions/ additions. Chromosomal rearrangements such as inversions and large deletions also occur. In these instances, it may not be possible to correct the endogenous disease causing mutation. An alternative strategy for the treatment of these diseases is to target a cassette carrying a transgene expressing the desired protein to a “safe harbor” site (one that, when altered, has no deleterious effects on the cell). Our lab is studying this aspect of gene targeting/ gene addition using the mouse model described above, with the ultimate goal of demonstrating that we can target transgenes (such as Factor VIII expression cassettes) to the mutated GFP site, and that these cells can be transplanted back into a recipient where they engraft and express the desired protein.

The future use of ZFN mediated gene targeting in the treatment of monogenic diseases looks promising. High rates of targeting have been demonstrated in a variety of

cell types by multiple groups including our own. The mouse model developed in this work will prove valuable in studying the applicability of gene targeting in numerous cell types, transplantation of targeted cells, and can further be used to study the safety of these processes as well.

REFERENCES

Aiuti, A., F. Cattaneo, et al. (2009). "Gene therapy for immunodeficiency due to adenosine deaminase deficiency." N Engl J Med **360**(5): 447-458.

Aiuti, A., S. Slavin, et al. (2002). "Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning." Science **296**(5577): 2410-2413.

Arnould, S., P. Chames, et al. (2006). "Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets." J Mol Biol **355**(3): 443-458.

Ashworth, J., J. J. Havranek, et al. (2006). "Computational redesign of endonuclease DNA binding and cleavage specificity." Nature **441**(7093): 656-659.

Ausubel, F. M., R. Brent, et al., Eds. (1996). Current Protocols in Molecular Biology. Boston, MA, John Wiles and Sons, Inc.

Bainbridge, J. W., A. J. Smith, et al. (2008). "Effect of gene therapy on visual function in Leber's congenital amaurosis." N Engl J Med **358**(21): 2231-2239.

Beumer, K., G. Bhattacharyya, et al. (2006). "Efficient gene targeting in Drosophila with zinc-finger nucleases." Genetics **172**(4): 2391-2403.

Bezzubova, O., A. Silbergleit, et al. (1997). "Reduced X-ray resistance and homologous recombination frequencies in a RAD54-/- mutant of the chicken DT40 cell line." Cell **89**(2): 185-193.

Bibikova, M., K. Beumer, et al. (2003). "Enhancing gene targeting with designed zinc finger nucleases." Science **300**(5620): 764.

Bibikova, M., D. Carroll, et al. (2001). "Stimulation of homologous recombination through targeted cleavage by chimeric nucleases." Mol Cell Biol **21**(1): 289-297.

Bibikova, M., M. Golic, et al. (2002). "Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases." Genetics **161**(3): 1169-1175.

Bitinaite, J., D. A. Wah, et al. (1998). "FokI dimerization is required for DNA cleavage." Proc Natl Acad Sci U S A **95**(18): 10570-10575.

Brenneman, M., F. S. Gimble, et al. (1996). "Stimulation of intrachromosomal homologous recombination in human cells by electroporation with site-specific endonucleases." Proc Natl Acad Sci U S A **93**(8): 3608-3612.

Buerstedde, J. M. and S. Takeda (1991). "Increased ratio of targeted to random integration after transfection of chicken B cell lines." Cell **67**(1): 179-188.

Burma, S., B. P. Chen, et al. (2006). "Role of non-homologous end joining (NHEJ) in maintaining genomic integrity." DNA Repair (Amst) **5**(9-10): 1042-1048.

Carroll, D. (2008). "Progress and prospects: zinc-finger nucleases as gene therapy agents." Gene Ther **15**(22): 1463-1468.
Cartier, N., S. Hacein-Bey-Abina, et al. (2009). "Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy." Science **326**(5954): 818-823.
Cathomen, T. and J. K. Joung (2008). "Zinc-finger nucleases: the next generation emerges." Mol Ther **16**(7): 1200-1207.
Cavazzana-Calvo, M., C. Lagresle, et al. (2005). "Gene therapy for severe combined immunodeficiency." Annu Rev Med **56**: 585-602.
Chandrasegaran, S. and J. Smith (1999). "Chimeric restriction enzymes: what is next?" Biol Chem **380**(7-8): 841-848.
Choo, Y. and A. Klug (1994). "Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions." Proc Natl Acad Sci U S A **91**(23): 11168-11172.
Choo, Y. and A. Klug (1994). "Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage." Proc Natl Acad Sci U S A **91**(23): 11163-11167.
Choulika, A., A. Perrin, et al. (1995). "Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccaromyces cerevisiae." Molecular and Cellular Biology **15**(4): 1968-1973.
Choulika, A., A. Perrin, et al. (1995). "Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae." Mol Cell Biol **15**(4): 1968-1973.
Doetschman, T., R. G. Gregg, et al. (1987). "Targetted correction of a mutant HPRT gene in mouse embryonic stem cells." Nature **330**(6148): 576-578.
Donoho, G., M. Jasin, et al. (1998). "Analysis of gene targeting and intrachromosomal homologous recombination stimulated by genomic double-strand breaks in mouse embryonic stem cells." Mol Cell Biol **18**(7): 4070-4078.
Doyon, Y., J. M. McCammon, et al. (2008). "Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases." Nat Biotechnol **26**(6): 702-708.
Durai, S., M. Mani, et al. (2005). "Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells." Nucleic Acids Res **33**(18): 5978-5990.
Gaspar, H. B., K. L. Parsley, et al. (2004). "Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector." Lancet **364**(9452): 2181-2187.
Geurts, A. M., G. J. Cost, et al. (2009). "Knockout rats via embryo microinjection of zinc-finger nucleases." Science **325**(5939): 433.

Greisman, H. A. and C. O. Pabo (1997). "A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites." Science **275**(5300): 657-661.

Guirouilh-Barbat, J., S. Huck, et al. (2004). "Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells." Mol Cell **14**(5): 611-623.

Hacein-Bey-Abina, S., F. Le Deist, et al. (2002). "Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy." N Engl J Med **346**(16): 1185-1193.

Hacein-Bey-Abina, S., C. von Kalle, et al. (2003). "A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency." N Engl J Med **348**(3): 255-256.

Hacein-Bey-Abina, S., C. Von Kalle, et al. (2003). "LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1." Science **302**(5644): 415-419.

Hanna, J., M. Wernig, et al. (2007). "Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin." Science **318**(5858): 1920-1923.

Hasty, P., J. Rivera-Perez, et al. (1991). "Target frequency and integration pattern for insertion and replacement vectors in embryonic stem cells." Mol Cell Biol **11**(9): 4509-4517.

Hickson, I., Y. Zhao, et al. (2004). "Identification and characterization of a novel and specific inhibitor of the ataxia-telangiectasia mutated kinase ATM." Cancer Res **64**(24): 9152-9159.

Hinnen, A., J. B. Hicks, et al. (1978). "Transformation of yeast." Proc Natl Acad Sci U S A **75**(4): 1929-1933.

Hockemeyer, D., F. Soldner, et al. (2009). "Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases." Nat Biotechnol **27**(9): 851-857.

Houtgraaf, J. H., J. Versmissen, et al. (2006). "A concise review of DNA damage checkpoints and repair in mammalian cells." Cardiovasc Revasc Med **7**(3): 165-172.

Hurt, J. A., S. A. Thibodeau, et al. (2003). "Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection." Proc Natl Acad Sci U S A **100**(21): 12271-12276.

Isalan, M., Y. Choo, et al. (1997). "Synergy between adjacent zinc fingers in sequence-specific DNA recognition." Proc Natl Acad Sci U S A **94**(11): 5617-5621.

Isalan, M., A. Klug, et al. (2001). "A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter." Nat Biotechnol **19**(7): 656-660.

Jamieson, A. C., S. H. Kim, et al. (1994). "In vitro selection of zinc fingers with altered DNA-binding specificity." Biochemistry **33**(19): 5689-5695.
Jasin, M., J. de Villiers, et al. (1985). "High frequency of homologous recombination in mammalian cells between endogenous and introduced SV40 genomes." Cell **43**(3 Pt 2): 695-703.
Karran, P. (2000). "DNA double strand break repair in mammalian cells." Curr Opin Genet Dev **10**(2): 144-150.
Kim, Y. G., J. Cha, et al. (1996). "Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain." Proc Natl Acad Sci U S A **93**(3): 1156-1160.
Kim, Y. G. and S. Chandrasegaran (1994). "Chimeric restriction endonuclease." Proc Natl Acad Sci U S A **91**(3): 883-887.
Kolachana, P., V. V. Subrahmanyam, et al. (1993). "Benzene and its phenolic metabolites produce oxidative DNA damage in HL60 cells in vitro and in the bone marrow in vivo." Cancer Res **53**(5): 1023-1026.
Li, L., L. P. Wu, et al. (1992). "Functional domains in Fok I restriction endonuclease." Proc Natl Acad Sci U S A **89**(10): 4275-4279.
Lin, Y. and A. S. Waldman (2001). "Promiscuous patching of broken chromosomes in mammalian cells with extrachromosomal DNA." Nucleic Acids Res **29**(19): 3975-3981.
Lloyd, A., C. L. Plaisier, et al. (2005). "Targeted mutagenesis using zinc-finger nucleases in Arabidopsis." Proc Natl Acad Sci U S A **102**(6): 2232-2237.
Lombardo, A., P. Genovese, et al. (2007). "Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery." Nat Biotechnol **25**(11): 1298-1306.
Maeder, M. L., S. Thibodeau-Beganny, et al. (2008). "Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification." Mol Cell **31**(2): 294-301.
Maguire, A. M., K. A. High, et al. (2009). "Age-dependent effects of RPE65 gene therapy for Leber's congenital amaurosis: a phase 1 dose-escalation trial." Lancet **374**(9701): 1597-1605.
Manno, C. S., A. J. Chew, et al. (2003). "AAV-mediated factor IX gene transfer to skeletal muscle in patients with severe hemophilia B." Blood **101**(8): 2963-2972.
Manno, C. S., G. F. Pierce, et al. (2006). "Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response." Nat Med **12**(3): 342-347.
Mao, Z., M. Bozzella, et al. (2008). "Comparison of nonhomologous end joining and homologous recombination in human cells." DNA Repair (Amst) **7**(10): 1765-1771.
Meng, X., M. B. Noyes, et al. (2008). "Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases." Nat Biotechnol **26**(6): 695-701.

Merrihew, R. V., K. Marburger, et al. (1996). "High-frequency illegitimate integration of transfected DNA at preintegrated target sites in a mammalian genome." Mol Cell Biol **16**(1): 10-18.
Miller, D. G., L. M. Petek, et al. (2004). "Adeno-associated virus vectors integrate at chromosome breakage sites." Nat Genet **36**(7): 767-773.
Miller, D. G., E. A. Rutledge, et al. (2002). "Chromosomal effects of adeno-associated virus vector integration." Nat Genet **30**(2): 147-148.
Miller, D. G., P. R. Wang, et al. (2006). "Gene targeting in vivo by adeno-associated virus vectors." Nat Biotechnol **24**(8): 1022-1026.
Miller, J. C., M. C. Holmes, et al. (2007). "An improved zinc-finger nuclease architecture for highly specific genome editing." Nat Biotechnol **25**(7): 778-785.
Moehle, E. A., J. M. Rock, et al. (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**(9): 3055-3060.
Morton, J., M. W. Davis, et al. (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**(44): 16370-16375.
Nakayama, C., N. Adachi, et al. (1998). "Bleomycin enhances random integration of transfected DNA into a human genome." Mutat Res **409**(1): 1-10.
O'Driscoll, M. and P. A. Jeggo (2006). "The role of double-strand break repair - insights from human genetics." Nat Rev Genet **7**(1): 45-54.
Olive, P. L. and J. P. Banath (1993). "Detection of DNA double-strand breaks through the cell cycle after exposure to X-rays, bleomycin, etoposide and 125IdUrd." Int J Radiat Biol **64**(4): 349-358.
Ott, M. G., M. Schmidt, 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**(4): 401-409.
Palmer, T. D., A. R. Thompson, et al. (1989). "Production of human factor IX in animals by genetically modified skin fibroblasts: potential therapy for hemophilia B." Blood **73**(2): 438-445.
Paques, F. and P. Duchateau (2007). "Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy." Curr Gene Ther **7**(1): 49-66.
Paques, F. and J. E. Haber (1999). "Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae." Microbiol Mol Biol Rev **63**(2): 349-404.
Pavletich, N. P. and C. O. Pabo (1991). "Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A." Science **252**(5007): 809-817.

Perez, E. E., J. Wang, et al. (2008). "Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases." Nat Biotechnol **26**(7): 808-816.
Pierce, A. J., P. Hu, et al. (2001). "Ku DNA end-binding protein modulates homologous repair of double-strand breaks in mammalian cells." Genes Dev **15**(24): 3237-3242.
Pierce, A. J. and M. Jasin (2001). "NHEJ deficiency and disease." Mol Cell **8**(6): 1160-1161.
Pierce, A. J., R. D. Johnson, et al. (1999). "XRCC3 promotes homology-directed repair of DNA damage in mammalian cells." Genes and Development **13**: 2633-2638.
Porteus, M. H. (2006). "Mammalian gene targeting with designed zinc finger nucleases." Mol Ther **13**(2): 438-446.
Porteus, M. H. and D. Baltimore (2003). "Chimeric nucleases stimulate gene targeting in human cells." Science **300**(5620): 763.
Porteus, M. H. and D. Carroll (2005). "Gene targeting using zinc finger nucleases." Nat Biotechnol **23**(8): 967-973.
Potts, P. R., M. H. Porteus, et al. (2006). "Human SMC5/6 complex promotes sister chromatid homologous recombination by recruiting the SMC1/3 cohesin complex to double-strand breaks." EMBO J **25**(14): 3377-3388.
Pruett-Miller, S. M., J. P. Connelly, et al. (2008). "Comparison of zinc finger nucleases for use in gene targeting in mammalian cells." Mol Ther **16**(4): 707-717.
Qiu, X., D. Lu, et al. (1996). "Implantation of autologous skin fibroblast genetically modified to secrete clotting factor IX partially corrects the hemorrhagic tendencies in two hemophilia B patients." Chin Med J (Engl) **109**(11): 832-839.
Raper, S. E., N. Chirmule, et al. (2003). "Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer." Mol Genet Metab **80**(1-2): 148-158.
Raper, S. E., M. Yudkoff, et al. (2002). "A pilot study of in vivo liver-directed gene transfer with an adenoviral vector in partial ornithine transcarbamylase deficiency." Hum Gene Ther **13**(1): 163-175.
Rassool, F. V., T. W. McKeithan, et al. (1991). "Preferential integration of marker DNA into the chromosomal fragile site at 3p14: an approach to cloning fragile sites." Proc Natl Acad Sci U S A **88**(15): 6657-6661.
Rebar, E. J. and C. O. Pabo (1994). "Zinc finger phage: affinity selection of fingers with new DNA-binding specificities." Science **263**(5147): 671-673.

Rijkers, T., J. Van Den Ouweland, et al. (1998). "Targeted inactivation of mouse RAD52 reduces homologous recombination but not resistance to ionizing radiation." Mol Cell Biol **18**(11): 6423-6429.

Ross, W., T. Rowe, et al. (1984). "Role of topoisomerase II in mediating epipodophyllotoxin-induced DNA cleavage." Cancer Res **44**(12 Pt 1): 5857-5860.

Roth, D. A., N. E. Tawa, Jr., et al. (2001). "Nonviral transfer of the gene encoding coagulation factor VIII in patients with severe hemophilia A." N Engl J Med **344**(23): 1735-1742.

Roth, D. B., G. N. Proctor, et al. (1991). "Oligonucleotide capture during end joining in mammalian cells." Nucleic Acids Res **19**(25): 7201-7205.

Roth, D. B. and J. H. Wilson (1986). "Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction." Mol Cell Biol **6**(12): 4295-4304.

Rouet, P., F. Smih, et al. (1994). "Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells." Proc Natl Acad Sci U S A **91**(13): 6064-6068.

Rouet, P., F. Smih, et al. (1994). "Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease." Mol Cell Biol **14**(12): 8096-8106.

Saleh-Gohari, N. and T. Helleday (2004). "Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells." Nucleic Acids Res **32**(12): 3683-3688.

Santiago, Y., E. Chan, et al. (2008). "Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases." Proc Natl Acad Sci U S A **105**(15): 5809-5814.

Sarkaria, J. N., E. C. Busby, et al. (1999). "Inhibition of ATM and ATR kinase activities by the radiosensitizing agent, caffeine." Cancer Res **59**(17): 4375-4382.

Sedivy, J. M. and P. A. Sharp (1989). "Positive genetic selection for gene disruption in mammalian cells by homologous recombination." Proc. Natl. Acad. Sci. USA **86**: 227-231.

Segal, D. J. and D. Carroll (1995). "Endonuclease-induced, targeted homologous extrachromosomal recombination in Xenopus oocytes." Proc Natl Acad Sci U S A **92**(3): 806-810.

Shukla, V. K., Y. Doyon, et al. (2009). "Precise genome modification in the crop species Zea mays using zinc-finger nucleases." Nature **459**(7245): 437-441.

Smih, F., P. Rouet, et al. (1995). "Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells." Nucleic Acids Res **23**(24): 5012-5019.
Smith, A. J. and P. Berg (1984). "Homologous recombination between defective neo genes in mouse 3T6 cells." Cold Spring Harb Symp Quant Biol **49**: 171-181.
Smith, J., M. Bibikova, et al. (2000). "Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains." Nucleic Acids Res **28**(17): 3361-3369.
Smithies, O., R. G. Gregg, et al. (1985). "Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination." Nature **317**(6034): 230-234.
Smithies, O., M. A. Koralewski, et al. (1984). "Homologous recombination with DNA introduced into mammalian cells." Cold Spring Harb Symp Quant Biol **49**: 161-170.
Sonoda, E., M. S. Sasaki, et al. (1999). "Sister chromatid exchanges are mediated by homologous recombination in vertebrate cells." Mol Cell Biol **19**(7): 5166-5169.
Sung, P. and H. Klein (2006). "Mechanism of homologous recombination: mediators and helicases take on regulatory functions." Nat Rev Mol Cell Biol **7**(10): 739-750.
Szczepek, M., V. Brondani, et al. (2007). "Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases." Nat Biotechnol **25**(7): 786-793.
Takata, M., M. S. Sasaki, et al. (1998). "Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells." EMBO Journal **17**(18): 5497-5508.
Todaro, G. J. and H. Green (1963). "Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines." J Cell Biol **17**: 299-313.
Townsend, J. A., D. A. Wright, et al. (2009). "High-frequency modification of plant genes using engineered zinc-finger nucleases." Nature **459**(7245): 442-445.
Urnov, F. D., J. C. Miller, et al. (2005). "Highly efficient endogenous human gene correction using designed zinc-finger nucleases." Nature **435**(7042): 646-651.
van Gent, D. C., J. H. Hoeijmakers, et al. (2001). "Chromosomal stability and the DNA double-stranded break connection." Nat Rev Genet **2**(3): 196-206.
W.H.O. (2010). "Genes and Human Disease." Retrieved 01/20/2010, 2010, from http://www.who.int/genomics/public/geneticdiseases/en/index2.html.

Wah, D. A., J. A. Hirsch, et al. (1997). "Structure of the multimodular endonuclease FokI bound to DNA." Nature **388**(6637): 97-100.
West, S. C., C. Chappell, et al. (2000). "Double-strand break repair in human cells." Cold Spring Harb Symp Quant Biol **65**: 315-321.
Willmore, E., S. de Caux, et al. (2004). "A novel DNA-dependent protein kinase inhibitor, NU7026, potentiates the cytotoxicity of topoisomerase II poisons used in the treatment of leukemia." Blood **103**(12): 4659-4665.
Wu, H., W. P. Yang, et al. (1995). "Building zinc fingers by selection: toward a therapeutic application." Proc Natl Acad Sci U S A **92**(2): 344-348.
Yamaguchi-Iwai, Y., E. Sonoda, et al. (1998). "Homologous recombination, but not DNA repair, is reduced in vertebrate cells deficient in RAD52." Mol Cell Biol **18**(11): 6430-6435.
Zambrowicz, B. P., A. Imamoto, et al. (1997). "Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells." Proc Natl Acad Sci U S A **94**(8): 3789-3794.
Zou, J., M. L. Maeder, et al. (2009). "Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells." Cell Stem Cell **5**(1): 97-110.