Improving Zinc Finger Nucleases - Strategies for Increasing Gene Editing Activities and Evaluating Off-Target Effects

Dissertation Advisor: Dr. J. Keith Joung Cherie Lynn Ramírez

Improving Zinc Finger Nucleases –

Strategies for Increasing Gene Editing Activities and Evaluating Off-Target Effects

Abstract

Zinc finger nucleases (ZFNs) induce double-strand DNA breaks at specific recognition sites. ZFNs can dramatically increase the efficiency of incorporating desired insertions, deletions, or substitutions in living cells. These tools have revolutionized the field of genome engineering in several model organisms and cell types including zebrafish, rats, and human pluripotent stem cells. There have been numerous advances in ZFN engineering and characterization strategies, some of which are detailed in this work.

The central theme of this dissertation is improving the activity and specificity of engineered zinc finger nucleases with the ultimate goal of increasing the safety and efficacy of these tools for human therapy. As a first step, I undertook a large-scale effort to demonstrate that the modular assembly method of ZFN synthesis has a significantly higher failure rate than previously reported in the literature. This strongly suggested that engineering of ZFNs should better account for context-dependent effects among zinc fingers.

The second advance reported in this dissertation is a method for biasing repair of zinc finger protein-induced DNA breaks toward homology-driven rather than error-prone repair in the presence of a donor template. Catalytically inactivating one monomer of a ZFN dimer results in a zinc finger nickase (ZFNickase) whose cleavage preference is directed at only one DNA strand. In human cell reporter assays, these ZFNickases exhibit a higher likelihood of repair by

homology-driven processes, albeit with reduced absolute rates of correction. With further optimization, zinc finger nickases could provide a safer alternative to ZFNs in the context of gene correction therapies.

Third, realizing there was no robust method for determining off-target cleavage sites of ZFNs in a genome-wide manner, I validated a collaborator's novel *in vitro* selection system in human cells by identifying eight new potential off-target cleavage sites for a ZFN pair currently being used in clinical trials. Although it is unlikely these low-frequency mutations would be deleterious to patients, these results demonstrated that ZFNs induced more off-target effects than had been appreciated by previous work in the field. Collectively, the findings of this dissertation have contributed to more robust strategies for designing and evaluating ZFNs.

Table of Contents

Abstract……...iii

Chapter 1. Introduction……1

Chapter 2. “Unexpected failure rates for modular assembly of engineered zinc-fingers”…….....24

Chapter 3. “Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects”...……...48

Chapter 4. “Revealing off-target cleavage specificities of zinc-finger nucleases by *in vitro* selection”……72

Chapter 5. Discussion and Conclusions..……125

Appendix 1. “A synthetic biology framework for programming eukaryotic transcription functions”……....135

Bibliography……176

Para Mami, Papi y Michy:

mis profesores y animadores favoritos

Acknowledgements

To my dearest colleagues, friends, family, and cheerleaders, many of whom I have named below: I would not have the privilege of writing these words without your boundless warmth and encouragement.

Foremost, I need to thank my PhD advisor J. Keith Joung, whom I have always looked up to for wise guidance and uncompromising scientific ideals. Many thanks to the members of the Joung lab for your helpfulness, kindness, and unlimited capacity for fun: James Angstman, Elizabeth Dahlborg, Jenn Foden, Jon Foley, Yanfang Fu, Mat Goodwin, Cyd Khayter, Chelsea Lane, Long Le, Sam Linder, Morgan Maeder, Maureen Reagan, Deepak Reyon, Jeff Sander, Stacey Thibodeau-Beganny, and Shengdar Q. Tsai.

My heartfelt gratitude also goes out to the members of the MGH Molecular Pathology Unit and MGH Cancer Center, past and present, who always made it a little more interesting and pleasant to be at lab, including Susan Smith, James Kim, Dave Langenau, David Louis, Ryan McMullin, Finola Moore, Myron Ignatius, Miguel Rivera, Eveline Schneeberger, and Dennis Sgroi.

To the kind ladies and gentlemen of MGH Police & Security, Vanderbilt Hall Security, MGH Environmental Services, and Boston Cab: My deepest thanks for taking care of me (especially at odd hours) and extending your friendship.

Many thanks to all my scientific collaborators, with particular gratitude to Mike Certo and Andy Scharenberg from the University of Washington, Vikram Pattanayak and David Liu from the Department of Chemistry at Harvard University, and Mo Khalil, Tim Lu, and Jim

Collins of Boston University. It was a pleasure working with all of you and advancing our common interests together.

For their consistently helpful comments and support, I fondly thank my Dissertation Advisory Committee: Dennis Brown, Ralph Scully, and Lee Zou. Thank you to the members of my Dissertation Examination Committee, chaired by Ralph Scully, for taking the time to read my dissertation and pose thoughtful questions during my defense: Martha Bulyk, Grace Gill, and Richard Mulligan.

To the scientists of Sangamo BioSciences: Both through their rigorous scientific investigations and promotion of healthy competition, they have advanced the engineered nuclease field immeasurably.

. . .

My parents, María del Rosario Fátima Matabuena Cascajares and Napoleón Bernardo Ramírez Minvielle, and my sister, Rosario Michelle Ramírez Matabuena, deserve my eternal gratitude for their unwavering support and love. The countless sacrifices, both great and small, they have made for me to achieve this honor are endlessly appreciated. To them I dedicate this work.

Among the dear friends and family I wish to acknowledge for lending a hand or a smile or a shoulder on so many occasions are David Grasso-Ortega, Melody Boeringer-Hartnup and her family, Chris French, Jon Aster, Mike Cameron, Dennis Ho, Jenny Sims, Giovanni Franzesi, James Baker, Alexandra Robinson, Michael Chin, Yousseff Bouhid, and the Gonzalez, Grasso, Hickman, Leahey, Matabuena, Ramirez, and Thurston families.

For your invaluable lessons and advice, I warmly thank my teachers, professors, mentors, and colleagues from St. James School, Celebration School, Valencia College, Rollins College,

Tufts University, and Harvard University, including the Biological & Biomedical Sciences Program, the Division of Medical Sciences, the Office of Diversity and Minority Affairs, the administration and staff of the Graduate School of Arts and Sciences, the GSAS Graduate Student Council, Harvard Integrated Life Sciences, and Dudley House.

To my colleagues and friends affiliated with the Leder Human Biology program: I am so happy to have spent the last few years in your company. Thank you for making the study of translational medicine so engaging and providing copious opportunities for personal and professional development.

I could not imagine Harvard without the wonderful SHURP family that Jocelyn Spragg so selflessly started. My most sincere thanks go to her for taking a chance on me and welcoming me to be part of it. Proud to avoid being a gatekeeper whenever possible, I will always admire her for being one of the most caring and forward-thinking facilitators I may ever know.

To my favorite co-author, Tyler Thomas Hickman: You have been an inspiration to me since I first knew of you. Growing with you and watching you triumph as a scientist has brought me immeasurable joy and pride. I look forward to a lifetime of happiness with you.

"You must be the change you wish to see in the world."

Mahatma Ghandi

Chapter One

Introduction

Parts of this chapter were reprinted with permission from Springer/Kluwer Academic Publishers and originally appeared in "Chapter 5: Engineered Zinc Finger Nucleases for Targeted Genome Editing." Cherie L. Ramirez & J. Keith Joung. *Site directed insertion of transgenes*. Editors: Sylvaine Renault & Philippe Duchateau. 2012.

Since its proposal in the 1970s, the idea of modifying genomic DNA to cure diseases caused by genetic mutations has remained a challenging goal (Wigler et al., 1979). Despite the multitude of advances recently made to improve delivery methods and minimize unwanted side-effects, gene therapy technologies are still moving slowly and hesitantly toward clinical applications (Jensen et al., 2011). Within the last two decades, the field has been galvanized by the discovery that introducing double-strand breaks (DSBs) in DNA at specific genomic locations can significantly stimulate gene repair machinery (Rouet et al., 1994). Using engineered nuclease technologies, natural cellular repair processes have been harnessed to either disrupt targeted genes as in a therapy currently in clinical trials as a treatment for HIV/AIDS (Holt et al., 2010) or to correct genetic defects to restore health as in a recent study to treat hemophilia in a mouse model (Li et al., 2011). In particular, zinc finger nuclease (ZFN) technology is emerging as a versatile tool for the realization of these therapeutic strategies as well as the advancement of several key areas of basic research, including the rapid generation of new transgenic organisms that facilitate the study of human disease (Urnov et al., 2010, Rahman et al., 2011).

Zinc Fingers: Structure and Function

Proteins with zinc finger domains comprise the most abundant class of DNA transcription factors in the human genome, with as many as 900 of ~20,000 predicted proteins bearing this DNA-binding motif (Miller et al., 1985, Tupler et al., 2001, Urnov, 2002). Study of the *Xenopus* transcription factor TFIIIA led to identification of the consensus sequence for a Cys_2-His_2 zinc finger domain: (F/Y)-X-C-X_{2-4}-C-X-X-X-(F/Y)-X-X-X-X-X-γ-X-X-H-X_{3-5}-H, where X is any amino acid and γ is a hydrophobic residue (Miller et al., 1985). Subsequent studies demonstrated that sequences in this configuration adopt a ββα fold that coordinates a zinc atom using the conserved cysteine and histidine pairs; this structure is particularly well-suited for DNA-binding (Parraga et al., 1988, Lee et al., 1989).

The concept of engineering Cys_2-His_2 zinc finger (hereafter zinc finger) domains with new DNA-binding specificities originated from analyzing the co-crystal structure of DNA-bound zinc fingers from the Zif268/Egr1 protein, a transcriptional regulator that determines thymus size in mouse (Figure 1.1a) (Pavletich and Pabo, 1991, Bettini et al., 2002). The DNA binding domain of Zif268 consists of three tandem zinc fingers that wrap around B-DNA. Primary contacts are made between residues -1, +2, +3, and +6 of each zinc finger's α-helix (or "recognition helix") and the edges of bases in the major groove from the G-rich strand of the Zif268 binding site (Figure 1.1a). Each finger binds partially overlapping 3-4 base pair (bp) "subsites": the amino-terminal finger (F1) binds to the 3'-most subsite, the middle finger (F2) binds to the middle subsite, and the carboxy-terminal finger (F3) binds to the 5'-most subsite within the target site. Interactions both among adjacent zinc finger domains within a multi-finger array and between the zinc finger protein and DNA are highly complex, setting the stage for the

formidable challenge of developing techniques for creating custom zinc finger proteins that bind novel DNA sequences.

Zinc Finger-FokI Chimeras: The First Zinc Finger Nucleases

In most restriction enzymes, the cleavage and recognition domains are inextricably linked (Pingoud et al., 2005). Zinc finger nuclease technology was made possible by the fortuitous discovery that the FokI restriction enzyme DNA cleavage domain is effectively independent

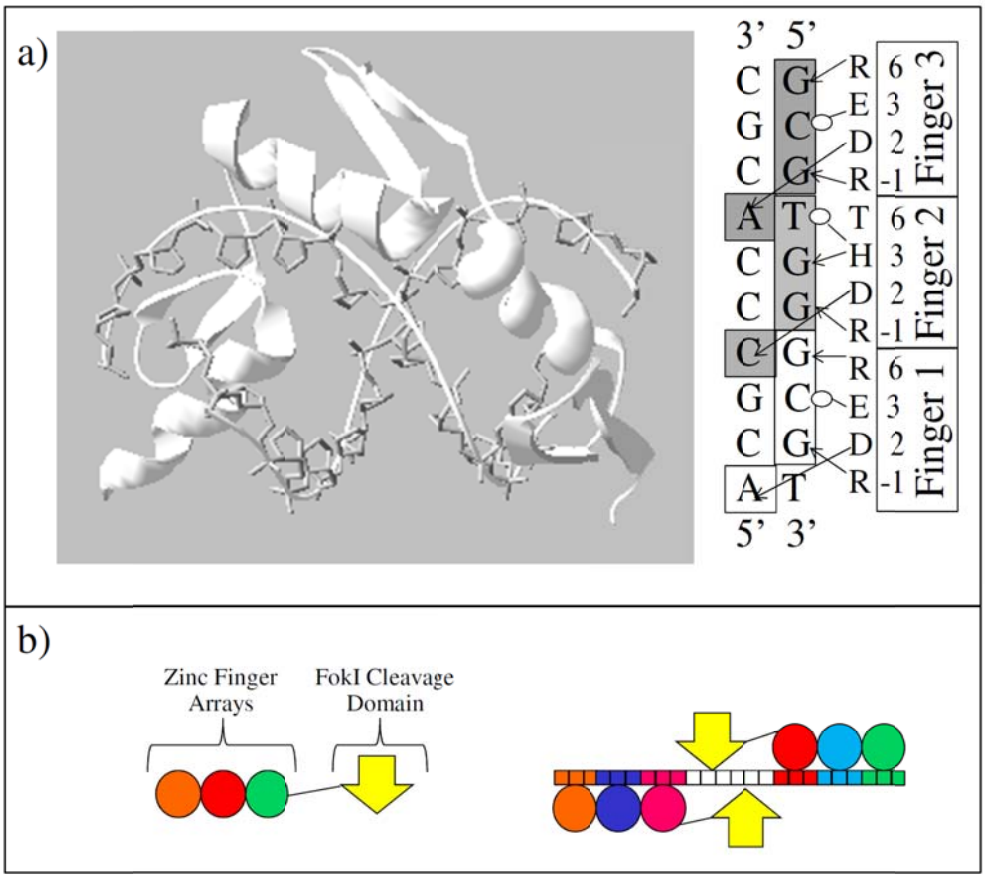


Figure 1.1: Zinc finger proteins and zinc finger nucleases
a) Structure of Zif268 Bound to DNA. In the panel to the right, contacts between the numbered amino acids in the recognition helix and bases in the target DNA site are summarized. Arrows and open circles indicate hydrogen bonds and hydrophobic interactions, respectively. Figure adapted from Wolfe et al. (Wolfe et al., 2000).
b) Zinc finger nucleases. A zinc finger nuclease is a fusion of an engineered zinc finger array (circles) and the non-specific FokI domain (arrow). Zinc finger nucleases bind and cleave their target sites as dimers.

from its natural DNA-binding domain (Li et al., 1992). The first zinc finger nucleases were assembled as chimeras of zinc finger domains for DNA-binding and the FokI cleavage domain (Kim et al., 1996, Bitinaite et al., 1998). The *Fok*I cleavage domain functions enzymatically as a dimer in the native protein (Kim et al., 1996, Bitinaite et al., 1998). However, it has been suggested that unlike the wild-type *Fok*I enzyme (Pingoud and Silva, 2007), ZFNs engineered as dimer pairs may have a small potential for cleavage when only one of the monomers is bound to DNA and the other is weakly bound to DNA or in solution (Bitinaite et al., 1998, Catto et al., 2008, Halford et al., 2011). Each artificial zinc finger monomer typically recognizes a 9 to 12 bp "half-site" separated by 5 to 7 bp of intervening sequence (within which the FokI typically cleaves), conferring sufficient specificity in dimer pairs to generate DSBs at sites long enough (typically 18 to 24 bp long) to be considered unique in complex genomes (Handel and Cathomen, 2011, Rahman et al., 2011).

Zinc Finger Engineering Platforms

Engineering Single Zinc Finger Domains with Novel DNA-Binding Specificities

One of the first effective methods for engineering individual single ZFs with novel DNA-binding specificities was phage-display of combinatorial zinc finger libraries comprised of recognition helices with randomized residues at positions expected to make sequence-specific DNA contacts. In this selection strategy, the randomized finger was positioned over a desired 3-4 bp subsite of interest by embedding it within a three-finger array between two fingers that were held constant and served as "anchors" that would bind well to a constant sequence. Phage display was used successfully to interrogate these libraries and identify individual fingers that bound to a wide variety of different 3-4 bp target sites (Choo and Klug, 1994, Jamieson et al., 1994, Rebar

and Pabo, 1994, Wu et al., 1995, Choo and Isalan, 2000). In later studies, a bacterial cell-based "two-hybrid" system was also used successfully to identify fingers from combinatorial libraries with novel binding specificities (Joung et al., 2000).

Various methods have been described for the comparatively more challenging task of creating engineered multi-finger arrays, including modular assembly, Oligomerized Pool Engineering (OPEN), Context-Dependent Assembly (CoDA), and proprietary commercial methods (Figure 1.2). These approaches are reviewed below.

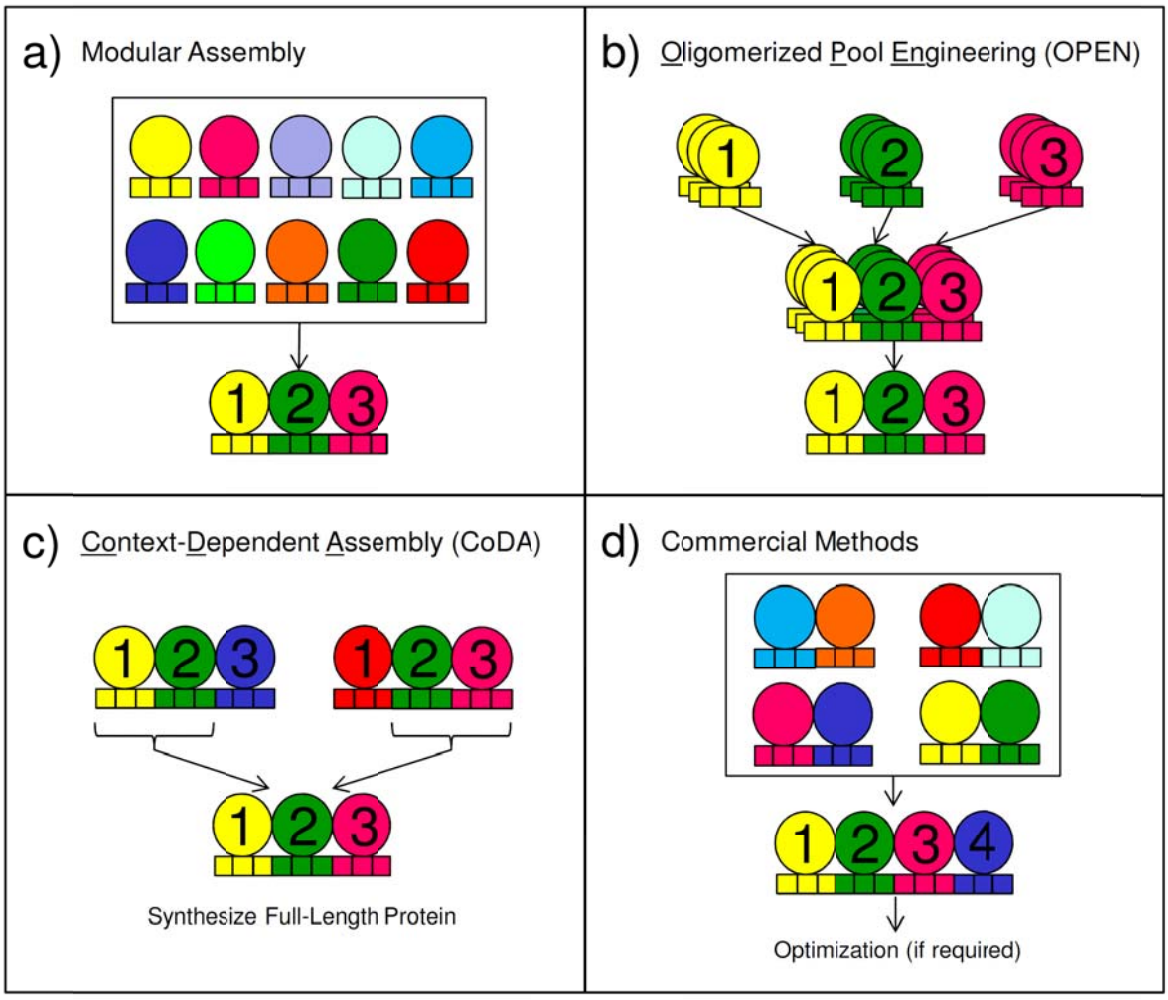


Figure 1.2: Overview of various zinc finger engineering strategies
a) Modular Assembly, b) Oligomerized Pool Engineering (OPEN), c) Context-Dependent Assembly (CoDA), and d) Commercial Methods. Zinc fingers are represented as circles, and individual base pairs of DNA are shown as small squares.

Modular Assembly

One of the first practical and simple methods for the rational design of artificial zinc finger arrays was modular assembly (Figure 1.2a) (Bhakta and Segal, 2010). In pioneering work by Klug and colleagues, three-finger arrays capable of binding to a sequence in the BCR-ABL translocation were engineered (Choo et al., 1994). During the years following that initial success, large collections of individual zinc finger domains with *in vitro* binding activity for particular DNA triplets were identified by three research groups: Barbas and colleagues at the Scripps Research Institute (Beerli et al., 1998, Segal et al., 2003), Sangamo BioSciences (Liu et al., 2002), and ToolGen, Inc. (Bae et al., 2003). Along with detailed protocols, these reagents were later made publicly available by the Zinc Finger Consortium (Wright et al., 2006). With access to the pre-made module archives, polydactyl zinc finger proteins could be assembled by repeating simple steps until the desired number of modules—typically three or four, as too few fingers have neither sufficient affinity nor specificity and too many fingers can reduce activity (Shimizu et al., 2011)—had been added to recognize a particular binding site.

Although modular assembly is straightforward to practice, the success rate of this method was empirically found to be low. A large-scale study by our group estimated that three-finger proteins made by modular assembly and tested for DNA binding activity in a bacterial two-hybrid transcriptional activation assay have on average a 76% failure rate, which was as high as 100% for sites lacking GNN subsites[1] (Ramirez et al., 2008). Another large-scale analysis by Kim and colleagues also revealed a similarly low success rate for modularly assembled ZFNs in human cells (Kim et al., 2009). The most likely reason for these high failure rates is that modular assembly largely ignores the well-documented context-dependent behaviors of zinc fingers that can occur within a multi-finger array (Elrod-Erickson et al., 1998, Wolfe et al., 1999, Wolfe et

[1] Refer to Chapter 2.

al., 2001). This limitation of the method may explain why modularly assembled proteins have in some cases been reported to have low affinities and specificities *in vitro* (Hurt et al., 2003) and low activities and high toxicities in cultured human cells (Cornu et al., 2008).

An alternative argument brought forth by Hughes and colleagues to explain the high failure rates reported by Ramirez & Foley *et al.* is that modular assembly is appropriate for engineering zinc finger proteins with specificity to degenerate motifs (a normal property of naturally-occurring transcription factors) as assessed by protein-binding microarray assays, but is an unrealistic choice for generating proteins with exclusive specificity to a particular 9 bp site (Lam et al., 2011). Direct comparisons of proteins generated by modular assembly and context-dependent assembly methods demonstrate that while proteins made by modular assembly are often not functional at their intended targets in bacterial- and human-cell-based assays, context-dependent methods can more reliably produce proteins capable of interacting with their intended targets in living cells (Maeder et al., 2008, Sander et al., 2011b).

Given these difficulties, some groups have suggested that one way of ensuring a higher likelihood of success is to only use modules that consistently perform well when assembled into arrays (Kim et al., 2009, Sander et al., 2009, Kim et al., 2010). However, an important consequence of using a smaller subset of modules is that the absolute probability of identifying a functional protein for a target of interest is reduced, which in turn limits the range of potential downstream applications possible (Joung et al., 2010).

Oligomerized Pool Engineering (OPEN)

Selection protocols that take into account the context-dependent effects of adjacent fingers on overall binding specificity have been successful at identifying multi-finger proteins with high activity (Greisman and Pabo, 1997, Hurt et al., 2003, Maeder et al., 2008). Building on

the work of Hurt *et al.*, Oligomerized Pool Engineering (OPEN) was developed by the Zinc Finger Consortium to simplify highly specialized techniques for interrogating partially randomized zinc finger libraries with greater than 10^8 members into a relatively rapid protocol accessible to academic scientists (Figure 1.2b) (Maeder et al., 2009). The key to this method is the archive of degenerate zinc finger pools for 66 currently available DNA triplets of a theoretical maximum of 192. Each pool contains up to 95 unique solutions that may function in many potential contexts. Targetable binding sites within user-defined sequences of interest can be identified with the free online tool ZiFiT (Sander et al., 2007, Fu et al., 2009, Sander et al., 2010a). Although practicing this method requires exquisite attention to the detailed protocol, 16 of 24 (~67%) zinc finger nuclease pairs were found to be active in zebrafish, plant, or human cells, and on average, it is possible to find approximately five potential zinc finger nuclease sites per kilobase (kb) of random sequence (Maeder et al., 2008).

Data derived from the outcomes of 135 successful and failed OPEN selections were used to create a machine learning algorithm to assign relative probabilities that a zinc finger nuclease pair will function with high accuracy and specificity. These confidence scores have been integrated into ZiFiT to facilitate the process of identifying the most promising zinc finger nuclease sites from user-specified sequences (Sander et al., 2010b). ZiFDB was developed as an online repository for investigator-annotated data on the outcomes of zinc finger engineering efforts (Fu et al., 2009). To further aid researchers interested in modifying endogenous loci in frequently studied organisms, ZFNGenome was developed as a tool dynamically linked to ZiFDB that is used for browsing genomes to visualize any of the 11.6 million potential target sites of OPEN-generated zinc finger nucleases (Reyon et al., 2011).

Context-Dependent Assembly (CoDA)

By analyzing the results of dozens of OPEN selections, the striking observation was made that certain F2 (middle) zinc fingers in three-finger arrays appear repeatedly between different first and third fingers, suggesting that these zinc fingers may remain functional under various contexts (Sander et al., 2011b). Using 18 fixed second-finger units, 319 first-finger units and 344 third-finger units were selected by OPEN and assembled into 181 three-finger arrays to assess the relative success of the method (Figure 1.2c). The targeting range of CoDA is approximately 1 in 500 bp of random sequence with a success rate of >50% (26 of 47 target sites) when tested as nucleases in zebrafish, *Arabidopsis*, and soybean (Curtin et al., 2011, Sander et al., 2011b). Independent groups have also used the method to generate active ZFNs (Osborn et al., 2011). The ZiFiT software was updated to output sequences encoding CoDA proteins if they are found within the investigator's region of interest.

Commercial Zinc Finger Engineering

In addition to the above-mentioned publicly available zinc finger engineering platforms, it is also possible to purchase zinc finger proteins through a commercial source, albeit at a significant cost and under restrictions posed by legal terms and conditions (Urnov et al., 2005, Pearson, 2008, Carbery et al., 2010). While these engineering methods and the reagents necessary to practice them are proprietary, some of the details have been disclosed in publications; using a previously published strategy originally described by Choo and colleagues (Isalan and Choo, 2001, Isalan et al., 2001, Moore et al., 2001, Jamieson et al., 2003), four- and six-finger arrays are likely assembled from validated two-finger modules (Doyon et al., 2008, Perez et al., 2008), derived from a proprietary archive owned by Sangamo (Figure 1.2d).

Depending on context, two-finger units can be joined together by canonical linkers (TGEKP) or by non-canonical "disrupted linkers" (e.g. TGSQKP) (Moore et al., 2001, Perez et al., 2008).

Modified Zinc Finger Nuclease Architectures

One problem with using the FokI cleavage domain is that instead of the two different monomers of a ZFN pair interacting as a heterodimeric pair as intended, each monomer could potentially form homodimers and cut at undesired sites, possibly causing genetic damage. Structural analysis and functional characterization of FokI by independent groups yielded two sets of heterodimer architectures in which hydrophobic and electrostatic interactions have been modified, including the EL:KK (also termed '-' and '+', respectively) (Miller et al., 2007) and DD:RR frameworks (Szczepek et al., 2007); these frameworks are virtually orthologous and therefore enable the concurrent introduction of two pairs of ZFNs into cells with minimal cytotoxicity (Sollu et al., 2010, Doyon et al., 2011), a capability that is useful for applications requiring simultaneous modification of two genomic loci. Unsurprisingly, ZFNs with FokI heterodimers have been shown to significantly reduce rates of off-target cleavage (particularly at loci with homodimeric half-sites) in human cells relative to the homodimeric FokI framework (Gabriel et al., 2011).

Although heterodimeric FokI cleavage domains exhibit reduced toxicity, they also possess less robust nuclease activity compared to the original homodimeric FokI domain. Novel mutations (N496D and H537R, numbered relative to the wild-type FokI sequence) that favor the formation of salt bridges to strengthen interactions between the aforementioned heterodimeric FokI domains directly address this problem. These mutations not only enhance ZFN activity to levels comparable with the activity of homodimeric FokI ZFNs, but also serve to further

suppress homodimerization (Doyon et al., 2011). An engineered hyperactive mutant of FokI ("Sharkey") has been shown to be compatible with these newest heterodimeric FokI domains, yielding an additive increase in activity (Guo et al., 2010, Doyon et al., 2011).

The length and sequence composition of linkers between zinc finger arrays and the FokI cleavage domain have also been varied and analyzed for function. These studies have yielded a repertoire of linkers that enable ZFN binding to target sites with 5 to 18 bp spacers between the half-sites (Bibikova et al., 2001, Handel et al., 2009).

General Types of Zinc Finger Nuclease-Inducible Modifications

Non-Homologous End-Joining-Mediated Modifications

In a cellular context, DNA DSBs produced by zinc finger nucleases create short (typically 4 to 5 bp) single-strand overhangs that are typically rejoined perfectly or

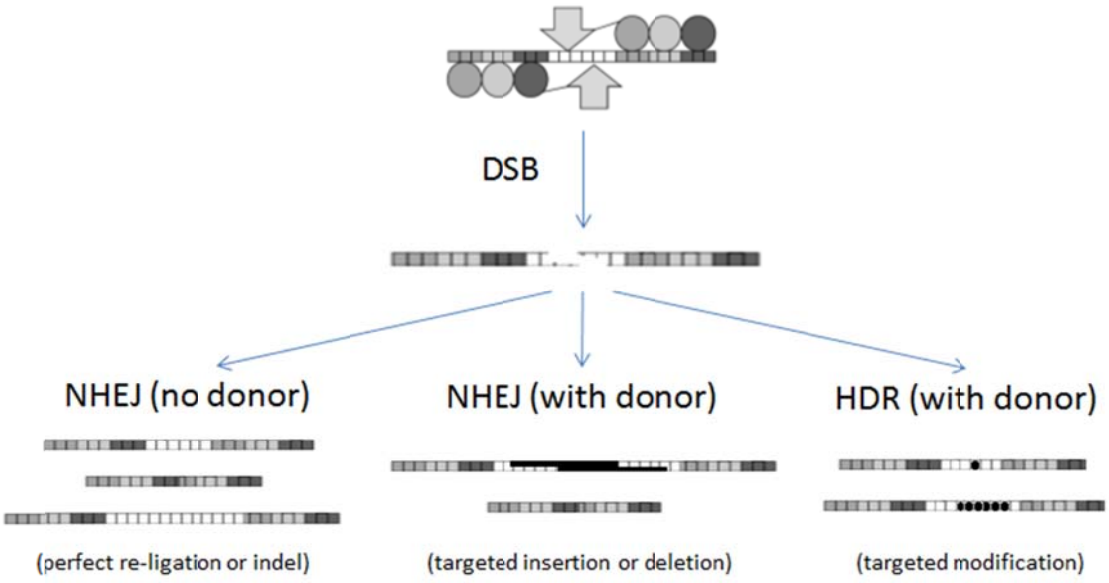


Figure 1.3: ZFN-mediated DSBs are repaired primarily by NHEJ or HDR
NHEJ proceeds in the absence of a donor and can result in perfect re-ligation of ends or the introduction of short (usually 4-5 bp) insertions or deletions. NHEJ can also drive incorporation of a donor with compatible ends at high efficiency. HDR requires a single or double-strand DNA donor template to stimulate precise genome modifications, including point mutations.

repaired with the introduction of short insertions or deletions (indels) of extraneous bases by an error-prone process known as non-homologous end joining (NHEJ; Figure 1.3 "NHEJ (no donor)") (Hartlerode and Scully, 2009, Orlando et al., 2010). NHEJ is the primary mechanism for correction in the event that a repair template for homology-directed repair (HDR) cannot be found (van Gent et al., 2001). Even in the absence of selection, gene disruption rates exceeding 50% can be achieved under optimal conditions (Doyon et al., 2011). Short indels from ZFN-mediated mutagenesis can be detected by mismatch-sensitive endonuclease assays (Cel-I/Surveyor or T7 endonuclease I) with a detection limit of about 1% (Oleykowski et al., 1998, Kim et al., 2009, Guschin et al., 2010). To identify lower-frequency mutations, DNA sequencing is required (Perez et al., 2008, Foley et al., 2009a, Foley et al., 2009b).

Longer deletions can also be introduced by ZFNs, but these would likely be missed by performing standard mutagenesis detection assays or by sequencing short stretches of DNA surrounding the intended target site (Morton et al., 2006). In one case, genomic deletions of several lengths were introduced using zinc finger nucleases targeted within and near the human *CCR5* gene; the longest was 15 kb of DNA and attributed to the simultaneous cleavage of cognate sites at neighboring genes, *CCR5* and *CCR2* (Lee et al., 2009). *CCR5* is a desirable therapeutic target due to the observation that individuals with mutations inactivating this gene are not only healthy, but possess a natural resistance to HIV infection (Hutter et al., 2009). Translocations can also be induced when ZFN-mediated DSBs are introduced into sites present on different chromosomes (Brunet et al., 2009, Simsek et al., 2011).

Not only can gene disruption be achieved by the NHEJ repair pathway, but this repair mechanism can also mediate the insertion of desired donor DNAs with greater than 10% success (Figure 1.3 "NHEJ (with donor)") (Orlando et al., 2010). The authors also report that

simultaneous cutting by two zinc finger nuclease pairs in close proximity can result in up to 50% donor insertion.

Stimulation of Homology-Directed Repair

Homologous recombination is a DNA repair pathway in which sequence similarity guides DNA repair. Under ordinary circumstances, the spontaneous frequency of homologous recombination in mammalian cells is on the order of 1 in 10^6, a level insufficient for therapeutically relevant applications (Vasquez et al., 2001, Paques and Duchateau, 2007). Site-specific DSBs can stimulate HDR by several orders of magnitude such that 1 in 100 or more cells undergo desired DNA repair events (Figure 1.3 "HDR (with donor)) (Rouet et al., 1994, Jasin, 1996). It is generally believed that ZFN-induced alterations or insertions should be within 100 bp of the DSB site for maximal activity due to the relatively short conversion tract length in mammals and the observation that the efficiency of gene conversion is inversely related to distance (Donoho et al., 1998, Elliott et al., 1998, Hartlerode and Scully, 2009).

While NHEJ promotes DNA repair throughout a cell's life, HDR is more active in the S and G2 phases of the cell cycle to counteract damage associated with genome replication (Takata et al., 1998, Mao et al., 2008, Hartlerode et al., 2011). This observation has been exploited to increase the rates of repair for ZFN-induced DSBs by treatment with cytostatic drugs such as vinblastine (Urnov et al., 2005) and nocodazole (Olsen et al., 2010), which arrest cultured cells in G2/M. Increasing levels of donor in the nucleus can also bias repair toward HDR (Lombardo et al., 2007, Gellhaus et al., 2010, Certo et al., 2011). If HDR is the desired outcome, this competition between pathways can pose a challenge to achieving high-efficiency error-free repair (Hartlerode and Scully, 2009).

The delivery method and configuration of the repair template donor can have dramatic

effects on rates of correction and fidelity (Figure 1.3 "HDR (with donor)). For instance, plasmid DNA with 1.5 kb of homology was used to introduce a silent point mutation in up to 18% of *IL2Rγ* loci in transformed human cells (Urnov et al., 2005). Mutations in the *IL2Rγ* gene are associated with X-linked severe combined immune deficiency (X-SCID, also known as "bubble boy" disease); this proof-of-concept could eventually be developed into a therapy to treat individuals with this condition. It was later demonstrated that sequences ranging from 12 bp epitope tags to 8 kb cassettes harboring multiple transgenes could be knocked-in with 16% and 6% efficiency, respectively, in the absence of selection (Moehle et al., 2007). Up to 50% endogenous gene correction or addition by HDR was achieved in human cells by delivery of ZFN and donor via integration-defective lentiviral (IDLV) vectors (Lombardo et al., 2007). Adeno-associated viral (AAV) vectors have even been delivered to whole organisms yielding partial phenotypic correction of hemophilia in a mouse model (Li et al., 2011).

Despite surprisingly little shared sequence homology compared to the aforementioned donor types, it has also been demonstrated that PCR-derived double-strand linear DNA donors with 50 bp homology arms can stimulate transgene addition in 5-10% of chromosomes (Orlando et al., 2010). Although initial reports suggested that short single stranded DNA oligonucleotide donors could only achieve less than 1% gene conversion (Radecke et al., 2006, Olsen et al., 2009, Radecke et al., 2010), rates as high as 57% for incorporation of insertions have been published for slightly longer single-stranded donors flanked by approximately 40 bp of homology on either side of the zinc finger nuclease cut site (Chen et al., 2011). It is not yet clear by which mechanism the gene conversion occurs, but there is evidence to suggest that the oligonucleotide donors serve as informational templates only that are not incorporated upon repair (Radecke et al., 2006, Radecke et al., 2010). In addition, appropriately designed single-

stranded donors can be used to mediate deletion of sequences of up to 100 kb (Chen et al., 2011). This represents a significant region potentially spanning multiple genes, which may allow for the creation of new models of disease.

Modification of Model Organisms and Cells with Zinc Finger Nucleases

Zinc finger nucleases have radically simplified and accelerated the generation of novel cell lines and model organisms, many of which could not formerly be manipulated genetically (Table 1.1). As is often the case, high efficiency rates allow for the identification of founders

Table 1.1: Summary of ZFN-mediated organismic and cell line modifications

Organism	***References***
Xenopus	(Bibikova et al., 2001, Young et al., 2011)
Drosophila	(Bibikova et al., 2002, Bibikova et al., 2003, Beumer et al., 2006, Beumer et al., 2008, Bozas et al., 2009)
Arabidopsis	(Lloyd et al., 2005, Zhang et al., 2010, Sander et al., 2011b)
Nematode	(Morton et al., 2006, Wood et al., 2011)
Tobacco	(Wright et al., 2005, Maeder et al., 2008)
Zebrafish	(Doyon et al., 2008, Meng et al., 2008, Foley et al., 2009b, Sander et al., 2011b)
Maize	(Shukla et al., 2009)
Rat	(Geurts et al., 2009a, Geurts et al., 2009b, Mashimo et al., 2010, Cui et al., 2011, Moreno et al., 2011)
Mouse	(Carbery et al., 2010, Goldberg et al., 2010, Meyer et al., 2010, Cui et al., 2011, Li et al., 2011)
Pig	(Watanabe et al., 2010, Whyte et al., 2011, Yang et al., 2011)
Sea Urchin	(Ochiai et al., 2010)
Soybean	(Sander et al., 2011b)
Silkworm	(Takasu et al., 2010)
Rabbit	(Flisikowska et al., 2011)
Cell Line	
Chinese hamster ovary (CHO) cells	(Santiago et al., 2008, Cost et al., 2009, Liu et al., 2010)
Immortalized or transformed human cell lines	(Porteus and Baltimore, 2003, Alwin et al., 2005, Urnov et al., 2005, Moehle et al., 2007, Szczepek et al., 2007, Maeder et al., 2008, Radecke et al., 2010)
Primary human T cells	(Urnov et al., 2005, Perez et al., 2008, Wilen et al., 2011)
Human induced pluripotent stem cells (iPSCs)	(Hockemeyer et al., 2009, Zou et al., 2009, Zou et al., 2011)
Human mesenchymal stromal cells	(Benabdallah et al., 2010)
Human hematopoietic stem cells (HSCs)	(Holt et al., 2010)

with new genetic mutations by screening a small number of organisms (Foley et al., 2009b). For loss-of-function studies, disrupting only one allele is often not enough to suppress activity. Fortunately, it has been shown that cell lines with gene disruptions at all alleles can be created at high efficiency (Cost et al., 2009, Liu et al., 2010). Despite this success, it is important to note that zinc finger nucleases may be less efficient in certain genomic contexts (e.g. regions of heterochromatin or DNA methylation), which can vary between cell types (Liu et al., 2001, Maeder et al., 2008). As an indirect solution, the introduction of genes into putative "safe harbor" loci may be a viable therapeutic strategy (Benabdallah et al., 2010, Zou et al., 2010).

Safe harbors are loci at which introducing DNA sequence changes is not expected to have negative consequences for the host and transgene expression is predicted to be robust. Although it has previously been thought that safe harbor loci should be in gene "deserts" to avoid disruption of transcription, having the ability to characterize (and in many cases, through adequate design considerations, minimize) perturbations in genes neighboring transgene integration events may be preferable to altering a poorly understood genetic landscape with unknown consequences (Lombardo et al., 2011).

In proof-of-concept studies at safe harbor loci, zinc finger nuclease-stimulated gene integration was achieved in up to 50% of alleles surviving selection in human embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs) (Hockemeyer et al., 2009, Hockemeyer and Jaenisch, 2011). This strategy could one day be used to restore gene function in diseases such as hemophilia and cystic fibrosis, for which a particular genetic defect prevents adequate levels of the affected protein from being made. Employing autologous cell transplantation, a patient's own cells can be modified and then re-introduced, theoretically preventing immune rejection unless a hitherto unfamiliar gene product is generated that is recognized as an antigen.

Having the ability to directly modify the patient's DNA at the sequence level offers numerous advantages over competing gene therapy strategies, including the retention of an endogenous promoter to regulate expression of the corrected gene.

Cytotoxicity and Off-Target Effects

Assessing and Attenuating Zinc Finger Nuclease-Mediated Cytotoxicity

Having been explicitly designed to cleave DNA, zinc finger nucleases are by nature cytotoxic, the degree of which varies for each zinc finger nuclease pair and is determined in part by the specificity with which a zinc finger nuclease can bind to its intended unique target at the exclusion of alternative sites. Unsurprisingly, it has been demonstrated that obligate zinc finger nuclease heterodimers greatly reduce cytotoxicity (Miller et al., 2007, Szczepek et al., 2007, Doyon et al., 2011).

As excess protein concentration exacerbates cytotoxic effects, zinc finger nuclease expression should ideally be as low and transient as possible in order to achieve the desired modification without otherwise affecting genomic integrity and cellular viability (Beumer et al., 2006, Pruett-Miller et al., 2008). A variety of delivery options have been explored, including plasmids (Porteus and Baltimore, 2003, Holt et al., 2010), mRNAs (Doyon et al., 2008, Meng et al., 2008, Zou et al., 2010), and viral vectors—adeno-associated (Porteus et al., 2003), adeno (Perez et al., 2008), and integrase-deficient lenti (Cornu and Cathomen, 2007, Lombardo et al., 2007).

A multitude of assays have been used to evaluate cytotoxicity of zinc finger nucleases, for readouts including cell death (Alwin et al., 2005), integration of donor plasmid at off-target double-strand break sites (Olsen et al., 2010), cell survival (Porteus and Baltimore, 2003, Alwin

et al., 2005, Cornu et al., 2008, Maeder et al., 2008, Pruett-Miller et al., 2008, Sollu et al., 2010), and organismal developmental defects (Doyon et al., 2008, Meng et al., 2008, Foley et al., 2009b). Visualizing repair factors recruited to sites of DSBs (e.g. γ-H2AX foci staining) has been used to assess cytotoxicity in the past (Miller et al., 2007, Szczepek et al., 2007, Cornu et al., 2008). However, these assays only provide a snapshot of double-strand break repair activity and lack the sensitivity to determine whether DSBs are occurring at off-target sites over the background of intended DSBs and those occurring naturally in dividing cells (Perez et al., 2008, Ramalingam et al., 2010).

Profiling and Validation of Off-Target Cleavage Events

DNA-binding properties of zinc finger monomers have been evaluated using a variety of techniques, including Enzyme-Linked Immunosorbent Assays (ELISA) (Segal et al., 1999, Segal et al., 2003), microarrays (Bulyk et al., 2001), Electrophoretic Shift Assays (EMSA) (Hurt et al., 2003), Systematic Evolution of Ligands by Exponential Enrichment (SELEX) (Wolfe et al., 1999, Liu and Stormo, 2005, Perez et al., 2008), bacterial two-hybrid transcriptional activation assay (Wright et al., 2006), Cyclical Amplification and Selection of Targets (CAST) (Segal et al., 2003, Brayer et al., 2008), bacterial one-hybrid profiles (Meng et al., 2007, Gupta et al., 2011), Bind-n-Seq (Zykovich et al., 2009), fluorescent anisotropy (Sander et al., 2009), molecular modeling (Yanover and Bradley, 2011), and purely computational approaches (Sander et al., 2010a, Cradick et al., 2011). A subset of these various methods can be used to obtain DNA specificity profiles for zinc finger proteins; this information can then be used to computationally search for potential off-target sites in genomes of interest.

The greatest limitation of these computational approaches is that the DNA-binding profiles of each ZFN monomer must be used to extrapolate the cleavage specificity of the zinc

finger nuclease dimer. Given that dimerization of the FokI nuclease domains may lead to cooperative binding of zinc finger nuclease monomers in a dimer (Bitinaite et al., 1998), it has been postulated that using the monomeric DNA-binding specificities of zinc finger nucleases may not provide sufficient information to identify the full range of potential off-target cleavage events (Pattanayak et al., 2011). A novel and recently developed selection scheme for elucidating actual cleavage preferences of an active zinc finger nuclease dimer *in vitro* has enabled the identification of new potential zinc finger nuclease off-target sites in human cells. With this approach, eight previously unknown off-target sites for zinc finger nucleases targeted to the *CCR5* locus were identified in human K562 cells[2] (Pattanayak et al., 2011). In addition, an important finding of this study was that the position and number of mutations in one half-site can affect the probability that a ZFN monomer will be able to bind to the other half-site, suggesting a compensation model for cooperative ZFN binding and supporting the theory that monomeric binding data is insufficient to infer dimeric ZFN cleavage preferences.

A complementary approach by which ZFN off-target loci were identified was by mapping insertions of integration-defective lentiviruses at DSB sites, which led to the identification of eight additional off-target sites for the same *CCR5*-targeted zinc finger nucleases (Gabriel et al., 2011). Both of these methods identified off-target sites in cells that had not been predicted computationally from monomeric zinc finger nuclease binding data. Interestingly, these methods yielded non-overlapping lists of novel off-target cleavage sites for the *CCR5* zinc finger nucleases, strongly suggesting that neither approach comprehensively identifies off-target sites at the genome-wide level and that there is much work remaining to be done in mapping zinc finger nuclease specificity profiles.

[2] Refer to Chapter 4.

As the cost of DNA sequencing continues to drop, the possibility of subjecting clonal cell populations or small organisms such as nematodes to whole genome sequencing analysis will become an increasingly viable option (Rahman et al., 2011, Wood et al., 2011). However, in applications for which polyclonal cell populations will be used therapeutically, the need for comprehensive off-target characterization will continue to be highly relevant (Cathomen and Schambach, 2009). Given that chromosomal translocations are a potential risk with zinc finger nuclease treatment, karyotyping should be performed to screen for gross abnormalities (Brunet et al., 2009, Cathomen and Schambach, 2009, Hockemeyer et al., 2009, Simsek et al., 2011). Once zinc finger nuclease off-target cleavage profiles can be derived with greater accuracy and confidence and for less cost, these methods will enable routine identification of zinc finger nuclease off-target sites. An important question for future research will be to develop methods for further optimizing the cleavage profiles of zinc finger nucleases so as to minimize particularly deleterious off-target effects (Gupta et al., 2011).

Acknowledgements

We thank members of the Joung lab for helpful comments and suggestions, Victor Quesada for assistance with the development of Figure 1.1, and Corie Anastasia Schnittke for help copy-editing this chapter. C.L.R. is partially supported by the National Science Foundation and the Ford Foundation. J.K.J is supported by National Institutes of Health (NIH) R01 GM088040, an NIH Director's Pioneer Award (DP1 OD006862), and the Massachusetts General Hospital Jim and Ann Orr Research Scholar Award.

Introduction to the Thesis

The collected works presented in this dissertation describe strategies for making improvements to zinc finger nuclease engineering efforts on multiple fronts including design strategies (Chapter 2), reducing genotoxicity (Chapter 3), and assessing specificity (Chapter 4).

Chapter 2: "Unexpected failure rates for modular assembly of engineered zinc-fingers"

Before the results presented in Chapter 2 were made public, success rates for modularly assembled zinc finger proteins were reported in the range of 60% (Bae et al., 2003) to 100% (Segal et al., 2003) in the most comprehensive large-scale studies available in the published literature. Based on experimental failures far surpassing these estimates, we and others pooled our collective data on the effectiveness of modular assembly in a variety of *in vitro* and cell-based assays and supplemented it with a large-scale survey of 168 zinc-finger arrays designed for 104 DNA target sites of diverse compositions (Ramirez et al., 2008). We found that the failure rate in a bacterial-cell-based reporter assay for zinc finger proteins made by modular assembly was on average ~76%, which was much greater than previously thought, and higher still (~94%) if one considers the combined probabilities for two arrays used in a ZFN dimer pair.

Our study raised awareness of the hitherto underappreciated limitations of modular assembly not only within the zinc finger engineering community, but also within the community of non-specialist investigators who used these reagents as tools for attempting to modify model organisms and cells. We also established the standard against which future zinc finger engineering efforts would be measured, paving the way for more robust context-dependent methods such as OPEN (Maeder et al., 2008) and CoDA (Sander et al., 2011b).

Chapter 3: "Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects"

While ZFNs are powerful tools for modifying genomes as reviewed above, there is a strong bias for NHEJ-mediated gene repair over HDR in most cellular contexts encountered, which may be problematic when the goal of an intervention is gene correction. Realizing that nicks are less genotoxic to cells than DSBs and inspired by descriptions of a FokI mutant (D450A) that is incapable of cleaving DNA yet continues to promote dimerization (Waugh and Sauer, 1993, Bitinaite et al., 1998, Beumer et al., 2006), I hypothesized that ZFNs could make nicks instead of cuts in DNA if the D450A mutation were introduced into only one monomer of a ZFN pair.

While it was relatively straightforward to characterize the novel nicking activities of zinc finger nickases (ZFNickases) *in vitro*, demonstrating that they were active in cells and could preferentially mediate HDR over NHEJ was a significant challenge due to the relatively insensitive (detection limit: ~1%) assays traditionally used for detecting HDR events. Using human cell-based reporter assays—in particular the Traffic Light Reporter assay developed by Certo and colleagues, which made it possible to simultaneously measure NHEJ and HDR rates (Certo et al., 2011)—we were able to demonstrate that ZFNickases do in fact typically stimulate HDR more often than NHEJ, although absolute rates of HDR are significantly reduced relative to nucleases. We present these findings in Chapter 3 (Ramirez et al., 2012) and offer suggestions for increasing the activity of ZFNickases in Chapter 5.

Chapter 4: "Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection"

Zinc finger engineering strategies have become sufficiently reliable for creating ZFNs targeted to loci of interest that characterizing the full range of ZFN specificity has moved decidedly into the forefront of experimental concern in recent years. At the time the work described in Chapter 4 was begun, the gold standard for surveying genome-wide off-target ZFN-cleavage sites was limited to performing monomeric DNA binding assays and using the results to computationally extrapolate a list of about a dozen putative loci of interest to deep sequence (Perez et al., 2008); the ZFNs analyzed in this report were subjected to this level of scrutiny due to their intended use in clinical trials for conferring HIV-resistance to T-cells by mimicking a naturally-occurring mutation in the *CCR5* gene (NCT01044654, NCT00842634, NCT01252641).

Postulating that even the best available methods were likely missing potentially important off-target cleavage sites, we developed various approaches to examine the specificity of ZFNs more rigorously. Among these strategies, we validated putative off-target sites for the *CCR5* ZFNs guided by a method developed by Pattanayak and colleagues for assaying the cleavage preferences of dimeric zinc finger nucleases *in vitro* (Pattanayak et al., 2011). We identified evidence of NHEJ at eight previously undescribed *CCR5* ZFN off-target sites in human K562 cells, supporting our hypothesis that existing methods were not sufficiently robust and offering a feasible alternative that could be built upon for future studies of nuclease specificity.

The ongoing advancement of zinc finger nuclease and nickase engineering technologies and recent successes in modifying model organisms and cells with high efficiency and specificity are promising steps toward the realization of gene therapy as a commonplace therapeutic strategy.

Chapter Two

Unexpected failure rates for modular assembly of engineered zinc fingers

Cherie L. Ramirez*, Jonathan E. Foley*, David A. Wright, Felix Müller-Lerch, Shamim H. Rahman, Tatjana I. Cornu, Ronnie J. Winfrey, Jeffry D. Sander, Fengli Fu, Jeffrey A. Townsend, Toni Cathomen, Daniel F. Voytas, and J. Keith Joung

The thesis author and Jonathan Foley were joint first authors (denoted by "*") of this work, and together constructed the zinc finger arrays that were tested for the described bacterial-2-hybrid experiments, performed these experiments, and conducted data analysis. The thesis author oversaw these efforts and performed Western blots on all bacterial cell lysates. Jonathan Foley generated the reporter strains used for the bacterial-2-hybrid activity assays and re-sequenced all constructs. David A. Wright, Ronnie J. Winfrey, Jeffry D. Sander, Fengli Fu, and Jeffrey A. Townsend performed electrophoretic mobility shift assays (EMSA) and plant single-strand annealing (PSSA) assays. Episomal homologous recombination (HR) assays in human cells were performed by Felix Müller-Lerch, Shamim H. Rahman, and Tatjana I. Cornu. Jeffry D. Sander provided assistance in DNA and protein sequence data analysis.

Reference: "Unexpected failure rates for modular assembly of engineered zinc fingers." Ramirez CL, Foley JE, Wright DA, Müller-Lerch F, Rahman SH, Cornu TI, Winfrey RJ, Sander JD, Fu F,

Townsend JA, Cathomen T, Voytas DF, Joung JK. *Nature Methods*. 2008 May; 5(5):374-5.

Unexpected failure rates for modular assembly of engineered zinc-fingers

Cherie L. Ramirez[1,2,7], Jonathan E. Foley[1,7], David A. Wright[3], Felix Müller-Lerch[4], Shamim H. Rahman[4], Tatjana I. Cornu[4], Ronnie J. Winfrey[3], Jeffry D. Sander[3,5],Fengli Fu[3,5], Jeffrey A. Townsend[3], Toni Cathomen[4], Daniel F. Voytas[3], & J. Keith Joung[1,2,6]

[1] Molecular Pathology Unit, Center for Cancer Research, and Center for Computational and Integrative Biology, 149 13th Street, Room 7132, Massachusetts General Hospital, Charlestown, MA 02129 USA.
[2] Biological and Biomedical Sciences Program, Division of Medical Sciences, Harvard Medical School, Boston, MA 02115 USA.
[3] Department of Genetics, Development & Cell Biology, Iowa State University, 1035A Roy J. Carver Co-Lab, Ames, Iowa 50011, USA.
[4] Charité Medical School, Institute of Virology (CBF), Hindenburgdamm 27, D-12203 Berlin, Germany.
[5] Interdepartmental Graduate Program in Bioinformatics and Computational Biology, 2114 Molecular Biology Building, Iowa State University, Ames, Iowa 50011, USA.
[6] Department of Pathology, Harvard Medical School, Boston, MA 02115 USA.
[7] These authors contributed equally to this work.

To the Editor: Zinc-finger nucleases (ZFNs) consist of an engineered zinc-finger array fused to a nuclease domain. Dimers of ZFNs can create targeted double-strand DNA breaks which can stimulate highly efficient gene targeting in many cell types (Supplementary Figure 2.1a-b) (Porteus and Carroll, 2005). Here we show that the modular assembly method of engineering zinc-finger arrays has an unexpectedly higher failure rate than previously reported.

Modular assembly advocates linking individual zinc-fingers, each of which typically binds to a 3 bp "subsite" (Supplementary Figure 2.2). Two large-scale surveys have suggested that modular assembly is highly effective (*i.e.* 60-100% success rates) for making three-finger arrays designed to bind 9 bp target sites (Bae et al., 2003, Segal et al., 2003). The Zinc Finger

Consortium recently constructed an archive of 141 previously published finger modules (Liu et al., 2002, Bae et al., 2003, Mandell and Barbas, 2006) encoded on a standardized platform (Wright et al., 2006). However, our initial experiences using these reagents suggested that modular assembly was surprisingly inefficient (Supplementary Discussion and Supplementary Table 2.1).

To perform a larger-scale test, we assembled 168 zinc-finger arrays designed for 104 diverse target DNA sites (Supplementary Table 2.2 and Chapter 2 Supplementary Methods). These domains were tested for DNA-binding using a bacterial two-hybrid (B2H) assay (Wright et al., 2006) which accurately identifies arrays that lack activity as ZFNs in human cells (Supplementary Figures 2.3 and 2.4 and Chapter 2 Supplementary Discussion). For 79 of the 104 target sites, we failed to obtain a single three-finger array that scored positively in the B2H assay (overall failure rate of ~76%) (Figure 2.1a and Chapter 2 Supplementary Table 2). Strikingly, the method was far less effective for target sites composed of two, one, or no GXX subsites (where X is any base) compared with those composed of three GXX subsites (Figure 2.1a). In addition, since ZFNs function as dimers, failure rates for making a ZFN pair would be expected to be even higher (Chapter 2 Supplementary Discussion). Importantly, these values are all likely underestimates of actual failure rates because not all finger arrays that are positive in the B2H will be active as ZFNs in human cells (Supplementary Figure 2.4).

One reason for the apparent discrepancy between previously published studies (Bae et al., 2003, Segal et al., 2003) and this report is that the former primarily used 9 bp sites composed of two or three GXX subsites whereas we used sites with more varied GXX subsite composition (Figure 2.1b); this difference will critically affect the observed failure rate (Figure 2.1a). Our study represents a more meaningful evaluation because target sites containing one or no GXX

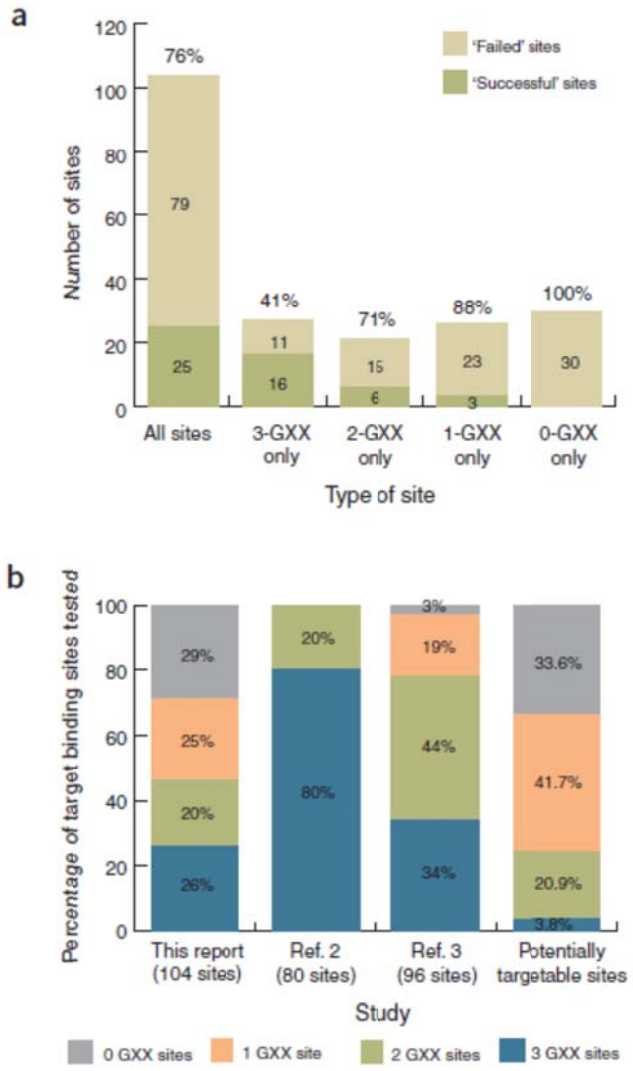


Figure 2.1: Large-scale evaluation of the modular assembly method for engineering zinc finger arrays

(a) Success and failure rates of modular assembly as judged by B2H assay for all 104 target DNA sites (see Chapter 2 Supplementary Discussion for definition of success threshold) and for subsets of target sites containing 3, 2, 1, or no GXX subsites. Light blue bars represent sites for which at least one successful zinc-finger array was identified and pink bars represent sites for which we failed to obtain a single successful array (see Supplementary Table 2.2 for details). Predicted failure rates for each set of target sites are shown in purple circles.

(b) Distributions of target DNA sites used in different modular assembly evaluation studies, including "Ref. 2" (Segal et al., 2003) and "Ref. 3" (Bae et al., 2003). Target sites were categorized according to the number of GXX subsites present (0, 1, 2, or 3 GXX subsites represented by light blue, yellow, pink, and purple, respectively). The distribution of all 107,011 potential 9 bp sites that can be targeted using 141 zinc-finger modules from the Zinc Finger Consortium Modular Assembly Kit 1.0 (Wright et al., 2006) is also shown (far right bar).

subsites (underrepresented in previous studies) encompass the majority (>75%) of the 107,011 potential 9 bp sites that can be targeted with existing modules (Figure 2.1b).

Our results strongly suggest that potential users of modular assembly should expect that it will fail to yield a functional multi-finger array for the majority of potentially targetable sites and to emphasize this we have modified content on the Zinc Finger Consortium website (http://www.zincfingers.org) (Chapter 2 Supplementary Methods). Highly effective but more labor-intensive selection-based methods for engineering zinc-finger arrays have been previously described (Chapter 2 Supplementary Discussion) and at present these are the only publicly available alternatives for academic researchers interested in using ZFN technology.

Acknowledgements

We thank D. Dobbs, P. Zaback and E. Unger-Wallace for contributions. J.K.J is supported by the NIH (R21 RR024189), the Cystic Fibrosis Foundation (MCCRAY07G0) and the MGH Department of Pathology. D.F.V is supported by NSF grant DBI 0501678. T.C. is supported by the German Research Foundation (SPP1230 CA311/2). C.L.R. is supported by a Ford Foundation Pre-Doctoral Diversity Fellowship and a National Science Foundation Graduate Research Fellowship. T.I.C. is supported by a fellowship from the Swiss Foundation for Grants in Biology and Medicine.

Chapter 2 Supplementary Materials:

Supplementary Figure 2.1: ZFNs induce highly efficient gene targeting events

Supplementary Figure 2.2: Schematic illustrating the "modular assembly" method of engineering multi-finger domains

Supplementary Figure 2.3: Schematic of the bacterial two-hybrid (B2H) reporter used to assess DNA-binding activities of zinc finger arrays

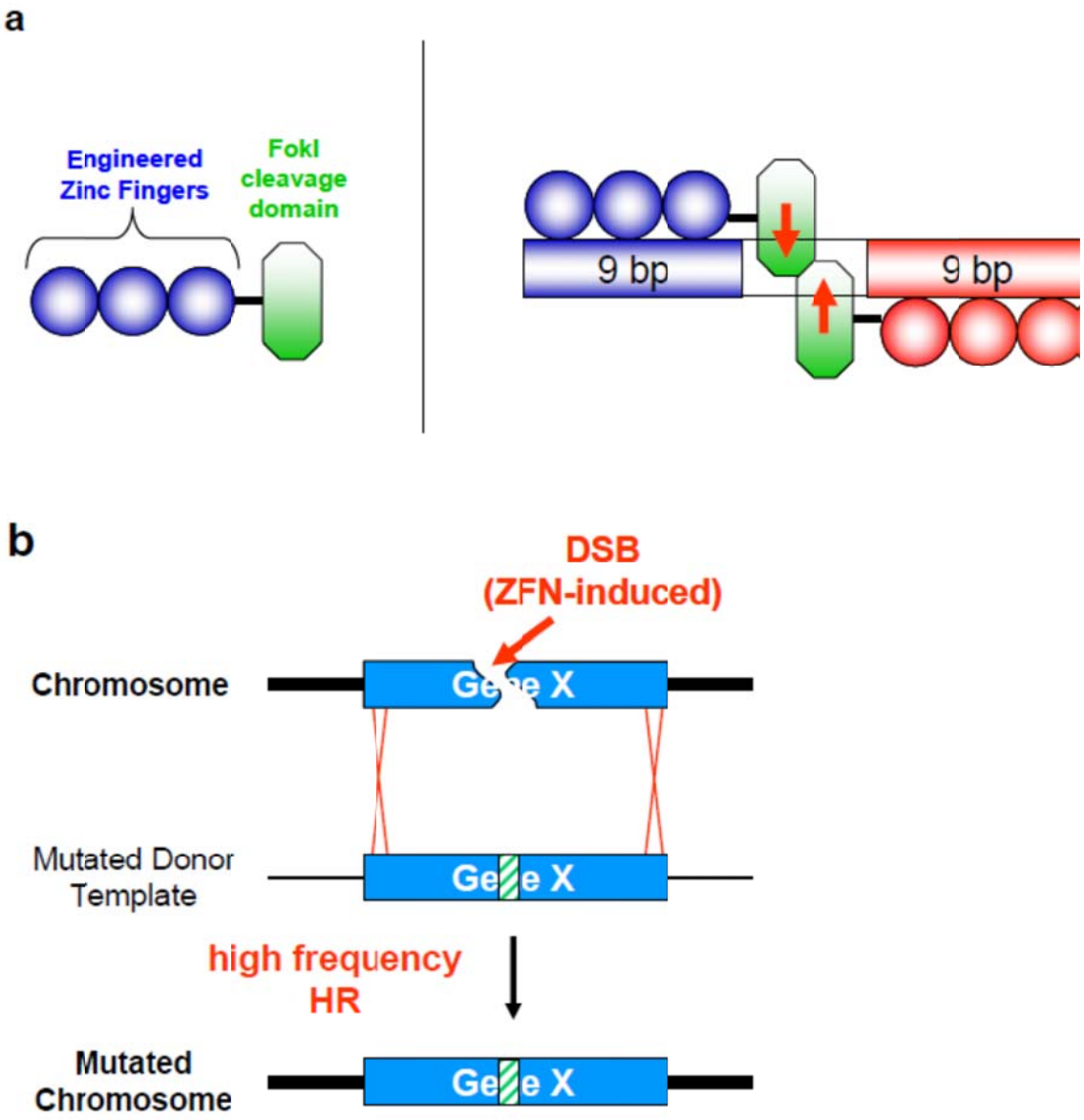
Supplementary Figure 2.4: The B2H assay identifies zinc-finger arrays that fail to show significant activity as ZFNs in human cells

Supplementary Table 2.1: Small-scale tests of modular assembly using various activity assays

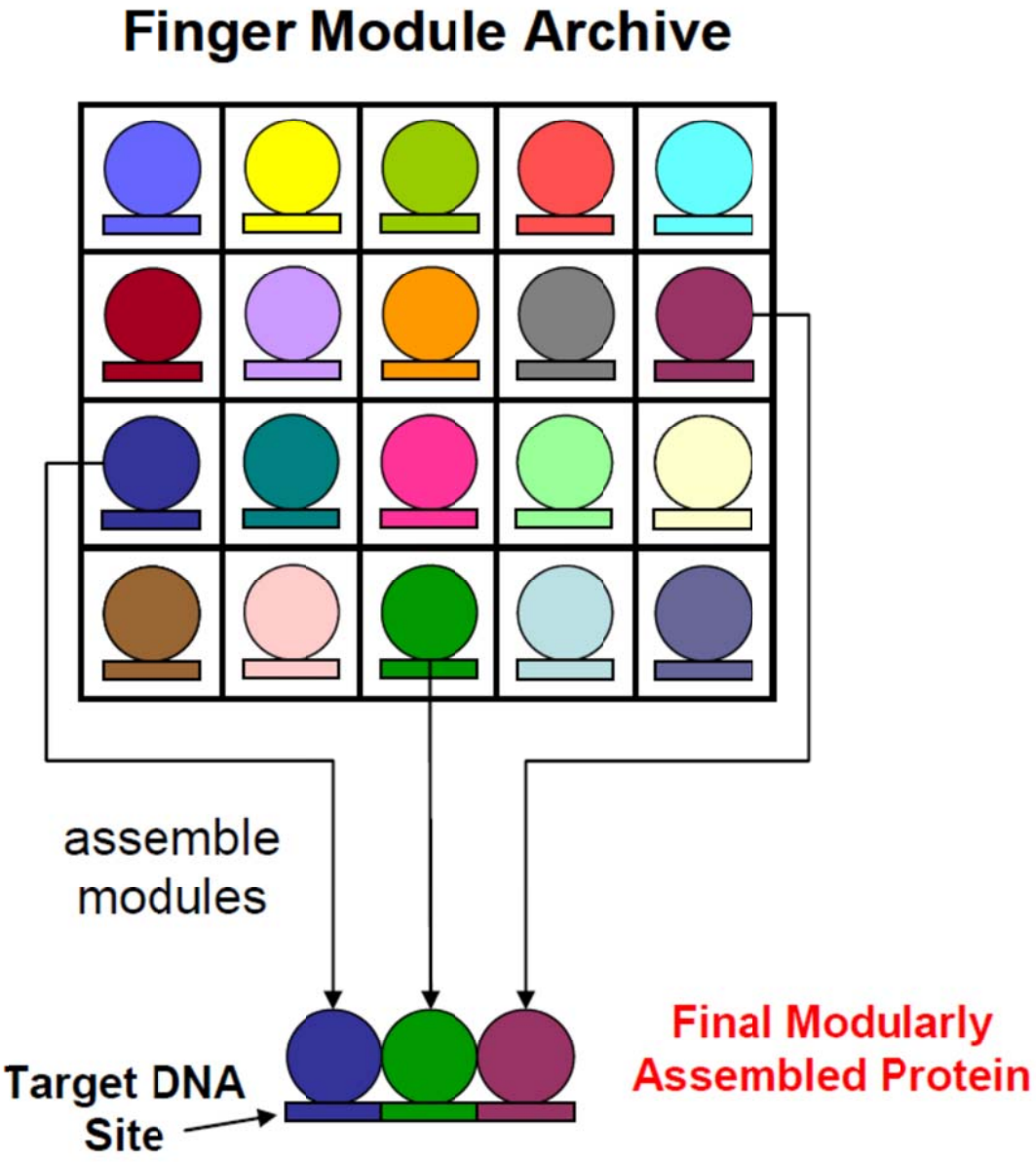
Supplementary Table 2.2: Modularly assembled zinc finger arrays, cognate target binding sites, and their activities in the B2H assay

Chapter 2 Supplementary Discussion

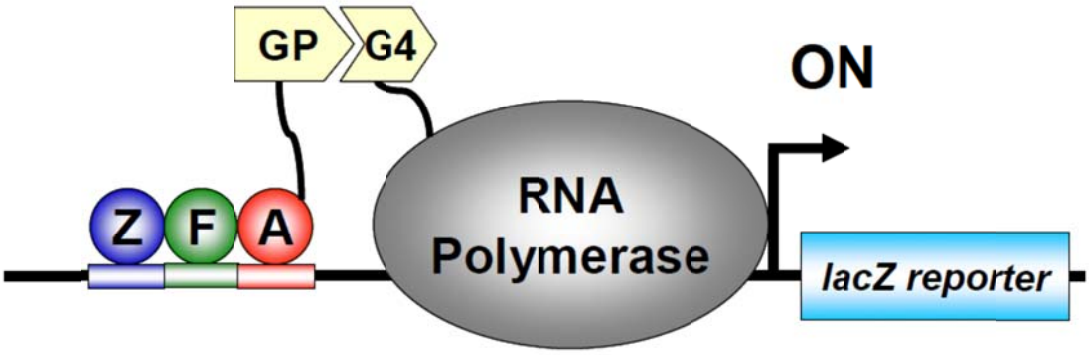
Chapter 2 Supplementary Methods



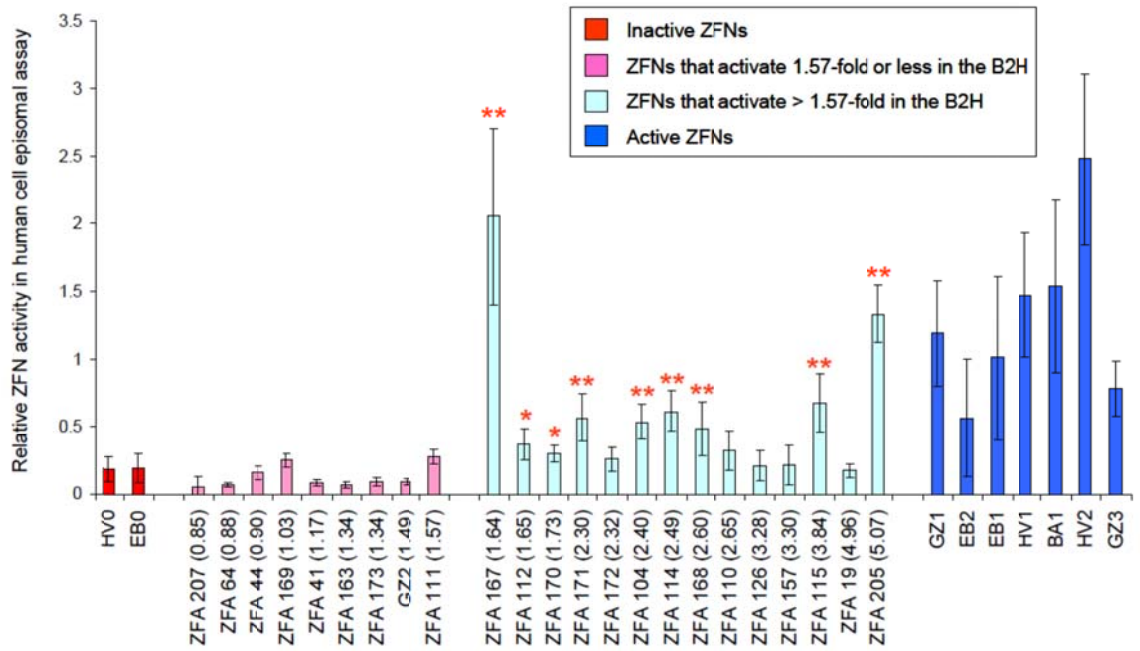
Supplementary Figure 2.1: ZFNs induce highly efficient gene targeting events
(a) Schematic of a ZFN monomer (left) and a pair of ZFNs cleaving DNA as a dimer (right). (b) Schematic representing ZFN-enhanced recombination-based genome manipulation. ZFNs introduce site-specific double-strand DNA breaks (DSBs) that can be harnessed to mediate gene targeting via homologous recombination with an exogenously introduced "donor template."



Supplementary Figure 2.2: Schematic illustrating the "modular assembly" method of engineering multi-finger domains
Fingers from archives of pre-characterized modules are joined together to create three finger domains capable of recognizing a 9 bp target DNA site. Single fingers (colored spheres) recognize their cognate 3 bp "subsites" (colored rectangles).



Supplementary Figure 2.3: Schematic of the bacterial two-hybrid (B2H) reporter used to assess DNA-binding activities of zinc finger arrays
In the B2H assay, the ability of a zinc finger array to bind a target DNA site is reflected as increased expression of a *lacZ* reporter gene encoding β-galactosidase.[1-3] The B2H requires the expression of two hybrid proteins: one a fusion between the test zinc finger array (ZFA) and a fragment of the yeast Gal11P protein (GP) and the other a hybrid of the RNA polymerase α-subunit and a fragment of the yeast Gal4 protein (G4). If the test ZFA binds to a target DNA site positioned upstream of a promoter, RNA polymerase complexes that have incorporated Rα hybrid proteins are recruited to the promoter and expression of the downstream *lacZ* reporter is activated.



Supplementary Figure 2.4: The B2H assay identifies zinc-finger arrays that fail to show significant activity as ZFNs in human cells

23 different zinc-finger arrays (pink and light blue bars) which activated transcription in the B2H assay to various levels (fold-activations are indicated in parentheses on the x-axis) were tested as homodimeric ZFNs using a previously described human cell-based episomal recombination (HR) assay (Cornu et al., 2008). The names of these 23 zinc-finger arrays are shown on the x-axis and details about their binding sites and the modules used to construct them can be found in Supplementary Tables 2.1 and 2.2. ZFN activity is reported as a percentage of EGFP-rescue relative to an I-*Sce*I control. In addition, the episomal repair activities of control ZFNs previously shown to lack or possess significant activity as ZFNs on chromosomal targets in human cells[1] are shown for comparison (red and blue bars, respectively). Names of the control ZFNs are shown on the x-axis and are as previously described.[1] Means of at least three independent experiments and standard deviations (error bars) are shown. ZFNs with mean episomal repair activities that are significantly higher than that of the inactive EB0 ZFN are indicated with one ($p<0.05$) or two ($p<0.01$) red asterisks. The expression levels of all ZFNs shown were verified by Western blot using an antibody to a HA tag present on all proteins (data not shown).

Supplementary Table 2.1: Small-scale tests of modular assembly using various activity assays

Thirty-six three-finger arrays (named "ZFA __" for zinc finger array) were assembled using modules based on the Zinc Finger Consortium Modular Assembly Kit 1.0 (Wright et al., 2006), and the identities of the three modules used to construct each three-finger array use the Consortium numbering scheme (F1, F2, and F3 are the amino-terminal, middle, and carboxy-terminal fingers, respectively). All target binding sites are written 5' to 3' (note that the F3 module binds to the 5' most triplet subsite while F1 binds to the 3' most triplet subsite). The number of GXX subsites in each target site is also indicated. For nine of the arrays, the amino acid sequence of the zinc finger backbone differs from the modules in the archive. These arrays are noted in the far left column by a double asterisk, and their complete amino acid sequences are available upon request. Activity of 27 arrays was tested by electrophoretic mobility shift assays (EMSA). A plus sign in the far right column indicates that the finger array caused a shift in mobility of a double-strand oligonucleotide corresponding to the intended target sequence on polyacrylamide gels. In all cases, binding could be competed away with an excess of the target oligonucleotide. Western blots were performed for all zinc finger arrays to ensure they were expressed. Plant single-strand annealing (PSSA) assays (see Chapter 2 Supplementary Methods) were used to assess function of seven arrays, five of which were also tested by EMSA. The PSSA assay tests the ability of the zinc finger array to function as a ZFN. A minus sign in the table indicates that activity was comparable to negative controls in which plant protoplasts were not transformed with the ZFN construct. An episomal recombination (HR) assay was used to assess the activity of nine ZFNs in human 293T cells (see Chapter 2 Supplementary Methods). A plus sign indicates that ZFN activity in stimulating HR was >40% as compared to the activity of a control I-*Sce*I meganuclease on the same target locus. EMSA, electrophoretic mobility shift assay; a positive value indicates DNA binding that could be competed away by an excess of an oligonucleotide corresponding to the target site. PSSA, plant single-strand annealing assay; a positive value indicates reporter gene function that is at least two-fold over background controls. HR, episomal recombination assay; a positive value indicates at least 40% activity as compared to I-*Sce*I on the same target template. For further information on the activity assays, see Chapter 2 Supplementary Methods.

**Finger arrays with backbone sequences that differ from the ZF finger archive.
***Finger arrays previously described (Alwin et al., 2005).

Supplementary Table 2.1 (continued):

ID#	F1	F2	F3	Target Sequence (F3-F2-F1)	# GXX subsites	Activity*	
						Assay	Results
ZFA 1	73	72	60	5′-GGT-GCT-GCC-3′	3	EMSA	–
ZFA 2	72	72	106	5′-TGG-GCT-GCT-3′	2	EMSA	–
ZFA 3	15	19	43	5′-GCT-GAC-GTG-3′	3	EMSA	–
ZFA 4	14	30	53	5′-AGG-GTC-GTC-3′	2	EMSA	–
ZFA 5	12	23	44	5′-GGA-GCG-GGT-3′	3	EMSA	–
ZFA 6	64	73	63	5′-GAA-GCC-GAT-3′	3	EMSA	–
ZFA 7	58	72	59	5′-GGA-GCT-GGG-3′	3	EMSA	–
ZFA 8	67	60	64	5′-GAT-GGT-GTA-3′	3	EMSA	–
ZFA 9	61	61	63	5′-GAA-GGC-GGC-3′	3	EMSA	–
ZFA 10	70	61	59	5′-GGA-GGC-GCG-3′	3	EMSA	–
ZFA 11	10	23	40	5′-GCA-GCG-GGC-3′	3	EMSA	–
ZFA 12**	15	19	43	5′-GCT-GAC-GTG-3′	3	EMSA	–
ZFA 13**	4	21	39	5′-GAT-GAT-GAT-3′	3	EMSA	–
ZFA 14**	15	31	132	5′-AGG-GTG-GTG-3′	2	EMSA	–
ZFA 15	61	70	71	5′-GCA-GCG-GGC-3′	3	EMSA	+
ZFA 16	62	68	63	5′-GAA-GTT-GAG-3′	3	EMSA	+
ZFA 17	60	70	59	5′-GGA-GCG-GGT-3′	3	EMSA	+
ZFA 18	67	66	65	5′-GAC-GTG-GTA-3′	3	EMSA	+
ZFA 19	66	66	84	5′-AGG-GTG-GTG-3′	2	EMSA	+
ZFA 20	62	62	62	5′-GAG-GAG-GAG-3′	3	EMSA	+
ZFA 21**	67	31	65	5′-GAC-GTG-GTA-3′	3	EMSA	–
						PSSA	–
ZFA 22**	8	24	57	5′-TGG-GCT-GCT-3′	2	EMSA	–
						PSSA	–
ZFA 23	15	31	49	5′-GTG-GTG-GTG-3′	3	EMSA	–
						PSSA	–
ZFA 24**	8	24	43	5′-GCT-GCT-GCT-3′	3	EMSA	–
						PSSA	–
ZFA 25**	15	31	84	5′-AGG-GTG-GTG-3′	2	EMSA	–
						PSSA	–
ZFA 26**	15	31	84	5′-AGG-GTG-GTG-3′	2	PSSA	–
ZFA 27**	13	31	37	5′-GAC-GTG-GTA-3′	3	PSSA	–
ZFA 28***	61	65	65	5′-GAC-GAC-GGC-3′	3	HR	–
						EMSA	–
ZFA 29	100	96	75	5′-AAC-CCT-CGT-3′	0	HR	–
ZFA 30	61	80	61	5′-GGC-ACG-GGC-3′	2	HR	–
ZFA 31	91	104	85	5′-AGT-TAG-CAG-3′	0	HR	–
ZFA 32	83	103	104	5′-TAG-CTT-AGC-3′	0	HR	–
ZFA 33	83	68	74	5′-AAA-GTT-AGC-3′	1	HR	–
ZFA 34	100	104	72	5′-GCT-TAG-CGT-3′	1	HR	–
ZFA 35	17	25	119	5′-AGA-GGA-AGA-3′	1	HR	–
ZFA 36***	60	64	63	5′-GAA-GAT-GGT-3′	3	HR	+
						EMSA	+

Supplementary Table 2.2: Modularly assembled zinc finger arrays, cognate target binding sites, and their activities in the B2H assay

See legend to Supplementary Table 2.1 for descriptive detail pertaining to each of this table's columns. In addition, the original source of the modules used to construct each array (B = Barbas group (Mandell and Barbas, 2006), S = Sangamo BioSciences (Liu et al., 2002, Wright et al., 2006), and T = Toolgen, Inc. (Bae et al., 2003)) is indicated. Because the Zinc Finger Consortium Modular Assembly Kit 1.0 includes more than one module for certain subsites (Wright et al., 2006), we were able to construct 168 zinc finger arrays that could potentially recognize the 104 different target sites. Additional details regarding the construction of the B2H zinc finger expression plasmids and the B2H binding site reporter plasmids are provided in Chapter 2 Supplementary Methods. Fold-activation of transcription in the B2H assay was determined for each zinc finger array using β-galactosidase assays as previously described (Cornu et al., 2008). β-galactosidase assays for each zinc finger array were performed a minimum of four times and means and standard errors of the mean are shown. The expression of all zinc finger arrays that failed to activate transcription by more than 1.57-fold in the B2H assay were verified by Western blot using a monoclonal antibody which detects a FLAG epitope present on all arrays (data not shown).

Supplementary Table 2.2 (continued):

ID#	F1	F2	F3	Module Source	Target Sequence (F3-F2-F1)	# GXX subsites	B2H fold activation (mean)	SEM
ZFA 37	75	75	74	B	5'-AAA-AAC-AAC-3'	0	1.152	0.030
ZFA 38	91	78	74	B	5'-AAA-ACA-CAG-3'	0	1.097	0.030
ZFA 39	73	91	74	B	5'-AAA-CAG-GCC-3'	1	1.325	0.095
ZFA 40	100	101	74	B	5'-AAA-CTA-CGT-3'	0	1.347	0.037
ZFA 41	83	68	74	B	5'-AAA-GTT-AGC-3'	1	1.166	0.106
ZFA 42	91	93	75	B	5'-AAC-CCA-CAG-3'	0	0.974	0.075
ZFA 43	68	93	75	B	5'-AAC-CCA-GTT-3'	1	0.980	0.010
ZFA 44	100	96	75	B	5'-AAC-CCT-CGT-3'	0	0.900	0.105
ZFA 45	64	86	77	B	5'-AAT-ATA-GAT-3'	1	1.606	0.037
ZFA 46	90	75	78	B	5'-ACA-AAC-CAC-3'	0	0.850	0.022
ZFA 47	85	75	79	B	5'-ACC-AAC-AGT-3'	0	1.001	0.028
ZFA 48	86	83	79	B	5'-ACC-AGC-ATA-3'	0	1.036	0.028
ZFA 49	83	100	79	B	5'-ACC-CGT-AGC-3'	0	1.235	0.016
ZFA 50	63	59	79	B	5'-ACC-GGA-GAA-3'	2	1.386	0.068
ZFA 51	91	83	80	B	5'-ACG-AGC-CAG-3'	0	0.955	0.027
ZFA 52	72	91	80	B	5'-ACG-CAG-GCT-3'	1	1.236	0.007
ZFA 53	70	95	80	B	5'-ACG-CCG-GCG-3'	1	1.266	0.040
ZFA 54	100	65	80	B	5'-ACG-GAC-CGT-3'	1	1.799	0.060
ZFA 55	71	73	80	B	5'-ACG-GCC-GCA-3'	2	1.295	0.031
ZFA 56	82	59	82	B	5'-AGA-GGA-AGA-3'	1	1.287	0.021
ZFA 57	119	117	119	T	5'-AGA-GGA-AGA-3'	1	0.911	0.012
ZFA 58	119	114	119	T	5'-AGA-GGA-AGA-3'	1	1.928	0.082
ZFA 59	119	129	119	T	5'-AGA-GGA-AGA-3'	1	1.974	0.066
ZFA 60	119	118	119	T	5'-AGA-GGA-AGA-3'	1	2.371	0.047
ZFA 61	64	69	83	B	5'-AGC-GTC-GAT-3'	2	1.206	0.019
ZFA 62	103	88	85	B	5'-AGT-ATT-CTT-3'	0	1.015	0.022
ZFA 64	91	104	85	B	5'-AGT-TAG-CAG-3'	0	0.879	0.092
ZFA 65	85	93	86	B	5'-ATA-CCA-AGT-3'	0	0.989	0.017
ZFA 66	78	81	87	B	5'-ATG-ACT-ACA-3'	0	0.956	0.043
ZFA 67	92	90	87	B	5'-ATG-CAC-CAT-3'	0	1.122	0.068
ZFA 68	78	96	88	B	5'-ATT-CCT-ACA-3'	0	1.064	0.135
ZFA 69	95	72	88	B	5'-ATT-GCT-CCG-3'	1	1.256	0.016
ZFA 70	63	98	90	B	5'-CAC-CGC-GAA-3'	1	1.099	0.011
ZFA 71	60	67	90	B	5'-CAC-GTA-GGT-3'	2	1.476	0.023
ZFA 72	86	98	92	B	5'-CAT-CGC-ATA-3'	0	0.985	0.027
ZFA 74	90	100	93	B	5'-CCA-CGT-CAC-3'	0	1.173	0.034
ZFA 75	73	64	93	B	5'-CCA-GAT-GCC-3'	2	1.085	0.006
ZFA 76	103	85	94	B	5'-CCC-AGT-CTT-3'	0	1.376	0.153
ZFA 77	66	100	94	B	5'-CCC-CGT-GTG-3'	1	1.431	0.069
ZFA 78	80	79	95	B	5'-CCG-ACC-ACG-3'	0	1.090	0.020
ZFA 79	74	91	95	B	5'-CCG-CAG-AAA-3'	0	1.123	0.009
ZFA 80	100	98	95	B	5'-CCG-CGC-CGT-3'	0	1.146	0.074
ZFA 81	86	63	95	B	5'-CCG-GAA-ATA-3'	1	0.907	0.048
ZFA 82	91	80	96	B	5'-CCT-ACG-CAG-3'	0	1.013	0.037
ZFA 83	90	81	96	B	5'-CCT-ACT-CAC-3'	0	1.113	0.042
ZFA 84	70	87	96	B	5'-CCT-ATG-GCG-3'	1	1.239	0.104
ZFA 85	58	67	96	B	5'-CCT-GTA-GGG-3'	2	1.504	0.009

Supplementary Table 2.2 (continued):

ZFA 86	101	92	98	B	5'-CGC-CAT-CTA-3'	0	1.052	0.052
ZFA 87	88	81	101	B	5'-CTA-ACT-ATT-3'	0	0.884	0.013
ZFA 88	83	59	101	B	5'-CTA-GGA-AGC-3'	1	1.303	0.089
ZFA 89	80	83	103	B	5'-CTT-AGC-ACG-3'	0	1.037	0.086
ZFA 90	94	88	103	B	5'-CTT-ATT-CCC-3'	0	1.031	0.018
ZFA 91	91	60	103	B	5'-CTT-GGT-CAG-3'	1	1.044	0.032
ZFA 92	69	70	63	B	5'-GAA-GCG-GTC-3'	3	2.309	0.129
ZFA 93	14	23	36	S	5'-GAA-GCG-GTC-3'	3	1.620	0.059
ZFA 94	109	130	124	T	5'-GAA-GCG-GTC-3'	3	1.339	0.165
ZFA 95	109	130	122	T	5'-GAA-GCG-GTC-3'	3	2.279	0.123
ZFA 96	87	94	65	B	5'-GAC-CCC-ATG-3'	1	1.074	0.034
ZFA 97	72	61	65	B	5'-GAC-GGC-GCT-3'	3	1.118	0.045
ZFA 98	8	26	37	S	5'-GAC-GGC-GCT-3'	3	0.930	0.039
ZFA 101	4	24	38	S	5'-GAG-GCT-GAT-3'	3	1.327	0.126
ZFA 102	112	139	113	T	5'-GAG-GCT-GAT-3'	3	1.142	0.091
ZFA 103	112	139	136	T	5'-GAG-GCT-GAT-3'	3	1.185	0.087
ZFA 104	112	139	115	T	5'-GAG-GCT-GAT-3'	3	2.399	0.120
ZFA 105	66	72	62	B	5'-GAG-GCT-GTG-3'	3	4.553	0.295
ZFA 106	15	24	38	S	5'-GAG-GCT-GTG-3'	3	3.038	0.153
ZFA 107	138	139	136	T	5'-GAG-GCT-GTG-3'	3	0.959	0.026
ZFA 108	138	139	113	T	5'-GAG-GCT-GTG-3'	3	1.004	0.037
ZFA 109	138	139	115	T	5'-GAG-GCT-GTG-3'	3	3.277	0.153
ZFA 110	68	59	64	B	5'-GAT-GGA-GTT-3'	3	2.647	0.158
ZFA 111	16	25	39	S	5'-GAT-GGA-GTT-3'	3	1.571	0.111
ZFA 112	111	118	112	T	5'-GAT-GGA-GTT-3'	3	1.645	0.071
ZFA 113	111	129	112	T	5'-GAT-GGA-GTT-3'	3	2.013	0.083
ZFA 114	111	117	112	T	5'-GAT-GGA-GTT-3'	3	2.489	0.136
ZFA 115	111	114	112	T	5'-GAT-GGA-GTT-3'	3	3.840	0.308
ZFA 116	58	60	64	B	5'-GAT-GGT-GGG-3'	3	2.390	0.065
ZFA 117	11	28	39	S	5'-GAT-GGT-GGG-3'	3	0.959	0.098
ZFA 118	131	140	112	T	5'-GAT-GGT-GGG-3'	3	1.985	0.067
ZFA 119	133	140	112	T	5'-GAT-GGT-GGG-3'	3	2.444	0.103
ZFA 120	134	140	112	T	5'-GAT-GGT-GGG-3'	3	6.665	0.290
ZFA 121	60	78	71	B	5'-GCA-ACA-GGT-3'	2	1.388	0.014
ZFA 122	100	80	71	B	5'-GCA-ACG-CGT-3'	1	1.274	0.031
ZFA 123	81	90	71	B	5'-GCA-CAC-ACT-3'	1	0.851	0.023
ZFA 124	59	73	71	B	5'-GCA-GCC-GGA-3'	3	1.574	0.099
ZFA 125	9	33	40	S	5'-GCA-GCC-GGA-3'	3	0.953	0.111
ZFA 126	71	70	71	B	5'-GCA-GCG-GCA-3'	3	3.276	0.111
ZFA 127	5	23	40	S	5'-GCA-GCG-GCA-3'	3	0.887	0.068
ZFA 128	58	72	71	B	5'-GCA-GCT-GGG-3'	3	1.186	0.045
ZFA 129	11	24	40	S	5'-GCA-GCT-GGG-3'	3	1.157	0.039
ZFA 130	65	69	71	B	5'-GCA-GTC-GAC-3'	3	0.867	0.046
ZFA 131	2	30	40	S	5'-GCA-GTC-GAC-3'	3	0.897	0.052
ZFA 132	72	69	71	B	5'-GCA-GTC-GCT-3'	3	1.365	0.049
ZFA 133	8	30	40	S	5'-GCA-GTC-GCT-3'	3	1.116	0.092
ZFA 134	61	69	71	B	5'-GCA-GTC-GGC-3'	3	1.077	0.024
ZFA 135	10	30	40	S	5'-GCA-GTC-GGC-3'	3	0.999	0.039
ZFA 136	70	66	71	B	5'-GCA-GTG-GCG-3'	3	1.915	0.050
ZFA 137	7	31	40	S	5'-GCA-GTG-GCG-3'	3	0.819	0.051
ZFA 138	95	79	73	B	5'-GCC-ACC-CCG-3'	1	1.000	0.032
ZFA 139	59	87	73	B	5'-GCC-ATG-GGA-3'	2	0.974	0.043
ZFA 140	103	63	73	B	5'-GCC-GAA-CTT-3'	2	1.137	0.057

Supplementary Table 2.2 (continued):

ZFA 141	66	71	73	B	5'-GCC-GCA-GTG-3'	3	1.061	0.036
ZFA 142	15	22	41	S	5'-GCC-GCA-GTG-3'	3	0.914	0.023
ZFA 143	59	61	73	B	5'-GCC-GGC-GGA-3'	3	1.182	0.049
ZFA 144	9	26	41	S	5'-GCC-GGC-GGA-3'	3	0.814	0.021
ZFA 145	132	116	130	T	5'-GCG-GAA-AGG-3'	2	1.342	0.134
ZFA 146	132	124	130	T	5'-GCG-GAA-AGG-3'	2	1.736	0.063
ZFA 147	132	122	130	T	5'-GCG-GAA-AGG-3'	2	1.759	0.054
ZFA 148	132	123	130	T	5'-GCG-GAA-AGG-3'	2	2.136	0.064
ZFA 149	102	59	70	B	5'-GCG-GGA-CTG-3'	2	1.522	0.088
ZFA 150	59	66	70	B	5'-GCG-GTG-GGA-3'	3	1.751	0.124
ZFA 151	9	31	42	S	5'-GCG-GTG-GGA-3'	3	2.774	0.116
ZFA 152	114	138	130	T	5'-GCG-GTG-GGA-3'	3	1.029	0.046
ZFA 153	129	138	130	T	5'-GCG-GTG-GGA-3'	3	1.272	0.039
ZFA 154	118	138	130	T	5'-GCG-GTG-GGA-3'	3	1.384	0.059
ZFA 155	117	138	130	T	5'-GCG-GTG-GGA-3'	3	1.953	0.050
ZFA 156	60	66	70	B	5'-GCG-GTG-GGT-3'	3	2.229	0.208
ZFA 157	12	31	42	S	5'-GCG-GTG-GGT-3'	3	3.298	0.266
ZFA 158	70	106	70	B	5'-GCG-TGG-GCG-3'	2	2.713	0.181
ZFA 159	7	34	42	S	5'-GCG-TGG-GCG-3'	2	0.856	0.034
ZFA 160	78	86	72	B	5'-GCT-ATA-ACA-3'	1	0.988	0.024
ZFA 161	62	100	72	B	5'-GCT-CGT-GAG-3'	2	1.703	0.012
ZFA 163	100	104	72	B	5'-GCT-TAG-CGT-3'	1	1.343	0.143
ZFA 164	61	74	59	B	5'-GGA-AAA-GGC-3'	2	1.392	0.084
ZFA 165	92	90	59	B	5'-GGA-CAC-CAT-3'	1	1.182	0.016
ZFA 166	80	101	59	B	5'-GGA-CTA-ACG-3'	1	1.085	0.034
ZFA 167	66	70	59	B	5'-GGA-GCG-GTG-3'	3	1.640	0.043
ZFA 168	15	23	44	S	5'-GGA-GCG-GTG-3'	3	2.599	0.092
ZFA 169	138	130	117	T	5'-GGA-GCG-GTG-3'	3	1.027	0.068
ZFA 170	138	130	118	T	5'-GGA-GCG-GTG-3'	3	1.732	0.041
ZFA 171	138	130	129	T	5'-GGA-GCG-GTG-3'	3	2.296	0.121
ZFA 172	138	130	114	T	5'-GGA-GCG-GTG-3'	3	2.316	0.046
ZFA 173	61	80	61	B	5'-GGC-ACG-GGC-3'	2	1.341	0.109
ZFA 174	63	95	61	B	5'-GGC-CCG-GAA-3'	2	1.171	0.030
ZFA 175	66	73	61	B	5'-GGC-GCC-GTG-3'	3	1.116	0.072
ZFA 176	15	33	45	S	5'-GGC-GCC-GTG-3'	3	0.868	0.017
ZFA 177	73	58	61	B	5'-GGC-GGG-GCC-3'	3	1.203	0.035
ZFA 178	6	27	45	S	5'-GGC-GGG-GCC-3'	3	1.258	0.053
ZFA 180	61	70	58	B	5'-GGG-GCG-GGC-3'	3	2.191	0.105
ZFA 181	10	23	51	S	5'-GGG-GCG-GGC-3'	3	6.162	0.147
ZFA 182	73	72	58	B	5'-GGG-GCT-GCC-3'	3	1.363	0.053
ZFA 183	6	24	51	S	5'-GGG-GCT-GCC-3'	3	1.494	0.054
ZFA 184	108	139	134	T	5'-GGG-GCT-GCC-3'	3	2.022	0.253
ZFA 185	79	88	60	B	5'-GGT-ATT-ACC-3'	1	1.078	0.016
ZFA 186	72	63	60	B	5'-GGT-GAA-GCT-3'	3	1.589	0.094
ZFA 187	8	18	46	S	5'-GGT-GAA-GCT-3'	3	0.944	0.081
ZFA 188	139	122	140	T	5'-GGT-GAA-GCT-3'	3	1.176	0.125
ZFA 189	139	124	140	T	5'-GGT-GAA-GCT-3'	3	1.800	0.124
ZFA 190	139	123	140	T	5'-GGT-GAA-GCT-3'	3	2.189	0.234
ZFA 191	139	116	140	T	5'-GGT-GAA-GCT-3'	3	2.431	0.090
ZFA 192	98	103	67	B	5'-GTA-CTT-CGC-3'	1	1.274	0.034
ZFA 193	61	68	67	B	5'-GTA-GTT-GGC-3'	3	1.056	0.098
ZFA 194	10	32	47	S	5'-GTA-GTT-GGC-3'	3	0.854	0.028
ZFA 195	91	101	69	B	5'-GTC-CTA-CAG-3'	1	1.132	0.035

Supplementary Table 2.2 (continued):

ZFA 196	81	65	66	B	5'-GTG-GAC-ACT-3'	2	1.387	0.028
ZFA 197	73	61	66	B	5'-GTG-GGC-GCC-3'	3	1.104	0.036
ZFA 198	6	26	49	S	5'-GTG-GGC-GCC-3'	3	0.984	0.055
ZFA 199	66	82	68	B	5'-GTT-AGA-GTG-3'	2	1.683	0.046
ZFA 200	138	119	111	T	5'-GTT-AGA-GTG-3'	2	2.872	0.107
ZFA 201	12	20	50	S	5'-GTT-GAG-GGT-3'	3	0.671	0.015
ZFA 202	140	136	111	T	5'-GTT-GAG-GGT-3'	3	3.477	0.076
ZFA 203	140	113	111	T	5'-GTT-GAG-GGT-3'	3	4.735	0.489
ZFA 204	101	71	68	B	5'-GTT-GCA-CTA-3'	2	1.009	0.043
ZFA 205	73	70	68	B	5'-GTT-GCG-GCC-3'	3	5.065	0.290
ZFA 206	6	23	50	S	5'-GTT-GCG-GCC-3'	3	1.232	0.014
ZFA 207	83	103	104	B	5'-TAG-CTT-AGC-3'	0	0.849	0.100
ZFA 208	70	59	104	B	5'-TAG-GGA-GCG-3'	2	2.185	0.068
ZFA 209	3	20	56	S	5'-TGA-GAG-GAG-3'	2	1.704	0.068
ZFA 210	86	105	105	B	5'-TGA-TGA-ATA-3'	0	1.001	0.030

Chapter 2 Supplementary Discussion:

Small-scale surveys suggest a low success rate for modular assembly

Motivated by a desire to use ZFNs for genome modification, our labs constructed the 36 zinc finger arrays listed in Supplementary Table 2.1. These arrays are highly biased for GXX subsites, and collectively include 22 3-GXX arrays, 8 2-GXX arrays, 2 1-GXX arrays and 4 arrays without GXX subsites. Using three different assays to test for function (see legend to Supplementary Table 2.1), seven arrays were deemed functional. Six of these sites were composed of three GXX subsites and one was composed of two GXX subsites. Since six of the seven arrays were only tested by EMSA, it is difficult to extrapolate how many of these arrays would function *in vivo* when challenged with the diverse sequence targets present in a genome.

The B2H assay accurately identifies zinc finger arrays that fail to show ZFN function in human cells

In previously published studies from our groups, we observed a general correlation between failure to activate in the B2H assay and failure to show ZFN function in human cells (Cornu et al., 2008). To further assess this correlation, we tested the activities of 23 ZFNs using a human cell-based episomal homologous recombination (HR) assay (Chapter 2 Supplementary Methods). The zinc finger arrays in these 23 ZFNs each activated transcription to different levels in the B2H system ranging from 0.85- to 5.07-fold. As shown in Supplementary Figure 2.4, we found that zinc finger arrays which activated transcription by 1.57-fold or less in the B2H system all failed to show significant activity as ZFNs in human cells (nine out of nine ZFNs tested; pink bars in Supplementary Figure 2.4). Conversely, many (although not all) of the zinc finger arrays which activated transcription by >1.57-fold in the B2H system showed significant activity as ZFNs in the episomal repair assay (light blue bars with red asterisks in Supplementary Figure

2.4). We conclude that zinc finger arrays that activate transcription by 1.57-fold or less in the B2H system are unlikely to function as ZFNs in human cells and we used this threshold level to interpret the results of our experiments.

Predicted failure rates for identifying zinc finger arrays needed to engineer a ZFN pair

Analysis of our results shows that modular assembly is far less effective for target sites composed of two, one, or no GXX subsites (71%, 88% and 100% failure rates, respectively; Figure 2.1a) compared with those that contain three GXX subsites (41% failure rate; Figure 2.1a). Because ZFNs function as dimers, the failure rates for making a ZFN pair will therefore be ~65%, ~92%, ~99%, and 100% for ZFN targets composed of pairs of 3-GXX, 2-GXX, 1-GXX, and 0-GXX 9 bp "half-sites", respectively. These failure rates are calculated by multiplying estimated success rates for each monomer in the ZFN pair and subtracting this percentage from 100%.

Alternative selection-based strategies for engineering multi-finger arrays

A fundamental flaw with the modular assembly method (and a likely cause of its low success rate) is its assumption that zinc finger domains behave as independent modular units. A number of studies have shown that zinc fingers can cross over and interact with adjacent fingers and neighboring finger binding sites (Elrod-Erickson et al., 1996, Isalan et al., 1997, Isalan et al., 1998, Wolfe et al., 2001). Various engineering strategies have been described in the literature that account for these context-dependent effects on zinc finger DNA-binding. Greisman and Pabo first described a sequential optimization strategy in which combinations of fingers that work well together are identified using serial selections from randomized libraries (Greisman and Pabo, 1997). Isalan, Klug, and Choo described a "bipartite" optimization strategy in which "halves" of a three-finger domain are first optimized by randomization and selection and then

joined together to create a final protein (Isalan et al., 2001). Finally, Joung and colleagues described a domain shuffling approach in which pools of fingers are first identified for each subsite in parallel and then shuffled together to create a recombined library for use in a final stringent selection (Hurt et al., 2003). A limitation of all of these different approaches is that they require specialized expertise in the construction of multiple very large randomized libraries and in the interrogation of these libraries using selection methods.

Chapter 2 Supplementary Methods

Construction of B2H zinc finger expression vectors and reporter strains

104 B2H reporter strains each harboring a single copy target site-reporter plasmid were constructed as previously described (Wright et al., 2006). B2H expression vectors encoding different three finger arrays were constructed essentially as previously described using the Zinc Finger Modular Assembly Kit 1.0 which includes 141 modules made by the Barbas group, Sangamo BioSciences, and Toolgen, Inc. (Wright et al., 2006). For each target site, we made three-finger arrays using modules from the Barbas, Sangamo, and Toolgen collections but we did not mix and match modules across these different sets. We chose not to make "cross-platform" arrays because: (1) the Barbas website software does not advocate use of their modules with others (Mandell and Barbas, 2006), (2) the Toolgen group discovered that their human zinc fingers worked best with one another and not as well with other engineered modules (Bae et al., 2003), and (3) the Sangamo modules are position-specific and have linkers joining them that differ from the canonical TGEKP linker used by the Barbas and Toolgen modules (Liu et al., 2002). 15 of the arrays we constructed were toxic when expressed in the B2H assay and these proteins were not included in our analysis. After setting aside these toxic proteins, we

characterized the remaining 168 zinc finger arrays. All expression and reporter plasmids were confirmed to be correct by DNA sequencing prior to use in B2H assays.

B2H assays, verification of protein expression, and re-verification of DNA sequences

The activities of three-finger arrays were tested in the B2H assay by co-transforming a target site reporter strain with a B2H zinc finger array-Gal11P expression vector and a compatible plasmid encoding an *E. coli* RNA polymerase alpha-Gal4 hybrid protein (Wright et al., 2006). Fold-activation was calculated as the ratio of the level of β-galactosidase (*lacZ*) reporter activity in the presence and absence of the zinc finger array as previously described (Thibodeau et al., 2004). All β-galactosidase measurements were performed at least four independent times. The expression of each zinc finger array that failed to activate transcription by more than 1.57-fold in the B2H assay was verified by performing Western blot on cell extracts from the β-galactosidase assays using a monoclonal antibody against the FLAG epitope present on all zinc finger array-Gal11P fusion proteins. For zinc finger arrays that exhibited less than 1.57-fold transcriptional activation in the B2H assay, we re-sequenced the zinc finger coding sequences and the reporter plasmid binding sites from the same cells used for β-galactosidase assays to re-confirm the identities of the zinc finger recognition helices and binding sites tested in these cells. We performed this additional re-sequencing control on >80% of our samples (94 of the 115 zinc finger array/binding site combinations) that were negative in the B2H assay and found that all helices and binding sites were correct as expected.

Plant single-strand annealing (PSSA) assays

Zinc finger arrays were fused to the *FokI* nuclease domain and the resulting ZFNs were transiently expressed in tobacco protoplasts. Plasmids encoding these ZFNs were co-transformed with a target plasmid carrying a non-functional β-glucuronidase (*GUS*) gene. The *GUS* gene was

rendered non-functional by a direct duplication of part of its coding sequence. Between the direct repeats, a recognition site was inserted for the given ZFN being tested. After expression of the ZFN and cleavage of the *GUS* reporter, repair by single strand annealing can restore GUS function. This can be measured by standard GUS activity assays. We routinely observe GUS activity 25-fold above background when the assay is performed with a ZFN based on the transcription factor Zif268.

Episomal recombination (HR) assay in human cells

For the episomal HR assay, 293T cells in 24-well plates were transfected by calcium phosphate precipitation with 20 ng of the respective target plasmid, 1 μg of the donor plasmid (pUC.Zgfp/REx), and 100 ng of a CMV-driven endonuclease expression vector encoding the ZFNs, I-SceI (pRK5.LHA-Sce1) or a control vector (pCMV.Luc). Two days after transfection, 50,000 cells were analyzed by flow cytometry (FACSCalibur, BD Bioscience) to determine the percentage of EGFP-positive cells. The number of DsRedExpress (REx)-positive cells was used to normalize for transfection efficiency. All experiments were repeated at least three times. ZFN activity is indicated as the fraction of EGFP-positive cells relative to the number of EGFP-positive cells measured in the presence of I-*Sce*I.

Revisions to the Zinc Finger Consortium website

We have previously described ZiFiT, a web-based software program that enables users to rapidly identify potential zinc finger target sites within genes of interest (Sander et al., 2007). ZiFiT identifies target sites for which zinc finger arrays might be made using modular assembly and the Zinc Finger Consortium Modular Assembly Kit 1.0 (Wright et al., 2006). Given the results of this current study, we have revised the ZiFiT program by adding the following section to its instructions:

***Scoring*:** To assess the effectiveness of modular assembly, we generated a large number of zinc finger arrays and tested their activity using a bacterial cell-based two-hybrid (B2H) assay (Ramirez et al., 2008). In the B2H assay, productive interaction of a zinc finger array with its cognate binding site activates transcription of a linked reporter gene. A total of 168 three finger arrays were assembled that recognize 104 different target DNA sites that varied in their GXX, AXX, CXX, and TXX subsite composition. Transcriptional activation of >1.57-fold over negative controls was determined to be indicative of target site recognition by a given protein. After measuring the activity of the 168 zinc finger arrays, it was found that the subsite composition of binding sites critically affected success rates. Modular assembly was far less effective for target sites composed of one or no GXX subsites (12% and 0% success rates, respectively) compared with those that contain three or two (59% and 29% success rates, respectively). Based on these results, we provide users of ZiFiT with an approximation of the likelihood that a given three finger protein will recognize its target. For example, three finger proteins comprised of three GXX subsites will receive a score of 0.59, reflecting the success rate observed in our survey. Most users will be interested in the likelihood that two arrays will function together as a ZFN. This can be approximated by multiplying the success rate of individual arrays. For example, the likelihood of success that a ZFN with all GXX subsites will recognize its target is 0.59 X 0.59 = 0.35 or 35%. Because multiple modules are often available for given subsites, investigators may increase their likelihood of success by making multiple different proteins against a single target. One note of caution: the scoring function is a prediction for the activity of a given three finger array in the B2H assay; the protein may behave differently when tested for activity in a eukaryotic cell. We have also revised the output that ZiFiT currently provides to users to include guidance about the likelihood of success for different target DNA binding sites.

Chapter Three

Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects

Cherie L. Ramirez*, Michael T. Certo*, Claudio Mussolino, Mathew J. Goodwin, Thomas J. Cradick, Anton P. McCaffrey, Toni Cathomen, Andrew M. Scharenberg, and J. Keith Joung

The thesis author and Michael Certo were joint first authors (denoted by "*") of this work. The thesis author hypothesized the method for generating nickases, optimized conditions for *in vitro* protein activity assays and coordinated the efforts of the collaborating authors. Mathew Goodwin performed *in vitro* assays in replicate and assisted with data analysis. With regard to human cell experiments, Michael Certo designed and carried out Traffic Light Reporter assays and Claudio Mussolino performed GFP reconstitution assays. Thomas Cradick provided assistance with determining initial conditions for *in vitro* protein activity assays.

Reference: "Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects." Ramirez CL, Certo MT, Mussolino C, Goodwin MJ, Cradick TJ, McCaffrey AP, Cathomen T, Scharenberg AM, Joung JK. Nucleic Acids Research. 2012 Feb 28. Published online ahead of print. Reprinted by permission of Oxford University Press. The article has been adapted from its published version for presentation in the dissertation.

Engineered Zinc Finger Nickases Induce Homology-Directed Repair with Reduced Mutagenic Effects

Cherie L. Ramirez[1,2,9], Michael T. Certo[3,4,9], Claudio Mussolino[5], Mathew J. Goodwin[1], Thomas J Cradick[6], Anton P. McCaffrey[6], Toni Cathomen[5], Andrew M. Scharenberg[4,7], and J. Keith Joung[1,2,8]

[1] Molecular Pathology Unit, Center for Cancer Research, and Center for Computational and Integrative Biology, Massachusetts General Hospital, Charlestown, Massachusetts, 02129, USA.
[2] Biological and Biomedical Sciences Program, Harvard Medical School, Boston, Massachusetts, 02115, USA.
[3] Program in Molecular and Cellular Biology, University of Washington, Seattle, Washington, 98195, USA.
[4] Center of Immunity and Immunotherapies, Seattle Children's Research Institute, Seattle, Washington, 98105, USA.
[5] Institute of Experimental Hematology, Hannover Medical School, Hannover, 30625, Germany.
[6] University of Iowa School of Medicine, Department of Internal Medicine, Iowa City, Iowa, 52245, USA.
[7] Department of Pediatrics, University of Washington, Seattle, Washington, 98105, USA.
[8] Department of Pathology, Harvard Medical School, Boston, Massachusetts, 02115, USA.
[9] The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint first authors.

Abstract

Engineered zinc finger nucleases (ZFNs) induce DNA double-strand breaks at specific recognition sequences and can promote efficient introduction of desired insertions, deletions, or substitutions at or near the cut site via homology-directed repair (HDR) with a double- and/or single-stranded donor DNA template. However, mutagenic events caused by error-prone non-

homologous end-joining (NHEJ)-mediated repair are introduced with equal or higher frequency at the nuclease cleavage site. Furthermore, unintended mutations can also result from NHEJ-mediated repair of off-target nuclease cleavage sites. Here we describe a simple and general method for converting engineered ZFNs into zinc finger nickases (ZFNickases) by inactivating the catalytic activity of one monomer in a ZFN dimer. ZFNickases show robust strand-specific nicking activity *in vitro*. In addition, we demonstrate that ZFNickases can stimulate HDR at their nicking site in human cells, albeit at a lower frequency than by the ZFNs from which they were derived. Finally, we find that ZFNickases appear to induce greatly reduced levels of mutagenic NHEJ at their target nicking site. ZFNickases thus provide a promising means for inducing HDR-mediated gene modifications while reducing unwanted mutagenesis caused by error-prone NHEJ.

Introduction

Zinc finger nucleases (ZFNs) are chimeras of engineered zinc finger domains fused to the non-specific nuclease domain of the restriction enzyme FokI (Kim et al., 1996). Dimers of ZFNs generate site-specific DNA double-strand breaks (DSBs) with each ZFN monomer cutting one DNA strand (Mani et al., 2005). Obligate heterodimeric versions of the FokI nuclease domain have been engineered that minimize homodimeric interactions between ZFN monomers within a pair (Miller et al., 2007, Szczepek et al., 2007, Sollu et al., 2010, Doyon et al., 2011).

ZFNs, as well as engineered homing endonucleases and transcription activator-like effector nucleases (TALENs), can be used to improve the efficiency of homology-directed repair (HDR) in a variety of different organisms and cell types (Arnould et al., 2011, Bogdanove and Voytas, 2011, Handel and Cathomen, 2011). Repair of a nuclease-induced DSB mediated by an

exogenous “donor template” can be exploited to introduce sequence alterations or insertions at or near the site of the break. Although nuclease-induced HDR is highly efficient, repair of a DSB can also occur by the non-homologous end-joining (NHEJ) pathway. NHEJ-mediated repair of nuclease-induced DSBs has been shown to be error-prone, leading to insertion or deletion mutations (indels) at the site of the break (Bibikova et al., 2002) or the formation of chromosomal translocations (Brunet et al., 2009). NHEJ and HDR are believed to be competing pathways (Hartlerode and Scully, 2009). Thus, although HDR-mediated alterations can be efficiently introduced using engineered nucleases, alleles can also acquire NHEJ-mediated mutations (e.g. (Maeder et al., 2008, Zou et al., 2011)). Unwanted alterations at other off-target genomic sites can also be introduced by NHEJ-mediated repair; for example, two recent reports have shown that ZFNs introduce a greater spectrum of off-target DSBs (and therefore NHEJ-mediated mutations) than previously appreciated (Gabriel et al., 2011, Pattanayak et al., 2011).

Given the potential undesirable consequences of introducing DSBs in living cells, we hypothesized that it might be possible to induce DNA repair with single strand breaks (SSBs or nicks) as a less mutagenic alternative to DSBs. Thousands of SSBs naturally occur per day in human cells, generally without deleterious consequences (Holmquist, 1998). The concept of harnessing the benign nature of nicks for stimulation of homologous recombination has been previously suggested in the context of theoretical models and RAG1-induced recombination (Meselson and Radding, 1975, Lee et al., 2004, Weinstock and Jasin, 2006, Holliday, 2007). In addition, homing endonucleases have been demonstrated to stimulate HDR when converted to nickases (McConnell Smith et al., 2009, van Nierop et al., 2009, Certo et al., 2011, Chan et al., 2011, Davis and Maizels, 2011, Metzger et al., 2011). However, conferring novel DNA binding specificities to this class of enzymes without disrupting catalytic activity has proven to be

challenging because the domains for DNA recognition and cleavage are not structurally independent (Arnould et al., 2011) as they are for ZFNs and TALENs.

Here we describe a general method for creating site-specific zinc finger nickases (ZFNickases). To do this, we employed obligate heterodimeric ZFNs (Miller et al., 2007) and introduced a mutation that had previously been described to inactivate FokI cleavage activity (D450A) (Waugh and Sauer, 1993, Bitinaite et al., 1998, Beumer et al., 2006) into one monomer, thereby directing a break to only one strand, as recently shown with the native FokI enzyme (Sanders et al., 2009). We demonstrate that ZFNickases can generate DNA single strand breaks efficiently *in vitro* and can also induce targeted HDR in cultured human cells with significantly lower rates of associated NHEJ-mediated mutation at the nicking site. ZFNickases provide an important additional tool for performing highly precise genome editing with reduced levels of NHEJ-mediated mutagenesis.

Methods

Qualitative in vitro analysis of ZFN and ZFNickase activities

For *in vitro* protein expression experiments, we constructed vectors derived from pMLM290 and pMLM292 (Addgene plasmids 21872 and 21873, respectively (Maeder et al., 2008)). Zinc finger domains in the ZFNs HX735, VF2468, and VF2471 have been described previously (Maeder et al., 2008) and were cloned as XbaI/BamHI fragments into pMLM290 and pMLM292 vectors modified to contain 3xFLAG instead of 1xFLAG epitopes. Previously reported zinc finger domains designed to bind the human CCR5 gene (Lombardo et al., 2007, Perez et al., 2008) were assembled from overlapping oligonucleotides generated by DNAWorks (Hoover and Lubkowski, 2002) and cloned into the 3xFLAG-pMLM290 and 3xFLAG-

pMLM292 vectors. The nuclease inactivating D450A mutation (numbered relative to the native FokI enzyme) (Bitinaite et al., 1998) was introduced by QuikChange Lightning Site-Directed Mutagenesis (Agilent). Protein lysates were prepared following manufacturer's instructions for the T7 TnT Quick Coupled Transcription/Translation System (Promega) using 1 μg plasmid template per 50 μl lysate and incubating for 90 minutes at 30°C.

To generate target sites for *in vitro* analysis, annealed oligonucleotides with compatible overhangs were cloned into the BsaI restriction sites of pBAC-lacZ (using oligonucleotides OC152/OC153 for CCR5; Table 3.1) as described previously (Maeder et al., 2009) or the BglII/SpeI sites of pCP5 (a gift from Daniel Voytas; using oligonucleotides OC665/OC666 for VF2468, OC667/OC668 for VF2471, and OC671/OC672 for HX735; Table 3.1) as described previously (Townsend et al., 2009). Primers labeled with 6–carboxyfluorescein (6-FAM) were used to amplify DNA fragments for cleavage assays with the Expand High Fidelity PCR System (Roche Applied Science). Primers OC213/OC215 were used for pBAC-lacZ-derived targets and OC417/OC418 used for pCP5-derived targets (see Table 3.1 for sequences of primers).

Table 3.1: Sequences of oligonucleotides used in this study

Name	5' to 3' Sequence
OC152	ATCTGTCATCCTCATCCTGATAAACTGCAAAAG
OC153	TAATCTTTTGCAGTTTATCAGGATGAGGATGAC
OC213	GACGCCCGCCATAAACTG
OC215	GCACGTTCCTTATATGTAGCTTTCG
OC417	CGGTGTGAAATACCGCACAG
OC418	CAGCAGCAGCAGACCATTTTC
OC665	GATCGAGCAGCGTCTTCGAGAGTGAGGAC
OC666	CTAGGTCCTCACTCTCGAAGACGCTGCTC
OC667	GATCCAGCGTCTTCGAGAGTGAGGACGTGT
OC668	CTAGACACGTCCTCACTCTCGAAGACGCTG
OC671	GATCGAGACTCCCACGGCCGGGGAAGAGT
OC672	CTAGACTCTTCCCCGGCCGTGGGAGTCTC
ams1228	GGTCGAGCAGCGTCTTCGAGAGTGAGGACGTGTA
ams1229	CTAGTACACGTCCTCACTCTCGAAGACGCTGCTCGACCTGCA
ams1230	GGGGTCATCCTCATCCTGATAAACTGCAAAAGGA
ams1231	CTAGTCCTTTTGCAGTTTATCAGGATGAGGATGACCCCTGCA

Cleavage reactions were performed under light-protected conditions using opaque black tubes in 100 µl volumes with 10 µl protein lysate and 80 ng 6-FAM-labeled cleavage substrate in 1x NEBuffer 4 (New England Biolabs). Reactions were incubated at 37°C for 1 hour, purified using a Minelute PCR Purification kit (Qiagen) according to the manufacturer's instructions with final elution into 20 µl 0.1X buffer EB, and submitted to the DNA Core Facility at Massachusetts General Hospital (Cambridge, MA) for denaturing capillary electrophoresis with fluorescent detection. Analysis of the resulting data was performed using Peak Scanner Software v1.0 (Applied Biosystems).

Chromosomal EGFP repair assay in U2OS cells

For the generation of the reporter cell lines used in the chromosomal EGFP repair assay, the cleavage site for the HX735 ZFNs (Maeder et al., 2008) was cloned between the *lacZ* ORF and the 5'-truncated *EGFP* (∂GFP) gene in plasmid pLV.LacZ∂GFP (Alwin et al., 2005). The donor plasmid used for these experiments harbors a 5'-truncated lacZ (∂lacZ) gene followed by the corrected *EGFP* ORF and also lacks a promoter (pUC. ∂LacZ-GFP) (Cornu et al., 2008).

To generate ZFNs and ZFNickases for mammalian expression, zinc finger domains were cloned into a previously described dual expression plasmid in which the CMV promoter drives expression of two ZFN monomers separated by a self-cleaving T2A peptide (Sollu et al., 2010). The D450A mutation was introduced into one or both of the FokI subunits in this vector as needed by subcloning or using the QuikChange Lightning Site-Directed Mutagenesis (Agilent).

For the chromosomal EGFP repair assay, U2OS-based reporter cell lines containing the LacZ-HX735-∂GFP target locus were generated by lentiviral transduction (LV.CMV.LacZ-HX735-∂GFP) with a viral dose that rendered <1% of cells resistant to geneticin-sulfate (0.4 mg/ml), thus preferentially generating reporter cells with a single copy target locus (Alwin et al.,

2005). Reporter cells, cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin (Invitrogen), were seeded at a density of 50,000 cells/well in 24-well plates. Twenty-four hours later, transfection was performed using X-tremeGENE HP DNA Transfection Reagent (Roche Applied Science) following the manufacturer's instructions. Transfection cocktail included 150 ng of ZFN or ZFNickase expression plasmids, 600 ng of donor plasmid (pUC.∂LacZ-GFP) and 100 ng of a plasmid encoding for mCherry (Mussolino et al., 2011) to identify transfected cells. The extent of gene targeting was assessed after 8 days using flow cytometry (FACSCalibur; BD Biosciences) to determine the percentage of EGFP-positive cells in the fraction of mCherry-positive cells. Experiments were performed at least twice independently in duplicate.

Traffic Light Reporter assay in HEK293 cells

Traffic Light Reporter (TLR) experiments were performed as described in Certo *et al.* (Certo et al., 2011). Briefly, oligonucleotides each harboring one or two ZFN or ZFNickase target sites (oligonucleotides ams1228/ams1229 for the VF2468 and VF2471 sites and oligonucleotides ams1230/ams1231 for the CCR5 site) were cloned into the *Sbf*I and *Spe*I restriction sites of the TLR2.1 plasmid. Note that the VF2468 and VF2471 targets overlap significantly so only one TLR reporter encompassing both sites was created. Cell lines were generated by transduction of HEK293T cells with limiting amounts of a lentivirus containing a target site of interest cloned into the Traffic Light Reporter, followed by selection in 1 μg/ml puromycin. The puromycin-resistant population was then bulk sorted by FACS to isolate a polyclonal population of EGFP-negative, mCherry-negative cells. These cells were cultured in glutamine-free Dulbecco's modified Eagle's medium supplemented with 2 mM L-glutamine, 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin (Invitrogen). For transfections, 1

x 10^5 reporter cells were plated per 24-well plate and transfected 24 hours later with 0.5 μg of ZFN- or ZFNickase-encoding plasmids and 0.5 μg Donor-T2A-BFP plasmid (Addgene plasmid 31485) using Fugene6 reagent according to the manufacturer's protocol (Roche Applied Science). Cells were split into a six-well plate 24 hours post-transfection, and analyzed using a flow cytometer (LSRII or FACSAria; BD Biosciences) 72 hours post-transfection. Transfection efficiency was controlled for by gating on 10^3 to 10^4 BFP-positive cells prior to HDR and NHEJ analysis. Experiments were performed independently three times.

Results

In vitro *enzymatic activities of ZFNs and ZFNickases*

Based on recent work demonstrating that mutational inactivation of one monomer in a FokI dimer can convert this nuclease into a nickase (Sanders et al., 2009), we reasoned that a similar strategy might be used to convert a ZFN dimer into a nickase. To test this possibility, we used four previously described ZFN pairs targeted to sites in three endogenous human genes; one to the *HOXB13* gene (HX735), two to the *VEGF-A* gene (VF2468 and VF2471), and one to the *CCR5* gene (CCR5) (Lombardo et al., 2007, Maeder et al., 2008, Perez et al., 2008). For each ZFN pair, we arbitrarily designated one of the monomers as the "Left monomer" and the other as the "Right monomer" (Table 3.2) and generated variants of each monomer harboring a previously described mutation (D450A) that inactivates the catalytic activity of the FokI nuclease domain (Bitinaite et al., 1998).

To test whether inactivation of one monomer in a ZFN pair might result in generation of a zinc finger nickase (ZFNickase), we developed a qualitative version of an *in vitro* assay similar to one recently described by Halford and colleagues (Sanders et al., 2009) that allowed us to

Table 3.2: ZFN/ZFNickase target sites

Target site	Full site	LEFT half-site (5'to3')	RIGHT half-site (5'to3')
CCR5	gGTCATCCTCATCctgatAAACTGCAAAAGg cCAGTAGGAGTAGgactaTTTGACGTTTTCc	gGATGAGGATGACc	tAAACTGCAAAAGg
HX735	gAGACTCCCAcggccGGGGAAGAGt cTCTGAGGGTgccggCCCCTTCTCa	gTGGGAGTCTc	cGGGGAAGAGt
VF2468	gAGCAGCGTCttcgaGAGTGAGGAc cTCGTCGCAGaagctCTCACTCCTg	aGACGCTGCTc	aGAGTGAGGAc
VF2471	cAGCGTCTTCgagagtGAGGACGTGt gTCGCAGAAGctctcaCTCCTGCACa	cGAAGACGCTg	tGAGGACGTGt

assess the introduction of breaks into either strand of a double-stranded DNA fragment. In this system, a target site for a ZFN is positioned asymmetrically within a DNA fragment that is fluorescently labeled on the 5' ends of both DNA strands (Figure 3.1). This DNA is then incubated with different combinations of active and inactive Left/Right monomers that have been co-expressed *in vitro* using a coupled transcription/translation system (Cradick et al., 2010). Following this incubation, fluorescently labeled DNA strands of various sizes are generated depending upon whether the top or bottom strands are cut or not cut (Figure 3.1). These fluorescently labeled products can be analyzed under denaturing conditions using capillary electrophoresis, which separates DNA molecules based on size and enables visualization of 6-FAM-labeled DNA strands.

Using this qualitative *in vitro* assay, we tested the effects of various combinations of active/inactive monomers for the HX735, VF2468, VF2471, and CCR5 ZFNs on their target DNA sites. As expected, with all four active Left/active Right pairs, both strands in the target sites were cleaved and with all inactive Left/inactive Right pairs, no cleavage of either strand was observed (Figure 3.2a-d, top and bottom panels). However, when pairs of active/inactive monomers were tested, we observed preferential cleavage of only one DNA strand, the strand with which the active ZFN monomer is expected to make its primary DNA base contacts (Figure

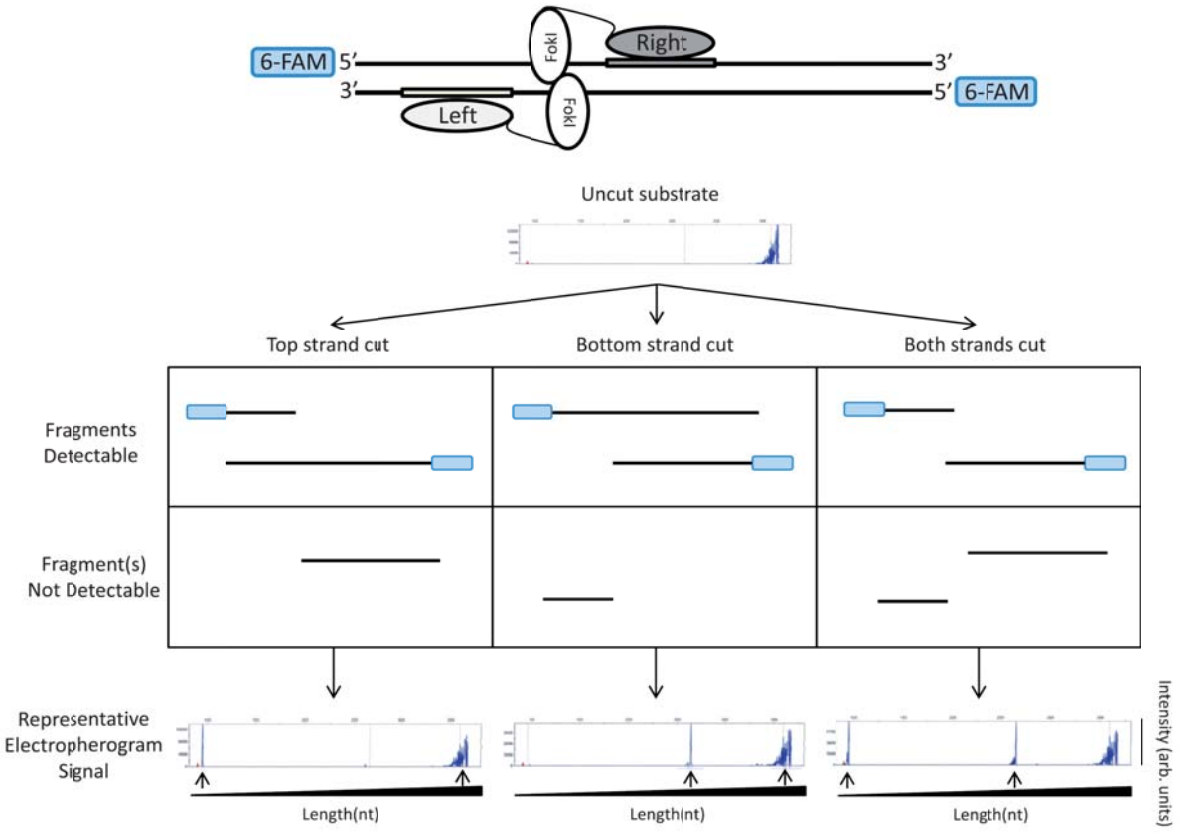


Figure 3.1: A qualitative *in vitro* assay to detect cleavage and nicking by ZFNs and ZFNickases

An asymmetrically positioned full ZFN target site is placed within a DNA fragment that has been labeled on both its 5' ends with 6-FAM fluorescent dye (depicted in blue). Only 6-FAM labeled strands will be detected in the denaturing capillary electrophoresis assay. In the example shown, the ZFN target site is positioned toward the left end of the DNA fragment. In this configuration, if nicking of the top strand occurs, this results in generation of one short and one full-length 6-FAM labeled product. If nicking of the bottom strand occurs, this results in generation of one medium-length and one full-length 6-FAM labeled product. Cleavage of both strands results in the generation of short and medium-length products. Sample electropherograms are shown with arbitrary intensity units on the y-axis and DNA strand length on the x-axis. DNA strands expected from nicking or cleavage reactions are designated by black arrows. Note that full-length DNA strands due to incomplete enzyme reactions may be present in addition to the expected products.

3.2a-d, middle panels). Analysis of the electropherograms in Figure 3.2 reveals that the cleavage positions for all of our ZFNickases are either identical or within 1 nt of the cleavage positions of their matched parental ZFNs (data not shown). These results demonstrate that introduction of an

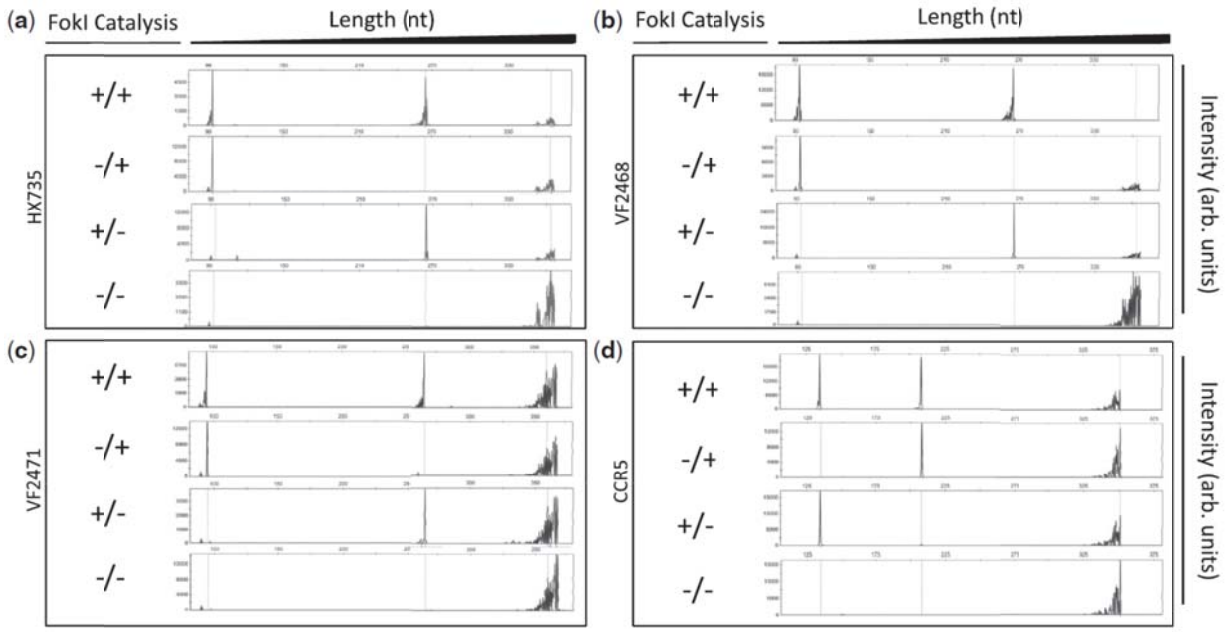


Figure 3.2: Site-specific nicking of DNA *in vitro* by ZFNickases
Substrates labeled with 6-FAM fluorescent dye harboring (a) HX735, (b) VF2468, (c) VF2471, or (d) CCR5 binding sites were incubated with active Left/active Right (+/+), inactive Left/active Right (-/+), active Left/inactive Right (+/-), and inactive Left/inactive Right (-/-) ZFN monomers. Cleavage products were subjected to denaturing capillary electrophoresis. Axes are arbitrary intensity units (y-axis) and DNA strand length (x-axis). The y-axis is differentially scaled for each plot, while the x-axis is scaled uniformly for all plots. Representative electropherograms are shown, but all experiments were performed in triplicate (data not shown). Note that the HX735, VF2468, and VF2471 targets were cloned into the pCP5 vector which results in asymmetric placement left of center within the substrate similar to the configuration depicted in Figure 1. However, the CCR5 target is cloned into pBAC-lacZ, which results in binding site placement right of center relative to the substrate; when the top strand is cleaved in this configuration, the fragment generated is longer than when the bottom strand is cleaved.

inactivating FokI mutation into one monomer in a ZFN obligate heterodimer pair provides a general method for converting ZFNs into ZFNickases and that nicking activity can be preferentially directed to one particular strand of the DNA.

Testing ZFNickase activities using a human cell-based chromosomal EGFP reporter assay

Having established the activities of ZFNickases *in vitro*, we next wished to test whether these nickases, like their nuclease counterparts, could induce HDR at their target sites in human cells. To do this, we used a previously described, human cell-based reporter gene assay for

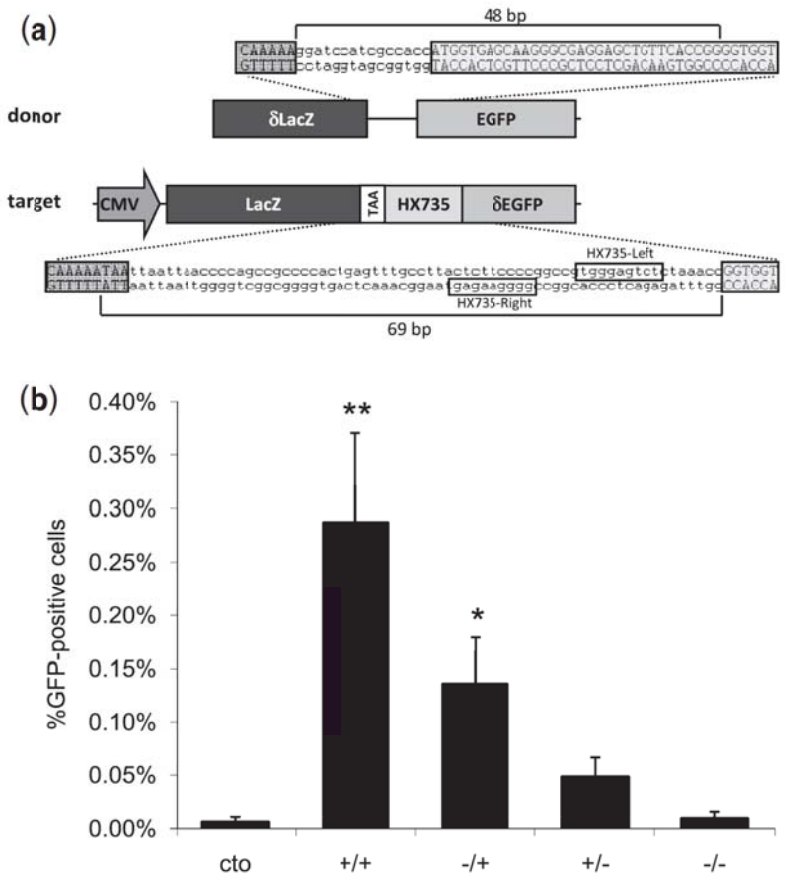


Figure 3.3: Assessment of ZFNickase-mediated HDR using a human cell-based chromosomal EGFP reporter assay

(a) A schematic of the U2OS.LacZ-HX735-∂GFP reporter construct integrated in a U2OS cell line. Note that the orientation of the binding site in the reporter is inverted relative to the configuration at the HX735 endogenous locus, for which the Left and Right designations were originally (but arbitrarily) assigned.

(b) ZFN and ZFNickase-mediated HDR in a U2OS EGFP reporter line. Cells were co-transfected with the donor plasmid and plasmids encoding HX735 ZFN pairs composed of active Left/active Right (+/+), inactive Left/active Right (-/+), active Left/inactive Right (+/-), and inactive Left/inactive Right (-/-) FokI domains. The graph shows the percentage of EGFP-positive cells 8 days following transfection. Statistically significant differences in HDR-based gene correction relative to donor-only control (cto) are indicated by * (P<0.05) or ** (P<0.01).

monitoring HDR (Alwin et al., 2005). In this assay, a 5' truncated *EGFP* reporter gene bearing a ZFN target site of interest is chromosomally integrated using a viral-based vector (Figure 3.3a). HDR with an appropriate donor construct leads to restoration of an intact *EGFP* reporter gene. Thus, the percentages of EGFP-positive cells arising after the co-delivery of ZFNs or

ZFNickases and the donor plasmid reflect the abilities of ZFNs to stimulate HDR at the target locus.

We tested ZFNs and corresponding ZFNickases targeted to the HX735 site for their abilities to induce HDR in the EGFP reporter assay. ZFNs targeted to the HX735 locus were able to stimulate gene repair, inducing EGFP expression in 0.29% of transfected cells, a statistically significant increase relative to the 0.01% of cells that expressed EGFP upon transfection with a catalytically inactive ZFN (Figure 3.3b). Interestingly, expression of corresponding HX735 ZFNickases designed to nick one strand or the other restored EGFP expression in 0.14% and 0.05% of cells, although only the former increase in HDR was statistically significant relative to the level observed with the catalytically inactive ZFN. Thus, these results suggest that a ZFNickase can induce HDR in this EGFP reporter gene assay albeit at a lower level than that observed with its parental ZFN.

Assessment of HDR and NHEJ induced by ZFNs and ZFNickases in human cells

To further test the ability of ZFNickases to induce HDR and to simultaneously assess the rate of NHEJ-mediated mutagenesis at the same target site, we used the recently described "Traffic Light Reporter" (Certo et al., 2011). In this assay, the reporter harbors a nuclease target site (or sites) of interest positioned within a defective EGFP coding sequence that is followed by a mCherry coding sequence joined out of frame to the *EGFP* gene via a T2A peptide sequence. HDR with an exogenously provided donor template reconstitutes a functional EGFP coding sequence, turning cells green, whereas NHEJ-induced indels can create frameshifts that place the downstream mCherry protein in-frame, turning cells red (Figure 3.4a). Thus, with these reporter cells, the extent of nuclease-induced HDR and NHEJ can be monitored simultaneously for a given target site using flow cytometry.

We derived polyclonal HEK293T cell lines harboring the Traffic Light Reporter with targets for either the VF2468 and VF2471 ZFNs (a single cell line with both overlapping targets present, just as they occur in the endogenous locus (Maeder et al., 2008)) or the CCR5 ZFNs. For each ZFN pair, we transfected combinations of plasmids encoding active and/or inactive ZFN monomers together with a donor template for correcting the *EGFP* gene. Flow cytometry was then used to determine HDR and NHEJ rates by quantifying the percentages of EGFP-positive and mCherry-positive cells, respectively (Figure 3.4b-e). For all three target sites, ZFNs tested showed robust activities, inducing high percentages of EGFP-positive cells (indicative of HDR events) and even higher percentages of mCherry-positive cells (indicative of NHEJ events) in transfected cells (Figure 3.4b, left column). All but one of the six ZFNickases tested for the three target sites induced significantly higher levels of EGFP-positive cells compared with negative controls (Figure 3.4b-e). The percentage of EGFP-positive cells for the VF2468 and CCR5 ZFNickases are three- to ten-fold lower than what was observed with their corresponding ZFNs (Figure 3.4c and e, compare second and third white-colored columns with the first white-colored column), suggesting that HDR is induced by ZFNickases but again at a lower rate than is observed with the parental ZFNs. The activities of the VF2471 ZFNickases were detectable but quite low despite the high activity of the VF2471 ZFNs (Figure 3.4d). However, for all three target sites, the ZFNickases consistently induce lower percentages of mCherry-positive cells relative to their matched ZFNs, suggesting that fewer mutagenic NHEJ-mediated events are occurring with the nickases compared with the nucleases (Figure 3.4c-e). In addition, the ratio of the percentage of EGFP-positive cells to the percentage of mCherry-positive cells is higher for 5 of the 6 ZFNickases compared with the parental ZFN (Figure 3.4f-h). We also found, in accordance with previous studies conducted with homing endonucleases and nickases, that

Figure 3.4: Assessment of ZFNickase-mediated HDR and NHEJ using a human cell-based Traffic Light Reporter assay

(a) Schematic of the "Traffic Light Reporter." HDR-mediated correction of the EGFP gene with a co-transfected donor template results in EGFP-positive cells. Mutagenic NHEJ events at the nuclease target site results in mCherry-positive cells.

(b) Representative flow cytometry plots showing percentages of EGFP-positive and mCherry-positive cells following transfection of Traffic Light Reporter cell lines with plasmid encoding the indicated ZFNs and ZFNickases and the donor template. In the experiments shown, cells have been gated for BFP expression (encoded by the plasmid harboring the donor template) to normalize for transfection efficiencies.

(c-e) Bar graphs showing mean percentages of EGFP-positive and mCherry-positive cells for experiments performed with the VF2468, VF2471, and CCR5 ZFNs and ZFNickases. Results were derived from 3 independent experiments with SEM shown. Statistically significant differences in HDR and mutagenic NHEJ rates relative to donor-only control (-/-) are indicated by * ($P<0.05$) or ** ($P<0.01$). (f-h) Ratios of percentage of EGFP-positive cells to percentage of mCherry-positive cells for the VF2468, VF2471, and CCR5 ZFNs and ZFNickases using the data from (c-e). Data used to create Figure 3.4c-h is available in Table 3.3.

Figure 3.4 (continued):

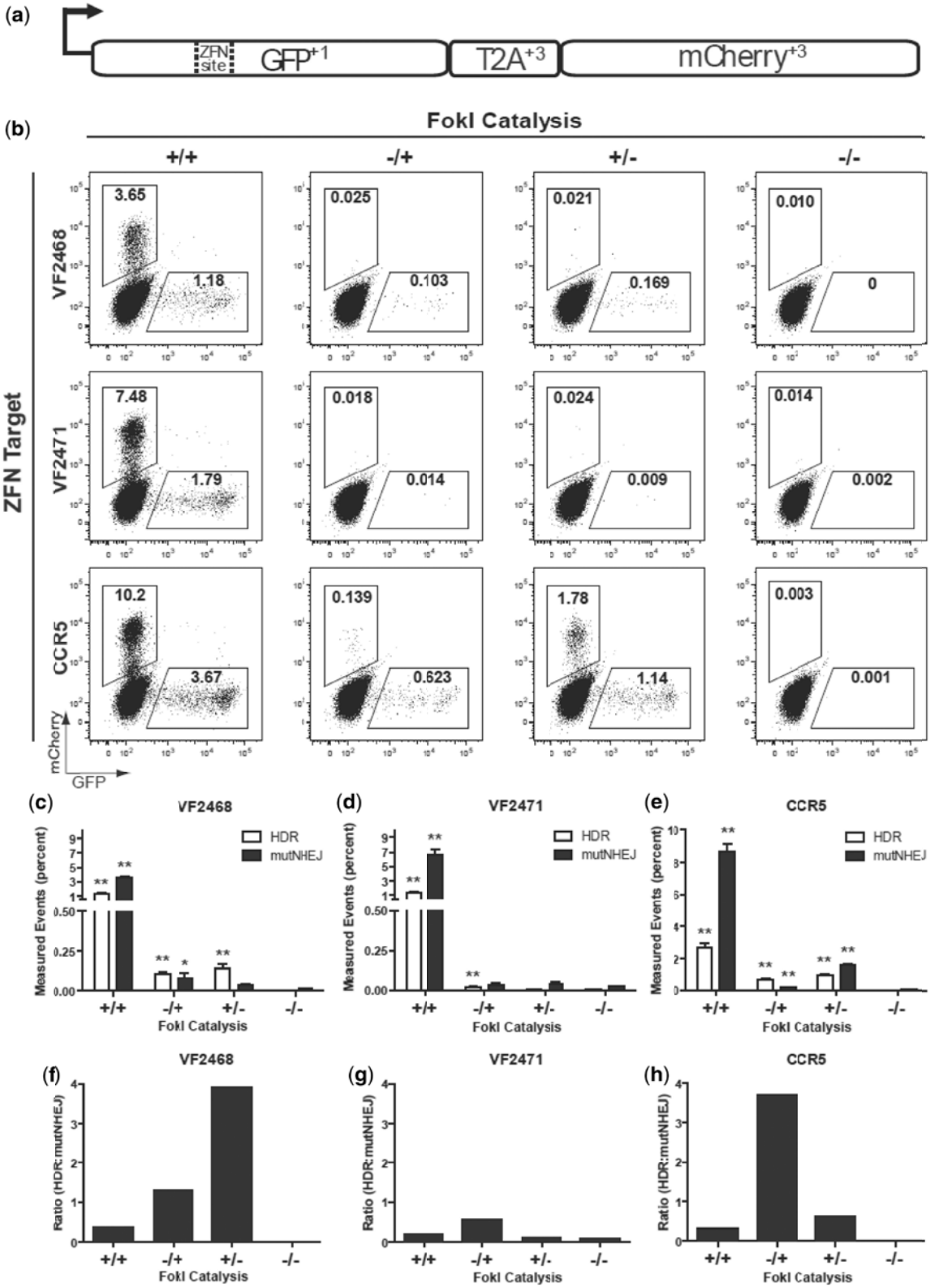


Table 3.3: Numerical data used to create Figure 4c-h.

VF2468			
FokI Catalysis	Avg HDR (%)	Avg mutNHEJ (%)	Ratio (HDR:mutNHEJ)
+/+	1.279	3.620	0.353
-/+	0.112	0.087	1.289
+/-	0.142	0.036	3.913
-/-	0.000	0.017	0.000

VF2471			
FokI Catalysis	Avg HDR (%)	Avg mutNHEJ (%)	Ratio (HDR:mutNHEJ)
+/+	1.280	6.684	0.192
-/+	0.019	0.035	0.544
+/-	0.005	0.040	0.117
-/-	0.002	0.024	0.080

CCR5			
FokI Catalysis	Avg HDR (%)	Avg mutNHEJ (%)	Ratio (HDR:mutNHEJ)
+/+	2.696	8.660	0.311
-/+	0.641	0.174	3.694
+/-	0.963	1.572	0.612
-/-	0.000	0.019	0.015

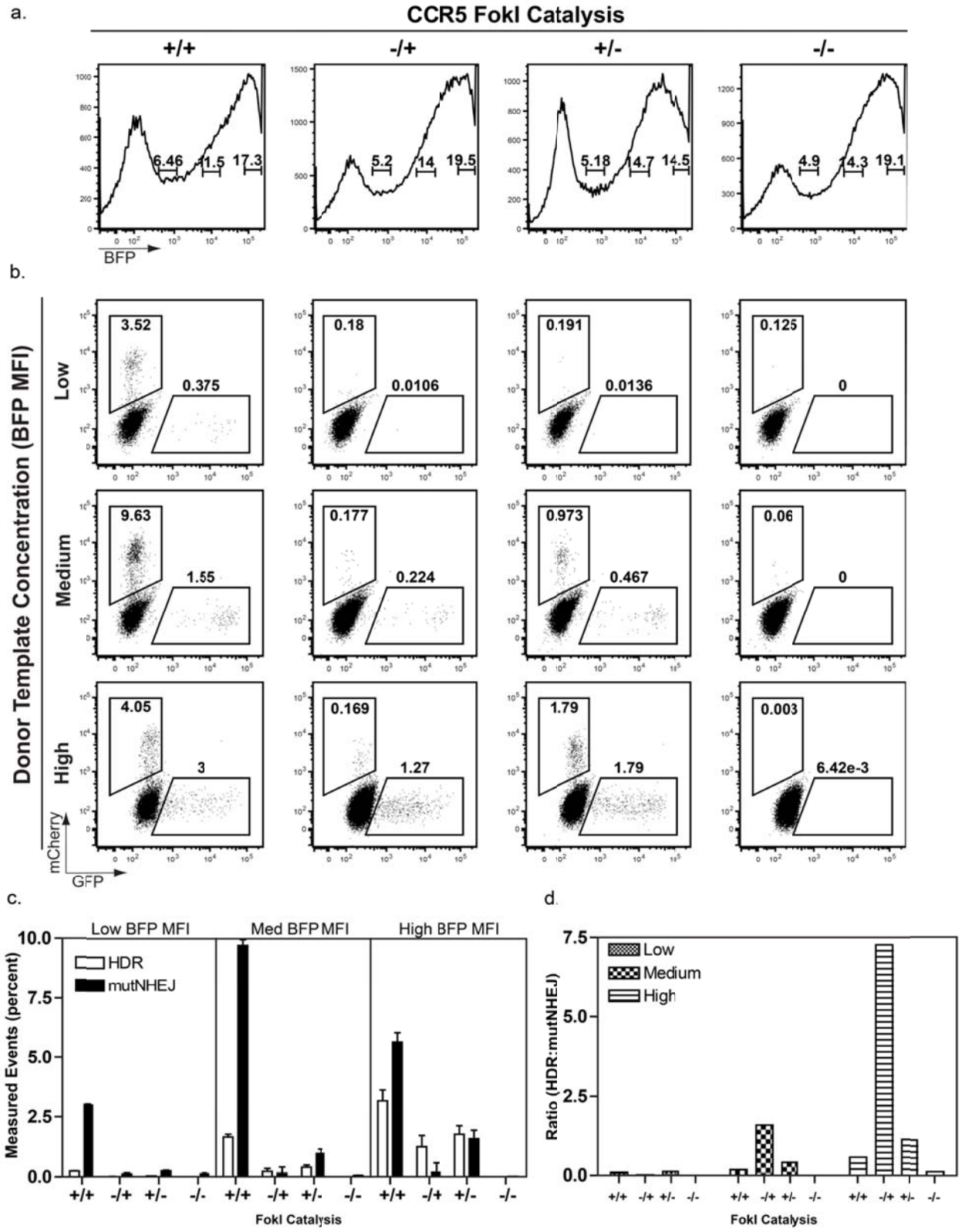
increased donor template concentrations were associated with increased nuclease- and nickase-induced HDR frequencies (Figure 3.5 and data not shown) (Certo et al., 2011, Davis and Maizels, 2011, Metzger et al., 2011). Interestingly, the effect of donor template concentration appears to be more dramatic on nickase-induced HDR than on nuclease-induced HDR (Figure 3.5 and data not shown). These results suggest that ZFNickases induce higher levels of HDR events relative to NHEJ events compared with ZFNs. Our results indicate that ZFNickases may offer the benefit of significantly reduced NHEJ rates albeit with a reduction in HDR activity.

Discussion

In this report, we describe a general and simple method for converting ZFNs into ZFNickases. Introduction of a previously described D450A mutation into one monomer of a

Figure 3.5: Effect of donor template concentration on ZFNickases (illustrated with the CCR5 site) as assessed by gating on different mean fluorescence intensities (MFI) of the BFP-linked donor template.
(a) Cell populations were gated to analyze cells with low, medium, and high BFP MFIs for each combination of active and/or inactive CCR5 monomers. A set of representative plots is shown.
(b) As indicators of mutagenic NHEJ and HDR rates, respectively, the percent of mCherry positive (y-axis) and EGFP-positive (x-axis) cells are noted for each cell population analyzed in (a).
(c) Bar graphs show mean percentages of EGFP-positive and mCherry-positive cells, with results derived from 3 independent experiments. Error bars indicate SEM.
(d) Ratios of percentage of EGFP-positive cells to percentage of mCherry-positive cells for the CCR5 ZFNs and ZFNickases using the data from (c).

Figure 3.5 (continued):



ZFN pair can generate a ZFNickase. This result parallels recent work from Halford and colleagues in which they used a similar approach to convert the wild-type FokI restriction enzyme into a nickase. Our qualitative *in vitro* data demonstrate that each ZFNickase preferentially cuts one DNA strand at a position either identical or within 1 nt of the cut positions of its matched ZFN. Furthermore, the data show that each ZFN monomer cuts the DNA strand to which it makes most of its DNA base contacts, providing direct experimental support for the model of binding and cleavage illustrated in Figure 3.1.

Testing in two different human cell-based reporter systems revealed that ZFNickases can induce HDR-mediated repair, albeit at lower levels than matched ZFNs from which they were derived. Of the eight ZFNickases we tested (two pairs each derived from ZFNs targeted to four different target sites; data presented in Figures 3.3 and 3.4), six induced statistically significant levels of HDR. The levels of HDR we observed with the ZFNickases ranged widely, from between two-fold to over 100-fold lower than those observed with the corresponding ZFNs from which they were derived. However, for at least some of the ZFNickases we tested (e.g. HX735 -/+, VF2468 -/+ and +/-, and CCR5 -/+ and +/-), the levels of HDR induced were of sufficiently high frequency (0.1% or higher) to be useful for research applications and some potential therapeutic strategies. Our observations that ZFNickases can induce HDR events and that HDR efficiency is positively correlated with the concentration of donor present in cells are consistent with the findings of others using homing endonucleases engineered to induce nicks (McConnell Smith et al., 2009, van Nierop et al., 2009, Certo et al., 2011, Chan et al., 2011, Davis and Maizels, 2011, Metzger et al., 2011). However, to our knowledge, our findings are the first to report that nickases derived from ZFNs can be used to induce HDR events.

Although absolute rates of HDR were lower for ZFNickases than ZFNs in our human cell-based reporter assays, we also observed a consistent reduction in mutagenic NHEJ rates in the Traffic Light Reporter assay. This reduction is not entirely surprising given that nicks are typically repaired without causing mutations (Holmquist, 1998). However, we do not know the origin of the residual NHEJ-mediated events we observed with some of the ZFNickases we tested. Possible explanations include conversion of a nick into a DSB that may occur with replication fork collapse (see below) or weak residual homodimerization of the active ZFNickase monomer that may lead to cleavage at the intended target site. Use of improved second-generation FokI heterodimer variants (Doyon et al., 2011) may reduce activity due to the latter mechanism (we used first-generation FokI heterodimer variants (Miller et al., 2007) for this study).

Importantly, for five of the six ZFNickases we tested in the Traffic Light Reporter assay, the ratio of HDR to NHEJ events was increased compared with the three matched ZFNs from which they were derived. These results demonstrate that ZFNickases can induce HDR events with relatively lower rates of NHEJ-mediated mutations created at the nick site. We do not currently know the mechanism of the ZFNickase-mediated HDR or the improved HDR:NHEJ ratios we observe. One possibility for the improved HDR:NHEJ ratios is that a nick in the path of a DNA replication fork may be converted to a DSB leading to fork collapse, repair of which would be expected to lead to repair by either NHEJ or HDR. A potential hypothesis for why we observe a preferential shift from NHEJ to HDR with ZFNickases may be the more frequent repair of nick-induced replication fork collapse by HDR (Saleh-Gohari et al., 2005), in part due to the availability of repair factors for homologous recombination during DNA replication in S-phase (Hartlerode et al., 2011). Interestingly, for every target site we tested in our human cell-

based assays, one ZFNickase combination consistently outperformed the other with respect to absolute HDR rates and, for those assayed using the Traffic Light Reporter assay, improved HDR:NHEJ ratios. This reproducible difference does not appear to be correlated with whether the nicked strand is transcribed or not, and there were no strand cleavage preferences discernible from the *in vitro* data. It is possible that strand-dependent differences in HDR activity arise due to different DNA-binding affinities of zinc finger domains in each monomer and how this may affect asymmetric accessibility to the break by cellular repair machinery. Regardless of the precise mechanism, our results suggest that testing both potential ZFNickases for a given target site is worthwhile to identify the most active nickase possible.

Our work demonstrates that ZFNickases with predictable strand nicking activities can be easily derived from ZFNs and that these enzymes can be used in cells to induce HDR with improved HDR:NHEJ ratios. It will be of interest in future experiments to test whether ZFNickase-induced HDR rates can be further increased by using improved FokI heterodimer frameworks and hyperactive FokI variants (Guo et al., 2010, Doyon et al., 2011). Our observation of reduced mutagenic NHEJ events at the target nicking site suggest that ZFNickases will also likely induce fewer mutations at potential off-target sites elsewhere in the genome, a prediction that can easily be tested for ZFNs with known off-target sites (Gabriel et al., 2011, Pattanayak et al., 2011). In addition, site-specific nickases may generally be of interest for the study of biological phenomena such as replication fork dynamics. Our results suggest ZFNickases may provide a means to induce HDR with reduced mutagenesis caused by NHEJ and that additional optimization of this platform should be an important goal for future investigation.

Acknowledgements

We thank Fabienne Lütge (Hannover Medical School) for technical assistance, Daniel Collette (Massachusetts General Hospital DNA Core) for help with capillary electrophoresis data analysis, Steve Halford and Eva Vanamee for helpful suggestions in the early stages of this project, and Ralph Scully and Lee Zou for helpful discussions regarding DNA repair. This work was supported by the National Institutes of Health (NIH) [R01 GM088040 (J.K.J.), an NIH Director's Pioneer Award DP1 OD006862 (J.K.J.), RL1 CA133832 (A.M.S.), UL1 DE019582 (A.M.S.), R01 AI068885 (A.P.M)]; the Jim and Ann Orr Massachusetts General Hospital Research Scholar award (J.K.J.); and the European Commission [PERSIST–222878 (T.C.)]. C.L.R. was supported by a National Science Foundation Graduate Research Fellowship and a Ford Foundation Predoctoral Fellowship. M.T.C. was supported in part by NIH T32 GM07270. Funding for open access charge: NIH Director's Pioneer Award DP1 OD006862 (J.K.J).

Chapter Four

Revealing off-target cleavage specificities of zinc-finger nucleases by *in vitro* selection

Vikram Pattanayak, Cherie L. Ramirez, J. Keith Joung, and David R. Liu

Vikram Pattanayak developed the selection methodology described in this chapter, performed the selection experiments, and carried out analysis for the selections and human cell deep sequencing data. The thesis author designed the strategy for examining putative off-target sites determined from *in vitro* selection results, optimized vectors and conditions for CCR5-224 and VF2468 ZFN expression in K562 cells, and performed deep sequencing experiments in K562 cells to interrogate potential off-target loci.

Reference: "Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection." Pattanayak V, Ramirez CL, Joung JK, Liu DR. *Nature Methods*. 2011 Aug 7; 8(9):765-70. The article has been adapted from its published version for presentation in the dissertation.

Revealing off-target cleavage specificities of zinc-finger nucleases by *in vitro* selection

Vikram Pattanayak[1], Cherie L. Ramirez[2,3], J. Keith Joung[2,3] & David R. Liu[1]

[1] Department of Chemistry and Chemical Biology, and Howard Hughes Medical Institute, Harvard University, Cambridge, Massachusetts, USA.
[2] Molecular Pathology Unit, Center for Cancer Research, and Center for Computational and Integrative Biology, Massachusetts General Hospital, Charlestown, Massachusetts, USA.
[3] Department of Pathology and Biological and Biomedical Sciences Program, Harvard Medical School, Boston, Massachusetts, USA.

Abstract

Engineered zinc-finger nucleases (ZFNs) are promising tools for genome manipulation, and determining off-target cleavage sites of these enzymes is of great interest. We developed an *in vitro* selection method that interrogates 10^{11} DNA sequences for cleavage by active, dimeric ZFNs. The method revealed hundreds of thousands of DNA sequences, some present in the human genome, that can be cleaved *in vitro* by two ZFNs: CCR5-224 and VF2468, which target the endogenous human *CCR5* and *VEGFA* genes, respectively. Analysis of identified sites in one cultured human cell line revealed CCR5-224–induced changes at nine off-target loci, though this remains to be tested in other relevant cell types. Similarly, we observed 31 off-target sites cleaved by VF2468 in cultured human cells. Our findings establish an energy compensation model of ZFN specificity in which excess binding energy contributes to off-target ZFN cleavage and suggest strategies for the improvement of future ZFN design.

Introduction

Zinc-finger nucleases (ZFNs) are enzymes engineered to recognize and cleave desired target DNA sequences. A ZFN monomer consists of a zinc-finger DNA-binding domain fused with a non-specific *Fok*I restriction endonuclease cleavage domain (Kim et al., 1996). As the FokI nuclease domain must dimerize and bridge two DNA half-sites to cleave DNA (Vanamee et al., 2001), ZFNs are designed to recognize two unique sequences flanking a spacer sequence of variable length and to cleave only when bound as a dimer. ZFNs have been used for genome engineering in many organisms including mammals (Urnov et al., 2005, Maeder et al., 2008, Perez et al., 2008, Santiago et al., 2008, Hockemeyer et al., 2009, Zou et al., 2009, Cui et al., 2011) by stimulating either non-homologous end joining or homologous recombination. In addition to providing powerful research tools, ZFNs also have potential as gene therapy agents. Clinical trials have recently started with two ZFNs: one as part of an anti-HIV therapeutic approach (CCR5-224 clinical trials NCT00842634, NCT01044654 and NCT01252641) and the other to modify cells used as anticancer therapeutics (NCT01082926).

DNA cleavage specificity is a crucial feature of ZFNs. The imperfect specificity of some designed zinc-fingers domains has been linked to cellular toxicity (Cornu et al., 2008) and therefore determining the specificities of ZFNs is of significant interest. ELISA assays (Segal et al., 1999), microarrays (Bulyk et al., 2001), a bacterial one-hybrid system (Meng et al., 2007), SELEX (Segal et al., 2003, Liu and Stormo, 2005), and Rosetta-based computational predictions(Yanover and Bradley, 2011) have all been used to characterize the DNA-binding specificity of monomeric zinc-finger domains in isolation. As a result, information about the specificity of zinc-finger nucleases to date has been based on the unproven assumptions that (*i*) dimeric zinc-finger nucleases cleave DNA with the same sequence specificity with which

isolated monomeric zinc-finger domains bind DNA; and (*ii*) the binding of one zinc-finger domain does not influence the binding of the other zinc-finger domain in a given ZFN. The DNA-binding specificities of monomeric zinc-finger domains have been used to predict potential off-target cleavage sites of dimeric ZFNs in genomes (Perez et al., 2008, Gupta et al., 2011, Soldner et al., 2011), but to our knowledge no study to date has reported a method for determining the broad DNA cleavage specificity of active, dimeric ZFNs.

Here we present an *in vitro* selection method to broadly examine the DNA cleavage specificity of active ZFNs. We coupled our selection with high-throughput DNA sequencing to evaluate two obligate heterodimeric ZFNs, CCR5-224 (Perez et al., 2008, Holt et al., 2010) currently in clinical trials, and VF2468 (Maeder et al., 2008) which targets the human *VEGFA* promoter, for their abilities to cleave each of 10^{11} potential target sites. We identified 37 sites and 2,652 sites in the human genome that can be cleaved *in vitro* by the ZFNs CCR5-224 and VF2468, respectively, and hundreds of thousands of *in vitro*–cleavable sites for both ZFNs that are not present in the human genome. We examined 34 sites or 90 sites for evidence of ZFN-induced mutagenesis in cultured human K562 cells expressing the CCR5-224 or VF2468 respectively. Ten of the CCR5-224 sites and 32 of the VF2468 sites we tested had DNA sequence changes consistent with ZFN-mediated cleavage in human cells, although we anticipate that cleavage is likely to be dependent on cell type and ZFN concentration. One CCR5-224 off-target site lies in the promoter of the malignancy-associated *BTBD10* gene.

Our results, which could not have been obtained by determining binding specificities of monomeric zinc-finger domains alone, indicate that excess DNA-binding energy results in increased off-target ZFN cleavage activity and suggest that ZFN specificity can be enhanced by designing ZFNs with decreased binding affinity, by lowering ZFN expression and by choosing

target sites that differ by at least three base pairs from their closest sequence relatives in the genome.

Results

In Vitro *Selection for ZFN-Mediated DNA Cleavage*

Libraries of potential cleavage sites were prepared as double-stranded DNA using synthetic primers and PCR. Each partially randomized position in the primer was synthesized by incorporating a mixture containing 79% wild-type phosphoramidite and 21% of an equimolar mixture of all three other phosphoramidites. Library sequences therefore differed from canonical ZFN cleavage sites by 21% on average, distributed binomially. We used a blunt ligation strategy to create a 10^{12}-member minicircle library. Using rolling-circle amplification with phi29 DNA polymerase, $>10^{11}$ members of this library were both amplified and concatenated into high molecular weight (>12 kb) DNA molecules.

The CCR5-224 or VF2468 DNA cleavage site library was incubated at a total cleavage site concentration of 14 nM with two-fold dilutions ranging from 0.5 nM to 4 nM of crude *in vitro*-translated CCR5-224 or VF2468, respectively (Figure 4.1). Following digestion, the resulting DNA molecules (Figure 4.2) were subjected to *in vitro* selection for DNA cleavage and analyzed by paired-end high-throughput DNA sequencing. Briefly, three selection steps (Figure 4.3) enable the separation of sequences that are cleaved from those that are not. First, only sites that have been cleaved contain 5' phosphates, which are required for the ligation of adapters required for sequencing. Second, after PCR, a gel purification step enriches the smaller, cleaved library members. Finally, a computational step after sequencing counts only sequences that have

filled- in, complementary 5' overhangs on both ends, the hallmark of cleavage of a concatemerized target site (Table 4.1).

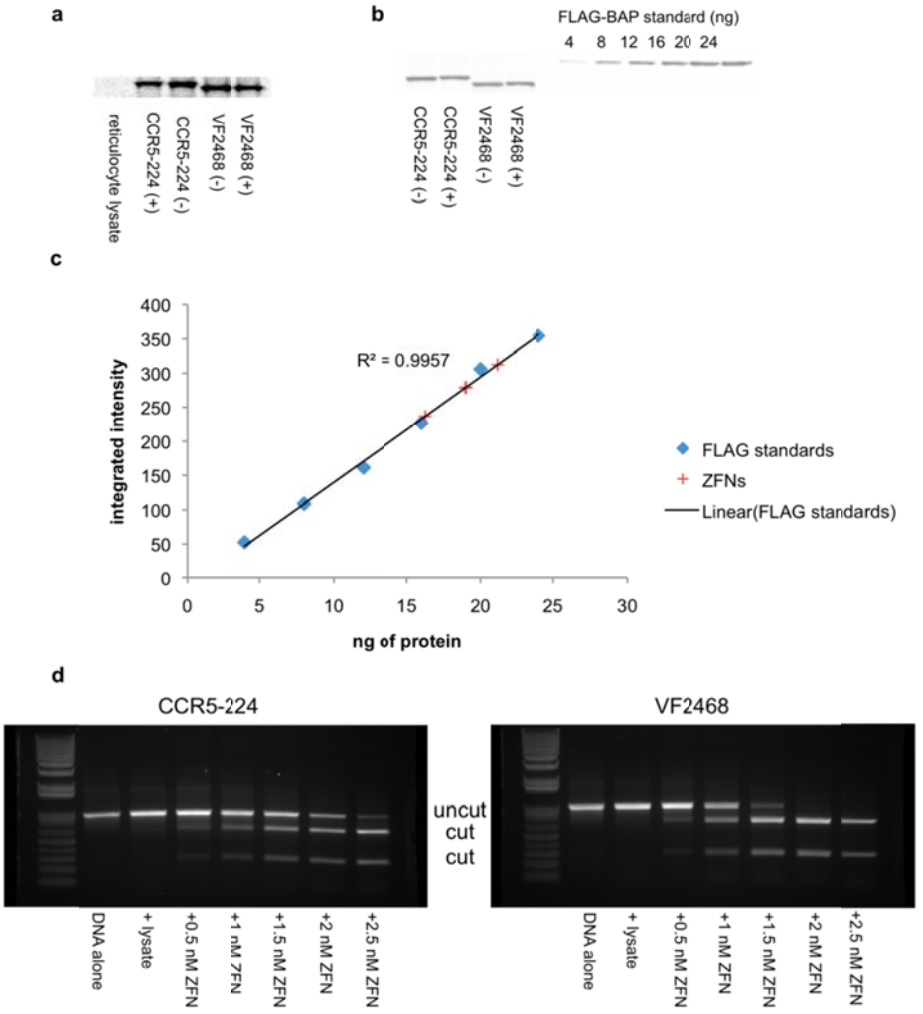


Figure 4.1: Expression and quantification of ZFNs
Western blots for CCR5-224 and VF2468 are shown (a) for the ZFN samples used in the *in vitro* selection, and (b) for quantification. (c) Known quantities of N-terminal FLAG-tagged bacterial alkaline phosphatase (FLAG-BAP) were used to generate a standard curve for ZFN quantification. Blue diamonds represent the intensities of FLAG-BAP standards from the Western blot shown in (b), red plus signs represent the intensities of bands of ZFNs, and the black line shows the best-fit curve of FLAG-BAP standards that was used to quantify ZFNs. (d) Gels are shown of activity assays of CCR5-224 and VF2468 on an 8 nM linear substrate containing one target cleavage site. The ZFNs were each incubated with their respective substrate for 4 hours at 37°C. DNA in the "+ lysate" lane was incubated with an amount of *in vitro* transcription/translation mixture equivalent to that used in the 2.5 nM ZFN reaction. ZFN-mediated cleavage results in two linear fragments approximately 700 bp and 300 bp in length. 2 nM CCR5-224 and 1 nM VF2468 were the amounts required for 50% cleavage of the linear substrate.

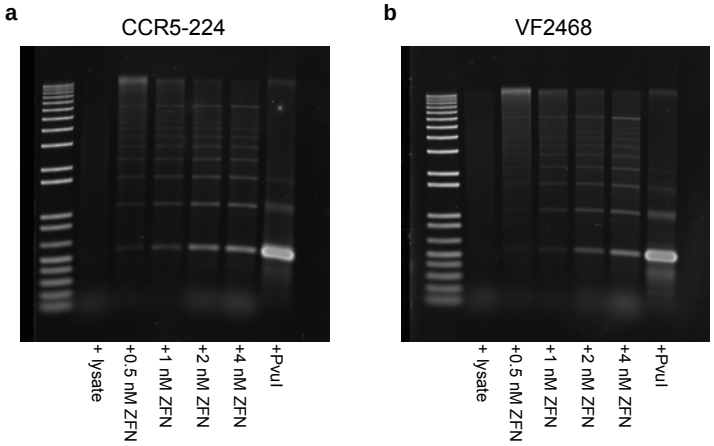


Figure 4.2: Library cleavage with ZFNs
Cleavage of 1 μg of concatemeric libraries of CCR5-224 (a) or VF2468 (b) target sites are shown with varying amounts CCR5-224 or VF2468, respectively. The lane labeled "+ lysate" refers to pre-selection concatemeric library incubated with the volume of *in vitro* transcription/translation mixture contained in the samples containing 4 nM CCR5-224 or 4 nM of VF2468. The lane labeled "+*Pvu*I" is a digest of the pre-selection library at *Pvu*I sites introduced adjacent to library members. The laddering on the gels result from cleavage of pre-selection DNA concatamers at more than one site. There is a dose-dependent increase in the amount of the bottom band, which corresponds to cleavage at two adjacent library sites in the same pre-selection DNA molecule. This bottom band of DNA was enriched by PCR and gel purification before sequencing.

Off-target cleavage is dependent on ZFN concentration

As expected, each enzyme cleaved only a subset of library members. The preselection libraries for CCR5-224 and VF2468 had means of 4.56 and 3.45 mutations per complete target site (two half-sites), respectively, whereas 'postselection' libraries exposed to the highest concentrations of ZFN (4 nM CCR5-224 and 4 nM VF2468) during the *in vitro* selection had means of 2.79 and 1.53 mutations per target site, respectively (Figure 4.4). We note that this selection strategy will most likely not recover all cleaved sequences owing to limitations in sequencing throughput.

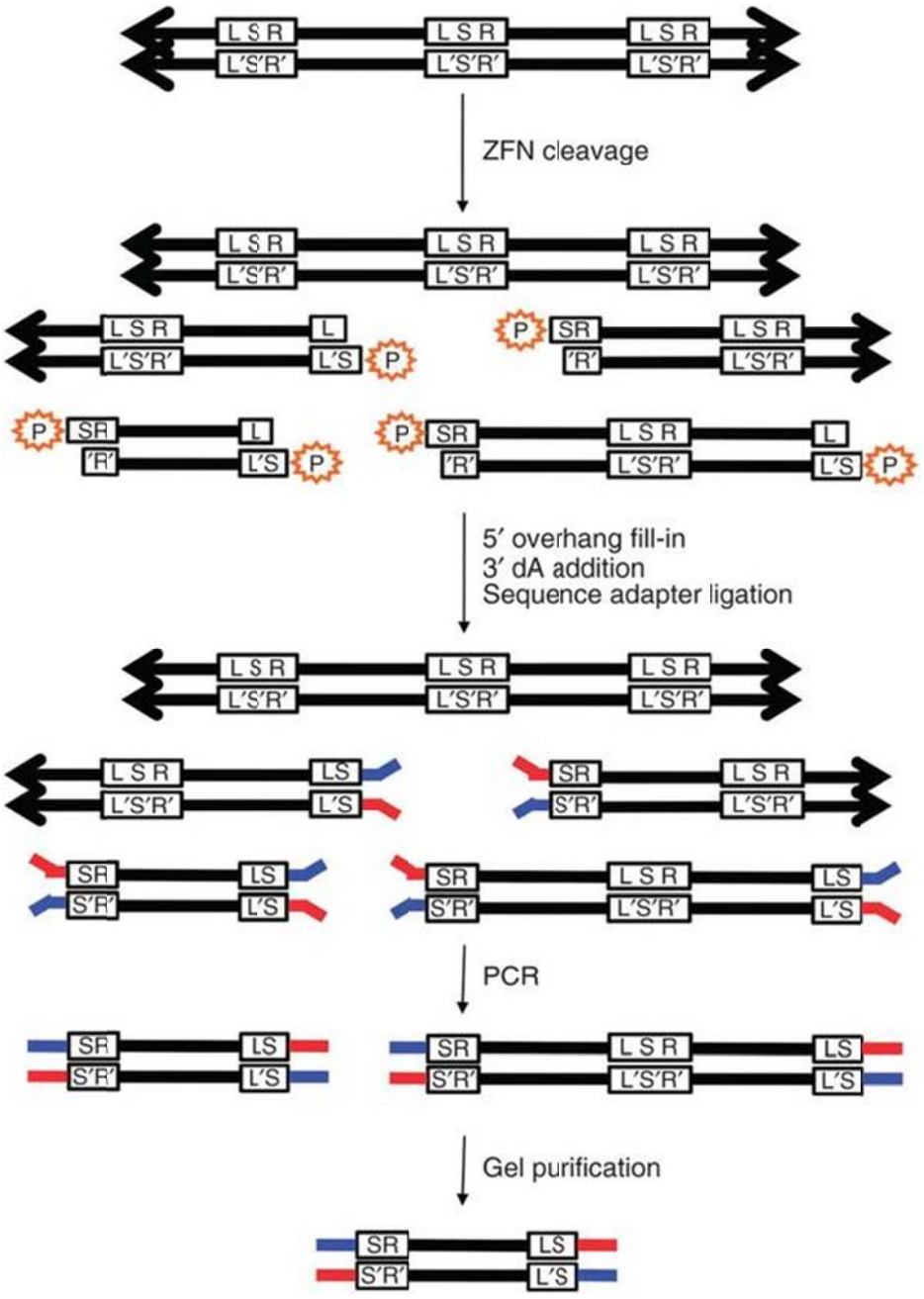


Figure 4.3: *In vitro* selection for ZFN-mediated cleavage
Preselection library members are concatemers (represented by arrows) of identical ZFN target sites lacking 5′ phosphates (P). L, left half-site; R, right half-site, S, spacer; and L′, S′ and R′ are sequences complementary to L, S and R, respectively. ZFN cleavage reveals a 5′ phosphate, which is required for ligation of sequencing adapters (red and blue). The only sequences that can be amplified by PCR using primers complementary to the red and blue adapters are sequences that have been cleaved twice and have adapters on both ends. DNA cleaved at adjacent sites is purified via gel electrophoresis and sequenced. A computational screening step after sequencing ensures that the filled-in spacer sequences (S and S′) are complementary and therefore from the same molecule.

Table 4.1. Sequencing statistics
The total number of interpretable sequences ("total sequences") and the number of analyzed sequences for each *in vitro* selection condition are shown. Analyzed sequences are non-repeated sequences containing no ambiguous nucleotides that, for post-selection sequences, contained reverse complementary overhang sequences of at least four bases, a signature used in this study as a hallmark of ZFN-mediated cleavage. "Incompatible overhangs" refer to sequences that did not contain reverse complementary overhang sequences of at least four bases. The high abundance of repeated sequences in the 0.5 nM, 1 nM, and 2 nM selections indicate that the number of sequencing reads obtained in those selections, before repeat sequences were removed, was larger than the number of individual DNA sequences that survived all experimental selection steps.

			Rejected Sequences		
	Total Sequences	Analyzed Sequences	Incompatible Overhangs	Repeated Sequences	Uncalled Bases in Half-Sites
CCR5-224 Pre-Selection	1,426,442	1,392,576	0	33,660	206
CCR5-224 0.5 nM	649,348	52,552	209,442	387,299	55
CCR5-224 1 nM	488,798	55,618	89,672	343,442	66
CCR5-224 2 nM	1,184,523	303,462	170,700	710,212	149
CCR5-224 4 nM	1,339,631	815,634	352,888	170,700	159
Total	5,088,742	2,619,842	822,702	1,645,563	635
VF2468 Pre-Selection	1,431,372	1,393,153	0	38,128	91
VF2468 0.5 nM	297,650	25,851	79,113	192,671	15
VF2468 1 nM	148,556	24,735	19,276	104,541	4
VF2468 2 nM	1,362,058	339,076	217,475	805,433	74
VF2468 4 nM	1,055,972	397,573	376,364	281,991	44
Total	4,295,608	2,180,388	692,228	1,422,764	228

As ZFN concentration decreased, both ZFNs exhibited less tolerance for off-target sequences. At the lowest concentrations (0.5 nM CCR5-224 and 0.5 nM VF2468), cleaved sites contained an average of 1.84 and 1.10 mutations, respectively. We placed a small subset of the identified sites in a new DNA context and incubated the constructs *in vitro* with 2 nM CCR5-224 or 1 nM VF2468 for 4 h at 37 °C (Figure 4.5). We observed cleavage for all tested sites, and those sites emerging from the more stringent (low ZFN concentration) selections were cleaved more efficiently than those from the less stringent selections. All of the tested sequences contained several mutations, yet some were cleaved *in vitro* more efficiently than the designed target.

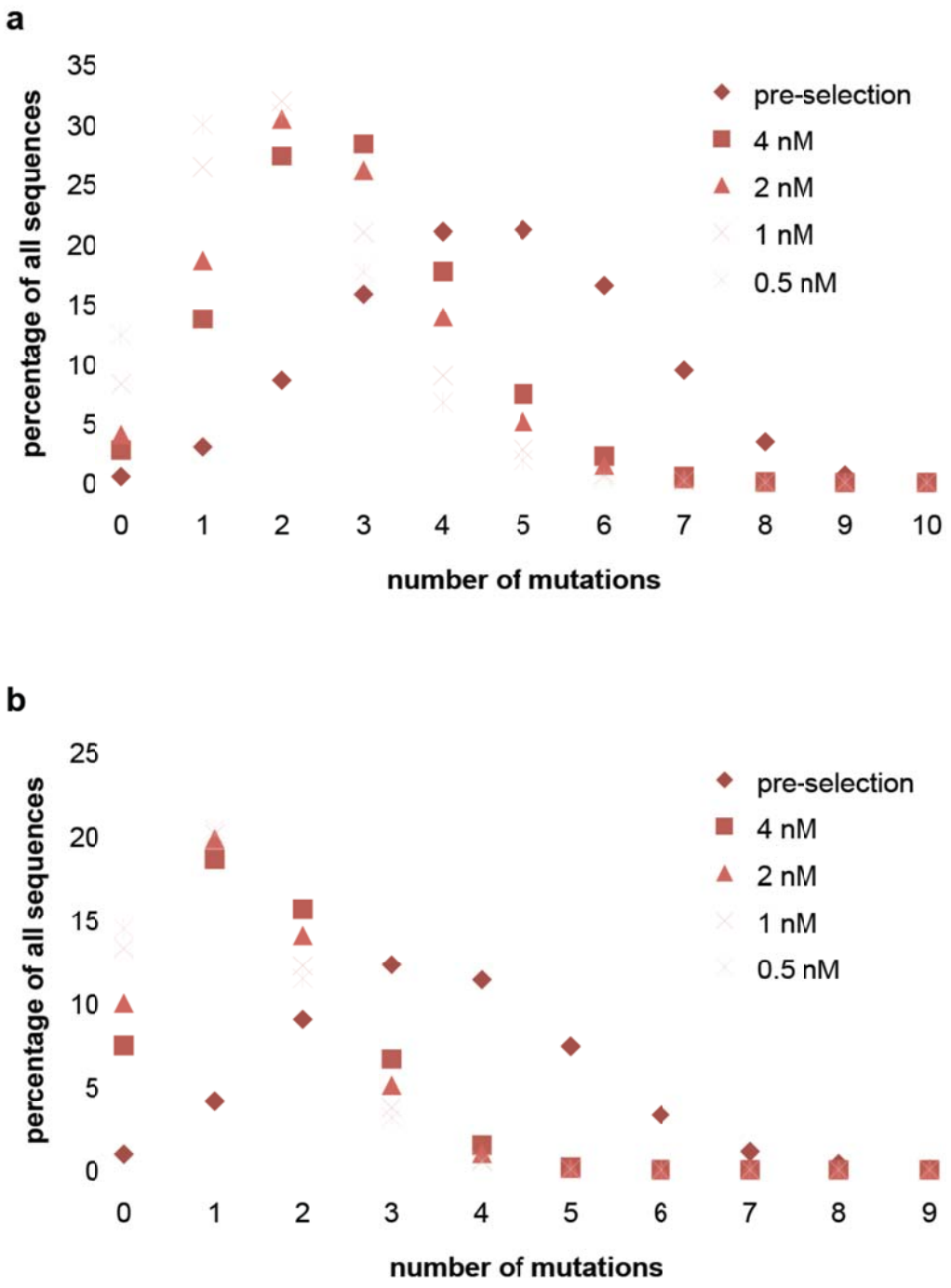
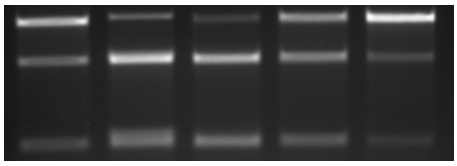


Figure 4.4: ZFN off-target cleavage is dependent on enzyme concentration
For both (a) CCR5-224 and (b) VF2468 the distribution of cleavable sites revealed by *in vitro* selection shifts to include sites that are less similar to the target site as the concentration of ZFN increases. Both CCR5-224 and VF2468 selections enrich for sites that have fewer mutations than the pre-selection library. For comparisons between preselection and post-selection library means for all combinations of selection stringencies, *P*-values are 0 with the exception of the comparison between 0.5 nM and 1 nM VF2468 selections, which has a *P*-value of 1.7 x 10-14.

a

CCR5-224

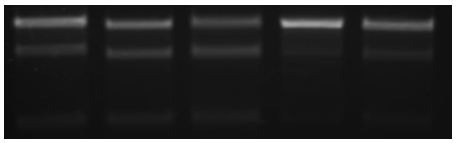


wt 5-4 5-3 5-2 5-1

	(+) site (5'->3')	(-) site (5'->3')	4 nM	2 nM	1 nM	0.5 nM
wt	GAT GAG GAT GAC	AAA CTG CAA AAG	X	X	X	X
5-4	GAT GAG Ggg cga	AAA CTG CAA AAG	X	X	X	X
5-3	GAT GAG Gca cga	AAA CTG CAA AAG	X	X	X	
5-2	GcT GAG GAT aAC	AAA aTG gAA cAG	X	X		
5-1	GAT aca GAT GAC	AAA CTG gAA AAa	X			

b

VF2468



wt 4-4 4-3 4-2 4-1

	(+) site (5'->3')	(-) site (5'->3')	4 nM	2 nM	1 nM	0.5 nM
wt	GAG TGA GGA	GAC GCT GCT	X	X	X	X
4-4	GAG TGA aac	GAC Gt T GCT	X	X	X	X
4-3	GAG TGA Gt c	GAC Gt T aCT	X	X	X	
4-2	Gt G TGA aaA	GAC Gt T GCT	X	X		
4-1	GAG TGA GGA	GAC Gaa aCc	X			

Figure 4.5: Cleavage efficiency of individual sequences is related to selection stringency
In vitro DNA digests were performed on sequences identified in selections of varying stringencies (marked with 'X's). 2 nM CCR5-224 (a) or 1 nM VF2468 (b) was incubated with 8 nM of linear substrate containing the sequence shown. The 1 kb linear substrate contained a single cleavage site with the spacer sequence found in the genomic target of CCR5-224 ("CTGAT") or VF2468 ("TCGAA") and the indicated (+) and (-) half-sites. Mutant base pairs are represented with lowercase letters. CCR5-224 sites and VF2468 sites that were identified in the highest stringency selections (0.5 nM ZFN) are cleaved most efficiently, while sites that were identified only in the lowest stringency selections (4 nM ZFN) are cleaved least efficiently.

The DNA-cleavage specificity profile of the dimeric CCR5-224 ZFN (Figure 4.6a, Figure 4.7a,b) was notably different from DNA-binding specificity profiles of the CCR5-224 monomers previously determined by SELEX (Perez et al., 2008). For example, some positions, such as the fifth position recognized in the (+) site, canonically containing an adenine (denoted (+)A5) and (+)T9, exhibited tolerance for off-target base pairs in our cleavage selection that had not been predicted by the SELEX study. VF2468, which had not been previously characterized with respect to either DNA-binding or DNA-cleavage specificity, revealed two positions, (−)C5 and (+)A9, that exhibited limited sequence preference, suggesting that they were poorly recognized by the ZFNs (Figure 4.6b, Figure 4.7c,d). We note that 2 nM CCR5-224 and 1 nM VF2468 cleave with similar efficiencies *in vitro* (Figure 4.1d), and therefore we show these data together in Figure 4.6.

Compensation between half-sites affects DNA recognition

Our results revealed that ZFN substrates with mutations in one half-site are more likely to have additional mutations in nearby positions in the same half-site compared to the preselection library and are less likely to have additional mutations in the other half-site. Although this effect was greatest when the most strongly specified base pairs were mutated (Figure 4.8), we observed this compensatory phenomenon for all specified half-site positions for both the *CCR5* and *VEGFA*-targeting ZFNs (Figure 4.9 and 4.10). For a minority of nucleotides in cleaved sites, such as VF2468 target site positions (+)G1, (–)G1, (–)A2 and (–)C3, mutation led to decreased tolerance of mutations in base pairs in the other half-site and also a slight decrease, rather than an increase, in mutational tolerance in the same half-site. When two of these mutations, (+)G1 and (–)G1, were included at the same time, mutational tolerance at all other positions decreased

(Figure 4.11). Thus, tolerance of mutations at one half-site is influenced by DNA recognition at the other half-site.

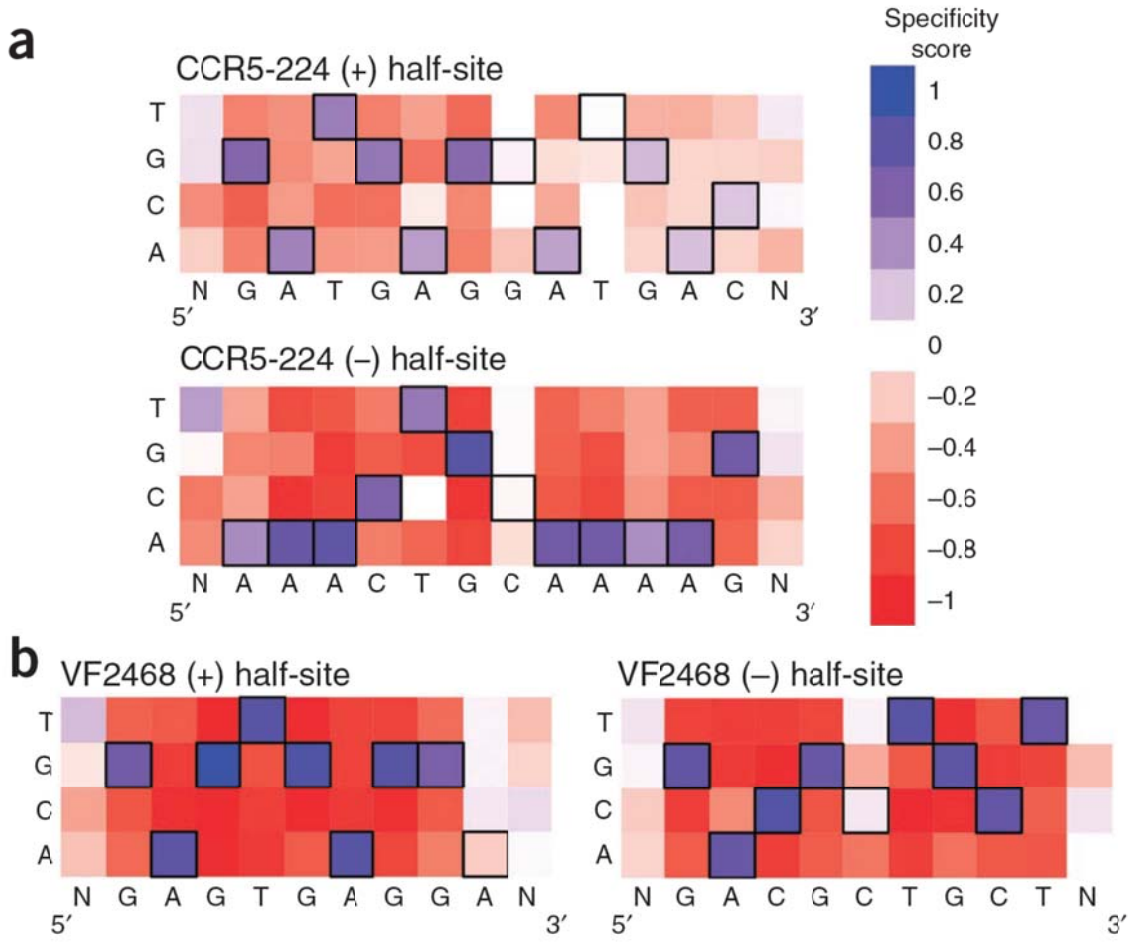


Figure 4.6: DNA cleavage sequence specificity profiles for CCR5-224 and VF2468 ZFNs
The heat maps show specificity scores for the cleavage of 14 nM of DNA library with (a) 2 nM CCR5-224 or (b) 1 nM VF2468. The target DNA sequence is shown below each half-site. Black boxes indicate target base pairs. Specificity scores were calculated by dividing the change in frequency of each base pair at each position in the post-selection DNA pool compared to the pre-selection pool by the maximal possible change in frequency of each base pair at each position. Blue boxes indicate enrichment for a base pair at a given position, white boxes indicate no enrichment, and red boxes indicate enrichment against a base pair at a given position. The darkest blue shown in the legend corresponds to absolute preference for a given base pair (specificity score = 1.0), while the darkest red corresponds to an absolute preference against a given base pair (specificity score = -1.0).

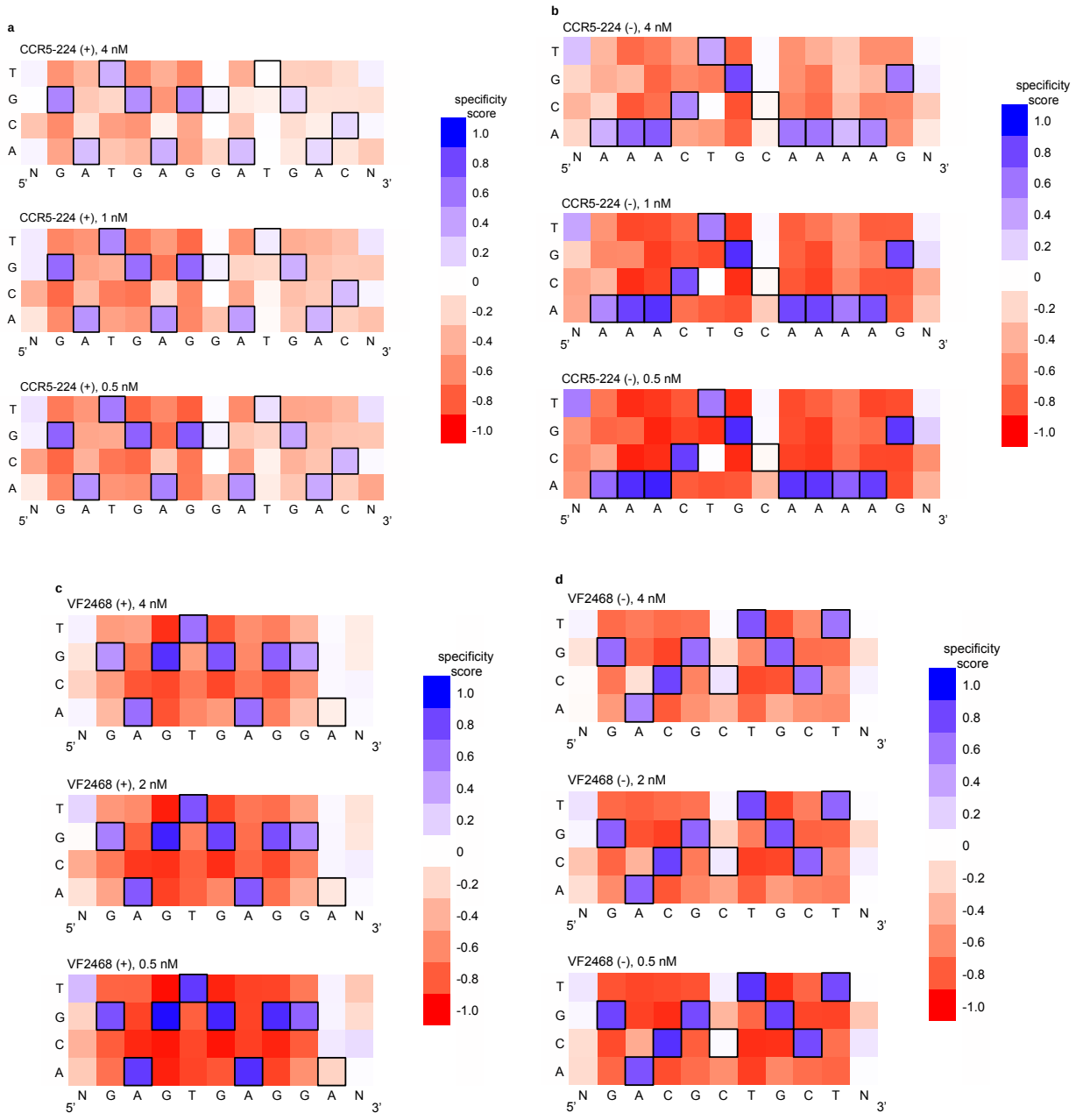


Figure 4.7: Concentration-dependent sequence profiles for CCR5-224 and VF2468 ZFNs The heat maps show specificity scores for the cleavage of 14 nM of total DNA library with varying amounts of (a) CCR5-224 or (b) VF2468. The target DNA sequence is shown below each half-site. Black boxes indicate target base pairs. Specificity scores were calculated by dividing the change in frequency of each base pair at each position in the post-selection DNA pool compared to the pre-selection pool by the maximal possible change in frequency of each base pair at each position. Blue boxes indicate specificity for a base pair at a given position, white boxes indicate no specificity, and red boxes indicate specificity against a base pair at a given position. The darkest blue shown in the legend corresponds to absolute preference for a given base pair (specificity score = 1.0), while the darkest red corresponds to an absolute preference against a given base pair (specificity score = -1.0).

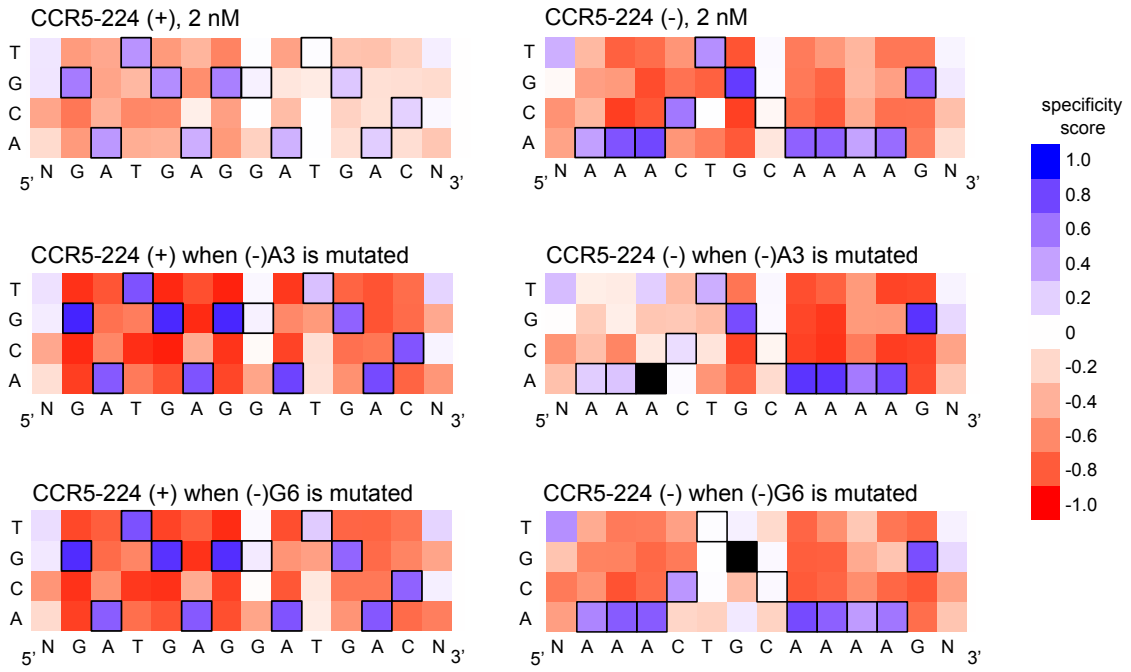
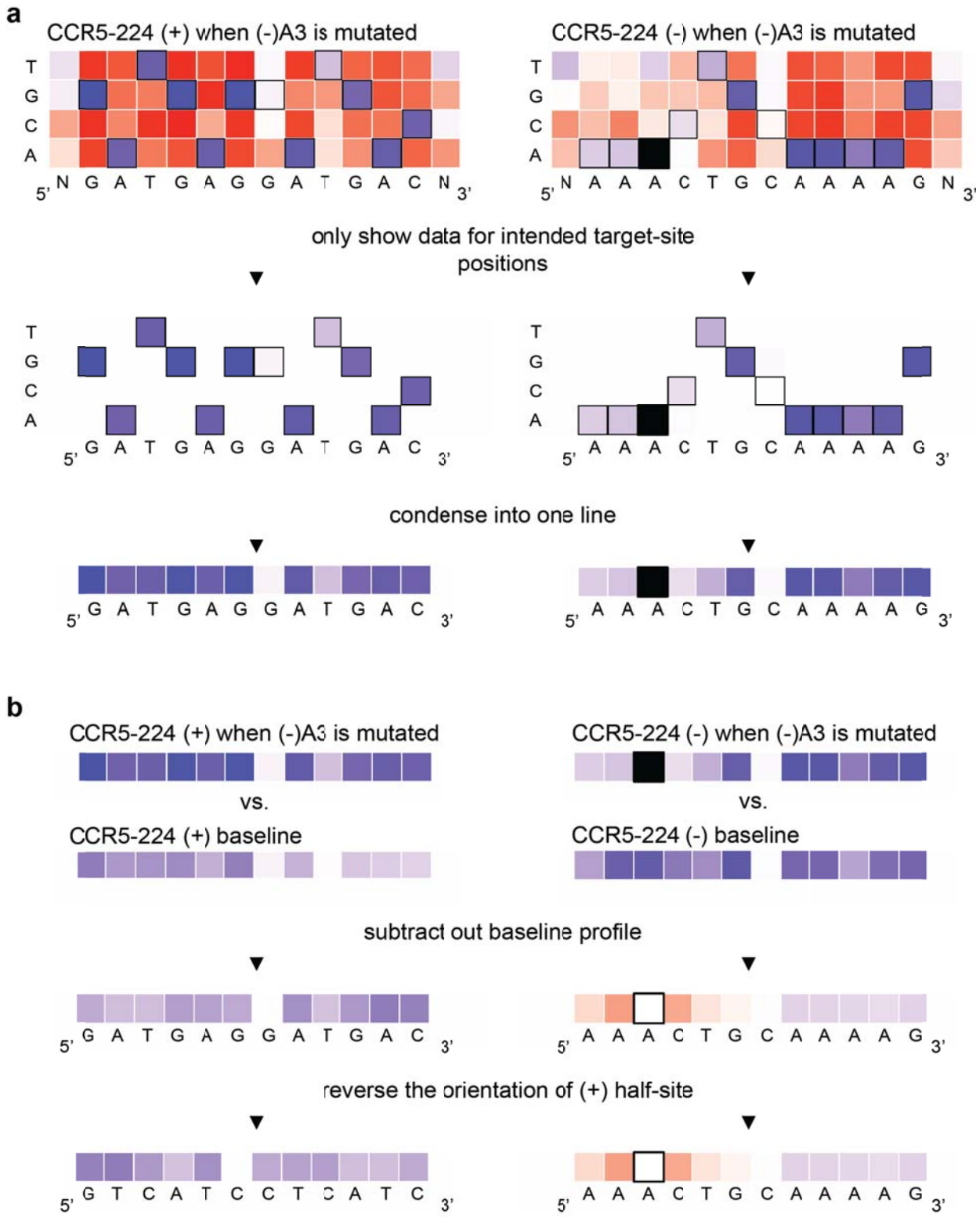


Figure 4.8: Stringency at the (+) half-site increases when CCR5-224 cleaves sites with mutations at highly specified base pairs in the (-) half-site

The heat maps show specificity scores for sequences identified in the *in vitro* selection with 2 nM CCR5-224. For (-)A3 and (-)G6, indicated by filled black boxes, both pre-selection library sequences and post-selection sequences were filtered to exclude any sequences that contained an A at position 3 in the (-) half-site or G at position 6 in the (-) half-site, respectively, before specificity scores were calculated. For sites with either (-) half-site mutation, there is an increase in specificity at the (+) half-site. Black boxes indicate target base pairs. Specificity scores were calculated by dividing the change in frequency of each base pair at each position in the post-selection DNA pool compared to the pre-selection pool by the maximal possible change in frequency of each base pair at each position. Blue boxes indicate specificity for a base pair at a given position, white boxes indicate no specificity, and red boxes indicate specificity against a base pair at a given position. The darkest blue shown in the legend corresponds to absolute preference for a given base pair (specificity score = 1.0), while the darkest red corresponds to an absolute preference against a given base pair (specificity score = -1.0).

Figure 4.9: Data processing steps used to create mutation compensation difference maps
The steps to create each line of the difference map in Figure 3 are shown for the example of a mutation at position (-)A3. (**a**) Heat maps of the type described in Supplementary Figure S7 are condensed into one line to show only the specificity scores for intended target site nucleotides (in black outlined boxes in Supplementary Figure S7). (**b**) The condensed heat maps are then compared to a condensed heat map corresponding to the unfiltered baseline profile from Figure 2, to create a condensed difference heat map that shows the relative effect of mutation at the position specified by the white box with black outline on the specificity score profile. Blue boxes indicate an increase in sequence stringency at positions in cleaved sites that contain mutations at the position indicated by the white box, while red boxes indicate a decrease in sequence stringency and white boxes, no change in sequence stringency. The (+) half-site difference map is reversed to match the orientation of the (+) half-site as it is found in the genome rather than as it is recognized by the zinc finger domain of the ZFN.

Figure 4.9 (continued):



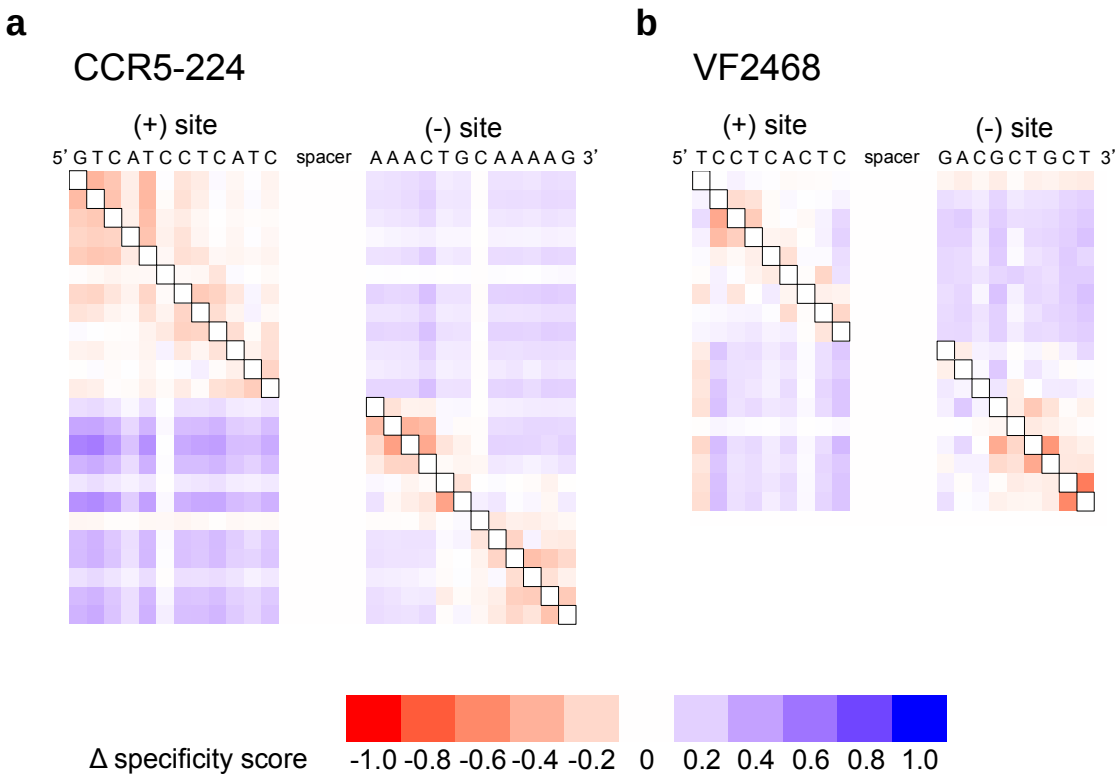


Figure 4.10: Evidence for a compensation model of ZFN target site recognition
Specificity scores calculated from the set of all sites identified in selections containing (a) 2 nM CCR5-224 or (b) 1 nM VF2468 selections were subtracted from specificity scores calculated from the set of sites containing mutations in the positions specified by black boxes. Shades of blue indicate increased specificity score (more stringency) when the boxed position is mutated and shades of red indicate decreased specificity score (less stringency). Sites are listed in their genomic orientation; the (+) half-site of CCR5-224 and the (+) half-site of VF2468 are therefore listed as reverse complements of the sequences found in Figure 4. Positions near off-target base pairs are generally recognized with decreased stringency, while positions in non-mutated half-sites are generally recognized with increased stringency.

This compensation model for ZFN site recognition applies not only to non-ideal half-sites but also to spacers with non-ideal lengths. In general, the ZFNs cleaved at characteristic locations in the spacers (Figure 4.12), and preferentially cleaved 5 bp and 6 bp spacers over 4 bp and 7 bp spacers (Figure 4.13 and 4.14). However, cleaved sites with 5 bp or 6 bp spacers had greater sequence tolerance at the flanking half-sites than sites with 4 bp or 7 bp spacers (Figure 4.15). Therefore, spacer imperfections, similar to half-site mutations, lead to more stringent *in vitro* recognition of other regions of the DNA substrate.

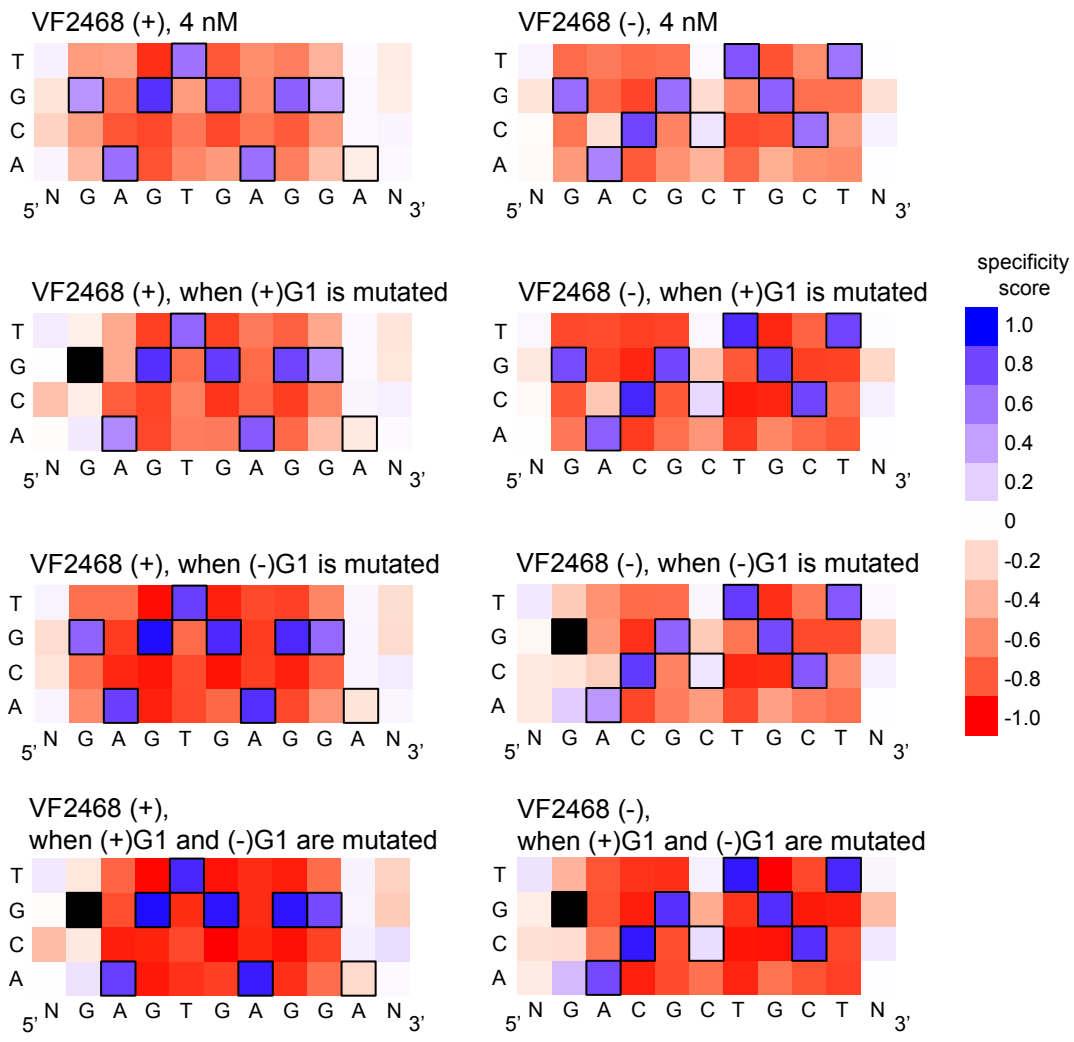
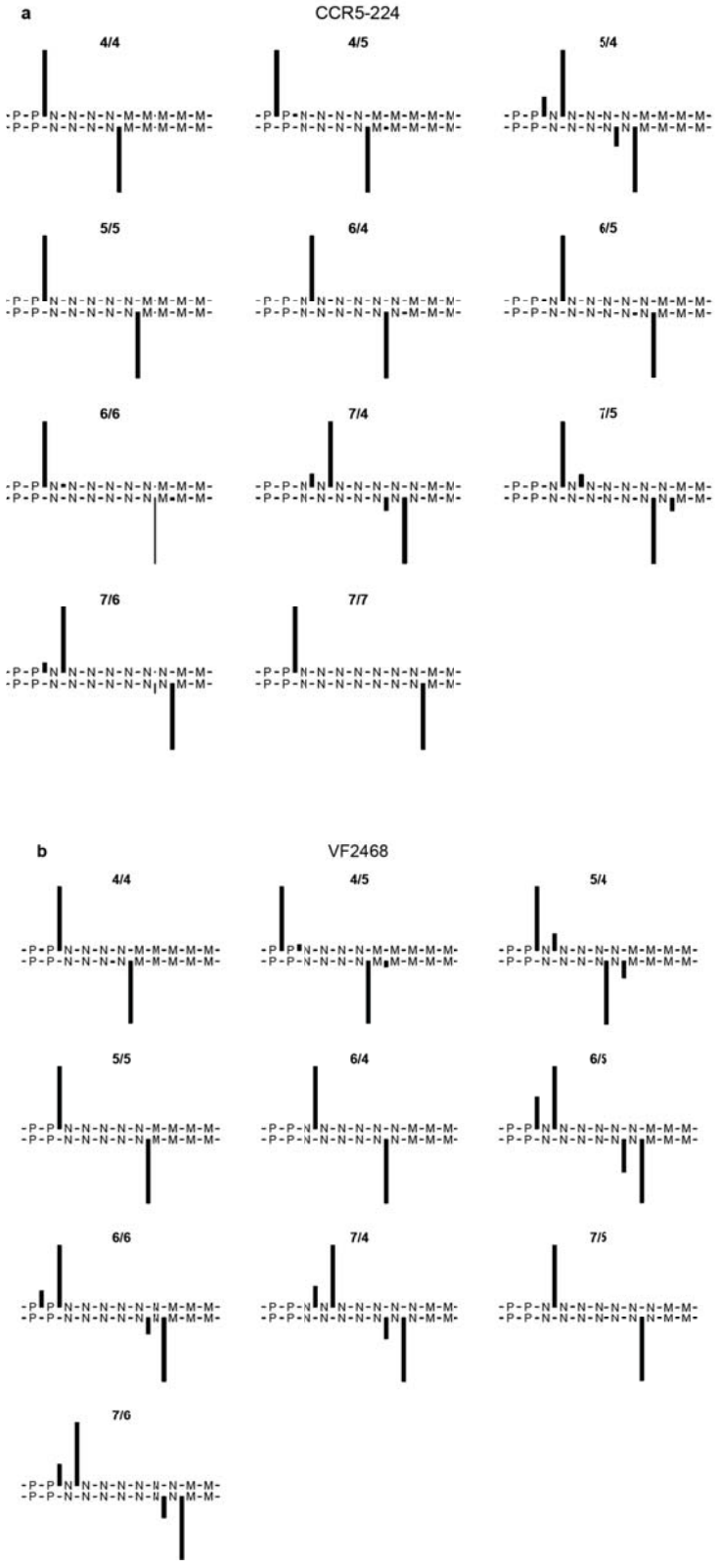


Figure 4.11: Stringency at both half-sites increases when VF2468 cleaves sites with mutations at the first base pair of both half-sites

The heat maps show specificity scores for sequences identified in the *in vitro* selection with 4 nM VF2468. For (+)G1, (-)G1, and (+)G1/(-)G1, indicated by filled black boxes, both pre-selection library sequences and post-selection sequences were filtered to exclude any sequences that contained an G at position 1 in the (+) half-site and/or G at position 1 in the (-) half-site, before specificity scores were calculated. For sites with either mutation, there is decrease in mutational tolerance at the opposite half-site and a very slight decrease in mutational tolerance at the same half-site. Sites with both mutations show a strong increase in stringency at both half-sites. Black boxes indicate on-target base pairs. Specificity scores were calculated by dividing the change in frequency of each base pair at each position in the post-selection DNA pool compared to the pre-selection pool by the maximal possible change in frequency of each base pair at each position. Blue boxes indicate specificity for a base pair at a given position, white boxes indicate no specificity, and red boxes indicate specificity against a base pair at a given position. The darkest blue shown in the legend corresponds to absolute preference for a given base pair (specificity score = 1.0), while the darkest red corresponds to an absolute preference against a given base pair (specificity score = -1.0).

Figure 4.12: ZFN cleavage occurs at characteristic locations in the DNA target site
The plots show the locations of cleavage sites identified in the *in vitro* selections with (a) 4 nM CCR5-224 or (b) 4 nM VF2468. Figure 4.16 shows results for 4 nM VF2468. The cleavage site locations show similar patterns for both ZFNs except in the case of five-base pair spacers with four-base overhangs. The titles refer to the spacer length/overhang length combination that is plotted (a site with a six base-pair spacer and a four base overhang is referred to as "6/4"). The black bars indicate the relative number of sequences cleaved for each combination of spacer length and overhang length. 'P' refers to nucleotides in the (+) target half-site, 'M' refers to nucleotides in the (-) target half-site, and 'N' refers to nucleotides in the spacer. There were no "7/7" sequences from the 4 nM VF2468 selection. Only sequences with overhangs of at least 4 bases were tabulated.

Figure 4.12 (continued):



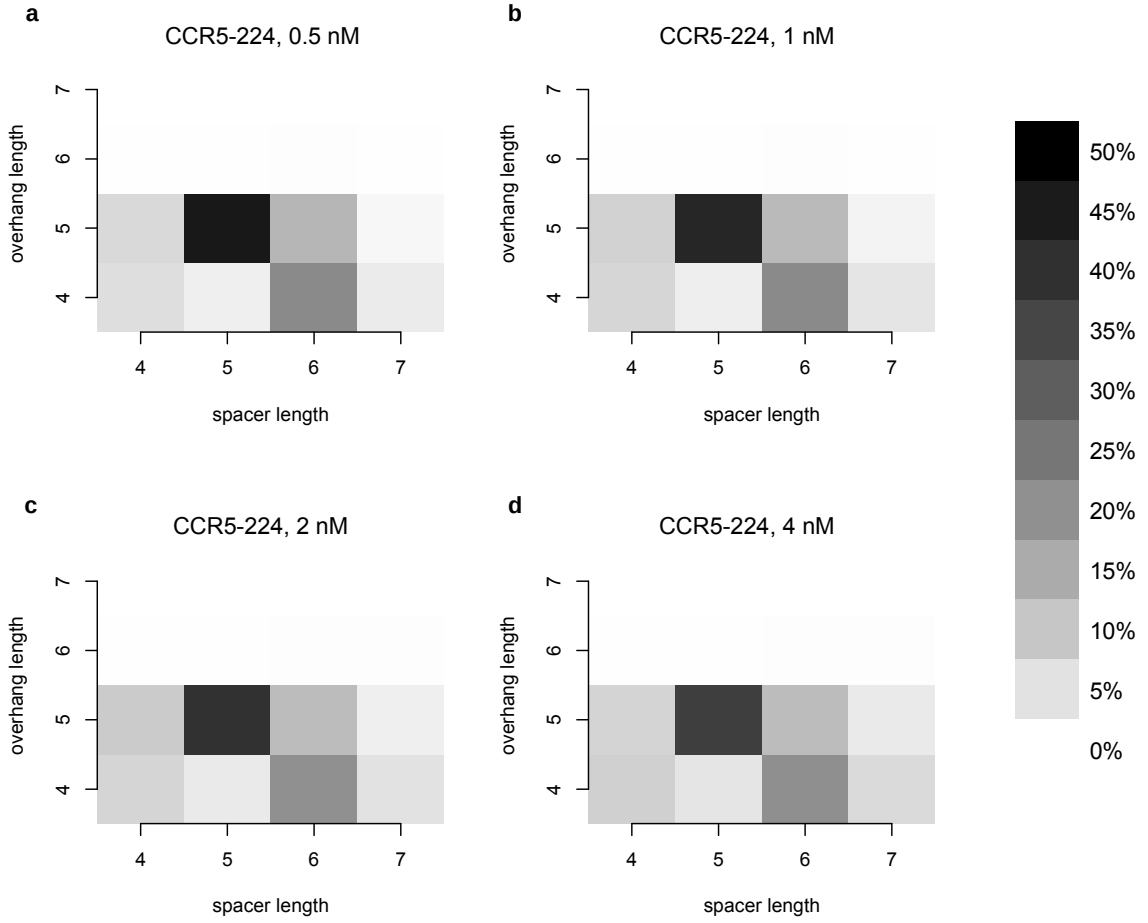


Figure 4.13: CCR5-224 preferentially cleaves five- and six-base pair spacers and cleaves five-base pair spacers to leave five-nucleotide overhangs
The heat maps show the percentage of all sequences surviving each of the four CCR5-224 *in vitro* selections (a-d) that have the spacer and overhang lengths shown.

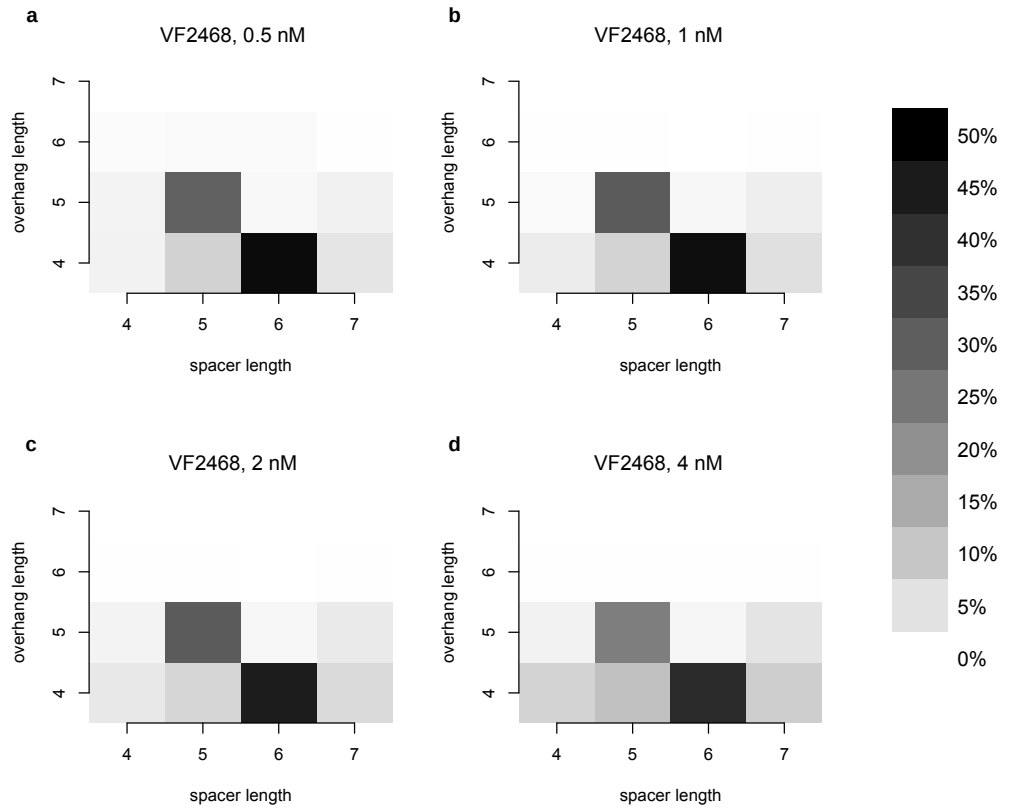


Figure 4.14: VF2468 preferentially cleaves five- and six-base pair spacers, cleaves five-base pair spacers to leave five-nucleotide overhangs, and cleaves six-base pair spacers to leave four-nucleotide overhangs

The heat maps show the percentage of all sequences surviving each of the four VF2468 *in vitro* selections (a-d) that have the spacer and overhang lengths shown.

Figure 4.15: ZFNs show spacer length-dependent sequence preferences
Both CCR5-224 (a-c) and VF2468 (d-f) show increased specificity for half-sites flanking four- and seven-base pair spacers than for half-sites flanking five- and six-base pair spacers. For both ZFNs, one half-site has a greater change in mutational tolerance than the other, and the change in mutational tolerance is concentration dependent.

Figure 4.15 (continued):

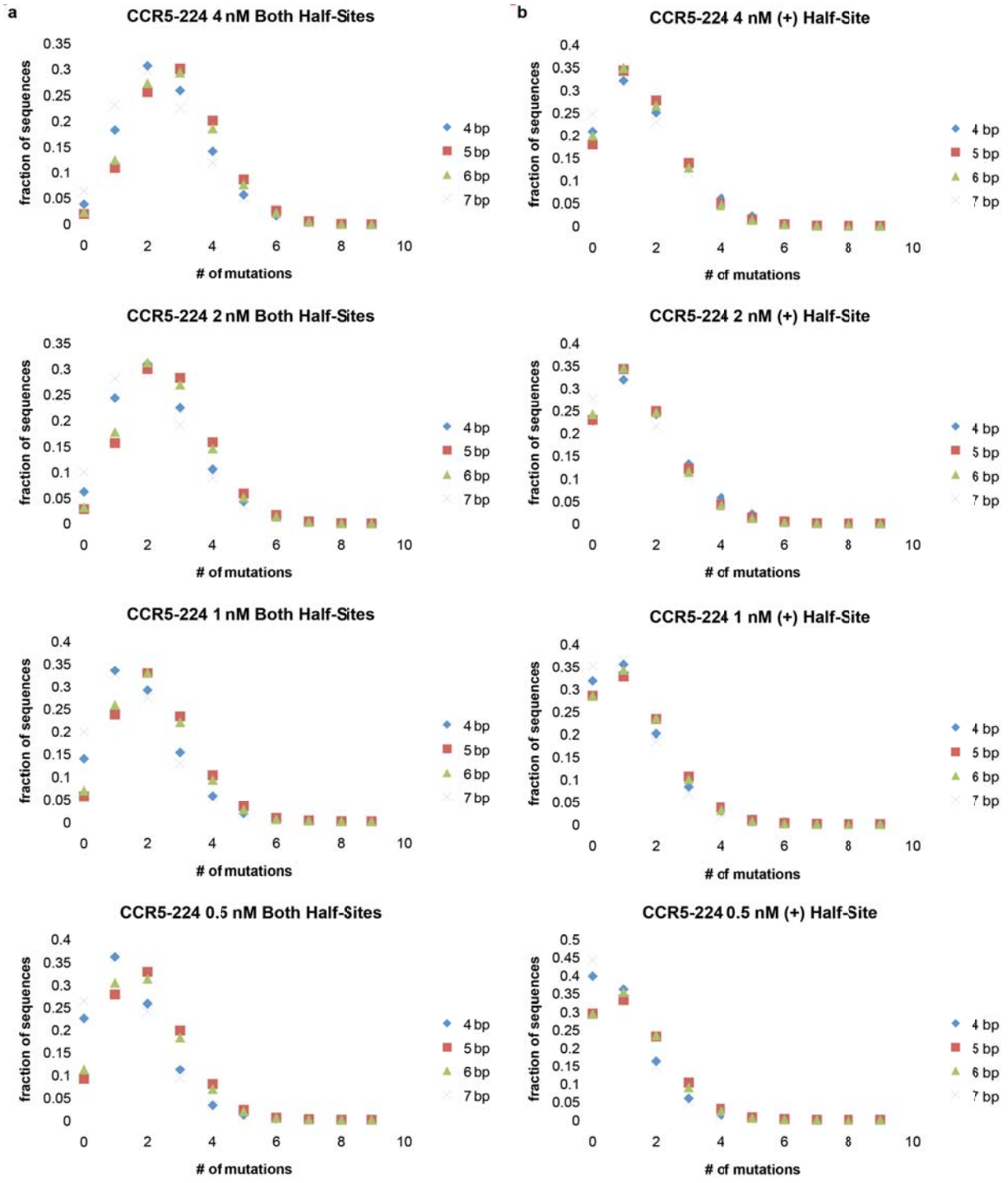


Figure 4.15 (continued):

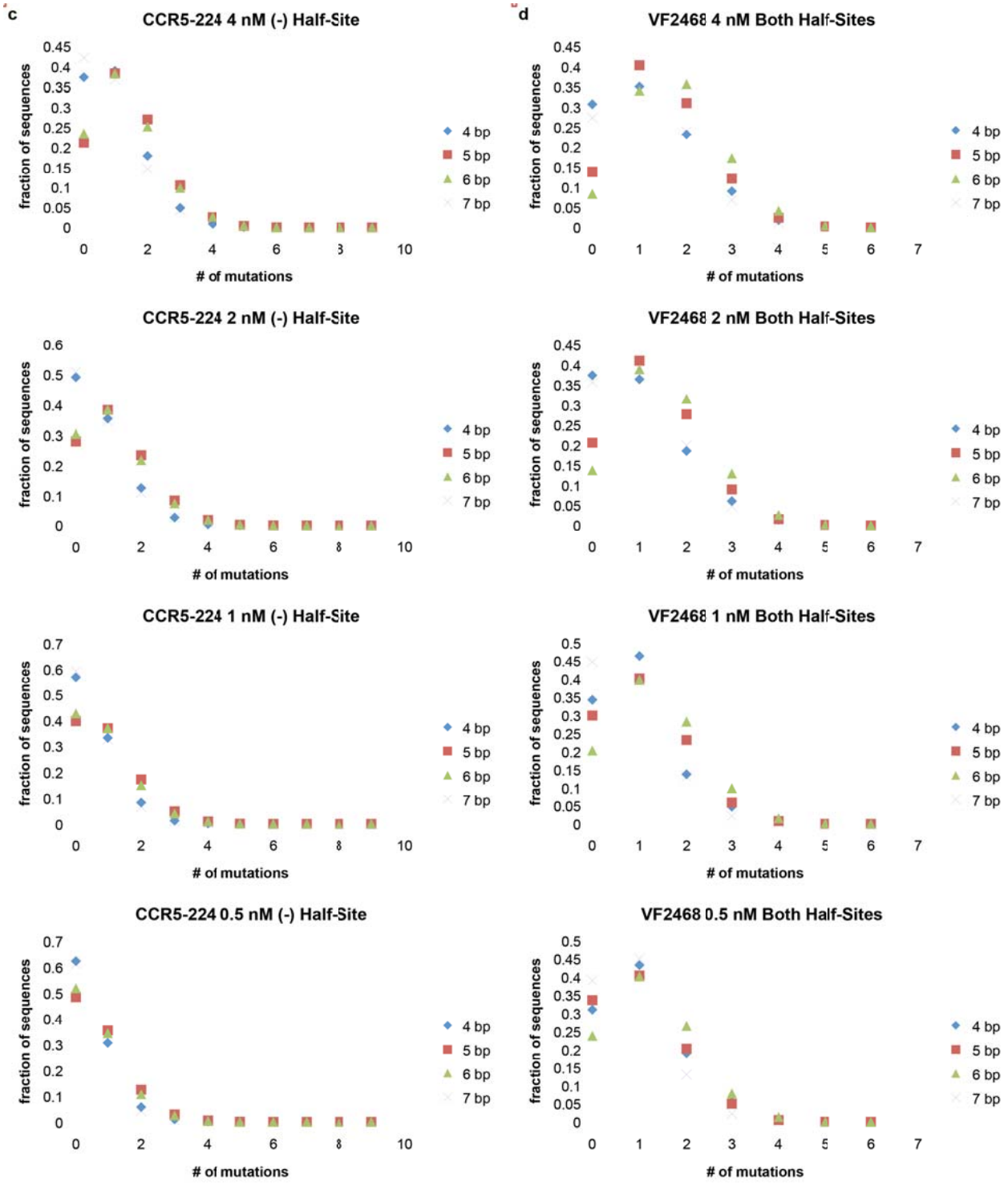
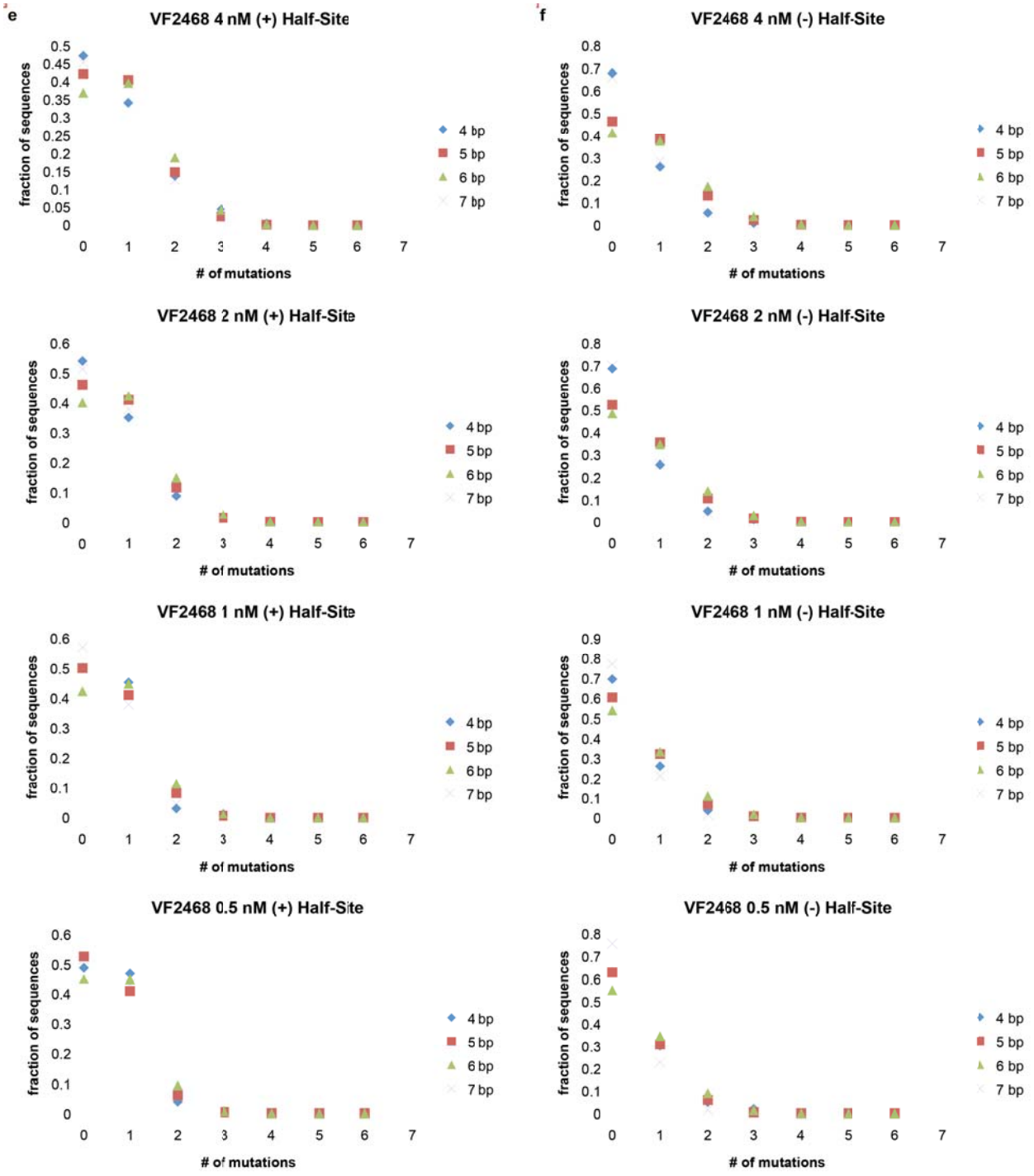


Figure 4.15 (continued):



ZFNs can cleave many sequences with up to three mutations

We calculated enrichment factors for all sequences containing three or fewer mutations by dividing each sequence's frequency of occurrence in the postselection libraries by its frequency of occurrence in the preselection libraries. Among sequences enriched by cleavage (enrichment factor > 1), CCR5-224 could cleave all unique single-mutant sequences, 93% of all unique double-mutant sequences and half of all possible triple-mutant sequences (Figure 4.16a and Table 4.2a) at the highest enzyme concentration used. VF2468 could cleave 98% of all unique single-mutant sequences, half of all unique double-mutant sequences and 17% of all triple-mutant sequences (Figure 4.16b and Table 4.2b).

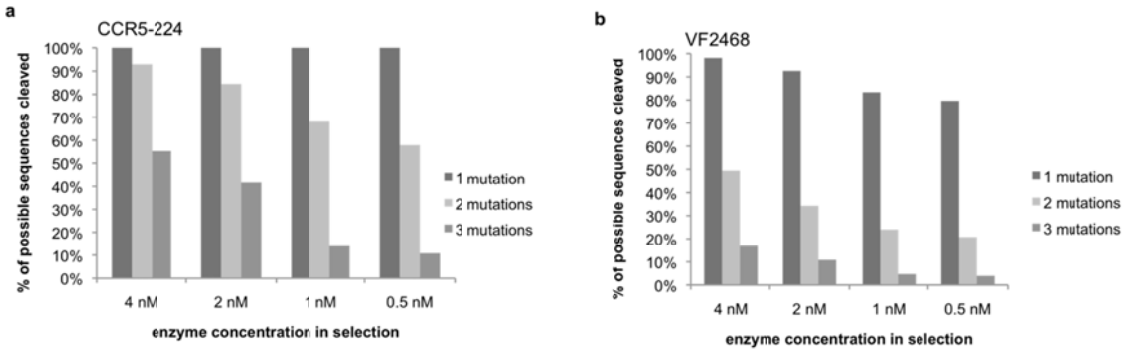


Figure 4.16: Both ZFNs tested have the potential to cleave a large fraction of target sites with three or fewer mutations
The percentages of the sequences with one, two, or three mutations that are enriched for cleavage (enrichment factor > 1) by the (a) CCR5-224 ZFN and (b) VF2468 ZFN are shown. Enrichment factors are calculated for each sequence identified in the selection by dividing the observed frequency of that sequence in the post-selection sequenced library by the frequency of that sequence in the pre-selection library.

Table 4.2: Both ZFNs tested have the ability to cleave a large fraction of target sites with three or fewer mutations

The percentage of the set of sequences with 1, 2, or 3 mutations (muts) that can be cleaved by (a) the CCR5-224 ZFN and (b) the VF2468 ZFN is shown. Enrichment factors (EFs) were calculated for each sequence identified in the selection by dividing the observed frequency of that sequence in the post-selection sequenced library by the observed frequency of that sequence in the pre-selection library. The enrichment factors for the wild-type sequence (wt EFs) calculated for each *in vitro* selection stringency are shown in the first row of the table.

a

CCR5-224	4 nM (wt EF = 5.48)			2 nM (wt EF = 8.11)			1 nM (wt EF = 16.6)			0.5 nM (wt EF = 24.9)		
	1 mut	2 muts	3 muts	1 mut	2 muts	3 muts	1 mut	2 muts	3 muts	1 mut	2 muts	3 muts
EF > 0	100%	99.96%	76%	100%	99%	49%	100%	83%	14%	100%	75%	11%
EF > 1	100%	93%	55%	100%	84%	42%	100%	68%	14%	100%	58%	11%
EF > 2	100%	78%	37%	100%	70%	31%	99%	55%	14%	96%	46%	11%
EF > (.5 x wt EF)	100%	63%	28%	93%	40%	17%	51%	15%	8%	31%	8%	4%
EF > wt EF	14%	9%	10%	8%	6%	6%	3%	2%	3%	6%	1%	2%

b

VF2468	4 nM (wt EF = 16.7)			2 nM (wt EF = 22.5)			1 nM (wt EF = 30.2)			0.5 nM (wt EF = 33.1)		
	1 mut	2 muts	3 muts	1 mut	2 muts	3 muts	1 mut	2 muts	3 muts	1 mut	2 muts	3 muts
EF > 0	100%	95%	38%	100%	92%	26%	100%	47%	5%	100%	44%	4%
EF > 1	98%	49%	17%	93%	34%	11%	83%	24%	5%	80%	21%	4%
EF > 2	89%	31%	10%	83%	23%	7%	74%	17%	5%	61%	14%	4%
EF > (.5 x wt EF)	57%	15%	4%	30%	10%	2%	11%	6%	1%	9%	5%	1%
EF > wt EF	7%	1%	1%	7%	1%	0.4%	7%	1%	0.4%	7%	1%	0.3%

Because our approach assays active ZFN dimers, it revealed the complete sequences of ZFN sites that can be cleaved. Ignoring the sequence of the spacer, the selection revealed 37 sites in the human genome with 5 bp or 6 bp spacers that CCR5-224 can cleave *in vitro* (Table 4.3), and 2,652 sites in the human genome that VF2468 can cleave *in vitro*. Among the genomic sites that VF2468 cleaved *in vitro*, 1,428 site (excluding the spacer sequence). Despite greater discrimination against single-, double- and triple-mutant sequences by VF2468 compared to CCR5-224 (Figure 4.16, Table 4.2), the larger number of *in vitro*–cleavable VF2468 sites reflects the difference in the number of sites in the human genome that are three or fewer mutations away from the VF2468 target site (3,450 sites) versus those that are three or fewer mutations away from the CCR5-224 target site (8 sites) (Tables 4.4 and 4.5).

Table 4.3: CCR5-224 off-target sites in the human genome

Sites marked with an 'X' were found in the *in vitro* selection dataset. 'T' refers to the total number of mutations in the site, and '(+)' and '(-)' to the number of mutations in the (+) and (-) half-sites, respectively. Chromosomal coordinates from build 36 of the human genome are listed. Mutation frequency for each site is the percentage of sequences with insertions or deletions (indels) in the sequenced DNA from cultured K562 cells expressing active CCR5-224. Bolded red sites have significantly enriched indel percentages in the active nuclease sample compared to cells containing empty vector. The sequences of the sites are listed as 5' (+) half-site/spacer/(-) half-site 3', therefore the (+) half-site is listed in the reverse sense as it is in the sequence profiles. Indels and totals are not shown for sites that were not tested. *P*-values shown are for the one-sided alternative hypothesis that the indel frequency is greater for active ZFN treated cells than for cells not expressing ZFN.

	mutations								*in vitro* selection stringency				empty vector			active CCR5-224			
	T	(+)	(-)	gene	build 36 coordinates	(+) half-site	spacer	(-)half-site	4 nM	2 nM	1 nM	0.5 nM	indels	total	mutation frequency	indels	total	mutation frequency	*p*-value
CCR5-224 1	**0**	**0**	**0**	**CCR5**	**chr3:46389548-46389576**	G'CATCCTCATC	CTGAT	AAACTGCAAAAG	X	X	X	X	**1**	**226676**	**0.00044%**	**105619**	**240966**	**44%**	**0**
CCR5-224 2	**2**	**1**	**1**	**CCR2**	**chr3:46374209-46374237**	G'CgTCCTCATC	TTAAT	AAACTGCAAAAa	X	X	X	X	**0**	**114904**		**12856**	**130496**	**10%**	**0**
CCR5-224 3	**3**	**2**	**1**	**BTBD10 (promoter)**	**chr11:13441738-13441766**	G'ttTCCTCATC	AAAGC	AAACTGCAAAAt	X	X			**1**	**283015**	**0.00035%**	**155**	**224000**	**0.070%**	**0**
CCR5-224 4	4	0	4		chr10:29604352-29604380	G'CATCCTCATC	AGAGA	AAACTGgctAAt	X	X			2	297084	0.00067%	3	245078	0.0012%	0.26
CCR5-224 5	4	3	1	SLC4A8	chr12:50186653-50186682	taaATCCTCATC	TCTATA	AAAaTGCAAAAG	X	X			0	147246		0	138979		
CCR5-224 6	3	2	1	Z83955 RNA	chr12:33484433-33484462	G'CATCCcaATC	GAAGAA	AAACTGaAAAAG	X			X	0	147157		1	146283	0.00068%	0.16
CCR5-224 7	3	1	2	DGKK	chrX:50149961-50149989	c'CATCCTCATC	CATGC	AcAaTGCAAAAG	X				0	316468		0	313981		
CCR5-224 8	3	1	2	GALNT13	chr2:154567664-154567692	G'CATCCTCAgC	ATGGG	AAACaGCAgAAG	X				0	136684		0	94657		
CCR5-224 9	**3**	**1**	**2**		**chr17:61624429-61624457**	G'CATCtTCATC	AAAAG	gAACTGCAAAAc	X				**0**	**178692**		**52**	**146525**	**0.035%**	**2.7E-13**
CCR5-224 10	4	0	4		chrX:145275453-145275481	G'CATCCTCATC	CAATA	AAAgaaCAAAgG	X				0	296730		0	276961		
CCR5-224 11	**4**	**1**	**3**	**TACR3**	**chr4:104775175-104775203**	G'CATCtTCATC	AGCAT	AAACTGtAAAgt	X				**0**	**273436**		**1045**	**308726**	**0.34%**	**0**
CCR5-224 12	4	1	3	PIWIL2	chr822191670-22191698	G'CATCCTCATa	CATAA	AAACTGCcttAG	X										
CCR5-224 13	4	1	3		chr976194351-76194379	a'CATCCTCATC	CATCC	AAtgTtCAAAAG	X				0	168244		1	171618	0.00058%	0.16
CCR5-224 14	**4**	**3**	**1**		**chr6:52114315-52114343**	G'CcTgCTCAgC	AAAAG	AAACTGaAAAAG	X				**0**	**66317**		**35**	**138728**	**0.025%**	**1.6E-09**
CCR5-224 15	**4**	**3**	**1**	**KCNB2**	**chr8:73899370-73899398**	a'gtTCCTCATC	TCCCG	AAACTGCAAAtG	X				**1**	**427161**	**0.00023%**	**280**	**393899**	**0.071%**	**0**
CCR5-224 16	**4**	**3**	**1**		**chr8:4865886-4865914**	G'CtTCCTgATg	CTACC	AAACTGgAAAAG	X				**0**	**190993**		**32**	**171160**	**0.019%**	**7.7E-09**
CCR5-224 17	4	3	1		chr914931072-14931100	aaCATCCaCATC	ATGAA	AAACTGCAAAAa	X				0	163704		0	146176		
CCR5-224 18	6	3	3		chr13:65537268-65537286	a'CtTCCTCATt	ACAGG	AAAaTGtAAtAG	X				0	109939		0	100948		
CCR5-224 19	6	4	2	CUBN	chr10:17044849-17044877	GgCtTCCTgAcC	CACGG	AAACTGtAAAtG	X										
CCR5-224 20	6	5	1	NID1	chr1234244827-234244855	G'ttTgCaCATt	TCAAT	tAACTGCAAAAG	X				0	114743		0	120169		
CCR5-224 21	**3**	**2**	**1**		**chr9:80584200-80584229**	G'CAaCCTCAaC	ACCTAC	AgACTGCAAAAG	X				**0**	**188149**		**127**	**213248**	**0.060%**	**0**
CCR5-224 22	4	1	3	WWOX	chr16:77185306-77185335	G'CATCCTCcTC	CAACTC	cAAtTGCtAAAG	X				0	366156		0	354878		
CCR5-224 23	4	2	2	AMBRA1	chr1':46422800-46422829	G'CtTCCTCcTC	TGCACA	tcACTGCAAAAG	X				0	237240		0	227568		
CCR5-224 24	4	2	2		chr199456616-99456645	G'gATaCTCATC	ATCAGC	AAtCTGCAtAAG	X				0	129468		0	144274		
CCR5-224 25	**4**	**2**	**2**	**WBSCR17**	**chr7:70557254-70557283**	G'tATCCTCAgC	AAACTA	AAACTGgAAcAG	X				**0**	**172543**		**486**	**417198**	**0.12%**	**0**
CCR5-224 26	4	2	2	ITSN	chr2l:34098210-34098239	c'CATgCTCATC	ATTTGT	tAACTGCAAAAt	X				0	267772		0	308093		
CCR5-224 27	4	4	0		chr9106457399-106457428	G;CAgtCTCAgC	ATGGTG	AAACTGCAAAAG	X				0	350592		0	335281		
CCR5-224 28	4	4	0		chr1':49929141-49929170	c'CATtCTgtTC	ATGAAA	AAACTGCAAAAG	X				0	105012		0	99968		
CCR5-224 29	5	3	2		chr15:96714952-96714981	GaagTCCTCATC	CCGAAG	AAACTGaAAgAG	X				0	355674		0	338910		
CCR5-224 30	5	3	2	ZNF462	chr9108684858-108684887	G'CtTCCTCtTt	CACATA	AAACcGCAAAtG	X				0	173646		1	152744	0.00065%	0.16
CCR5-224 31	5	4	1		chr5101113644-101113673	a'aATCCTttTC	TGTTTA	AAACaGCAAAAG	X				1	245650	0.00041%	0	185572		0.84
CCR5-224 32	5	4	1		chr1':43908810-43908839	GaCATCCaaATt	ACATGG	AAACTGaAAAAG	X				0	482635		2	413317	0.00048%	0.079
CCR5-224 33	5	5	0	SDK1	chr73446932-3446961	G'CtTgCTgtTg	CACCTC	AAACTGCAAAAG	X				0	237791		0	200398		
CCR5-224 34	4	1	3	SPTB(coding)	chr14:64329872-64329901	G'CATCCgCATC	GCCCTG	gAACTGgAAAAa		X			0	180783		0	167885		
CCR5-224 35	4	2	2		chr10:54268729-54268758	a'CATCCTCAaC	AAACTA	AAACaGgAAAAG		X									
CCR5-224 36	4	4	0	KIAA1680	chr492322851-92322880	GgaATgCcCATC	ACCACA	AAACTGCAAAAG		X			0	165657		2	153995	0.0013%	0.079
CCR5-224 37	5	5	0		chr5114708142-114708171	G'ttTgCTCcTg	TACTTC	AAACTGCAAAAG		X			0	152083		0	183305		

Table 4.4: Potential VF2468 off-target sites in the human genome

The human genome was searched for DNA sequences surviving *in vitro* selection for VF2468 cleavage. 2,652 sites were found, and are binned in the table by number of mutations.

	VF2468
	# of genomic sites identified
# of mutations	in *in vitro* selection
0	1
1	3
2	243
3	1,181
4	1,077
5	143
6	4

Table 4.5: There are many more similar genomic sequences to the VF2468 site than to the CCR5-224 site

The human genome was computationally searched for sites up to nine mutations away from the canonical CCR5-224 target site and up to six mutations away from the canonical VF2468 target site. The number of occurrences of sites containing five or six base pair spacers in the genome, including repeated sequences, is listed in the table.

CCR5-224	
# of mutations	# of sites in genome
0	1
1	0
2	1
3	6
4	99
5	964
6	9,671
7	65,449
8	372,801
9	1,854,317

VF2468	
# of mutations	# of sites in genome
0	1
1	3
2	245
3	3,201
4	35,995
5	316,213
6	2,025,878

Identified sites are cleaved by ZFNs in human cells

We tested whether CCR5-224 could cleave at sites identified by our selections in human cells by expressing CCR5-224 in K562 cells and examining 34 potential target sites in the human genome for evidence of ZFN-induced mutations using PCR and high-throughput DNA sequencing. We defined sites with evidence of ZFN-mediated cleavage as those with insertion or deletion mutations (indels) characteristic of non-homologous end-joining repair (Table 4.6) that were significantly enriched ($P < 0.05$) in cells expressing active CCR5-224 compared to control cells containing an empty vector. We obtained ~100,000 sequences for each site analyzed, which enabled us to detect that sites were modified at frequencies of at least ~1 in 10,000. Our analysis identified ten such sites: the intended target sequence in *CCR5*, a previously identified sequence in *CCR2* and eight other off-target sequences (Tables 4.3 and 4.6), one of which was in the promoter of the *BTBD10* gene. The eight newly identified off-target sites were modified at fre-

Table 4.6: Sequences of CCR5-224-mediated genomic DNA modifications identified in cultured human K562 cells

Sequences with insertions (blue) and deletions (red) identified after sequencing potential CCR5-224 off-target sites from cultured K562 cells expressing CCR5-224 are shown. The numbers of occurrences are shown to the right of each sequence. The unmodified site is listed under the gene name or coordinates (build 36), and the spacer sequence is underlined.

	# of sequences			# of sequences
BTBD10 (promoter)			chr6:52114315-52114343	
ATTTTGCAGTTT GCTTT GATGAGGAAAAC			CTTTTTCAGTTT CTTTT GCTGAGCAGGAC	
ATTTTGCAGTTT GCTTT GATGAGGAAAAC	63		CTTTTTCAGTTT CTTTT GCTGAGCAGGAC	35
ATTTTGCAGTTT GCTTTGCTTT GATGAGGAAAAC	86			
ATTTTGCAGTTT GgTTTGCTTT GATGAGGAAAAC	1			
ATTTTGCAGTTT GCTTTGCTTT GgTGAGGAAAAC	1		KCNB2	
gTTTTGCAGTTT GCTTTGCTTT GATGAGGAAAAC	1		CATTTGCAGTTT CGGGA GATGAGGAACAT	
cTTTTGCAGgTT GCTTTGCTTT GATGAGGAAAAC	1			
ATTTTGCAGTTT GCTTTGCTTT GATGtGGAAAAC	1		CATTTGCAGTTT CGGGAGA GATGAGGAACAT	158
ATTTTGCAGTTT GCTTT GATGAGGAAAAC	1		CATTTGCAGTTa CGGGAGA GATGAGGAACAT	1
			CATTTGCAGTTT CGGGAGA GATGAGGgACAT	1
			CATTTGacGcTT CGGGAGA GgTGAGGgACAT	1
chr17:61624429-61624457			CATTTGCAGTTT CGGGCGGGA GATGAGGAACAT	109
GTTTTGCAGTTC CTTTT GATGAAGATGAC			CATTTGCAGTTT CGGGCGGGA GATGcGGAACAT	1
			CATTTGCAGTTT CGGGCGGGc GATGAGGAACAT	1
GTTTTGCAGTTC CTTTT GATGAAGATGAC	51		CATTTGCAGTTT CGGGCGGGA GgTGAGGAACAT	1
GTTTTGCAGgTC CTTTT GATGAAGATGAC	1		CgTTTGCAGTTT CGGGCGGGA GATGAGGAACAT	2
			CATTTGCtGTTT CGGGCGGGA GATGAGGAACAT	1
			CATTTGCAGTTT CGGGCGGGA GATGAGGAcCAT	1
TACR3			CATTTGCAGTTT CGGGCGGGA GgTGAGGAACAT	1
ACTTTACAGTTT ATGCT GATGAAGATGAC			CcTTTGCAGTTT CGGGCGGGA GATGAGGAACAT	1
			CATTTGCAGTTg CGGGCGGGA GATGAGGAACAT	1
ACTTTACAGTTT ATGCT GATGAAGATGAC	5			
ACTTTACAGTTT ATGCT GATGAAGATGAC	169			
gCTTTACAGTTT ATGCT GATGAAGATGAC	1		chr8:4865886-4865914	
ACTTTACAGTTT ATGCT GATGAAGAatAC	1		GTCTTCCTGATG CTACC AAACTGGAAAAG	
ACTTTACAGTTT ATGCT GATGAAGATGtt	1			
ACTTTACAGTTT ATGCT GATGAAGATGAC	34		GTCTTCCTGATG CTACC AAACTGGAAAAG	30
ACTTTACgGTTT ATGCT GATGAAGATGAC	1		GTCTTCCTGATG CTACC AAACTtGAAAAG	1
ACTTTACAGTTT ATGCT GATGAAGATGAC	180		GTCTTCaTGATG CTACC AAACTGGAAAAG	1
ACTTTACAGTTT ATGCT GATGAAGATGcC	1			
ACTTTACAGTTT ATGCTATGCT GATGAAGATGAC	507			
gCTTTACAGTTT ATGCTATGCT GATGAAGATGAC	1		chr9:80584200-80584229	
ACTTTACgGTTT ATGCTATGCT GATGAAGATGAC	1		CTTTTGCAGTCT GTAGGT GTTGAGGTTGAC	
ACTTTACAGTTT ATGCTATGCT GATGAtGATGAC	1			
ACgTTACAGTTT ATGCTATGCT GATGAAGATGAC	1		CTTTTGCAGTCT GTAGGT GTTGAGGTTGAC	125
ACTTTACAGTTT ATGCT GATGAAGATGAC	140		CTTTTGCAGTCT GTAGGT GTTGAGGTTGAC	1
ACTTTACAGTTT ATGCT GATGAAGATGtC	1		CTTTTGCAGTCT GTAGGT GTTGAGGTTGAC	1
WBSCR17				
GTTATCCTCAGC AAACTA AAACTGGAACAG				
GTTATCCTCAGC AAACTAACTA AAACTGGAACAG	128			
GTTATCCTCAGC AAACTA AAACTGGAACAG	118			
GTTATCCTCAGC AAACTA AAACTGGgACAG	1			
GTTATCCTCAGC AAACTA AAACTGGAcCAG	1			
GTTATaCTCAGC AAACTA AAACTGGAACAG	1			
GTTATCCTCAGC AAACTA AAACTGGAACAG	116			
aTTATCCTCAGC AAACTA AAACTGGAACAG	1			
GTTATCCTtAGC AAACTA AAACTGGAACAG	1			
GTTATCCTCAGC AAACTA AAACTGGAACAG	118			
GaTATCCTCAGC AAACTA AAACTGGAACAG	1			

Table 4.7: **Potential VF2468 genomic off-target sites**
DNA for 90 out of 97 potential VF2468 genomic target sites were amplified by PCR from cultured K562 cells expressing active VF2468 ZFN or from cells containing empty expression vector. Bolded red sites have significantly enriched indel percentages in the active nuclease sample compared to cells not expressing nuclease. The sequences of the sites are listed as 5' (+) half-site/spacer/(-) half-site 3', therefore the (+) half-site is listed in the reverse sense as it is in the sequence profiles. Indels and totals are not shown for sites that were not tested. *P*-values shown are for the one-sided alternative hypothesis that the indel frequency is greater for active ZFN treated cells than for cells not expressing ZFN.

Table 4.7 (continued):

	mutations								*in vitro* selection stringency				empty vector			active VF2468			
	T	(+)	(-)	gene	build 36 coordinates	(+) half-site	spacer	(-) half-site	4 nM	2 nM	1 nM	0.5 nM	indels	total	mutation frequency	indels	total	mutation frequency	*p*-value
VF2468 1	0	0	0	VEGF-A (promoter)	chr6:43,845,393-43,845,415	AGCAGCGTC	TTCGA	GAGTGAGGA	X	X	X	X	125	147187	0.085%	27057	186785	14%	0
VF2468 2	1	0	1		chr1:168,832,650-168,832,672	AGCAGCGTC	AATAC	GAGTGAaGA	X	X	X	X	0	57855		1	62195	0.0016%	0.16
VF2468 3	1	1	0		chr1:242,574,122-242,574,144	AGCAGCcTC	TGCTT	GAGTGAGGA	X	X	X	X	0	167447		0	147340		
VF2468 4	1	1	0	ZNF683	chr1:26,569,668-26,569,690	AGCAGCGTt	GGGAG	GAGTGAGGA	X	X	X	X	0	111340		0	109365		
VF2468 5	2	0	2	GSG1L	chr16:27,853,984-27,854,006	AGCAGCGTC	AAAAA	cAGTGAGcA	X	X	X	X	0	80047		0	69080		
VF2468 6	2	0	2	C9orf98	chr9:134,636,934-134,636,956	AGCAGCGTC	GTGTG	GtGTGAGGt	X	X	X	X							
VF2468 7	2	0	2	EFHD1	chr2:233,205,384-233,205,407	AGCAGCGTC	GTTCTC	aAGTGgGGA	X	X	X	X	0	202694		0	204809		
VF2468 8	2	0	2		chr20:30,234,845-30,234,868	AGCAGCGTC	TAGGCA	GAGgGAaGA	X	X	X	X	0	160769		1	158886	0.00063%	0.16
VF2468 9	2	0	2	KIAA0841 (exon-intron)	chr19:40,800,797-40,800,820	AGCAGCGTC	TAGGGG	GAGgGAGGg	X	X	X	X	1	81164	0.0012%	445	79136	0.56%	0
VF2468 10	2	0	2	CES7	chr16:54,501,918-54,501,941	AGCAGCGTC	TCAAAA	GAGTGtGcA	X	X	X	X	1	168501	0.00059%	0	144701		0.84
VF2468 11	2	0	2	PTK2B	chr8:27,339,955-27,339,978	AGCAGCGTC	TCCCTT	GAGTGAtGg	X	X	X	X	0	179502		56	138649	0.040%	3.6E-14
VF2468 12	2	0	2		chr9:137,315,499-137,315,521	AGCAGCGTC	TGAAA	GAGTGAaaA	X	X	X	X	0	285536		165	254714	0.065%	0
VF2468 13	2	1	1		chr20:7,985,471-7,985,493	AGaaAGCGTC	ATCGA	GAGTGAGGt	X	X	X	X	0	166914		0	148547		
VF2468 14	2	1	1		chrY:8,461,018-8,461,041	AGCAaCGTC	AGATAG	GgGTGAGGA	X	X	X	X							
VF2468 15	2	1	1		chr1:53,720,668-53,720,690	AGCAaCGTC	ATATT	cAGTGAGGA	X	X	X	X	0	329599		145	290700	0.050%	0
VF2468 16	2	1	1		chrX:122,132,519-122,132,541	AGCAaCGTC	GTAGT	GAtTGAGGA	X	X	X	X	0	157651		0	136373		
VF2468 17	2	1	1	F4HA1	chr10:74,506,346-74,506,368	AGCAcCGTC	TTTTC	tAGTGAGGA	X	X	X	X							
VF2468 18	2	1	1	DFKZp686L07201	chrX:56,830,910-56,830,933	AGCAGaGTC	AGACTT	GAGTGAGGt	X	X	X	X	0	13660		13	12386	0.10%	0.00015
VF2468 19	2	1	1	TTC4	chr1:54,881,895-54,881,917	AGCAGaGTC	TCTGA	GAGTGAGGc	X	X	X	X	0	175808		163	191327	0.085%	0
VF2468 20	2	1	1		chr1:175,647,668-175,647,690	AGCAGCaTC	AGTGA	GAGTGAGGc	X	X	X	X	1	286818	0.00035%	3	343497	0.00087%	0.20
VF2468 21	2	1	1		chr1:50,490,333-50,490,356	AGCAGCcTC	TCCAAA	GAGTGAGGc	X	X	X	X	0	168032		0	183289		
VF2468 22	2	1	1		chr4:128,244,847-128,244,870	AGCAGCcTC	TGCATC	GAGTGAGGt	X	X	X	X	0	86347		0	87663		
VF2468 23	2	1	1		chr13:27,399,187-27,399,210	AGCAGCGaC	GCCTGG	GAGTGAGGt	X	X	X	X	0	23198		394	34455	1.1%	0
VF2468 24	2	1	1		chr16:62,603,303-62,603,326	AGCAGCGTa	TCACAT	GAGTGAGGg	X	X	X	X	0	57001		283	63841	0.44%	0
VF2468 25	2	1	1		chr11:69,063,501-69,063,523	AGCAGCGTg	CCCAA	GAGTGAGGc	X	X	X	X	0	181022		0	221989		
VF2468 26	2	1	1	TNR	chr1:173,885,442-173,885,465	AGCAGCttC	AGGGGA	GtGTGAGGA	X	X	X	X	0	132693		0	139071		
VF2468 27	2	1	1	PTPRM	chr18:8,320,310-8,320,332	AGCAGCttC	CTTTT	GAGTGAGaA	X	X	X	X	0	73084		0	100249		
VF2468 28	2	1	1		chr12:25,724,566-25,724,589	AGCAGCttC	TCCTGG	GAGTGAGGg	X	X	X	X	0	32323		1115	353441	0.32%	0
VF2468 29	2	1	1		chr13:82,039,140-82,039,163	AGCAGgGTC	AGGGCT	GAGTGAGGc	X	X	X	X	0	15624		439	168937	0.26%	0
VF2468 30	2	1	1		chr3:131,201,895-131,201,918	AGCAGtGTC	AGGCTG	GtGTGAGGA	X	X	X	X	0	77427		1960	92791	2.1%	0
VF2468 31	2	1	1		chr3:75,709,387-75,709,410	AGCAGtGTC	AGGCTG	GtGTGAGGA	X	X	X	X	0	34408		114	33070	0.34%	0
VF2468 32	2	1	1		chr11:3,556,299-3,556,322	AGCAGtGTC	AGGCTG	GtGTGAGGA	X	X	X	X	0	19630		19	17409	0.11%	6.5E-06
VF2468 33	2	1	1		chr3:126,970,762-126,970,785	AGCAGtGTC	AGGCTG	GtGTGAGGA	X	X	X	X	0	89679		2573	90901	2.8%	0
VF2468 34	2	1	1		chr11:71,030,884-71,030,907	AGCAGtGTC	AGGCTG	GtGTGAGGA	X	X	X	X	0	112446		231	150275	0.15%	0
VF2468 35	2	1	1	SBF2/U80769	chr11:9,884,211-9,884,234	AGCAGtGTC	CTAAGG	GgGTGAGGA	X	X	X	X	0	418083		699	532165	0.13%	0
VF2468 36	2	1	1	KRI1 (coding)	chr19:10,534,492-10,534,515	AGCAtCGTC	ATCAGA	cAGTGAGGA	X	X	X	X	0	141739		0	139368		
VF2468 37	2	1	1		chr6:112,421,476-112,421,499	AGCAtCGTC	TGAAGT	GAGTGAGGc	X	X	X	X	0	153897		1174	178559	0.66%	0
VF2468 38	2	1	1	MICAL3/KIAA1364	chr22:16,718,914-16,718,937	AGCAtCGTC	TTCTGT	GAGTGAGtA	X	X	X	X	0	267705		175	283796	0.062%	0
VF2468 39	2	1	1	MUC16 (exon-intron)	chr19:8,894,218-8,894,241	AGgAGCGTC	TCACCT	GAGTGAGGc	X	X	X	X	0	212038		0	219913		
VF2468 40	2	2	0		chr8:6,638,000-6,638,023	AaCAGCtTC	ATCTCG	GAGTGAGGA	X	X	X	X	0	132803		0	147070		
VF2468 41	2	2	0	PREX1	chr20:46,733,644-46,733,667	AaCAGCtTC	TCGGGA	GAGTGAGGA	X	X	X	X	0	204408		0	227091		
VF2468 42	2	2	0	CDH20	chr18:57,303,454-57,303,477	AaCAGCtTC	TCTGAG	GAGTGAGGA	X	X	X	X	1	313747	0.00032%	1	403382	0.00025%	0.57
VF2468 43	2	2	0		chr20:6,213,500-6,213,522	AGCAaaGTC	AAACA	GAGTGAGGA	X	X	X	X	1	154154	0.00065%	0	183644		0.84
VF2468 44	2	2	0		chr5:85,841,308-85,841,331	AGCAaCtTC	TGGAAA	GAGTGAGGA	X	X	X	X							
VF2468 45	2	2	0		chr8:20,481,270-20,481,292	AGCAcCtTC	AATTG	GAGTGAGGA	X	X	X	X	0	250890		0	297104		
VF2468 46	2	2	0		chr5:95,417,045-95,417,068	AGCAGCaTa	ATAGCA	GAGTGAGGA	X	X	X	X	0	274402		1	319493	0.00031%	0.16
VF2468 47	2	2	0	RORA	chr15:59,165,302-59,165,325	AGCAGCaTa	GATATG	GAGTGAGGA	X	X	X	X	0	270263		1	358704	0.00028%	0.16
VF2468 48	2	2	0		chr6:24,504,489-24,504,511	AGCAGCaTa	TCAGG	GAGTGAGGA	X	X	X	X	0	103878		0	176333		
VF2468 49	2	2	0		chr3:31,085,287-31,085,309	AGCAGCGag	AAAGA	GAGTGAGGA	X	X	X	X	0	542052		0	708517		
VF2468 50	2	2	0		chr6:27,579,690-27,579,712	AGCAGCGgt	CTTAG	GAGTGAGGA	X	X	X	X	0	177732		1	212250	0.00047%	0.16
VF2468 51	2	2	0		chr12:113,410,592-113,410,615	AGCAGgGTt	CTTCAA	GAGTGAGGA	X	X	X	X	0	294783		0	302167		
VF2468 52	2	2	0		chr1:11,399,534-11,399,556	AGCtGaGTC	CTAAA	GAGTGAGGA	X	X	X	X	0	482765		1	402831	0.00025%	0.16
VF2468 53	2	2	0	MCTP1	chr5:94,590,016-94,590,038	AGCtGaGTC	TTAAG	GAGTGAGGA	X	X	X	X	0	183510		1	202083	0.00049%	0.16
VF2468 54	2	2	0		chr1:13,394,902-13,394,924	AGCtGgGTC	ATGAG	GAGTGAGGA	X	X	X	X	0	88944		0	105879		
VF2468 55	2	2	0	PRAMEF20	chr1:13,615,741-13,615,763	AGCtGgGTC	ATGAG	GAGTGAGGA	X	X	X	X							
VF2468 56	2	2	0		chr20:59,154,784-59,154,806	AGCtGtGTC	CACAG	GAGTGAGGA	X	X	X	X	0	360710		0	351215		
VF2468 57	2	2	0		chr14:100,903,675-100,903,697	AGCtGtGTC	TTGGA	GAGTGAGGA	X	X	X	X	0	140671		0	157922		
VF2468 58	2	2	0		chrX:141,701,170-141,701,192	AGtAGCGgC	AAATT	GAGTGAGGA	X	X	X	X	0	196624		0	209781		
VF2468 59	2	2	0	GTF3C1	chr16:27,452,953-27,452,975	AGtgGCGTC	CCAGT	GAGTGAGGA	X	X	X	X	0	223714		0	246196		
VF2468 60	2	2	0	DNMBP/AK089111	chr10:101,688,961-101,688,983	gGCAGaGTC	CTAGA	GAGTGAGGA	X	X	X	X	0	302495		0	383303		
VF2468 61	2	2	0		chr6:137,852,455-137,852,478	tcCAGCGTC	CTCCCA	GAGTGAGGA	X	X	X	X	0	84153		0	113996		
VF2468 62	2	2	0	SARDH	chr9:135,592,239-135,592,262	tGCAGCGgC	GTAGGG	GAGTGAGGA	X	X	X	X	0	191187		138	212085	0.065%	0
VF2468 63	2	2	0		chr7:19,683,400-19,683,423	tGCAGCGTt	AAAATA	GAGTGAGGA	X	X	X	X	1	372808	0.00027%	2	438355	0.00046%	0.33
VF2468 64	2	2	0	ZNF628	chr19:60,683,246-60,683,269	tGCAGgGTC	GGGCAG	GAGTGAGGA	X	X	X	X	0	167551		0	185442		
VF2468 65	2	2	0		chr3:130,430,426-130,430,448	tGCAGtGTC	CACAA	GAGTGAGGA	X	X	X	X							
VF2468 66	3	1	2	GBF1	chr10:104,073,989-104,074,012	AGCAaCGTC	CATAGT	GtGTGAGaA	X	X	X	X	4	54554	0.00073%	3758	587393	0.64%	0
VF2468 67	3	1	2		chr14:96,561,728-96,561,751	AGCAaCGTC	TAACCC	GAGTGttGA	X	X	X	X	0	171147		0	203860		
VF2468 68	3	1	2	PDE9A	chr21:42,982,083-42,982,105	AGCAcCGTC	CCCCT	cAGTGAGGc	X	X	X	X	0	89976		391	137236	0.28%	0
VF2468 69	3	1	2	MTX2	chr2: 176,842,448-176,842,471	AGCAGCGgC	GGCTG	cAGTGAGGc	X	X	X	X	0	38342		153	50273	0.30%	0
VF2468 70	3	1	2		chr6:104,071,040-104,071,063	AGCAGCGgC	TTAAGG	CgGTGAGGt	X	X	X	X	0	252020		0	277262		
VF2468 71	3	1	2		chr3:32,220,862-32,220,885	AGCAGtGTC	TAAAAG	GAGTGAGat	X	X	X	X	0	178243		0	225921		
VF2468 72	3	2	1		chr2:11,429,195-11,429,218	AGaAaCGTC	GTGGAG	GAGTGAGGg	X	X	X	X	0	138844		94	221676	0.042%	0
VF2468 73	3	2	1	OPN5	chr6:47,891,415-47,891,438	AGCAaaGTC	TGTACT	GAGTGAGGg	X	X	X	X	0	182565		2808	212729	1.3%	0
VF2468 74	3	2	1		chr2:195,362,417-195,362,439	AGCAaCaTC	ATCTT	GAGTGAGGg	X	X	X	X	0	103739		1	130605	0.00077%	0.16
VF2468 75	3	2	1	MIPOL1	chr14:36,952,701-36,952,724	AGCAaCaTC	TGGTTG	GAGTGAGGg	X	X	X	X	1	300572	0.00033%	0	355283		0.84
VF2468 76	3	2	1		chr4:138,603,959-138,603,981	AGCAaCtTC	ATCTT	GAGTGAGGg	X	X	X	X	0	185773		0	185094		
VF2468 77	3	2	1	LRCH1	chr13:46,079,921-46,079,943	AGCAaCtTC	CTGGC	GAGTGAGGg	X	X	X	X	1	131239	0.00076%	0	133319		0.84
VF2468 78	3	2	1	KIAA0999	chr11:116,292,384-116,292,406	AGCAatGTC	AAAAA	GAGTGAGGc	X	X	X	X	0	212883		243	157434	0.15%	0
VF2468 79	3	2	1		chr12:13,353,629-13,353,651	AGCAcCGTg	GCTTC	GAGTGAGGc	X	X	X	X	0	112164		34	116695	0.029%	2.7E-09
VF2468 80	3	2	1	PLXNA4 (coding)	chr7:131,503,708-131,503,730	AGCAcgGTC	ATGAT	GAGTGAGGc	X	X	X	X	0	195577		77	188335	0.041%	0
VF2468 81	3	2	1	TSPEAR	chr21:44,817,259-44,817,282	AGCAGCagC	CCACAG	GAGTGAGGg	X	X	X	X	1	151845	0.00066%	0	144107		0.84
VF2468 82	3	2	1		chr18:74,634,790-74,634,812	AGCAGCagC	TAGGG	GAGTGAGGg	X	X	X	X							
VF2468 83	3	2	1		chr10:33,904,306-33,904,328	AGCAGCtcC	TCTCC	GAGTGAGGt	X	X	X	X	0	240952		1	233954	0.00043%	0.16
VF2468 84	3	2	1		chr6:170,226,156-170,226,178	AGCAGgGTg	GCGTG	GAGTGAGGc	X	X	X	X	0	191108		0	167663		
VF2468 85	3	2	1		chr3:118,594,878-118,594,901	AGCAtaGTC	TAGGCC	GAGTGAGGc	X	X	X	X	0	40564		209	355997	0.059%	0
VF2468 86	3	2	1	HRASLS	chr3:194,452,125-194,452,147	AGCAtgGTC	CCAAG	GAGTGAGGg	X	X	X	X	0	212942		393	247078	0.16%	0
VF2468 87	3	2	1		chr17:19,434,508-19,434,531	AGCAttGTC	TCATGT	GAGTGAGGt	X	X	X	X	0	4171		18	3024	0.60%	1.0E-05
VF2468 88	3	2	1		chr8:125,982,579-125,982,601	AGCAttGTC	TCCTG	GAGTGAGGg	X	X	X	X	0	115967		59	115258	0.051%	7.8E-15
VF2468 89	3	2	1	UHRF1BP1L	chr12:99,011,715-99,011,737	AGtAGCGTt	TTTAG	GAGTGAGGt	X	X	X	X	1	171873	0.00058%	0	207336		0.84
VF2468 90	3	2	1		chr1:14,762,405-14,762,427	AtCAGaGTC	TCTGG	GAGTGAGGc	X	X	X	X	2	193447	0.0010%	1	196665	0.00051%	0.72
VF2468 91	3	2	1	LAMB3 (promoter)	chr1:207,894,359-207,894,382	AtCAGtGTC	CCTCAG	GAGTGAGGc	X	X	X	X	0	107549		0	109933		
VF2468 92	3	2	1	BRUNOL4	chr18:33,160,009-33,160,032	gGCAGaGTC	AGGGCT	GAGTGAGGc	X	X	X	X	0	71298		0	77229		
VF2468 93	3	2	1		chr16:84,004,297-84,004,320	gGCAGCGgC	CGCTGT	GAGTGAGGt	X	X	X	X	0	99279		0	121284		
VF2468 94	3	2	1	DFKZp586E1619/BRD1	chr22:48,558,064-48,558,086	tGCAGCttC	ATGGT	GAGTGAGGc	X	X	X	X	0	152251		0	206428		
VF2468 95	3	3	0		chr7:22,054,784-22,054,807	AGCAtaGTt	ACCTGG	GAGTGAGGA	X	X	X	X	0	91338		0	134004		
VF2468 96	3	3	0		chr14:25,876,126-25,876,149	AGtAaaGTC	TAAGTA	GAGTGAGGA	X	X	X	X	0	245402		0	345728		
VF2468 97	4	3	1	CNNM2	chr10:104,716,593-104,716,615	tGCAGtcTC	CTTGG	GAGTGAGGt	X	X	X	X	0	76762		0	92742		

quencies of 1:300 to 1:5,300. We also expressed VF2468 in cultured K562 cells and performed the same analysis as above for 90 of the most highly cleaved sites identified by *in vitro* selection. Of the 90 VF2468 sites analyzed, 32 had indels consistent with ZFN-mediated targeting in K562 cells (Table 4.7). We did not obtain site-specific PCR amplification products for three CCR5-224 sites and seven VF2468 sites, and therefore could not analyze the occurrence of non-homologous end-joining at those loci. Taken together, these observations indicate that off-target sequences identified through the *in vitro* selection method include DNA sequences that can be cleaved by ZFNs in human cells.

Discussion

Using the method presented here we identified hundreds of thousands of sequences that can be cleaved by two active, dimeric ZFNs, including many that are present and can be cut in the genome of human cells.

One newly identified cleavage site for CCR5-224 is in the promoter of *BTBD10*. When downregulated, *BTBD10* has been associated with malignancy (Chen et al., 2004) and with pancreatic beta cell apoptosis (Wang et al., 2011). When upregulated, *BTBD10* has been shown to enhance neuronal cell growth (Nawa et al., 2008) and pancreatic beta cell proliferation through phosphorylation of Akt family proteins (Nawa et al., 2008, Wang et al., 2011). This potentially important off-target cleavage site, as well as seven others we observed in cells, had not been identified in a recent study6 that used *in vitro* monomer-binding data to predict potential CCR5-224 substrates. Although our selection results increased the number of known CCR5-224 off-target sites, the number of sequence reads obtained per selection in our study (approximately one million) is likely insufficient to cover all cleaved sequences in the

postselection libraries. It is therefore possible that additional off-target cleavage sites for CCR5-224 and VF2468 could be identified in the human genome as sequencing capabilities improve. It is also possible that the datasets generated by this method could be used to develop computational models to predict additional ZFN cleavage sites *in vitro* and in cells.

We have previously shown that ZFNs that can cleave at sites in one cell line may not necessarily function in a different cell line4 most likely because of local differences in chromatin structure. Therefore, it is likely that a different subset of the *in vitro*–cleavable off-target sites would be modified by CCR5-224 or VF2468 when expressed in different cell lines. Likewise, cell-line choice may influence purely cellular studies of endonuclease specificity, such as a recent study of homing endonuclease off-target cleavage that relied on adeno-associated virus integration into homing endonuclease–created double-strand breaks in cells (Petek et al., 2010). Whereas our *in vitro* method does not account for some features of cellular DNA, it provides general, cell type–independent information about endonuclease specificity and off-target sites that can inform subsequent studies performed in cell types of interest.

Although both ZFNs we analyzed had been engineered to a unique sequence in the human genome, both cleave several off-target sites in cells. This finding is particularly surprising for the four-finger CCR5-224 pair given that its theoretical specificity is 4,096-fold better than that of the three-finger VF2468 pair (CCR5-224 should recognize a 24 bp site that is 6 bp longer than the 18 bp VF2468 site). Examination of the CCR5-224 and VF2468 cleavage profiles and mutational tolerances of sequences with three or fewer mutations suggests different strategies may be required to engineer variants of these ZFNs with reduced off-target cleavage activities. The four-finger CCR5-224 showed a more diffuse range of positions with relaxed specificity and a higher tolerance of mutant sequences with three or fewer mutations than the three-finger

VF2468 ZFN. For VF2468, re-optimization of only a subset of fingers may enable a substantial reduction in undesired cleavage events. For CCR5-224, in contrast, a more extensive re-optimization of many or all fingers may be required to eliminate off-target cleavage events. Analysis of more three-finger and four-finger ZFNs will be required to determine whether these patterns of off-target cleavage activities are a general property of these respective frameworks. Not all four- and three-finger ZFNs will necessarily be as specific as the two ZFNs tested in this study. Both CCR5-224 and VF2468 had been engineered using methods designed to optimize the binding activity of the ZFNs. Previous work has shown that for both three-finger and four-finger ZFNs, the specific methodology used to engineer the ZFN pair can have a tremendous impact on the quality and specificity of nucleases (Hurt et al., 2003, Urnov et al., 2005, Meng et al., 2007, Ramirez et al., 2008). Therefore, it will be interesting and important to use a method such as the one described here to determine and compare the specificities of additional three-finger and four-finger ZFNs generated using various strategies.

Our findings have important implications for the design and application of ZFNs with increased specificity. Half or more of all potential substrates with one or two site mutations could be cleaved by ZFNs, suggesting that binding affinity between ZFN and DNA substrate is sufficiently high for cleavage to occur even with suboptimal molecular interactions at mutant positions. We also observed that ZFNs presented with sites that have mutations in one half-site exhibited higher mutational tolerance at other positions in the mutated half-site and lower tolerance at positions in the other half-site. These results suggest that to meet a minimum affinity threshold for cleavage, a shortage of binding energy from a half-site containing an off-target base pair must be energetically compensated by excess zinc finger: DNA binding energy in the other half-site, which demands increased sequence recognition stringency at the non-mutated half-site

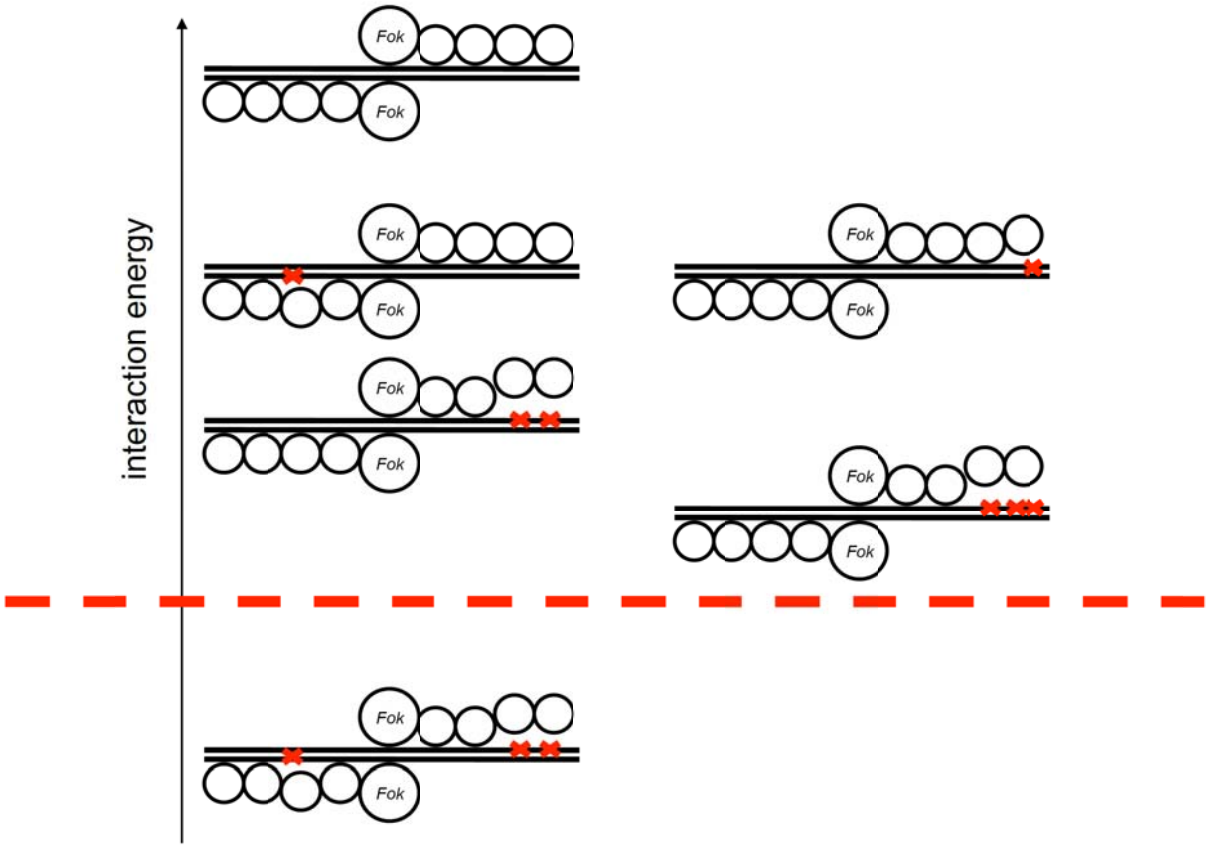


Figure 4.17: Model for ZFN tolerance of off-target sequences.
Our results suggest that some ZFNs recognize their intended target sites (top, black DNA strands with no red Xs) with more binding energy than is required for cleavage under a given set of conditions (red dotted line). Sequences with one or two mutations (one or two red Xs) are generally tolerated since they do not decrease the ZFN:DNA binding energy below the threshold necessary for cleavage. Some sequences with additional mutations can still be cleaved if the additional mutations occur in regions of the zinc-finger binding interface that have already been disrupted (three red Xs above the red dotted line), as long as optimal interactions present at other locations in the ZFN:DNA binding interface maintain binding energies above threshold values. Additional mutations that disrupt key interactions at other locations in the ZFN:DNA interface, however, result in binding energies that fall short of the cleavage threshold.

(Figure 4.17). Conversely, the relaxed stringency at other positions in mutated half-sites can be explained by the decreased contribution of that mutant half-site to overall ZFN binding energy.

This model also explains our observation that sites with suboptimal spacer lengths, which presumably were bound less favorably by ZFNs, were recognized with higher stringency than sites with optimal spacer lengths. *In vitro* spacer preferences do not necessarily reflect spacer

preferences in cells (Bibikova et al., 2001, Handel et al., 2009), but our results suggest that the dimeric FokI cleavage domain can influence ZFN target-site recognition. Consistent with this model, differences in the frequency of off-target events in zebrafish of two ZFNs with identical zinc-finger domains but different FokI domain variants have been reported (Gupta et al., 2011).

Collectively, our findings suggest that (i) ZFN specificity can be increased by avoiding the design of ZFNs with excess DNA binding energy; (ii) off-target cleavage can be minimized by designing ZFNs to target sites that do not have similar sites in the genome within three mutations; and (iii) ZFNs should be used at the lowest concentrations necessary to cleave the target sequence to the desired extent. Although in this study we focused on ZFNs, our method should be applicable to all sequence-specific endonucleases that cleave DNA *in vitro*, including engineered homing endonucleases and engineered transcription activator–like effector nucleases. This approach can provide important information when choosing target sites in genomes for sequence-specific endonucleases and when engineering these enzymes, especially for therapeutic applications.

Methods

Oligonucleotides and sequences

All oligonucleotides were purchased from Integrated DNA Technologies or Invitrogen and are listed in Table 4.8. Primers with degenerate positions were synthesized by Integrated DNA Technologies using hand-mixed phosphoramidites containing 79% of the indicated base and 7% of each of the other standard DNA bases.

Library construction

Libraries of target sites were incorporated into double-stranded DNA by PCR with Taq DNA Polymerase (NEB) on a pUC19 starting template with primers "N5-PvuI" and "CCR5-224-N4," "CCR5-224-N5," "CCR5-224-N6," "CCR5-224-N7," "VF2468-N4," "VF2468-N5," "VF2468-N6," or "VF2468-N7," yielding an approximately 545 bp product with a *Pvu*I restriction site adjacent to the library sequence, and purified with the Qiagen PCR Purification Kit. Library-encoding oligonucleotides were of the form 5' backbone-*Pvu*I site-NNNNNN-partially randomized half-site–N_{4-7}–partially randomized half site-N-backbone 3'. The purified oligonucleotide mixture (approximately 10 µg) was blunted and phosphorylated with a mixture of 50 units of T4 Polynucleotide Kinase and 15 units of T4 DNA polymerase (NEBNext End Repair Enzyme Mix, NEB) in 1x NEBNext End Repair Reaction Buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM ATP, 0.4 mM dATP, 0.4 mM dCTP, 0.4 mM dGTP, 0.4 mM dTTP, pH 7.5) for 1.5 hours at room temperature. The blunt-ended and phosphorylated DNA was purified with the Qiagen PCR Purification Kit according to the manufacturer's protocol, diluted to 10 ng/µL in NEB T4 DNA Ligase Buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM ATP, pH 7.5) and circularized by ligation with 200 units of T4 DNA ligase (NEB) for 15.5 hours at room temperature. Circular monomers were gel purified on 1% TAE-Agarose gels. 70 ng of circular monomer was used as a substrate for rolling-circle amplification at 30 °C for 20 hours in a 100 µL reaction using the Illustra TempliPhi 100 Amplification Kit (GE Healthcare). Reactions were stopped by incubation at 65 °C for 10 minutes. Target site libraries were quantified with the Quant-iT PicoGreen dsDNA Reagent (Invitrogen). Libraries with N_4, N_5, N_6, and N_7 spacer sequences between partially randomized half-sites were pooled in equimolar concentrations for both CCR5-224 and VF2468.

Table 4.8: Oligonucleotides used in this study

Oligonucleotides "[ZFN] [#] fwd/rev" were ordered from Invitrogen. All other oligonucleotides were ordered from Integrated DNA Technologies. 'N' refers to machine mixed incorporation of 'A','C','G', or 'T.' An asterisk indicates that the preceding nucleotide was incorporated as a mixture containing 79 mol % of that nucleotide and 7 mol % each of the other canonical nucleotides. "/5Phos/" denotes a 5' phosphate group installed during synthesis.

Oligonucleotide name	Oligonucleotide Sequence (5'->3')		
N5-PvuI	NNNNNCGATCGTTGGGAACCGGA		
CCR5-224-N4	NG*T*C*A*T*C*C*T*C*A*T*C*NNNNA*A*A*C*T*G*C*A*A*A*A*G*NCAGTGGAACGAA		
CCR5-224-N5	NG*T*C*A*T*C*C*T*C*A*T*C*NNNNNA*A*A*C*T*G*C*A*A*A*A*G*NCAGTGGAACGAAAACTCACG		
CCR5-224-N6	NG*T*C*A*T*C*C*T*C*A*T*C*NNNNNNA*A*A*C*T*G*C*A*A*A*A*G*NCAGTGGAACGAAAACTCACG		
CCR5-224-N7	NG*T*C*A*T*C*C*T*C*A*T*C*NNNNNNNA*A*A*C*T*G*C*A*A*A*A*G*NCAGTGGAACGAAAACTCACG		
VF2468-N4	NA*G*C*A*G*C*G*T*C*NNNNG*A*G*T*G*A*G*G*A*NCAGTGGAACGAAAACTCACG		
VF2468-N5	NA*G*C*A*G*C*G*T*C*NNNNNG*A*G*T*G*A*G*G*A*NCAGTGGAACGAAAACTCACG		
VF2468-N6	NA*G*C*A*G*C*G*T*C*NNNNNNG*A*G*T*G*A*G*G*A*NCAGTGGAACGAAAACTCACG		
VF2468-N7	NA*G*C*A*G*C*G*T*C*NNNNNNNG*A*G*T*G*A*G*G*A*NCAGTGGAACGAAAACTCACG		
test fwd	GCGACACGGAAATGTTGAATACTCAT		
test rev	CAGCGAGTCAGTGAGCGA		
adapter1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTT		
adapter1(AAT)	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAATT		
adapter1(ATA)	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATAT		
adapter1(TAA)	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAAT		
adapter1(CAC)	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCACT		
adapter2	/5Phos/AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG		
adapter2(AAT)	/5Phos/ATTAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG		
adapter2(ATA)	/5Phos/TATAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG		
adapter2(TAA)	/5Phos/TTAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG		
adapter2(CAC)	/5Phos/GTGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG		
PE1	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC		
PE2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT		
CCR5-224 1 fwd	ATACATCGGAGCCCTGCCAA	CCR5-224 1 rev	GGAAAAACAGGTCAGAGATGGC
CCR5-224 2 fwd	TCCTGCCTCCGCTCTACTCG	CCR5-224 2 rev	ACCCCAAAGGTGACCGTCCT
CCR5-224 3 fwd	TCCCACGTTTTCCCCTTGAC	CCR5-224 3 rev	GTCCCTCACGACGACCGACT
CCR5-224 4 fwd	GCACTGCCCCAGAAATATTGGT	CCR5-224 4 rev	TGGTTTGTTGGGGGATCAGG
CCR5-224 5 fwd	ATGCCACCCCTGCCAGATAA	CCR5-224 5 rev	GCCTACCTCAATGCAGGCAAA
CCR5-224 6 fwd	TCTGCTTCTGCCCTTCTGGA	CCR5-224 6 rev	GGAGGATCGCCAAGACCTGA
CCR5-224 7 fwd	CCCCAGTGCTTAACATAGTTCT	CCR5-224 7 rev	ACTCCCAGACAAACCCCGCT
CCR5-224 8 fwd	GGCACCAGAACTTACTCACTGC	CCR5-224 8 rev	TGTGAAGGCCCAAAACCCTG
CCR5-224 9 fwd	GTTTTGGGGGTCATGGCAAA	CCR5-224 9 rev	TGGGCAGCCCTAGGTCCTTT
CCR5-224 10 fwd	TTTCCCTGGTGATGCACTCCT	CCR5-224 10 rev	TGATGAGTAACTTGGGCGAAAA
CCR5-224 11 fwd	TTGGGGGAATGAGATTGGGA	CCR5-224 11 rev	GGAAAATCCAGCAAGGTGAAA
CCR5-224 13 fwd	CCTTCCCATGGTCACAGAGG	CCR5-224 13 rev	CAACTCTCTAACAGCAAAGTGGCA
CCR5-224 14 fwd	TCCTCCCGTTGAGGAAGCAC	CCR5-224 14 rev	GCCTCAAAAGCATAAACAGCA
CCR5-224 15 fwd	CAGACCGCTGCTGCTGAGAC	CCR5-224 15 rev	AGGGCGGACTCATTGCTTTG
CCR5-224 16 fwd	TGGGTTCCTCGGGTTCTCTG	CCR5-224 16 rev	GAAACCAGAAGTTCACAACAATGCTT
CCR5-224 17 fwd	AGGCATAAGCCACTGCACCC	CCR5-224 17 rev	TGGCAATGCCTAATCAGACCA
CCR5-224 18 fwd	GAGGATATTTTATTGCTGGCTC	CCR5-224 18 rev	GAGTTTGGGGAAAAGCCACTT
CCR5-224 20 fwd	GCTGAGGCCCACCTTTCCTT	CCR5-224 20 rev	TGCTCTGCCAACTGTGAGGG
CCR5-224 21 fwd	TGTTTTGGGTGCATGTGGGT	CCR5-224 21 rev	TCCAGGGAGTGAGGTGAAGACA
CCR5-224 22 fwd	CTGGGTCAGCTGGGCCATAC	CCR5-224 22 rev	TCACATCTCCGCCTCACGAT
CCR5-224 23 fwd	CCAGCCTTGGAAAAATGGACA	CCR5-224 23 rev	CTGACACAGTGGCCAGCAGC
CCR5-224 24 fwd	CATGGATGTAATGGGTTGTATC	CCR5-224 24 rev	GAGGGCAGAAGGGGGTGAGT

Table 4.8 (continued):

CCR5-224 25 fwd	AGGATGCATTGTCCCCCAGA	CCR5-224 25 rev	TGGAGTGACATGTATGAAGCCA
CCR5-224 26 fwd	CGTTGGCTTGCAGAAGGGAC	CCR5-224 26 rev	TGAACCCCGGATTTTTCAACC
CCR5-224 27 fwd	TGACCCAACTAAGTCTGTGACCC	CCR5-224 27 rev	TTGGGAAAGCTTTGATGCTGG
CCR5-224 28 fwd	TGGGTTGTGTTTTTGACTGACAGA	CCR5-224 28 rev	CCCTAGGGGTCACTGGAGCA
CCR5-224 29 fwd	CACCCCCATGCAGGAAAATG	CCR5-224 29 rev	TTGGCTGCTGGCATTTGGTA
CCR5-224 30 fwd	GGCCATTGGTTCTGGAGGAA	CCR5-224 30 rev	TCCGTTGCTTCATCCTTCCAA
CCR5-224 31 fwd	AGTCAGCAATGCCCCAGAGC	CCR5-224 31 rev	TGGAGAGGGTTTACTTTCCCAGA
CCR5-224 32 fwd	CCTGGGAGGGTGACTAGTTGGA	CCR5-224 32 rev	GCTCAGGGCCTGGCTTACAG
CCR5-224 33 fwd	TGGCAATTAGGATGTGCCAG	CCR5-224 33 rev	TCCACTCACAAATTTACCTTTCCAC
CCR5-224 34 fwd	TGCCCCACATCTTCACCAGA	CCR5-224 34 rev	CCGCATAAAGGAGGTGTCGG
CCR5-224 36 fwd	GTTGCATCTGCGGTCTTCCA	CCR5-224 36 rev	GGAGAGTCTTCCGCCTGTGTT
CCR5-224 37 fwd	TAGTGGCCCCAACATGCAAA	CCR5-224 37 rev	GCACATATCATGCACTGTGACTGTAA
VF2468 1 fwd	CCTTTCCAAAGCCCATTCCC	VF2468 1 rev	CAACCCCACACGCACACAC
VF2468 2 fwd	TTCACTGCCTTCAGGCCTCC	VF2468 2 rev	AATGGCCAGAAAATTCCCAAA
VF2468 3 fwd	CACAGGGACCCAGGACTGCT	VF2468 3 rev	TGACTGGAACCGTGCAGCAT
VF2468 4 fwd	GCACCAGGCTTCTCTGCCAT	VF2468 4 rev	TCGGGGGTCCATGGTATTTG
VF2468 5 fwd	CCAAGGCGAGGACATTGAGG	VF2468 5 rev	CCCCAAGTCAGACCCTGCAT
VF2468 7 fwd	ACCATAGTCCAGCGGGGTCA	VF2468 7 rev	TTCTCCCCAAGGAAGGCTGA
VF2468 8 fwd	AGAAAGGGTGGTCGGGGAAG	VF2468 8 rev	GCCACCATGCCCAGTCTACA
VF2468 9 fwd	TTCCCATGGGGTCTCAGCTC	VF2468 9 rev	ATGGCCTTCCCCAACTGTGA
VF2468 10 fwd	CAGCAAGGATGCCCTTCACC	VF2468 10 rev	CGTTGTGATTGAGGAGACGAGG
VF2468 11 fwd	GGCTTGAGCTGGAAGGACCA	VF2468 11 rev	TGGAGCAACTGAACATCTTGGG
VF2468 12 fwd	AACCGAGTTTGCACCGTCGT	VF2468 12 rev	CATAACCACCAGGACATCCGC
VF2468 13 fwd	TATCCTCCCCTTTCCCCTGA	VF2468 13 rev	TGTTGCCAGAAGTATCAGGTCCC
VF2468 15 fwd	AGAACCCGGAATCCCTTTGC	VF2468 15 rev	GCAGAGAAGGCAGCAGCACA
VF2468 16 fwd	GGTCTCTGCCATGCCCAACT	VF2468 16 rev	TGGAGGAAGCAGGAAAGGCAT
VF2468 18 fwd	CCCCTTGGGATCCTTGTCCT	VF2468 18 rev	TCAACAGGCAGCTACAGGGC
VF2468 19 fwd	CTAGGCCTGTGGGCTGAGGA	VF2468 19 rev	CAAATGTTGGGGTGTGGGTG
VF2468 20 fwd	TACCTGAAACCCCTGGCCCT	VF2468 20 rev	CAAGCTGGATGTGGATGCAGAG
VF2468 21 fwd	CGGGGGCCTGACATTAGTGA	VF2468 21 rev	GCCTGAAGATGCATTTGCCC
VF2468 22 fwd	TGCATTGGCTCAAGAATTGGG	VF2468 22 rev	TCACACAGTGGTAATGGACAGGAA
VF2468 23 fwd	GCGCTCCCTGTGTTCAGTACC	VF2468 23 rev	GCGCAAGTTCCCCTTTCTGA
VF2468 24 fwd	TGTTTGGGTTATGGGGGCAG	VF2468 24 rev	TCCAGCATCTGCTCCTGGTG
VF2468 25 fwd	AAGGAGACTTCTCAGGCCCCA	VF2468 25 rev	TGAAGGGAAGCCACAGCTCC
VF2468 26 fwd	CTTGGGGGCAGACAGCATCT	VF2468 26 rev	GCCATGGGATGGCAGTTAGG
VF2468 27 fwd	TGGCCTCAAGCAATCCTCCT	VF2468 27 rev	TTCCATGGCAGTGAAGGGTG
VF2468 28 fwd	CCAAAGAGCCTGGAGGAGCA	VF2468 28 rev	CAGAGGGTGTGGTGGTGTCG
VF2468 29 fwd	CCAGCCTGTGAAGCTGGAAGTAA	VF2468 29 rev	CCAGTGGGCTGAGTGGATGA
VF2468 30 fwd	CATCTGAATGCCCATGCTGC	VF2468 30 rev	CCGCCACACCCATTCCTC
VF2468 31 fwd	CCTCAAAGAAACGGCTGCTGA	VF2468 31 rev	GCCGCTCGAAAAGAGGGAAT
VF2468 32 fwd	CGGGCTCTCCTCCTCAAAGA	VF2468 32 rev	GGCCCCTTGAAAAGAGGGAA
VF2468 33 fwd	GGAATCGCATGACCTGAGGC	VF2468 33 rev	CGGGCTCTCCTCCTCAAAGA
VF2468 34 fwd	CCCGCCAGACACATTCCTCT	VF2468 34 rev	CATCTGAATGCCCATGCTGC
VF2468 35 fwd	CCGCACCTTTTTCCTATGTGGT	VF2468 35 rev	TCAGATGTGCTAGGACACAGATGAC
VF2468 36 fwd	GGTACATGGGCCGCACTTTC	VF2468 36 rev	GGACAGCTGGGAATTGGTGG
VF2468 37 fwd	TTACACCTGCTGGCAGGCAA	VF2468 37 rev	GCTGGTGTGAGCAAGAGGCA
VF2468 38 fwd	TGGCCAAGCCTGCCTAACTC	VF2468 38 rev	TGATCAGTTAGCCCTGGGGG
VF2468 39 fwd	CCCCTTCTGCTCCTGCTTCA	VF2468 39 rev	CCTTCCTTGCAGCTCAAACCC
VF2468 40 fwd	TGATTTTCAGCGTGGAGGGC	VF2468 40 rev	ACGGCAAAGCCAGAGCAAAG
VF2468 41 fwd	AAGCTGGCAGCCACTCTTCA	VF2468 41 rev	TCTCAGGGCTTCTGTGTGCG
VF2468 42 fwd	TCGATTCTCCATACACCATCAAT	VF2468 42 rev	GCAACCAACTCCCAACAGGG
VF2468 43 fwd	AGGTCCTGGCATTGTCTGGG	VF2468 43 rev	TGGTTGCCTGTTTCACACCC
VF2468 45 fwd	CTGGGAGGCAGCCAGTCAAG	VF2468 45 rev	GCCCTGTAAGCTGAAGCTGGA
VF2468 46 fwd	CAGGTGTGCATTTTGTTGCCA	VF2468 46 rev	GCCTGCCAGGTATTTCCTGTGT
VF2468 47 fwd	TGGCCCTGGTCATGTGAAAA	VF2468 47 rev	AACTGCAAGTGGCCTCCCAG
VF2468 48 fwd	TTGATAAGGGCGGTGCCACT	VF2468 48 rev	TAGAGGGAGGTGCTTGCCCA
VF2468 49 fwd	CATCCCCTTGACCAACAGGC	VF2468 49 rev	GCTTGGGCACTGATCCTGCT

Table 4.8 (continued):

VF2468 50 fwd	ACTGCCAATGGACCCTCTCG	VF2468 50 rev	GAGTTGCCCAGGTCAGCCAT
VF2468 51 fwd	GGGGAGCTAGAATGGTGGGC	VF2468 51 rev	CAAGGTACACAGCTGCCCAGG
VF2468 52 fwd	CCCATGCTGGTCCTGCTGTT	VF2468 52 rev	GGAGGCTCAGCGGAGAGGAT
VF2468 53 fwd	GGGGTCACCAGGGAAGGTTT	VF2468 53 rev	AGTTGCGGGGAGGTGCTACA
VF2468 54 fwd	TGCCCAGAGACCTTCCAAGC	VF2468 54 rev	TGGCCAAGGCCTCTCTAAGC
VF2468 56 fwd	GCCAATGTGCAATCGAGACG	VF2468 56 rev	TGCATGCCTCTGACTGATGCT
VF2468 57 fwd	TGACTTGAACTGGGTCCCCC	VF2468 57 rev	CTGGGGCTACAGCCCTCCTT
VF2468 58 fwd	CCCAATCCAGACACCACACG	VF2468 58 rev	TGCAGATTTTAGGGGTTGCCA
VF2468 59 fwd	GGTGAGGAAGGATGGGGGTT	VF2468 59 rev	GTAGGCTCTGCCACGCCAGT
VF2468 60 fwd	TGCCCATGTTGTTGCTCCAC	VF2468 60 rev	GACAAGTTAGACCATCCTAGCCCTCA
VF2468 61 fwd	TCACAGCTCCCCTTTCTCGG	VF2468 61 rev	TGTGCCTCCACTGACGCATT
VF2468 62 fwd	CCTAGGCACAGTGGGGGATG	VF2468 62 rev	GGGCTGACACACTGAGGGCT
VF2468 63 fwd	CCATGAGCACAATTGCCAAAA	VF2468 63 rev	TGAGTTATTTCGAAAGAGGAAACAGT
VF2468 64 fwd	CTGCCAAGAACAGGAGGGGA	VF2468 64 rev	AGCCCATCTACCATCCAGCG
VF2468 66 fwd	ATCGGGGCAGGGCTAGAGTC	VF2468 66 rev	CCCCTGGCATTCCCTACACA
VF2468 67 fwd	GCCGTTAGTGCATTTGCCTG	VF2468 67 rev	TCCCTTTCAACCCCTGTAGTGC
VF2468 68 fwd	GTTCCTCCCAGAGTGGGGCT	VF2468 68 rev	ACTGAGGGAGGCAGCACTGG
VF2468 69 fwd	AGGCCTGGCGGTAACCTTG	VF2468 69 rev	AAGCTCCAGCCCTGTACCCC
VF2468 70 fwd	GGGATCCTACAGGATGGGACAA	VF2468 70 rev	CAGCCCAGGACAAGGGTAGC
VF2468 71 fwd	GCCACCAAATGTCCACTGGTT	VF2468 71 rev	TTCCCCAAGCAGTCCAGCTC
VF2468 72 fwd	GCACCAGCCTCTTCGATGGT	VF2468 72 rev	CCTTTGGCAGACTGTGGCCT
VF2468 73 fwd	AATGGGGCAAAAGGCAAGAAA	VF2468 73 rev	CAGACCTCGTGGTGCATGTG
VF2468 74 fwd	TGGCGAGATAGGCTCTGCTACA	VF2468 74 rev	TGGACAGGGAATTACTCAGACCAG
VF2468 75 fwd	TGTGGGCATGAGACCACAGG	VF2468 75 rev	TTTGACTCCCCCGCATTGTT
VF2468 76 fwd	TCCTATTTTCAGATGCACTCGAACC	VF2468 76 rev	GTGCTCACTGAAGCCCACCA
VF2468 77 fwd	GGACCTTCTTGCCCTCATGATTC	VF2468 77 rev	GGGAACTGTGCCTTTGCGTC
VF2468 78 fwd	CCTTGCAAAGGCTTGCCTAAA	VF2468 78 rev	GGCAGGCACCTGTAGTCCCA
VF2468 79 fwd	TGGCTTGCAGAGGAGGTGAG	VF2468 79 rev	CAGGGAAGGGTGTTGGCTTG
VF2468 80 fwd	GCTTCAGCACATCAGTGGCG	VF2468 80 rev	TTCGCCCAGCTCATCAACAA
VF2468 81 fwd	GGTGAGGCCACTGTAAGCCAA	VF2468 81 rev	TGGGCTGCCATGACAAACAG
VF2468 83 fwd	GAGTTGAGCTGTCAGCGGGG	VF2468 83 rev	GAAGCCAACTGCCTTGTGAGC
VF2468 84 fwd	TGTTTTCTGCAGTTTTGCAGGG	VF2468 84 rev	GGCTCAGGGAGTTTGAGCCA
VF2468 85 fwd	GCTCTGGCACCAGGCACACT	VF2468 85 rev	GGGAGAGAACCATGAATTTCCCA
VF2468 86 fwd	GCCAAACCCTTTCCAGGGAG	VF2468 86 rev	CCCACCCTATGCACAGAGCC
VF2468 87 fwd	CCTCAGCCAGTTGGAATCGG	VF2468 87 rev	CAACGGTTTAGTTTAGTTCCGGTTT
VF2468 88 fwd	TGGGTGGTGAAAATGGGGTT	VF2468 88 rev	GGTGGGGTATGCACTGGTCA
VF2468 89 fwd	GGAATGTGTGGAACTCAATTTCTTT	VF2468 89 rev	TTGCTTGCAGGGTGTGGAAA
VF2468 90 fwd	CCACAAGGGTCATCTGGGGA	VF2468 90 rev	CGGAGGCATCATCCACTGAG
VF2468 91 fwd	CCTGGAGTGGTTTGGCTTCG	VF2468 91 rev	TGGAGCCCTGGAGTTCTTGG
VF2468 92 fwd	GGCTCCTGGGGTCATTTTCC	VF2468 92 rev	TGTGCTCCATCCTCCTCCCT
VF2468 93 fwd	GTGTGTTTCCGCACACCCTG	VF2468 93 rev	GCTCTTGGCTTCCCAACCCT
VF2468 94 fwd	CCATCGCCGTGTCTGAGTGT	VF2468 94 rev	CAGCAGGAACATCATCCCCC
VF2468 95 fwd	AGGCAATGGCACCAAAATGG	VF2468 95 rev	GCAGCCTTCACCATACCTGTGA
VF2468 96 fwd	TTTTGACTTTGAGAACCCCCTGA	VF2468 96 rev	CCTTGTCCTTTCTCAGTTAGACACA
VF2468 97 fwd	GCTGAGTGCAAAGCTCAGGGA	VF2468 97 rev	GGCAACACAGCAAGACCCCT

Zinc-finger nuclease expression and characterization

3xFLAG-tagged zinc finger proteins for CCR5-224 and VF2468 were expressed as fusions to FokI obligate heterodimers (Miller et al., 2007) in mammalian expression vectors (Maeder et al., 2008) derived from pMLM290 and pMLM292. Complete vector sequences are available upon request. 2 μg of ZFN-encoding vector was transcribed and translated *in vitro* using the TnT Quick Coupled rabbit reticulocyte system (Promega). Zinc chloride (Sigma-Aldrich) was added at 500 μM and the transcription/translation reaction was performed for 2 hours at 30 °C. Glycerol was added to a 50% final concentration. Western blots were used to visualize protein using the anti-FLAG M2 monoclonal antibody (Sigma-Aldrich). ZFN concentrations were determined by Western blot and comparison with a standard curve of N-terminal FLAG-tagged bacterial alkaline phosphatase (Sigma-Aldrich).

Test substrates for CCR5-224 and VF2468 were constructed by cloning into the *Hin*dIII/*Xba*I sites of pUC19. PCR with primers "test fwd" and "test rev" and Taq DNA polymerase yielded a linear 1 kb DNA that could be cleaved by the appropriate ZFN into two fragments of sizes ~300 bp and ~700 bp. Activity profiles for the zinc-finger nucleases were obtained by digestion of the 1 μg linear 1 kb DNA with varying amounts of ZFN in 1x NEBuffer 4 (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM dithiothreitol, pH 7.9) for 4 hours at 37 °C. 100 μg of RNase A (Qiagen) was added to the reaction for 10 minutes at room temperature to remove RNA from the *in vitro* transcription/translation mixture that could interfere with purification and gel analysis. Reactions were purified with the Qiagen PCR Purification Kit and analyzed on 1% TAE-agarose gels.

In vitro *selection*

ZFNs of varying concentrations, an amount of TnT reaction mixture without any protein-encoding DNA template equivalent to the greatest amount of ZFN used ("lysate"), or 50 units *Pvu*I (NEB) were incubated with 1 µg of rolling-circle amplified library for 4 hours at 37 °C in 1x NEBuffer 4 (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM dithiothreitol, pH 7.9). 100 µg of RNase A (Qiagen) was added to the reaction for 10 minutes at room temperature to remove RNA from the *in vitro* transcription/translation mixture that could interfere with purification and gel analysis. Reactions were purified with the Qiagen PCR Purification Kit. 1/10 of the reaction mixture was visualized by gel electrophoresis on a 1% TAE-agarose gel and staining with SYBR Gold Nucleic Acid Gel Stain (Invitrogen).

The purified DNA was blunted with 5 units DNA Polymerase I, Large (Klenow) Fragment (NEB) in 1x NEBuffer 2 (50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM dithiothreitol, pH 7.9) with 500 µM dNTP mix (Bio-Rad) for 30 minutes at room temperature. The reaction mixture was purified with the Qiagen PCR Purification Kit and incubated with 5 units of Klenow Fragment (3' exo^-) (NEB) for 30 minutes at 37 °C in 1x NEBuffer 2 (50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM dithiothreitol, pH 7.9) with 240 µM dATP (Promega) in a 50 µL final volume. 10 mM Tris-HCl, pH 8.5 was added to a volume of 90 µL and the reaction was incubated for 20 minutes at 75 °C to inactivate the enzyme before cooling to 12 °C. 300 fmol of "adapter1/2", barcoded according to enzyme concentration, or 6 pmol of "adapter1/2" for the *Pvu*I digest, were added to the reaction mixture, along with 10 ul 10x NEB T4 DNA Ligase Reaction Buffer (500 mM Tris-HCl, 100 mM $MgCl_2$, 100 mM dithiothreitol, 10 mM ATP). Adapters were ligated onto the blunt DNA ends with 400 units of T4 DNA ligase at room temperature for 17.5 hours and ligated DNA was purified away from unligated adapters

with Illustra Microspin S-400 HR sephacryl columns (GE Healthcare). DNA with ligated adapters were amplified by PCR with 2 units of Phusion Hot Start II DNA Polymerase (NEB) and 10 pmol each of primers "PE1" and "PE2" in 1x Phusion GC Buffer supplemented with 3% DMSO and 1.7 mM $MgCl_2$. PCR conditions were 98 °C for 3 min, followed by cycles of 98 °C for 15 s, 60 °C for 15 s, and 72 °C for 15 s, and a final 5 min extension at 72 °C. The PCR was run for enough cycles (typically 20-30) to see a visible product on gel. The reactions were pooled in equimolar amounts and purified with the Qiagen PCR Purification Kit. The purified DNA was gel purified on a 1% TAE-agarose gel, and submitted to the Harvard Medical School Biopolymers Facility for Illumina 36-base paired-end sequencing.

Data Analysis

Illumina sequencing reads were analyzed using programs written in C++. Algorithms are described in "Algorithms for Data Analysis" (below), and the source code is available on request. Sequences containing the same barcode on both paired sequences and no positions with a quality score of 'B' were binned by barcode. Half-site sequence, overhang and spacer sequences, and adjacent randomized positions were determined by positional relationship to constant sequences and searching for sequences similar to the designed CCR5-224 and VF2468 recognition sequences. These sequences were subjected to a computational selection step for complementary, filled-in overhang ends of at least 4 base pairs, corresponding to rolling-circle concatemers that had been cleaved at two adjacent and identical sites. Specificity scores were calculated with the formulae:

positive specificity score = (frequency of base pair at position[post-selection]-frequency of base pair at position[pre-selection])/(1-frequency of base pair at position[pre-selection])

negative specificity score = (frequency of base pair at position[post-selection]-frequency of base pair at position[pre-selection])/(frequency of base pair at position[pre-selection])

Positive specificity scores reflect base pairs that appear with greater frequency in the post-selection library than in the starting library at a given position; negative specificity scores reflect base pairs that are less frequent in the post-selection library than in the starting library at a given position. A score of +1 indicates an absolute preference, a score of -1 indicates an absolute intolerance, and a score of 0 indicates no preference.

Assay of genome modification at cleavage sites in human cells

CCR5-224 ZFNs were cloned into a CMV-driven mammalian expression vector in which both ZFN monomers were translated from the same mRNA transcript in stoichiometric quantities using a self-cleaving T2A peptide sequence similar to a previously described vector (Doyon et al., 2008). This vector also expresses enhanced green fluorescent protein (eGFP) from a PGK promoter downstream of the ZFN expression cassette. An empty vector expressing only eGFP was used as a negative control.

To deliver ZFN expression plasmids into cells, 15 μg of either active CCR5-224 ZFN DNA or empty vector DNA were used to Nucleofect $2x10^6$ K562 cells in duplicate reactions following the manufacturer's instructions for Cell Line Nucleofector Kit V (Lonza). GFP-positive cells were isolated by FACS 24 hours post-transfection, expanded, and harvested five days post-transfection with the QIAamp DNA Blood Mini Kit (Qiagen).

PCR for 37 potential CCR5-224 substrates and 97 potential VF2468 substrates was performed with Phusion DNA Polymerase (NEB) and primers "[ZFN] [#] fwd" and "[ZFN] [#]

rev" (Tables 4.8) in 1x Phusion HF Buffer supplemented with 3% DMSO. Primers were designed using Primer3 (Rozen and Skaletsky, 2000). The amplified DNA was purified with the Qiagen PCR Purification Kit, eluted with 10 mM Tris-HCl, pH 8.5, and quantified by 1K Chip on a LabChip GX instrument (Caliper Life Sciences) and combined into separate equimolar pools for the catalytically active and empty vector control samples. PCR products were not obtained for 3 CCR5 sites and 7 VF2468 sites, which excluded these samples from further analysis. Multiplexed Illumina library preparation was performed according to the manufacturer's specifications, except that AMPure XP beads (Agencourt) were used for purification following adapter ligation and PCR enrichment steps. Illumina indices 11 ("GGCTAC") and 12 ("CTTGTA") were used for ZFN-treated libraries while indices 4 ("TGACCA") and 6 ("GCCAAT") were used for the empty vector controls. Library concentrations were quantified by KAPA Library Quantification Kit for Illumina Genome Analyzer Platform (Kapa Biosystems). Equal amounts of the barcoded libraries derived from active- and empty vector- treated cells were diluted to 10 nM and subjected to single read sequencing on an Illumina HiSeq 2000 at the FAS Center for Systems Biology Core facility (Cambridge, MA).

Statistical analysis

In Tables 4.3 and 4.7, *p*-values were calculated for a one-sided test of the difference in the proportions of sequences with insertions or deletions from the active CCR5-224 sample and the empty vector control samples. The t-statistic was calculated as $t = (\text{p\_hat}_1 - \text{p\_hat}_2) / \text{sqrt}((\text{p\_hat}_1 \times (1 - \text{p\_hat}_1) / n_1) + (\text{p\_hat}_2 \times (1 - \text{p\_hat}_2) / n_2))$,where p_hat_1 and n_1 are the proportion and total number, respectively, of sequences from the active sample and p_hat_2 and n_2

are the proportion and total number, respectively, of sequences from the empty vector control sample.

Plots

All heat maps were generated in the R software package with the following command:

image([variable], zlim = c(-1,1), col = colorRampPalette(c("red","white","blue"),space="Lab")(2500)

Algorithms for Data Analysis

Quality score filtering and sequence binning

1) search each position of both pairs of sequencing read for quality score, reject if any position has quality score = 'B'
2) output to separate files all sequence reads where the first sequence in the pair start with barcodes ("AAT", "ATA", "TAA", "CAC","TCG") and count the number of sequences corresponding to each barcode

Filtering by ZFN ("AAT","ATA","TAA","CAC")

For each binned file,

1) accept only sequence pairs where both sequences in the pair start with the same barcode
2) identify orientation of sequence read by searching for constant regions
    - orientation 1 is identified by the constant region "CGATCGTTGG"
    - orientation 2 is identified by the constant region "CAGTGGAACG"
3) search sequences from position 4 (after the barcode) up to the first position in the constant region for the subsequence that has the fewest mutations compared to the CCR5-224 and VF2468 half-site that corresponds to the identified constant region

- search sequences with orientation 1 for "GATGAGGATGAC" (CCR5-224(+)) and "GACGCTGCT" (VF2468(-))
- search sequences with orientation 2 for "AAACTGCAAAAG" (CCR5-224(-)) and "GAGTGAGGA" (VF2468(+))

4) bin sequences as CCR5-224 or VF2468 by testing for the fewest mutations across both half-sites
5) the positions of the half-sites and constant sequences are used to determine the overhang/spacer sequences, the flanking nucleotide sequences, and the tag sequences
    - the subsequence between the half-site of orientation 1 and the constant region is the tag sequence
        - if there is no tag sequence, the tag sequence is denoted by 'X'
    - the overhang sequence is determined by searching for the longest reverse-complementary subsequences between the subsequences of orientation 1 and orientation 2 that start after the barcodes
    - the spacer sequence is determined by concatenating the reverse complement of the subsequence in orientation 1 that is between the overhang and the half-site (if any), the overhang, and the subsequence in orientation 2 that is between the overhang and the half-site
        - if there is overlap between the overhang and half-site, only the non-overlapping subsequence present in the overhang is counted as part of the spacer
6) to remove duplicate sequences, sort each sequence pair into a tree
    - each level of the tree corresponds to a position in the sequence

- each node at each level corresponds to a particular base (A, C, G, T, or X = not(A, C, G, or T)) and points to the base of the next position (A,C,G,T,X)
- the sequence pairs are encoded in the nodes and a subsequence consisting of the concatenation of the spacer sequence, flanking nucleotide sequence, and tag sequence is sorted in the tree
- at the terminal nodes of the tree, each newly entered sequence is compared to all other sequences in the node to avoid duplication

7) the contents of the tree are recursively outputted into separate files based on barcode and ZFN

Library filtering ("TCG")

1) accept only sequence pairs where both sequences in the pair start with the same barcode
2) analyze the sequence pair that does not contain the sequence "TCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGAC" (the other pair contains the library sequence)
3) search sequences for ZFN half-sites and bin by the ZFN site that has fewer mutations
    - search for "GTCATCCTCATC" and "AAACTGCAAAAG" (CCR5-224) and "AGCAGCGTC" and "GAGTGAGGA" (VF2468)
4) identify the spacer, flanking nucleotide, and nucleotide tag sequences based on the locations of the half-sites
5) use the tree algorithm in step 6 under "Filtering by ZFN" to eliminate duplicate sequences

Sequence profiles

1) analyze only sequences that contain no 'N' positions and have spacer lengths between 4 and 7
2) tabulate the total number of mutations, the spacer length, the overhang length, the nucleotide frequencies for the (+) and (-) half-sites, the nucleotide frequencies for spacers that are 4 bp, 5 bp, 6 bp, and 7 bp long, and the nucleotide frequencies for the flanking nucleotide and the tag sequence
3) repeat steps 1 and 2 for library sequences
4) calculate specificity scores at each position using positive specificity score = (frequency of base pair at position[post-selection]-frequency of base pair at position[pre-selection])/(1-frequency of base pair at position[pre-selection]) negative specificity score = (frequency of base pair at position[post-selection]-frequency of base pair at position[pre-selection])/(frequency of base pair at position[pre-selection])

Genomic matches

1) the human genome sequence was searched with 24 and 25 base windows (CCR5-224) and 18 and 19 base windows (VF2468) for all sites within nine mutations (CCR5-224) or six mutations (VF2468) of the canonical target site with all spacer sequences of five or six bases being accepted
2) each post-selection sequence was compared to the set of genomic sequences within nine and six mutations of CCR5-224 and VF2468, respectively

Enrichment factors for sequences with 0, 1, 2, or 3 mutations

1) for each sequence, divide the frequency of occurrence in the post-selection library by the frequency of occurrence in the pre-selection library

Filtered sequence profiles

1) use the algorithm described above in "Sequence profiles", except in addition, only analyze sequences with off-target bases at given positions for both pre- and post-selection data

Compensation difference map

1) use "Filtered sequence profiles" algorithm for mutation at every position in both half-sites
2) calculate: Δ(specificity score) = filtered specificity score – non-filtered specificity score

NHEJ search

1) identify the site by searching for exact flanking sequences
2) count the number of inserted or deleted bases by comparing the length of the calculated site to the length of the expected site and by searching for similarity to the unmodified target site (sequences with 5 or fewer mutations compared to the intended site were counted as unmodified)
3) inspect all sites other than CCR5, CCR2, and VEGF-A promoter by hand to identify true insertions or deletions

Chapter Five

Discussion and Conclusions

The landscape of the genome engineering field has been altered rapidly in the last few years. Interest both from protein engineers and investigators eager to modify model organisms and cells has converged not only to expand the repertoire of tools available for these efforts, but also to open the floodgates for elegant demonstrations of the efficiency with which virtually any genetic alteration—once thought impossible—can be carried out in record time.

The thesis author presents the findings of this dissertation in the context of their impact to the field and offers perspectives on future directions for study.

Modular Assembly—The Next Generation

The publication presented in Chapter 2 was met mostly with appreciation from members of the scientific community who had experienced previously unexplained frustration with making functional zinc finger proteins by modular assembly (Ramirez et al., 2008). As has been reiterated elsewhere (Joung et al., 2010), this is not to say that it is not possible to generate an active ZFN pair by modular assembly; nevertheless, given other options available for engineering ZFNs, this method offers at best a low probability of success even with GNN-rich target sites, severely limiting the targeting range and thus the broad applicability of this approach. Interestingly, despite some resistance to accept the validity of our findings from certain groups, their studies claiming to refute our conclusions in fact offer further support upon closer inspection: in another large-scale study of modular assembly, Kim and colleagues reported

that ~93% (294 of 315) of ZFN pairs failed to show activity in a human cell-based reporter assay, which is comparable to the ~94% ZFN failure rate predicted by Ramirez & Foley *et al.* (Ramirez et al., 2008, Kim et al., 2009, Joung et al., 2010, Kim et al., 2010).

A subset of the proteins described in Ramirez & Foley *et al.* were examined further by Hughes and colleagues (Lam et al., 2011). It was determined that these modularly assembled proteins resemble natural transcription factors in that they exhibit degenerate binding preferences in protein-binding microarray assays, sometimes without showing strong specificity for their intended 9 base pair target. Although this reasoning may help to explain why these proteins failed, it is clear from direct comparisons between modularly assembled proteins and those derived from context-dependent selection methods that the limited specificity and high failure rates of modularly assembled proteins in living cells are not a general property of all engineered zinc finger nucleases (Maeder et al., 2008, Sander et al., 2011b).

Substantial zinc finger engineering advances accessible to the scientific community have been made available in the intervening time since the publication of Ramirez & Foley *et al*., including Oligomerized Pool Engineering (OPEN) (Maeder et al., 2008) and Context-Dependent Assembly (CoDA) (Sander et al., 2011b). Described in more detail in Chapter 1, OPEN is a simplified context-dependent selection strategy for engineering three-finger arrays with a high success rate (~67%) as ZFNs in zebrafish, plant, and human cells (Maeder et al., 2008, Maeder et al., 2009).

A derivative of OPEN, Context-Dependent Assembly (CoDA) is a simpler method from a technical standpoint than modular assembly, yet it accounts for context-dependence of zinc fingers in a multi-finger array and has a much higher success rate (Sander et al., 2011b). Using data from dozens of OPEN selections, a subset of 18 “middle” fingers from selected three-finger

zinc finger arrays were found to appear in multiple protein contexts. ZFNs made by CoDA exhibit gene disruption rates of >50% reported by our group in zebrafish and plants (Sander et al., 2011b) and others in human cells (Osborn et al., 2011). The targeting range of this method is about 1 in 500 bp of random DNA sequence. The ease with which CoDA can be practiced and its higher success rate in direct comparisons with modularly assembled proteins (Sander et al., 2011b) support its recommendation over modular assembly for rapidly generating functional ZFNs (Segal, 2011).

These and other advances in ZFN nuclease engineering technologies provided a conceptual and experimental foundation for an even faster genesis of targeted nucleases amenable to a modular assembly-like design approach—Transcription Activator-Like Effector Nucleases (TALENs), fusions of TALE DNA-binding domains and the FokI nuclease domain that also function as dimers (Cermak et al., 2011, Huang et al., 2011, Miller et al., 2011, Sander et al., 2011a). Derived from highly repetitive proteins originally isolated from *Xanthomonas*, TALENs have emerged as strong competitors to ZFNs, due to both simplicity of design as well as generally having high activity in living cells (Clark et al., 2011). The sequence recognition elements of TALENs are small (33-35 amino acid) repeat domains that each confer binding specificity to one base of DNA, which is determined by the composition of hypervariable residues at positions 12 and 13, collectively known as repeat variable diresidues (RVDs) (Boch et al., 2009, Moscou and Bogdanove, 2009). TALE proteins with various numbers of repeat domains can be assembled rapidly (Cermak et al., 2011, Sander et al., 2011c) and have been shown to bind DNA effectively, with naturally occurring TALEs having about 18 repeats (Boch et al., 2009, Miller et al., 2011).

Although TALENs do generally have high success rates, the suggestion has been made in the literature that they may not always work, begging the question of what other parameters might affect activity. One possibility recently revealed by structural models is that TAL binding may be affected by DNA methylation (Deng et al., 2012, Mak et al., 2012). Functional studies support the notion that epigenetic modifications may affect TALEN binding in a cellular context (Bultmann et al., 2012). It has also been hypothesized that ZFNs may be similarly influenced by regions of heterochromatin and DNA methylation (Liu et al., 2001, Maeder et al., 2008); it will be interesting to see whether it is possible to design around these targeting limitations if needed using either platform. Yet another possibility is that the seemingly uniform TAL repeat domains have some subtle context-dependence or that certain module configurations can affect the strength of DNA binding affinity for the protein as a whole.

Additional questions remain in determining whether TALENs might entirely replace ZFNs for future applications, including whether TALENs would elicit an immune response if delivered systemically to higher organisms. Having been derived from a bacterial source and being much larger proteins than ZFNs, immune rejection may be more of a risk with TALENs than with ZFNs, which bear zinc finger motifs similar to those commonly found in transcription factors. Fortunately, many therapeutic strategies rely upon *ex vivo* treatment of cells and their subsequent reintroduction into the patient, so it is possible the transient expression of nucleases may not be a significant concern in these instances.

Another point to consider is how specifically TALENs recognize their intended binding sites at the exclusion of other sites. Particularly considering their long binding sites, extensive studies have not yet been performed on the number of mismatches that can be tolerated per monomer, assuming there might be a general trend that can be discerned. Initial efforts to

characterize the specificity of TALENs have been attempted (Hockemeyer et al., 2011, Miller et al., 2011, Wood et al., 2011), yet little is known about the degenerate sequence preferences of TALEN dimers; as was determined for ZFNs, current methods based on monomer binding assays may not be able to accurately predict the full range of potential off-target sites for TALEN pairs (Pattanayak et al., 2011). The dimeric nuclease selection described in Chapter 4 could conceivably be adapted for analysis of TALENs and may provide insights on the effects of cooperative binding that cannot be discerned from monomer binding data (Pattanayak et al., 2011).

Despite these open questions, it does seem in the short-term as though TALENs are a solid first-line option for generating targeted mutations in model organisms and cell lines due in large part to the simplicity with which novel proteins can be synthesized, the great amount of design flexibility their extensive targeting range provides, and the generally high success rates of these nucleases.

Engineering Genomes with Nickases

Engineered ZFNickases have been shown in most cases to promote higher rates of HDR relative to NHEJ in human cell reporter assays, but absolute rates of HDR are 2- to >100-fold lower than with nucleases (Ramirez et al., 2012). Soon after the publication of this work, evidence that ZFN-derived nicking enzymes can mediate gene correction at endogenous loci was reported (Kim et al., 2012, Wang et al., 2012). Despite their lower HDR rates relative to nucleases, the lower genotoxicity of ZFNickases is a promising indicator that they ought to be further optimized for therapeutic, basic research, or biotechnology applications. Other approaches including S/G2 cell cycle arrest with vinblastine (Urnov et al., 2005, Maeder et al.,

2008) and nocodazole (Olsen et al., 2010) have been shown to dramatically increase gene targeting efficiencies up to ~50% in cell culture models, in some cases driving HDR rates to exceed NHEJ rates; however, the trade-off for this benefit is even higher NHEJ rates relative to ZFN-treated cells not under arrest and much higher cell death. An important consideration is that treating humans with cytostatic drugs may limit the feasibility of a therapeutic strategy if there are significant side-effects associated with treatment. If this is the case, then nickases may be a more benign option to consider for downstream gene correction applications. Through the use of cytostatic drugs, nickases, or other approaches, it is critical to raise absolute HDR rates to therapeutically relevant levels while maintaining an acceptably low risk-to-benefit ratio for the patient.

There are several approaches that could be used to improve nickase activity. The most apparent is to use the newly-described optimized heterodimer ELD:KKR domains, which have been shown to increase cleavage activity and reduce the probability of homodimerization (Doyon et al., 2011). In addition, hyperactive FokI mutants compatible with this framework could be tested to attain more robust gene targeting rates (Guo et al., 2010). There is evidence to suggest that simultaneous nicks on both the repair target and donor may lead to significantly higher levels of gene conversion, so it is possible that increasing the nicking activity of ZFNickases may provide a non-linear improvements in rates of repair (Goncalves et al., 2011). An alternative approach to higher HDR rates would be to treat cells with drugs that inhibit nick repair (e.g. PARP-1 inhibitors), which could potentially allow for repair through other mechanisms, including HDR of the single-strand break or repair of a double strand break derived from a nick in the path of a replication fork (Bouchard et al., 2003). As HDR is most active in dividing cells during the S and G2 phases of the cell cycle to counteract damage associated with DNA

replication, converting nicks to double-strand breaks during this period could potentially leverage favorable cellular conditions toward the desired repair pathway (Takata et al., 1998, Mao et al., 2008, Hartlerode et al., 2011). There are currently PARP-1 inhibitors in clinical trials, so it is conceivable they could be used in combination with nickases for a future therapeutic strategy if they were found to be effective in preliminary experiments (Sessa, 2011).

It remains to be seen whether TAL nickases can be produced by the same method as ZFNickases. The relatively confined and reproducible cleavage patterns of ZFNs (which consistently leave 4-5 bp overhangs) is substantially different from the varied cleavage patterns of TALENs and the wide range of spacers between TALEN half-sites that can be tolerated. It has been demonstrated that certain TALEN architectures have more rigid spacing requirements than others, so it may be wisest to start with the former to avoid the risk of inadvertently cutting both strands despite only having one active subunit for cleavage (Miller et al., 2011). If it is not possible to create reliable TALEN-derived nickases, then this may be a unique engineering advantage that ZFNs have over TALENs.

Defining ZFN Specificity: A Work in Progress

A study with a complementary approach to examining off-target cleavage sites of ZFNs also appeared in print (Gabriel et al., 2011) at the same time as our manuscript presented in Chapter 3 was published (Pattanayak et al., 2011). In contrast to our method, in which an *in vitro* selection was used to identify potential off-target sites in ZFN-treated cells, the strategy of Gabriel & Lombardo *et al.* relied upon the integration of a GFP-encoding integration-deficient lentiviral vector at double-strand break sites in living cells; these breaks would include cleavage at the intended *CCR5* locus, off-target break sites, or other loci subject to ZFN-independent

double-strand breaks, likely as a natural consequence of normal cellular processes such as replication fork collapse during cell division.

It is striking that our study and the von Kalle study identified non-overlapping novel off-target sites for the CCR5 224 ZFNs being used in clinical trials (Cheng et al., 2011). One possible explanation is that the sensitivity levels for both assays are different. In the lentiviral integration method, at least two independent integration events within 500 bp are required to flag a locus as a potential ZFN target. As most of the double-strand breaks are likely to occur at the *CCR5* or *CCR2* loci or at non-specific locations in the genome that can vary with each cell division cycle, this leaves proportionally only a small fraction of the dataset's capacity to identify infrequent off-target cleavage events, probably biasing signal detection toward more frequently represented sites. This is the probable explanation for why the *in vitro* selection system successfully identified off-target sites below the detection limit of the lentiviral assay.

On the other hand, two drawbacks of the *in vitro* selection method are that all of the sequence space of the library cannot be sampled and only sites in the human genome directly derived from the selection were chosen for validation in human cells. For a subsequent iteration of this method, it would likely be fruitful to use the results of the *in vitro* selection to computationally predict additional off-target sites, as in its current state, the selection approach misses several key off-target sites that were frequent enough to appear in the lentiviral integration assay. Another way to improve on the *in vitro* selection method would be to construct a library from genomic DNA of the cell type of interest rather than use a combinatorial library of degenerate oligonucleotides. This would greatly reduce the theoretical library size and potentially allow for much more sensitive detection of off-target sites.

Moving forward, it is likely that combinations of approaches will be required to determine ZFN off-target cleavage specificity as extensively as possible. Without any significant adaptation, the lentiviral integration method may be better suited for determining off-target specificity of TALENs, as their long binding sites would even further limit the sampling of sufficient sequence space in the *in vitro* selection method. Alternatively, with deep sequencing prices continuing to drop and data yield continuing to grow, there may come a day when sequencing genomes of clonal cell populations may be a less tedious and costly than performing indirect methods to predict off-target sites. However, for approaches in which populations of patient cells need to be modified (and in which very low frequency but potentially deleterious off-target events may occur), approaches including those described herein may be necessary well into the future.

Closing Thoughts

We have demonstrated in this dissertation that modular assembly has significant limitations for generating functional ZFNs for genome engineering applications; the principle of this approach may be better suited for TALENs, whose modular recognition domains may lack substantial context-dependence.

Our method for generating zinc finger nickases provides a new tool for stimulating gene conversion preferentially over gene disruption, although additional work must be done to increase absolute gene conversion rates; it is not clear yet whether it is possible to generate TAL-derived nickases using a similar strategy.

We have also developed and validated a strategy for identifying ZFN off-target sites in human cells based on an *in vitro* selection assay. If the combinatorial library for this selection

can be adapted to be made from genomic DNA instead of degenerate oligonucleotides, it may be possible to sample the cleavage specificities of TALEN dimers in great detail using this method.

Taken together, these approaches represent our efforts to build on the rich legacy of zinc finger technology development strategies to move the genome engineering field ever toward the most straightforward, effective, and safe methods for modifying genomes of model organisms and human patients.

Appendix One

A Synthetic Biology Framework for Programming Eukaryotic Transcription Functions

Ahmad S. Khalil*, Timothy K. Lu*, Caleb J. Bashor*, Cherie L. Ramirez, Nora C. Pyenson, J. Keith Joung, and James J. Collins

Ahmad Khalil, Timothy Lu, and Caleb Bashor were joint first authors (designated by "*") of this work; the manuscript has been accepted for publication in *Cell* and has been reprinted with their permission. The thesis author performed the groundwork for characterizing zinc finger proteins with diminished non-specific contacts to the phosphate groups of DNA, including bacterial-2-hybrid activity assays included in this work and provided experimental suggestions as needed. Ahmad Khalil, Timothy Lu, Caleb Bashor, and Nora Pyenson performed all other experiments.

A Synthetic Biology Framework for Programming Eukaryotic Transcription Functions

Ahmad S. Khalil,[1,7] Timothy K. Lu,[2,7*] Caleb J. Bashor,[1,7] Cherie L. Ramirez,[3,4] Nora C. Pyenson,[1] J. Keith Joung[3,5] and James J. Collins[1,6,*]

[1]Howard Hughes Medical Institute, Department of Biomedical Engineering, and Center for BioDynamics, Boston University, Boston, MA, USA.
[2]Synthetic Biology Group, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA.
[3]Molecular Pathology Unit, Center for Cancer Research and Center for Computational and Integrative Biology, Massachusetts General Hospital, Charlestown, MA, USA.
[4]Biological and Biomedical Sciences Program, Division of Medical Sciences, Harvard Medical School, Boston, MA, USA.
[5]Department of Pathology, Harvard Medical School, Boston, MA, USA.
[6]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA.
[7]These authors contributed equally to this work.

Abstract

Eukaryotic transcription factors (TFs) perform a variety of complex and combinatorial functions within transcriptional networks. Here we present a synthetic framework for systematically constructing eukaryotic transcription functions by building from artificial zinc fingers, modular DNA-binding domains found within many eukaryotic TFs. Utilizing this platform, we construct a library of orthogonal synthetic transcription factor (sTF)-promoter pairs, and use these to wire synthetic transcriptional circuits in yeast. We design and interrogate complex functions, such as tunable output strength and transcriptional cooperativity, by rationally engineering a small, decomposed set of key component properties, e.g., DNA specificity, affinity, promoter design, protein-protein interactions. We show that subtle

perturbations to these properties can transform an individual sTF between distinct roles (activator, cooperative factor, inhibitory factor) within a transcriptional complex, thus drastically altering the signal processing behavior of multi-input systems. This platform provides new genetic components for synthetic biology, and enables bottom-up approaches to understanding the design principles of eukaryotic transcriptional complexes and networks.

Introduction

The genetic program of a living cell is governed by the faithful execution of a number of fundamental molecular functions by transcription factors (TFs). These include wiring specific connections to promoter regulatory elements, modulating the transcriptional output of a gene, tuning molecular noise, recruiting coactivator/repressor complexes and basal transcriptional machinery, cooperating with other TFs to regulate a gene, integrating an array of environmental signals, and even physically manipulating the geometrical configuration of chromosomes (Ptashne, 1986, Ptashne, 1988, Pedraza and van Oudenaarden, 2005, Rosenfeld et al., 2005, Hahn and Young, 2011). A tremendous amount of progress has been made toward understanding eukaryotic transcriptional regulation. Yet, there is still much to be learned about how eukaryotic TFs accomplish their fundamental tasks to bring about higher-order transcriptional behaviors (Hahn and Young, 2011). A synthetic approach, whereby minimal and insulated components and circuitry can be constructed to recapitulate eukaryotic transcription functions, would be valuable for studying how transcriptional regulatory complexes are assembled and how TFs are wired into networks.

A framework for eukaryotic transcription regulation would also be broadly valuable to synthetic biology efforts, which seek to uncover the design principles of gene regulatory

networks and program novel biological functions for a range of biotechnological and industrial applications (Elowitz and Leibler, 2000, Gardner et al., 2000, Hasty et al., 2002, Kramer et al., 2004b, Levskaya et al., 2005, Andrianantoandro et al., 2006, Ajo-Franklin et al., 2007, Bashor et al., 2008, Stricker et al., 2008, Win and Smolke, 2008, Mukherji and van Oudenaarden, 2009, Tigges et al., 2009, Bashor et al., 2010, Khalil and Collins, 2010, Lim, 2010, Tamsir et al., 2011). Transcriptional circuitry has been a major focus of the field dating back to its origins, and has been used to implement a variety of genetic behaviors, including memory, oscillations, logic operations, filtering, and noise propagation (Becskei and Serrano, 2000, Elowitz and Leibler, 2000, Gardner et al., 2000, Guet et al., 2002, Yokobayashi et al., 2002, Pedraza and van Oudenaarden, 2005, Rosenfeld et al., 2005). In these and virtually all other synthetic studies, small motifs and circuits have been constructed with a few, classical prokaryotic TFs; these "off-the-shelf parts" represent the extent of well-understood, reliable, and accessible transcriptional components. To date, the synthetic construction of transcriptional networks in eukaryotes has relied heavily upon these same, few bacterial TF-promoter pairs (Lu et al., 2009, Weber and Fussenegger, 2009).

This approach of porting bacterial transcriptional components has certain advantages. Bacterial TFs perform relatively simple transcriptional tasks (as compared with eukaryotic TFs), and therefore assembling and programming with them can be straightforward. Yet, for this reason and because they regulate transcription in fundamentally different ways than their eukaryotic counterparts, they are a poor starting point for engineering the complex transcriptional functions enumerated above. Furthermore, bacterial transcriptional components are severely limiting with respect to extensibility – they have been designed to bind a specific target and have integrated and coupled properties. This requires re-engineering schemes, such as

directed evolution, to generate an expanded set of components, connectivity, and behaviors. As a result, the use of bacterial transcriptional components is unlikely to scale to the more sophisticated circuitry needed for engineering higher-order behaviors.

We present an alternative approach to synthetic transcriptional components, control, and circuitry in eukaryotes. We model our framework on the functional subcomponents that enable eukaryotic TFs to bind DNA, activate/repress transcriptional machinery, bind other factors, etc. By deconstructing into these molecular subcomponents, we aim to create a modular decomposed design of a TF. Our design is based on zinc finger domains. TFs of virtually all eukaryotic taxa utilize Cys_2-His_2 zinc finger (ZF) domains to solve the combinatorial problem of DNA recognition and discrimination (Pabo et al., 2001). ZFs are small (~30 amino acid) domains that bind to approximately three bps of DNA (Pavletich and Pabo, 1991, Elrod-Erickson et al., 1998), and operate in tandem or in combination to dictate DNA specificity. Recent advances in ZF engineering have made it possible to purposefully re-engineer ZF DNA-binding specificities to recognize a wide variety of sequences and to covalently link them together into artificial, multi-finger arrays capable of recognizing longer DNA sequences (Pabo et al., 2001, Beerli and Barbas, 2002, Jamieson et al., 2003, Maeder et al., 2008, Maeder et al., 2009, Sander et al., 2011b). Notably, with Oligomerized Pool ENgineering (OPEN) (Maeder et al., 2008, Maeder et al., 2009) and other "context-dependent" engineering methods (Sander et al., 2011b), customized multi-finger arrays have been successfully generated with the predominant purpose of designing ZF nucleases (ZFNs) for targeted gene and genome modification (Maeder et al., 2008, Foley et al., 2009b, Townsend et al., 2009, Zou et al., 2009, Sebastiano et al., 2011).

The sequence-specific recognition of DNA elements by TFs is central to the initiation and regulation of transcription. The protein-DNA interaction specificity is therefore the core

component property that we wish to control as the basis for engineering synthetic transcriptional elements and circuitry. As ZFs represent modular domains underlying the structure-function of many eukaryotic TFs and versatile scaffolds for rational engineering, we sought to use artificial ZF domains as basic building blocks for eukaryotic synthetic transcription factors (sTFs) and gene circuitry. This allows us to generate libraries of interaction partners (artificial ZF proteins and target binding sites), and subsequently engineer components to meet functional criteria, such as activity within circuits and orthogonality from other synthetic components and native host machinery. Using this extensible platform, we construct a library of specific and orthogonal sTF-promoter pairs, and demonstrate that these pairs can be used to wire synthetic transcriptional cascades in *Saccharomyces cerevisiae*.

Constructing from artificial ZF domains enables a fully decomposed design of a sTF, in which the molecular component properties are accessible, modular, and tunable (Figure A1.1). We find that a few, key component properties made accessible by this decomposed design, e.g., DNA specificity, DNA affinity, promoter-operator design, protein interactions, can be rationally and independently adjusted to engineer complex transcriptional functions and behaviors (Figure A1.1). For example, we demonstrate the tuning of transcriptional output through the perturbation of multiple properties, notably generating weakly-activating sTFs by lowering the non-specific DNA affinity of a ZF. We engineer cooperative transcriptional systems by multimerizing the weakly-activating monomers using a modular protein-protein interaction. Finally, we construct a simple two-input promoter that recruits two individual sTFs to synthetically explore transcriptional signal integration. By systematically altering the architecture of the complex, through subtle changes to the sTF component properties, we can assign entirely different

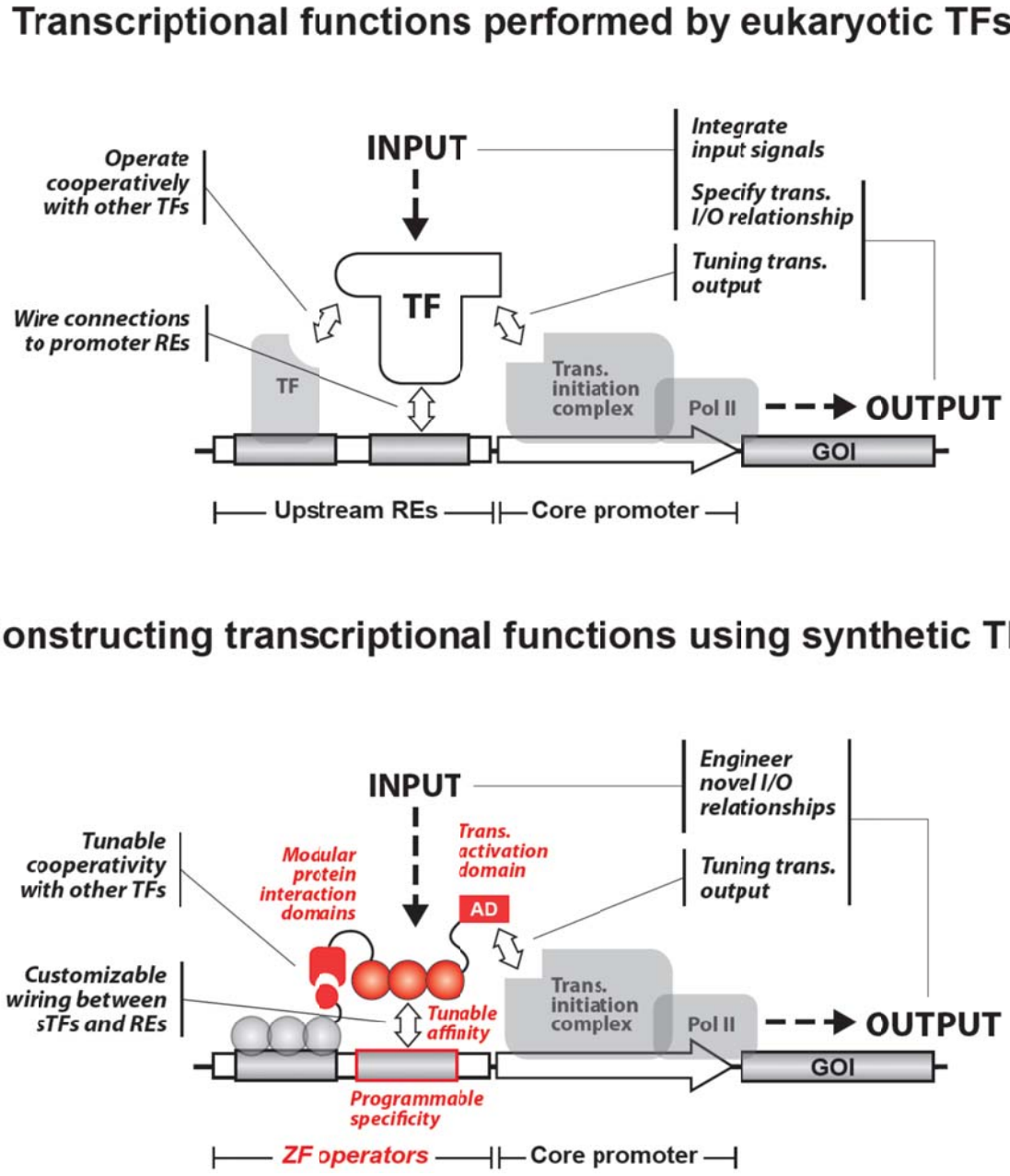


Figure A1.1: Synthetic construction of eukaryotic transcription functions
Eukaryotic transcription factors (TFs) perform a variety of molecular functions to control promoters and facilitate the operation of genetic networks (top panel). Zinc fingers (ZFs) are modular domains found in many eukaryotic TFs that make sequence-specific contact with DNA. Artificial ZF arrays were used as core building blocks for constructing synthetic TFs (sTFs) and gene circuitry in *S. cerevisiae* (bottom panel). The use of artificial ZF domains allows for a fully decomposed design of a sTF, for which the molecular component properties are accessible, modular, and tunable (red italicized). The independent control of these component properties enables the systematic construction and modulation of higher-order transcriptional functions and behaviors. Regulatory elements, REs; Transcriptional activation domain, AD.

transcriptional roles to an individual sTF and thus dramatically alter the signal processing output of the system.

Results

Wiring Specific and Orthogonal Transcriptional Connections with a Library of Synthetic TF-Promoter Pairs

Transcriptional networks, natural and synthetic, are wired together with sequence-specific protein-DNA interactions. We sought to program DNA-binding specificity, via artificial ZF proteins, in order to wire specific and orthogonal transcriptional connections in the eukaryote, *S. cerevisiae*. To do so, we first devised a platform by which ZF-based sTFs could be readily constructed and customized. The platform consists of a cassette, into which artificial three-finger arrays with engineered specificities are inserted to generate sTF species. The sTF cassette is paired with a synthetic promoter bearing ZF binding sequences that act as operators for the sTFs (Figure A1.2a).

Transcriptional activation is one of the most common mechanisms for the control of gene regulation and appears to be a universally conserved process in all eukaryotes, from fungi to metazoans (Fischer et al., 1988, Ma et al., 1988, Ma and Ptashne, 1988, Webster et al., 1988). We utilized the principle of activation by recruitment (Ma and Ptashne, 1988, Ptashne, 1988, Ptashne and Gann, 1997) to test our sTFs as minimal transcriptional activators. In our design, the engineered ZF array recapitulates the TF function of binding to a specific DNA site, in this case, to its cognate 9 bp operator in a synthetic promoter. The ZF protein is fused to a VP16 minimal activation domain (AD), which autonomously facilitates recruitment of the RNA polymerase II machinery for mRNA initiation (Ptashne, 1988). This scheme provides a decoupled, modular

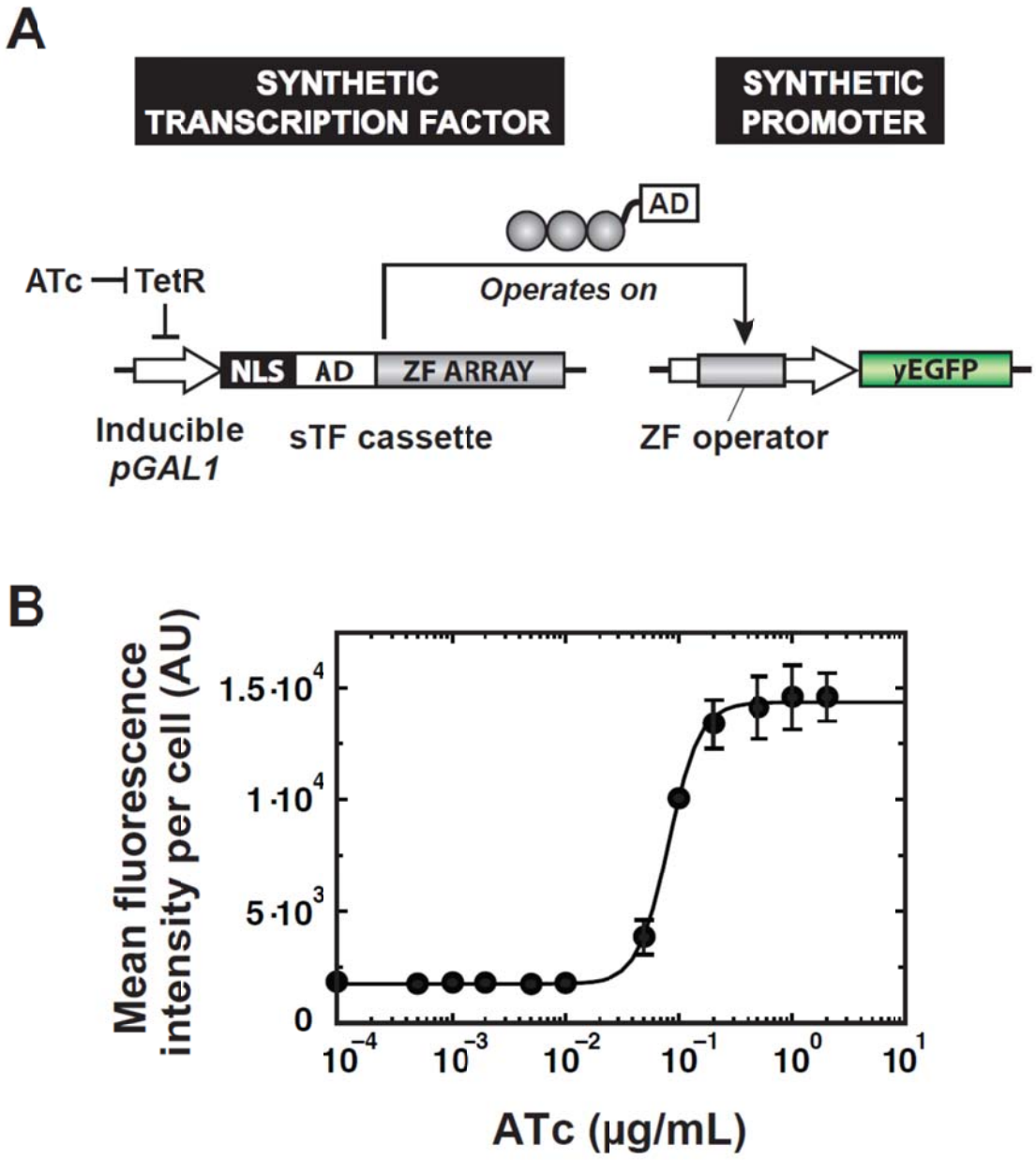


Figure A1.2: Artificial ZFs can be used to construct synthetic transcriptional activators
(a) Circuit design for synthetic transcriptional cascade. Synthetic transcription factors (sTFs) are expressed from an ATc-inducible *GAL1* promoter (*pGAL1*). sTF minimal activators are encoded by a cassette, into which artificial ZF arrays are inserted to generate fusions to a herpes simplex VP16 minimal activation domain (AD) and a nuclear localization sequence (NLS). Upon induction, sTFs operate on cognate synthetic promoters: minimal *CYC1* promoter engineered with ZF binding sequences directly upstream of the TATA box (reporter: yeast-enhanced green fluorescent protein (yEGFP)).
(b) sTF activator circuits built from artificial ZF arrays activate transcription from cognate synthetic promoters in a dose-dependent fashion (ZF 37-12 shown here). Points represent mean values for three experiments ± standard deviation.

approach to transcriptional activation, whereby TFs and the initiation machinery can be synthetically recruited in combinatorial ways. From these components, we constructed a synthetic transcriptional cascade and used it as a test bed for rationally customizing the properties of our transcriptional components to program *in vivo* behaviors; we chromosomally integrated the circuit into *S. cerevisiae* (Figure A1.2a). Within the circuit, sTF activators are first transcribed from a previously described TetR-controlled *GAL1* promoter (Murphy et al., 2007, Ellis et al., 2009), which is induced by anhydrotetracycline (ATc). Addition of ATc activates flux through the circuit to produce sTF activators, which in turn activate downstream transcription from cognate synthetic promoters to produce *yEGFP* expression (Figure A1.2b; Supplementary Figure A1.1). The resulting gene regulatory transfer function, which combines the effects of the TetR expression system and the operation of sTFs on their synthetic promoters, exhibits monotonic, dose-dependent production of yEGFP (Figure A1.2b). These results suggest that desired synthetic transcriptional connections can be made based on the specificity of engineered ZF proteins to their target sites.

With the OPEN selection system, we furthermore have the ability to rapidly alter the ZF-DNA interaction specificity to create a large library of interaction partners (i.e., engineered ZF proteins and corresponding target sites). We used artificial ZF arrays constructed by OPEN to generate a library of sTF-promoter pairs. In particular, we identified 19 three-finger arrays with binding specificities predicted to be orthogonal to one another (we predominately chose OPEN ZF arrays that had been engineered to bind sequences in orthologous genes found in plants, insects, and metazoans) (Figure A1.3a). The artificial arrays and cognate binding sequences were inserted into our framework, and the resulting library of sTF-promoter pairs were tested for activation by triggering our synthetic circuits. We found that the sTFs activated *yEGFP*

expression from cognate promoters by factors of 1.3–6.6 (compared to uninduced cells) (Figure A1.3b), showing that we could indeed make sequence-specific transcriptional connections with artificially designed ZF arrays. Notably, *yEGFP* expression levels in uninduced cells were mostly found to be similar to the basal expression levels of cells harboring only synthetic promoters (Supplementary Figure A1.2). Thus, in general, a signal was produced only when we induced expression of an sTF in the presence of a cognate promoter.

We next investigated whether the transcriptional connections made within our library of sTFs were indeed specific only to their cognate synthetic promoters. We selected a subset of 6 sTFs from our library that exhibited robust activation (> 2.5-fold) (Figure A1.3b, red stars), and crossed them with each of the other non-cognate promoters. Upon triggering the circuit, we observed no cross-activation in the subset of tested sTFs (Figure A1.3c) with one notable exception: the effect of $sTF_{43\text{-}8}$ on $Promoter_{21\text{-}16}$. Examination of the sequence just downstream of the ZF operator for $Promoter_{21\text{-}16}$ revealed the fortuitous creation of a sequence possessing significant similarity to the binding sequence of 43-8 (at 8 out of 9 bp) (Figure A1.3a, blue boxes). Thus, we attribute the observed cross-activation to the presence of this binding sequence within the non-cognate promoter. Overall, these results show that synthetic transcriptional connections can be designed to be orthogonal to one another by using the OPEN method to engineer the DNA-binding specificities of ZF arrays.

In the design of synthetic elements and gene circuitry, a further ‘orthogonality’ criterion is the degree to which the synthetic system interacts with pathways and machinery native to the cellular host. Ideally, insulated networks would interact with host pathways only at desired nodes and otherwise function independently. Using our synthetic yeast platform, we investigated one potential and rapid method for assessing sTF-host interactions. Specifically, we measured the

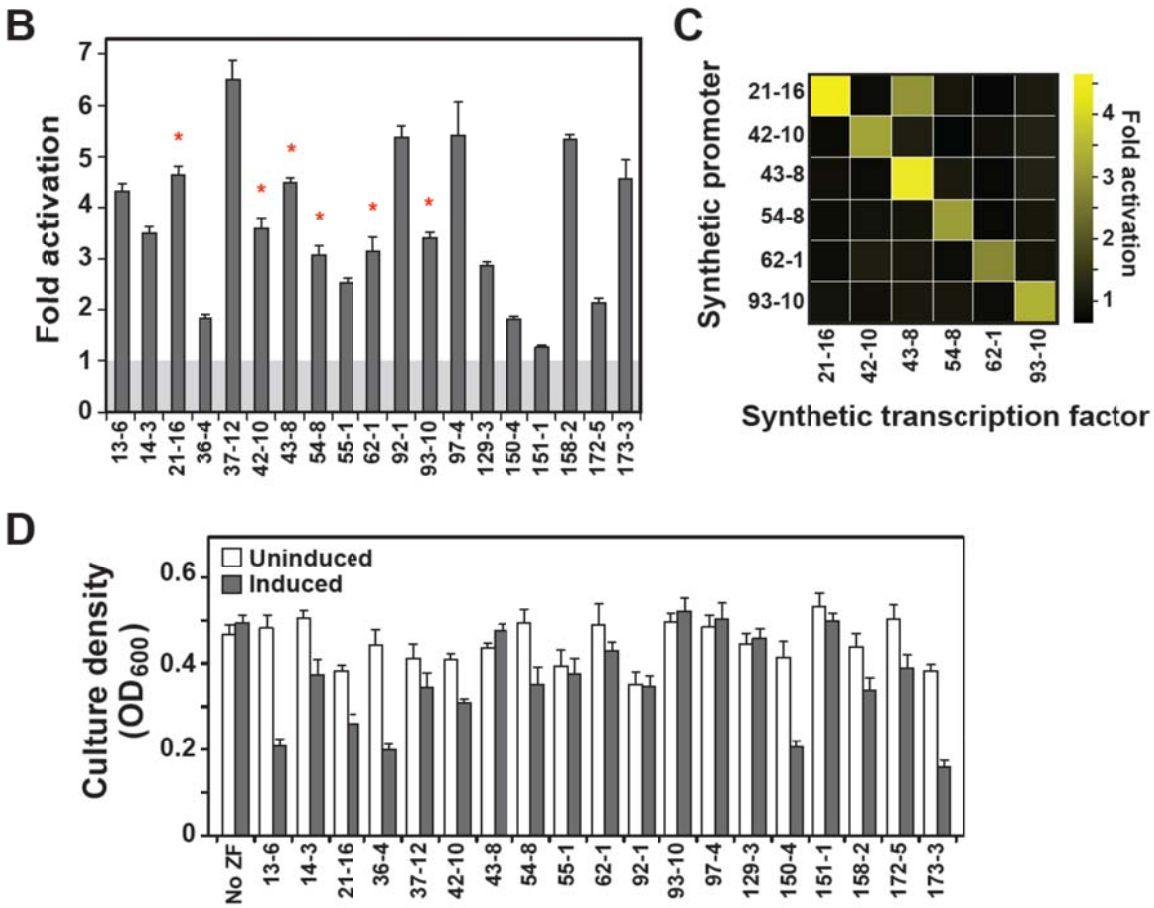
Figure A1.3: Wiring a library of specific and orthogonal transcriptional connections with engineered ZF arrays

(a) sTF-promoter pair library sequences. Amino acid residues of the recognition helices for the 19 OPEN-engineered three-finger arrays, and corresponding DNA binding sequences (ZF binding sequences were inserted between *EcoRI* and *BamHI* sites within synthetic promoters).
(b) sTFs activate transcription from cognate synthetic promoters. "Fold activation" values were calculated as the ratio of fluorescence values from induced cells (500 ng/mL ATc) to those from uninduced cell. Red stars denote the subset of 6 sTF-promoter pairs chosen to test for orthogonality.
(c) sTFs constructed from OPEN-engineered ZFs are orthogonal to one another. $sTF_{43\text{-}8}$ activated non-cognate $Promoter_{21\text{-}16}$ due to the fortuitous creation of a sequence that is significantly similar to the binding sequence of 43-8, when the downstream *BamHI* restriction site is considered (a, blue boxes).
(d) Fitness cost of sTF expression on host cell growth at 30 h after circuit induction ("No ZF" = strain integrated with synthetic promoter and sTF cassette lacking a ZF array). Error bars represent standard deviation of three experiments.

Figure A1.3 (continued):

A

	ZINC FINGER RESIDUES			SYNTHETIC PROMOTER OPERATORS		
sTF	Finger 1	Finger 2	Finger 3	EcoRI	Binding sequence	BamHI
13-6	TNQKLEV	YRHNLQR	QHPNLTR	GAATTC	a GAA GAT GGT g	GGATTC
14-3	APSKLDR	LGENLRR	DGGNLGR	GAATTC	g GAC GAC GGC a	GGATTC
21-16	RNFILQR	QGGNLVR	QQTGLNV	GAATTC	a TTA GAA GTG a	GGATTC
36-4	GRQALDR	DKANLTR	QRNNLGR	GAATTC	c GAA GAC GCT g	GGATTC
37-12	RNFILQR	DRANLRR	RHDQLTR	GAATTC	t GAG GAC GTG t	GGATTC
42-10	TGQILDR	YAHSLKR	DPSNLRR	GAATTC	a GAC GCT GCT c	GGATTC
43-8	RQDRLDR	QKEHLAG	RRDNLNR	GAATTC	a GAG TGA GGA c	GGATTC
54-8	NKTDLGR	RRDMLRR	RMDHLAG	GAATTC	a TGG GTG GCA t	GGATTC
55-1	DESTLRR	MKHHLGR	RSDHLSL	GAATTC	c TGG GGT GCC c	GGATTC
62-1	TGQRLRI	QNQNLAR	DKSVLAR	GAATTC	g GCC GAA GAT a	GGATTC
92-1	DSPTLRR	QRSSLVR	ERGNLTR	GAATTC	a GAT GTA GCC t	GGATTC
93-10	APSKLKR	HKSSLTR	QRNALSG	GAATTC	c TTT GTT GGC a	GGATTC
97-4	RQSNLSR	RNEHLVL	QKTGLRV	GAATTC	a GAC GCT GCT c	GGATTC
129-3	TAAVLTR	DRANLTR	RIDKLGD	GAATTC	a TTA TGG GAG a	GGATTC
150-4	KGERLVR	RMDNLST	RKDALNR	GAATTC	g GTG TAG GGG t	GGATTC
151-1	IPNHLAR	QSAHLKR	QDVSLVR	GAATTC	t GCA GGA GGT g	GGATTC
158-2	DKTKLRV	YRHNLTR	QSTSLQR	GAATTC	t GTA GAT GGA g	GGATTC
172-5	MKNTLTR	RQEHLVR	QKPHLSR	GAATTC	a GGA GGG GCT c	GGATTC
173-3	SAQALAR	QQTNLAR	VGSNLTR	GAATTC	a GAT GAA GCT g	GGATTC



growth of cells with and without the induction of sTFs, under the assumption that unwanted interactions with the host genome would impose a fitness cost on the cells. We observed no adverse or modest effects on growth in the great majority of sTFs from our library (Figure A1.3d, Supplementary Figure A1.3). Our scheme may thus represent a starting point for designing and screening sTFs with optimal functionality and orthogonality within a desired host.

These results show that engineered ZF arrays are effective building blocks for minimal sTF activators, and that DNA interaction specificity is a component property that can be programmed to mediate the construction of specific and orthogonal, synthetic transcriptional connections in yeast. Moreover, largely through this ability to engineer DNA specificity for many interaction partners, our platform is able to make meaningful predictions about orthogonality (among synthetic components and with host machinery), which remains a major unaddressed issue in synthetic biology.

Tuning Transcriptional Output

ZFs are well-studied structural motifs with crystallographic information providing blueprints for harnessing their structure-function relationship to program more complex transcriptional behaviors. We investigated how we could rationally engineer various component properties pertaining to the ZF-DNA interaction to tune transcriptional outputs in our synthetic eukaryotic system. For these studies, we focused on the sTF pair 42-10 and 43-8 ($sTF_{42\text{-}10}$ and $sTF_{43\text{-}8}$) because they activate transcription robustly to similar levels but show orthogonal activities to one another. In addition, these two activators show some distinct properties, e.g., 42-10 seemed to impose a fitness cost on the yeast host, while 43-8 did not (Figure A1.3d).

To tune up the level of transcriptional activation, we focused on alterations to the promoter architecture. We multimerized ZF binding sequences to create promoters with repeat

operators that would correspondingly recruit greater numbers of sTF interactions and thus ADs. With promoters harboring one, two, and eight tandem operators, we observed a corresponding increase in the transcriptional output of the system, confirming that we could tune up the level of activation (Figure A1.4a). Importantly, no cross-activation was observed between these sTFs and any of the tandem synthetic promoters (Supplementary Figure A1.4).

Eukaryotic promoters are known to integrate multiple inputs by binding to distinct TFs. In fact, transcriptional networks may even act as logic gates through such regulation schemes. Our synthetic promoters can similarly be designed to recruit distinct sTFs through architectures that include distinct operators. We constructed a two-input promoter with operators specific for $sTF_{43\text{-}8}$ and $sTF_{42\text{-}10}$. We then directed the expression of $sTF_{43\text{-}8}$ and $sTF_{42\text{-}10}$ under the independent and respective control of TetR- and LacI-controlled *GAL1* promoters, which could in turn be induced by the chemical inputs, ATc and isopropyl-β-D-1-thiogalactopyranoside (IPTG). Upon induction of either or both of the sTF species, we observed transcriptional activation over the uninduced case (Figure A1.4b), confirming that our promoter design can indeed integrate distinct transcriptional signals in Boolean OR-like fashion.

Promoter architecture can be designed to alter the number of sTFs recruited and thus tune transcriptional output strength. An alternative approach is to regulate the ZF-DNA interaction through structure-guided mutation of the ZF backbone to alter non-specific DNA affinity. Along these lines, we targeted four arginine residues outside of the DNA recognition helices that are known, based on structural studies, to mediate non-specific interactions of a three-finger array with the DNA phosphate backbone (Pavletich and Pabo, 1991, Elrod-Erickson et al., 1996). The first arginine residue (position 2) is located upstream of the first β-strand of the amino-terminal finger, while the remaining three (positions 11, 39, 67) are found within the β-sheets of each of

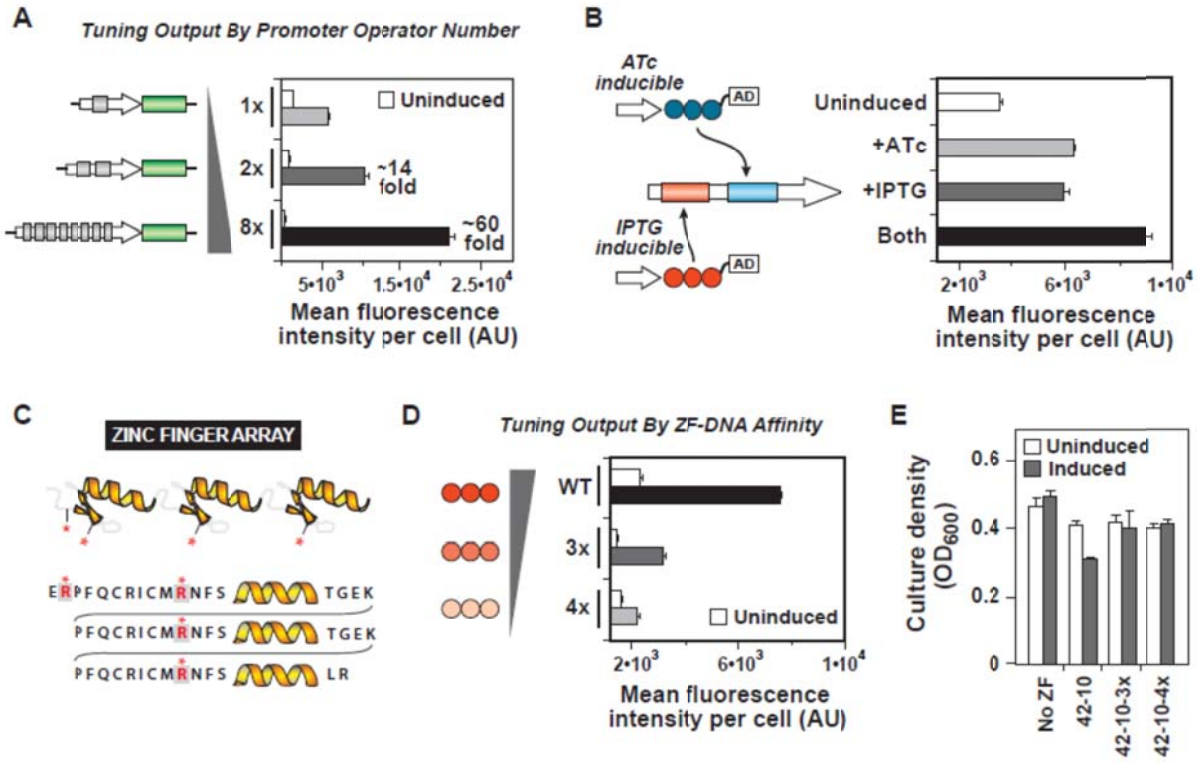


Figure A1.4: Tuning transcriptional outputs by rationally engineering multiple component properties
(a) Tuning up output strength by increasing ZF operator number in synthetic promoter ($sTF_{43\text{-}8}$). (b) Integrating two distinct sTFs at a single synthetic promoter. $sTF_{43\text{-}8}$ and $sTF_{42\text{-}10}$ were expressed independently from ATc- and IPTG-inducible *GAL1* promoters. (c) Schematic representation of the canonical Cys_2-His_2 ZF protein (top). Each finger is composed of two β-strands and a recognition helix, which makes sequence-specific contacts to three DNA bps. Four arginine residues in the ZF framework that mediate non-specific interactions with the phosphate backbone were targeted for mutation to alanine residues (grey boxes and highlighted in red) in order to alter the affinity of the ZF for its cognate binding sequence. (d) Tuning down activation output by engineering ZF affinity variants in $sTF_{42\text{-}10}$ (3x: R2A/R39A/R67A, 4x: R2A/R11A/R39A/R67A). Horizontal axis begins at basal (promoter-only) fluorescence level (B and D). (e) Phosphate backbone mutants of 42-10 rescue the fitness cost of $sTF_{42\text{-}10}$ on host cell growth.

the three fingers, immediately upstream of each recognition helix (Figure A1.4c). The arginine residues mediate non-specific interactions, in part, through their positive charge; thus, we altered each of these to alanine residues.

We screened the DNA-binding activities of ZF arrays possessing various combinations of these four phosphate backbone mutations using a previously described bacterial-two-hybrid

(B2H) system (Wright et al., 2006, Maeder et al., 2008). Single residue alanine substitution mutations yielded modest effects, whereas sets of mutations revealed a step-wise decrease in DNA-binding activity (Supplementary Figure A1.5). We next incorporated the phosphate backbone mutations into $sTF_{42\text{-}10}$, and tested the transcriptional activity of the resulting variants in our synthetic eukaryotic system. We found that transcriptional output could be analogously tuned down as the number of mutations was increased from zero ($sTF_{42\text{-}10}$) to three ($sTF_{42\text{-}10\text{-}3x}$) to four ($sTF_{42\text{-}10\text{-}4x}$), in effect creating weaker-activating sTF variants from the lower-affinity variants (Figure A1.4d). We additionally investigated the effects the weaker-activating sTFs have on the yeast host. In the fitness growth assay, we found that the phosphate backbone mutations were able to systematically rescue the growth inhibition observed with $sTF_{42\text{-}10}$ (Figure A1.4d; Supplementary Figure A1.3). We presume that this effect occurs because the mutant ZF proteins are less able to mediate off-target DNA interactions that may be at the root of the fitness cost.

Taken together, these results demonstrate that the rational engineering of ZF binding sites in the promoter architecture and the ZF-DNA binding interaction — two component properties of our synthetic system — provide effective strategies for tuning transcriptional output.

Engineering Cooperative Transcriptional Systems From Weakly-Interacting Components

The assembly of TFs into multimeric complexes is a mechanism for achieving cooperativity and shaping input-output responses to regulate transcription. Inspired by natural cooperative systems, we next sought to assemble sTFs into multimeric complexes that could achieve synergistic transcriptional behaviors (Ptashne and Gann, 2002). To do so, we harnessed PDZ interaction domains from metazoan cells. These domains are naturally responsible for organizing intracellular signaling complexes, so we explored whether they could be utilized to

assemble and organize our synthetic factors in transcriptional applications. Because these domains are modular, they provide an additional tunable component property to our framework and allow for generalizable designs for multimerization. Furthermore, canonical PDZ domains are extremely rare in non-metazoans (Harris and Lim, 2001), and are therefore unlikely to interact with endogenous yeast machinery.

The weakly-activating sTFs represent ideal components with which to demonstrate multimerization and the synthetic construction of transcriptional cooperativity. These sTFs were built from ZF mutants with lower non-specific DNA binding activity. We therefore investigated whether their assembly could stabilize a protein-DNA complex to better initiate transcription, presumably by slowing the off-rate of each component from the bound promoter. The well-studied PDZ domain from the mammalian protein □1-syntrophin (Craven and Bredt, 1998, Harris et al., 2001) was fused to the C-terminus of 43-8-4x, and its cognate peptide ligand (GSGS-VKESLV), to which it binds with low micromolar affinity, was fused to the C-terminus of 42-10-4x. For these studies, only $sTF_{43\text{-}8\text{-}4x}$ carried an AD, making it a single locus for the recruitment of transcription initiation machinery. We did not attach an AD to the dimeric partner ($sTF_{42\text{-}10\text{-}4x}$) in order to test whether this additional (non-activating) factor could stabilize the complex through dimerization and thus aid in the cooperative activation of transcription by $sTF_{43\text{-}8\text{-}4x}$. In addition, we generated a non-binding partner by fusing a mutated form of the cognate ligand (GSGS-VKEAAA) to 42-10-4x. Expression of each 'monomeric' sTF was driven by independently inducible *GAL1* promoters. Upon induction, the sTFs operate on a synthetic 'dimeric' promoter to activate downstream transcription (Figure A1.5).

We titrated expression of the PDZ-harboring $sTF_{43\text{-}8\text{-}4x}$, both in the presence and absence of its ligand-carrying partner. In the resulting dose responses, we observed a significant

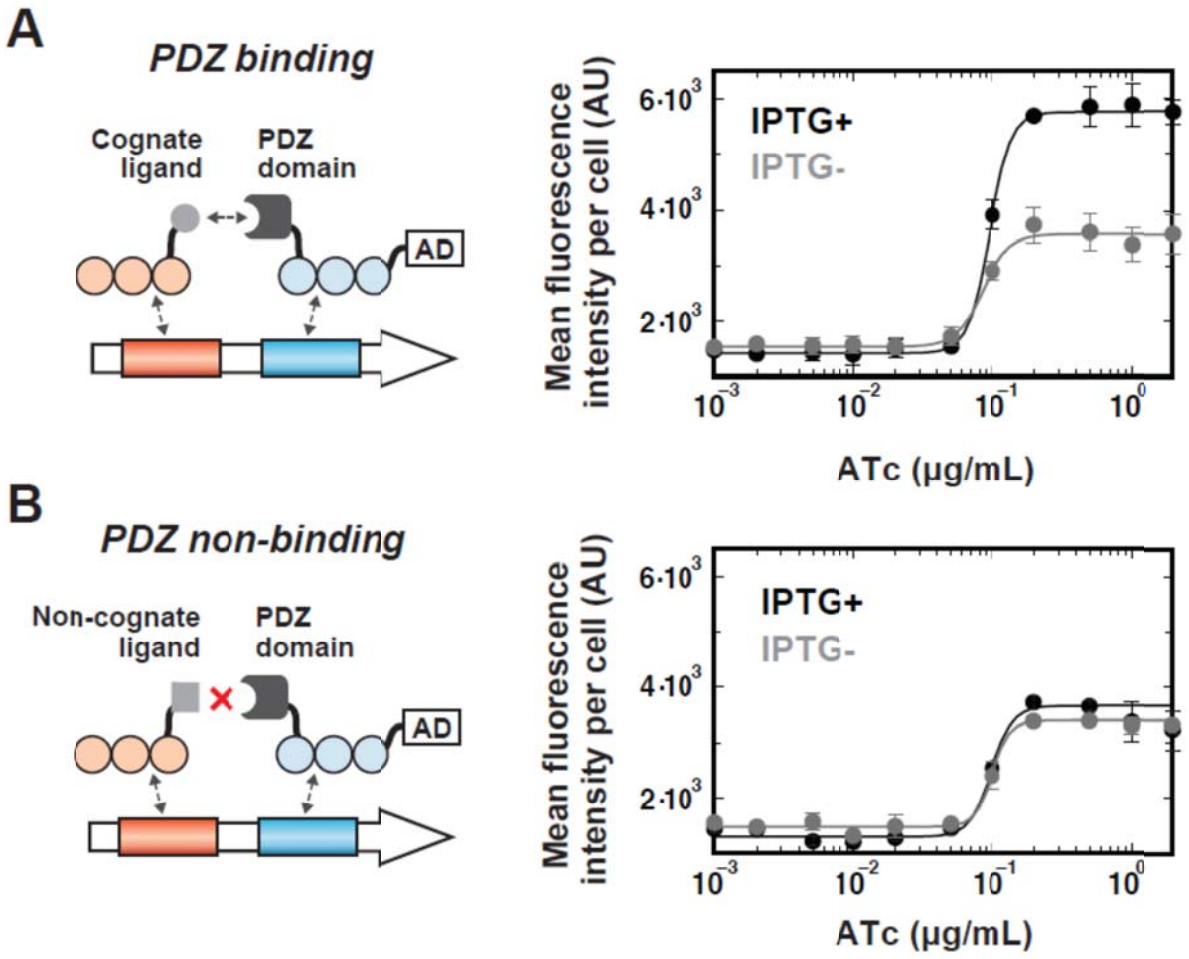


Figure A1.5: Transcriptional cooperative systems can be engineered from weakly-activating sTF 'monomers' that are dimerized with a PDZ interaction domain
(a) The dimerization interaction promotes cooperative behavior in transcriptional activation. Syntrophin PDZ domain (dark grey) was fused to the C-terminal of ZF affinity mutant 43-8-4x, and the resulting AD-carrying sTF 'monomer' was expressed from ATc-inducible *pGAL1*. The heterologous ligand (light grey) was fused to the C-terminal of 42-10-4x, and the resulting AD-less factor was expressed from IPTG-inducible *pGAL1*. The factors assemble at a synthetic 'dimeric' promoter to cooperatively activate downstream transcription ("IPTG+" = full induction with 20 mM IPTG). (b) Disruption of the dimerization interaction abolishes cooperative behavior in transcriptional activation. A non-binding ligand variant (GSGS-VKEAAA) was instead fused to 42-10-4x.

synergistic effect on the transcriptional output of $sTF_{43\text{-}8\text{-}4x}$ when the partner factor was present, as compared to when it was not present (Figure A1.5a). Critically, we observed no cooperative effect on the transcriptional output of the system in the analogous dose response with the non-binding ligand partner (Figure A1.5b). These results suggest that the monomeric sTFs, aided by

engineered protein-protein interactions, cooperate to recruit and stabilize one another at the synthetic promoter, thereby increasing the promoter occupancy and the resulting transcriptional activity of the complex.

Engineering Diverse Two-Input Signal Processing Behavior

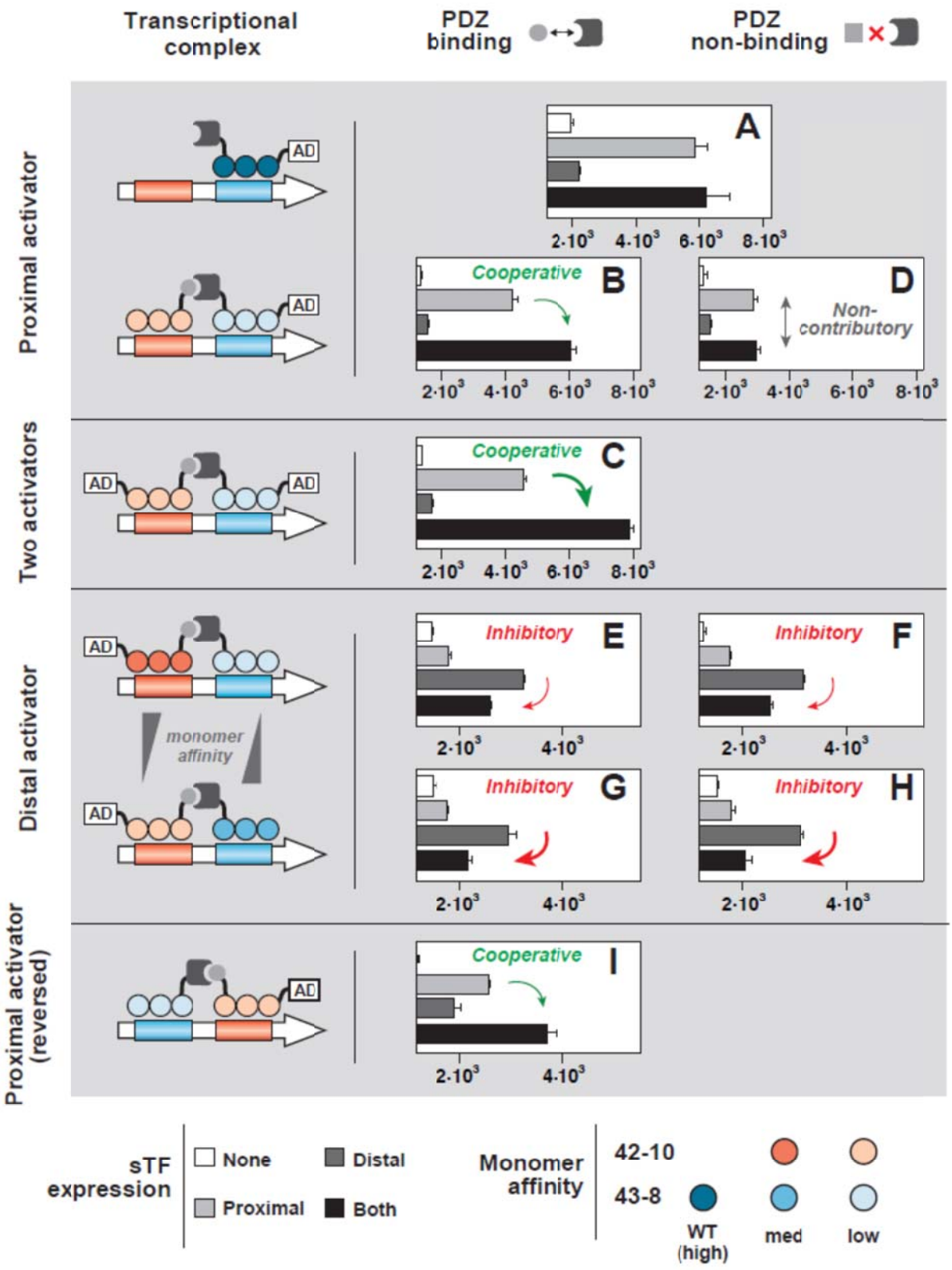
With this synthetic framework, we can construct and study cooperative transcriptional assemblies. Additionally, we can explore cooperativity and other complex transcriptional behaviors within the context of signal integration, as a result of having synthetic control over combinatorial inputs and components. We thus harnessed the various control 'knobs' of our framework to engineer and explore signal processing behavior for synthetic two-input promoters.

These studies were enabled by the design of 'dimeric' promoters harboring distinct operators for ZFs 43-8 and 42-10, and the independently-controlled expression of two customizable sTF cassettes by the chemical inputs ATc and IPTG (Figure A1.6). Induction of a single sTF (PDZ-carrying $sTF_{43\text{-}8}$) by the addition of either the input controlling its expression or both inputs resulted in robust and equal levels of transcriptional output from the two-input promoter (Figure A1.6a). We utilized this system to engineer a variety of transcriptional input combinations. Our previous cooperativity results (Figure A1.5) established an interesting starting point for investigating how a pair of transcriptional signals can be integrated. So, we first used the dimerizing sTFs, constructed from ZF mutants 43-8-4x and 42-10-4x. The $sTF_{43\text{-}8\text{-}4x}$ activator was directed to the operator closest to the downstream gene's start codon (proximal position), and the AD-less partner monomer to the distal position (Figure A1.6, "Proximal activator" architecture). When the distal monomer was engineered to carry the heterologous PDZ ligand, we observed cooperative-like amplification in transcriptional output in the presence of both inputs (Figure A1.6b). In this case, the distal monomer participates in binding to and stabilizing

Figure A1.6: Synthetic ZF-based transcription framework can be used to engineer diverse two-input behaviors

(a) The transcriptional operation of a single $sTF_{43\text{-}8}$ (carrying a PDZ domain) at the proximal position of a two-input promoter. (b) Cooperative two-input synergy engineered with PDZ-carrying $sTF_{43\text{-}8\text{-}4x}$ as the proximal activator and cognate ligand-carrying $sTF_{42\text{-}10\text{-}4x}$ as the distal partner. (c) Cooperative two-input synergy further enhanced by the addition of an AD onto the distal partner to create a two-activator system. (d) A 'null' two-input system engineered by abolishing the dimerization interaction with a PDZ non-binding ligand on the distal partner, thus rendering it non-contributory. (e) Inhibitory two-input behavior engineered by reversing the activator location (from proximal to distal) and using either PDZ binding or (f) non-binding ligands. (g) Inhibition by the proximal monomer can be further increased by increasing the proximal ZF affinity to DNA (43-8-4x to 43-8-3x) and decreasing the distal ZF affinity to DNA (42-10-3x to 42-10-4x) in both PDZ binding and (h) non-binding cases. (i) By reversing the orientation of the operators, $sTF_{43\text{-}8\text{-}4x}$ is converted from an inhibitor to a cooperative factor to, once again, obtain cooperative transcriptional synergy in the two-input behavior. All sTFs were expressed from either ATc- or IPTG-inducible *pGAL1* (500 ng/mL ATc and/or 20 mM IPTG). Horizontal axes correspond to "Mean fluorescence intensity per cell (AU)" and begin at basal (promoter-only) fluorescence level.

Figure A1.6 (continued):



the proximal sTF activator at the promoter to enhance transcription. Interestingly, the two-component complex achieved transcriptional output levels similar to those of the single WT activator (Figure A1.6a), but only through the addition of both inputs and a total DNA operating specificity of 18 bp rather than 9 bp. Furthermore, we found that we could boost this effect by adding an AD onto the distal monomer, thus engineering a two-activator system and providing another source for transcriptional machinery recruitment at the promoter (Figure A1.6c).

The PDZ-mediated sTF dimerization therefore serves as a key component property for enabling this type of synergistic two-input behavior. By simply modifying the ligand to abolish the binding interaction (i.e., mutating it to the non-cognate GSGS-VKEAAA), we rendered the distal monomer transcriptionally non-contributory in the proximal activator scenario, and subsequently engineered a different two-input behavior: one that shows equal output levels in the presence of either both inputs or the input directing the proximal activator (Figure A1.6d). In other words, we created a two-input system with a 'null' effect when both inputs are present.

We next sought to reverse the monomer roles by simply switching the placement of the AD. We loaded the AD onto ligand-carrying $sTF_{42\text{-}10\text{-}3x}$ and removed it from PDZ-carrying $sTF_{43\text{-}8\text{-}4x}$, while directing the two monomers at the same 'dimeric' promoter (Figure A1.6, "Distal activator" architecture). Strikingly, we did not observe transcriptional output synergy. Instead, we observed an inhibition of the output signal in the presence of both inputs (Figure A1.6e). Furthermore, the inhibitory behavior was conserved in both PDZ-binding and non-binding cases (Figure A1.6f). These results suggest that, with this particular combination of components in the distal activator scenario, the proximal monomers take on an inhibitory as opposed to a cooperative role. If this were the case, then we reasoned that we should be able to further strengthen the inhibition by increasing the ZF affinity of the inhibitory monomer (43-8-

4x) to its operator and decreasing the ZF affinity of the distal activator (42-10-3x). Indeed, when we replaced ZF mutants 43-8-4x and 42-10-3x with 43-8-3x and 42-10-4x, respectively, we observed a commensurate decrease in transcriptional output in the presence of both inputs down to near baseline levels (Figure A1.6g-h). These results further suggest that, through this slight change to the complex architecture (proximal activator to distal activator scenarios), the AD-less partner monomer has shifted its transcriptional function from cooperative to inhibitory, in effect completely transforming the logical behavior of the system.

Finally, we expected that flipping the orientation of the operators, such that the 42-10 operator was placed in the proximal position, could 'rescue' the cooperative behavior (Figure A1.6, 'Proximal activator (reversed)' architecture). Indeed, with a reversed dimeric promoter, we once again observed a cooperative enhancement in the system's output in the presence of both inputs as compared to that of the single inputs (Figure A1.6i). In effect, this change served to transform the transcriptional role of the 43-8-based species from inhibitor to a cooperative factor.

Taken together, our results demonstrate that sTF monomers can be customized to different roles (e.g., activating, cooperative, non-participatory, inhibitory) within a simple two-input system through the rational perturbation of component properties made accessible by our synthetic framework. These distinct roles can differentially shape the signal-processing behavior at a promoter. The results also highlight the importance in how a promoter's geometry couples TF recruitment and binding to a downstream transcriptional behavior (Ptashne, 1986).

Discussion

Synthetic approaches to understand, rewire, and construct higher-order transcriptional networks, particularly in eukaryotes, have been severely hindered by a lack of reliable

components and a framework for designing and assembling them. We have developed an extensible synthetic biology framework for regulating eukaryotic transcription, whereby artificial ZF proteins are used as core building blocks from which to construct complex transcription functions and circuitry. The use of a context-dependent ZF selection scheme allows us to rapidly alter and program the ZF-DNA interaction specificity, and identify orthologous pairs of ZF arrays-DNA sites that can be engineered into sTFs for wiring networks within yeast. This work brings new forms and levels of connectivity to synthetic transcriptional circuits, beyond that which is achievable with the few, classical prokaryotic TF-promoter pairs. Using our methodology, one should be able to create a virtually unlimited number of sTF-promoter pairs, with which to make transcriptional circuit connections. In this regard, we note that three-finger arrays have been engineered for more than 500 different nine bp sites using the OPEN (Maeder et al., 2008) and Context-Dependent Assembly (CoDA) (Sander et al., 2011b) methods (J.K. Joung and colleagues, unpublished data).

Engineered Transcription Activator-Like (TAL) effectors have recently emerged as an important alternative to customized ZFs for programming DNA specificity. Naturally occurring TAL effectors encoded by *Xanthamonas* bacteria bind to target DNA sequences using arrays of highly conserved 33-35 amino acid repeat domains. A single TAL effector repeat binds to one nucleotide of DNA with specificity of binding associated with the identities of amino acids at two positions known as Repeat Variable Di-residues (RVDs). TAL effector repeats bearing different RVDs have been described for specific binding to each of the four possible DNA nucleotides, and these repeats can be simply joined together to create arrays capable of binding to extended target DNA sequences (Bogdanove and Voytas, 2011). The simplicity and modularity of the TAL effector repeats suggests that nearly any DNA sequence should be

targetable, and recent work has demonstrated that engineered TAL effector nuclease (TALEN) fusions can be robustly generated for a wide variety of different sequences (Reyon et al., 2012). Nonetheless, engineered ZFs have several important advantages, including their considerably smaller size, their less repetitive coding sequence (potentially important for packaging into viral vectors), and a greater understanding of their biochemical properties, structure, and function, which is important for creating variations in affinity, specificity, and tunability.

The use of artificial ZF domains enables a fully decomposed design of a sTF, for which the molecular component properties are accessible, modular, and tunable. With these new components, we constructed a synthetic transcriptional cascade in yeast and used it as a test bed for rationally customizing the component properties to program *in vivo* behaviors. Specifically, we showed that systematic construction of complex transcriptional functions can be achieved by the independent control of a set of key component properties. For instance, we tuned the strength of transcriptional outputs through modifications of ZF binding sites in the promoter architecture as well as through structure-guided modifications to the ZF protein to alter ZF-DNA interactions.

Our framework additionally provides the ability to engineer and tune transcriptional cooperativity. To date, there exists no simple way of building cooperative transcriptional systems, even though their importance is well-documented in both natural and synthetic gene regulation. As a result, in most synthetic studies, researchers have used TFs with integrated, cooperative properties. In contrast, our framework establishes a modular technique for constructing cooperative transcriptional activation schemes *de novo*, through the multimerization of weakly-activating ZF-TFs using low-affinity protein interaction domains (i.e., PDZ domains). This has important consequences for constructing higher-order complexes that more accurately mimic eukaryotic transcriptional regulation schemes, lead to sharper switch-like responses, and

modulate cooperativity within circuits. Indeed, multimerization and cooperativity are ubiquitous molecular regulation schemes that underlie complex gating and decision-making in cells. For example, the yeast *GAL1* promoter is able to integrate coactivator proteins in specific temporal order by utilizing the cooperativity of certain interactions to gate subsequent recruitment events (Bryant and Ptashne, 2003). It is of great interest to understand how activators function cooperatively to assemble specific initiation complexes and regulate transcription. Our bottom-up and modular approach to transcriptional cooperativity could be used to synthetically recapitulate such phenomena so as to study these fundamental mechanisms of regulation. This type of approach has been used to understand transcriptional synergy in prokaryotes (Joung et al., 1993, Joung et al., 1994), and our platform should now enable this strategy to be used to model more combinatorially complex eukaryotic promoters. Furthermore, our extensible and modular framework for multimeric and cooperative transcriptional systems may allow for the implementation of expanded computational operations in eukaryotes, such as logic devices with vastly more input possibilities.

Previous reports have described various frameworks for creating dimeric ZF proteins. In all of these studies, elements derived from naturally occurring TFs (Pomerantz et al., 1998, Wolfe et al., 2000) or ones selected from combinatorial peptide libraries (Wang and Pabo, 1999, Wang et al., 2001) were used to dimerize two-finger units. A disadvantage of this strategy is that a dimerized two-finger complex would have a maximum specificity of 12 bp (assuming that each of the four fingers in the dimer specifies three bp). Our approach differs by utilizing three-finger monomers that have had their binding activities reduced by mutagenesis of non-specific phosphate-contacting residues. This strategy creates dimeric proteins that can have specificities as high as 18 bps, a sequence long enough to be potentially unique in a mammalian genome.

Furthermore, the use of modular protein-protein interaction domains for multimerization is advantageous for various reasons. For example, the interaction is orthogonal and tunable, allowing us to 'match' affinities of all the component interactions making up the sTF complex, and it provides extensibility for further expanding the complex architecture and dynamically increasing or decreasing the DNA-substrate specificity of the complexes.

We also showed that complex signal processing functions can arise when control of TF cooperativity is combined with the ability to engineer promoters of multimerized ZF binding sites. Cellular signal processing is a mechanism by which environmental and other signals are integrated to modulate transcription and thus critical cellular processes, such as growth and stress responses. We constructed a simple, synthetic two-input transcriptional system that allowed us to decompose contributions from the sTF component properties to the system's processing behavior. We showed that, with the same two core TFs and promoter operator sites, a cell could process and integrate signals in a variety of ways. For example, subtle changes, such as reversing promoter operators and disrupting protein-protein interactions, can have striking effects on the output of the system. This led to the construction of not just varied digital logic behavior, but a range of analog tunability. In an inhibitory system (Figure A1.6e-h), we arrived at an interesting Boolean logic gate that produced a positive signal only in the presence of a single input: A>B (A does not imply B). A broad observation from our studies was that specific perturbations to an sTF's component properties (DNA affinity, multimerization with other species, location of operator, etc.) could allow it to convert between different transcriptional roles within the complex, such as activator, cooperative factor, non-contributory, and inhibitor. This synthetic approach could be utilized to explore the diversity of behaviors that can be programmed by even just a few transcriptional components; furthermore, our findings provide simple strategies for

reprogramming the signal processing behavior of a cell. Similar strategies are undoubtedly employed naturally, where there are many examples of individual proteins that can take on either activating or repressing roles depending on the cellular and environmental states (Ma and Ptashne, 1988, Rubin-Bejerano et al., 1996, Chandarlapaty and Errede, 1998, Maxon and Herskowitz, 2001, Kassir et al., 2003).

Given that TFs containing ZFs play a central role in eukaryotic promoter regulation (Pabo et al., 2001), our system represents a promising means for synthetic recapitulation of eukaryotic promoter function, and thus will significantly enhance our ability to construct synthetic gene networks in mammalian cells, an area of tremendous potential in synthetic biology (Weber and Fussenegger, 2009). Indeed, yeast may serve as a well-characterized testbed for the design and validation of synthetic gene circuits that can be subsequently ported to higher organisms. Synthetic transcriptional regulators based on the sTFs described here can be used to create classifier circuits to identify cell state (Nissim and Bar-Ziv, 2010, Xie et al., 2011), memory devices to record cellular events, and logic gates for cellular processing (Kramer et al., 2004a), which can aid in the study and control of stem cell differentiation, therapeutics, and complex human diseases. Additionally, ZF-based proteins have been shown to be powerful targeting elements of endogenous genomic loci in many mammalian cells, including cancer, immune, and stem cells. These proteins include ZF nucleases (Zou et al., 2009), which are being tested in therapeutic applications for modifying/disrupting disease-causing alleles and genes, and artificial TFs (Blancafort et al., 2005), which can be used to modify the expression of native genes for reprogramming purposes. Together with these ZF technologies, our work may lead to the construction of integrated gene circuits – artificial circuitry that seamlessly and specifically

integrates into endogenous gene networks and/or leads to the modification of endogenous genes – for more dynamical and sophisticated genetic control in cell-based therapeutic applications.

Synthetic biology is helping us to understand how organisms behave and develop through the forward engineering of molecular circuitry with well-understood genetic components (Elowitz and Lim, 2010). The present work expands the synthetic biology toolkit with new genetic components, beyond re-purposed bacterial transcriptional components, to program eukaryotic cells. Additionally, it provides a bottom-up framework for exploring the complexity of eukaryotic promoters and their combinatorial regulation by TF complexes and circuitry. This framework can be a starting point for determining the transcriptional components, modules, and circuitry needed to implement the sophisticated behaviors that control the development and function of eukaryotic cells.

Methods

Strains and Media

S. cerevisiae YPH500 (*α, ura3-52, lys2-801, ade2-101, trp1Δ63, his3Δ200, leu2Δ1*) (Stratagene) was used as the host strain in all yeast experiments, and plasmid chromosomal integrations were specifically targeted to the redundant *ura-52* locus. Culturing, genetic transformation, and verification of transformation were done as previously described (Murphy et al., 2007), using either the *URA3*, *TRP1*, or *LEU2* genes as selectable markers.

Plasmid Construction

Synthetic promoter plasmids were constructed from integrative plasmid pRS406 (Stratagene) by cloning ZF binding sequences (BS) directly upstream of the *CYC1* minimal

promoter (*pCYC1*) TATA box. The corresponding ZF-activated promoter drives the expression of a yeast enhanced green fluorescent protein (yEGFP) (Cormack et al., 1997), which is preceded by a Kozak consensus sequence.

sTF circuit plasmids were constructed from the previously described yeast integrative plasmids pTPG1 (TX: TetR-regulated control promoter) and pLOG1 (LX: LacI-regulated control promoter) (Ellis et al., 2009). Briefly, these plasmids consist of a *GAL1* upstream activation signal (UAS) followed by either a TetR- or LacI-regulated wild-type *GAL1* promoter, which drives the expression of our sTF cassettes; the strong constitutive *TEF1* promoter also directs the expression of yeast codon-optimized versions of TetR (Tn*10.*B tetracycline repressor) and the *E. coli* Lac inhibitor (LacI). Constitutive expression of the repressors ensures low basal levels of expression of our sTF cassettes from the engineered *GAL1* promoter, which can be relieved by the respective addition of the chemical inputs, anhydrotetracycline (aTc) and isopropyl-β-D-1-thiogalactopyranoside (IPTG), to the medium.

The sTF circuit cassette, which was synthesized (DNA2.0) and cloned as a *KpnI/XhoI* fragment into pTPG1 and pLOG1, enables the molecular customization and expression of synthetic ZF-TFs. It consists of an open cloning site (*NheI*/*BglII* or *XbaI/BamHI*) for engineered ZF arrays. Upon insertion of a ZF gene, the resulting minimal sTF becomes (N- to C-terminal): 3xFLAG – nuclear localization sequence (NLS) – VP16 minimal activation domain (AD) (Beerli et al., 1998) – ZF array. All ZF genes were codon-optimized, individually synthesized (IDT), and cloned as *XbaI*/*BamHI* fragments into the cassette. Protein-protein interaction domains, namely, syntrophin PDZ domain and peptide ligands, were added as C-terminal fusions to the sTF, separated by a short GSGS linker, and cloned from synthesized, codon-optimized gene fragments (DNA2.0).

All plasmids were constructed and used to transform *E. coli* to harvest DNA for yeast transformations, as previously described (Murphy et al., 2007).

Determination of ZF Protein Activity by Bacterial-2-Hybrid (B2H) Reporter Assay

Activity of ZF proteins was quantified using the B2H reporter assay as described previously (Wright et al., 2006, Maeder et al., 2009). Briefly, ZF protein monomers were cloned as *XbaI*/*BamHI* fragments into the B2H expression plasmid (pGP-FF, Addgene plasmid 13480) to generate ZF-Gal11P activation domain fusion proteins. Each ZF-Gal11P fusion was introduced along with a low-copy alpha-Gal4 hybrid protein into a bacterial reporter strain harboring a single-copy reporter (pBAC-lacZ; Addgene plasmid 13422), modified to encode the corresponding zinc-finger binding site upstream of a weak promoter controlling *lacZ* expression as previously described (Maeder et al., 2009). The DNA-binding activities of ZF proteins were determined by measuring the expression level of β-galactosidase as described previously (Thibodeau-Beganny and Joung, 2007). The results were normalized by dividing the number of units measured in the presence of the zinc finger monomer by those in its absence, which is referred to as the "Fold activation relative to Gal11P" value.

Induction Experiments

Single yeast colonies for each strain were picked and used to inoculate 500 µL of SD-Glu (synthetic drop-out media containing 2% glucose with selectable amino acid mixtures) in Costar 96-well assay blocks (V-bottom; 2mL max volume; Fisher Scientific). The cultures were grown at 30°C with 900 rpm shaking for 24–48 h. A triplicate set of 500-µL YEP-Gal (yeast extract peptone media containing 2% galactose) cultures, with and without appropriate inducers (e.g., ATc and/or IPTG) (Ellis et al., 2009), were inoculated to an OD_{600} of ~0.08–0.1 and grown at

30°C with 900 rpm shaking for ~14–16 h. Cells were then treated with cycloheximide to inhibit protein synthesis, and then assayed for *yEGFP* expression by flow cytometry.

Flow Cytometry and Data Analysis

For all data, we acquired 5,000–10,000 events using a BD LSRFortessa equipped with a High Throughput Sampler (BD Biosciences), running samples on a medium flow rate. Events were gated by forward and side scatter. The geometric means of the fluorescence distributions were calculated. The autofluorescence value of *S. cerevisiae* YPH500 cells harboring no genomic integrations was subtracted from these values to give the fluorescence values reported in this study. "Fold activation" values were calculated as the ratio of fluorescence values from induced cells to those from uninduced cells.

Growth Assays

Growth assays were performed similarly to induction experiments, except that experimental cultures were inoculated to an OD_{600} of ~0.03–0.05 and grown at 30°C with 900 rpm shaking for 30 h. OD_{600} measurements were taken using a SpectraMax M5 fluorescence microplate reader (Molecular Devices) using culture volumes of 100 µL. A "No ZF" control – a strain engineered with synthetic promoter and sTF cassette lacking a ZF array – was assayed in parallel.

Acknowledgements

We thank members of the Collins Lab for helpful discussions, Carl O. Pabo for helpful suggestions on reducing non-specific affinity of zinc fingers, and Katie M. Flynn for help with artwork. This work was supported by the Howard Hughes Medical Institute (J.J.C.), NIH Director's Pioneer Awards DP1 OD003644 (J.J.C.) and DP1 OD006862 (J.K.J.), an Office of

Naval Research Multidisciplinary University Research Initiative (MURI) grant (T.K.L.), a Defense Advanced Research Projects Agency grant (DARPA-BAA-11-23), and a National Science Foundation Graduate Research Fellowship (C.L.R.).

Appendix 1 Supplementary Materials:

Supplementary Figure A1.1: Demonstration of yEGFP fluorescence measurements as a reliable indicator of steady-state transcriptional circuit output

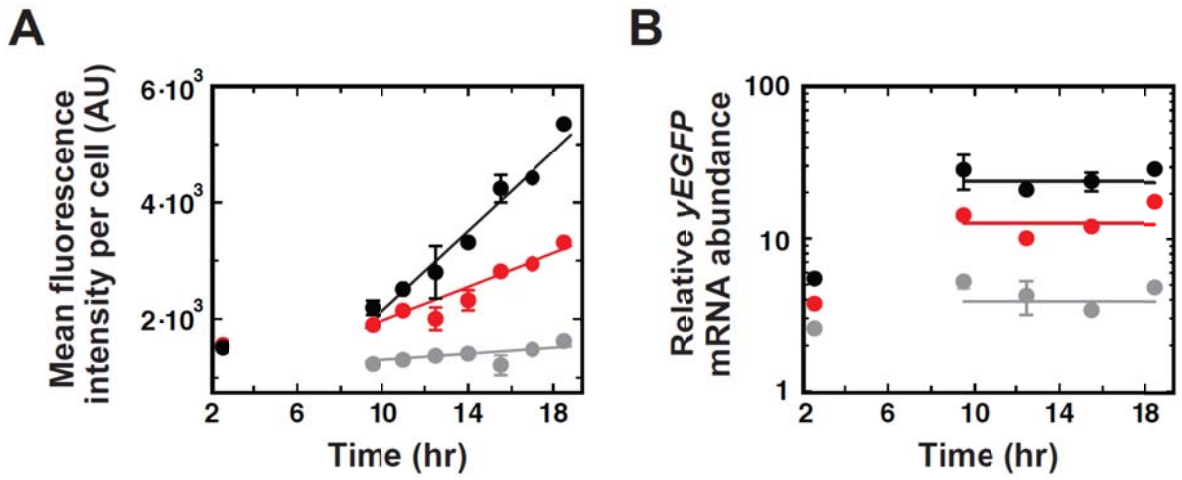
Supplementary Figure A1.2: Flow cytometry data of transcriptional activation by a library of synthetic transcription factor (sTF)-promoter pairs

Supplementary Figure A1.3: Growth assays for a representative set of sTF circuit strains

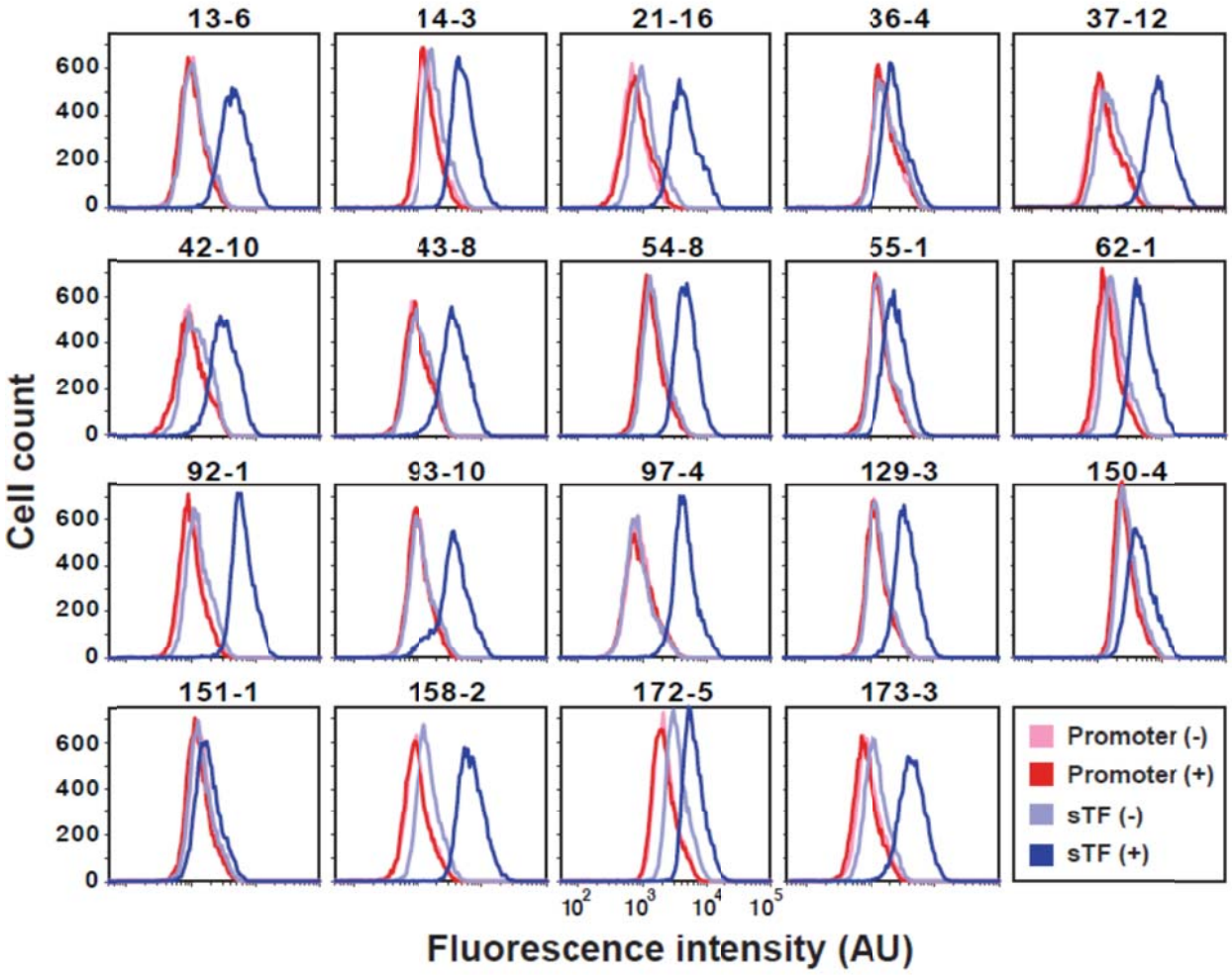
Supplementary Figure A1.4: sTFs constructed from OPEN ZFs operate orthogonally to one another

Supplementary Figure A1.5: DNA-binding activities of engineered ZF proteins possessing various combinations of non-specific affinity mutations

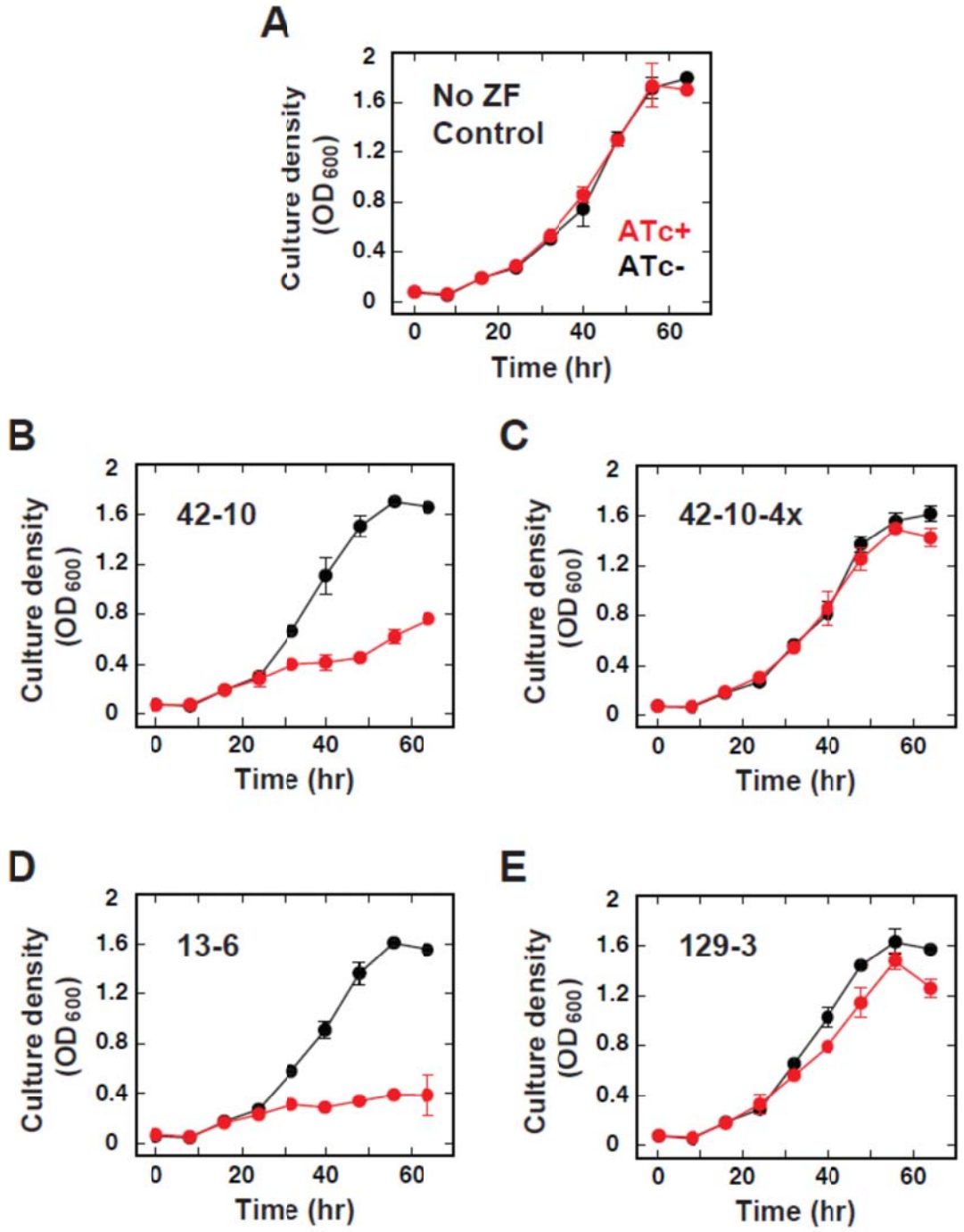
Appendix 1 Supplementary Methods



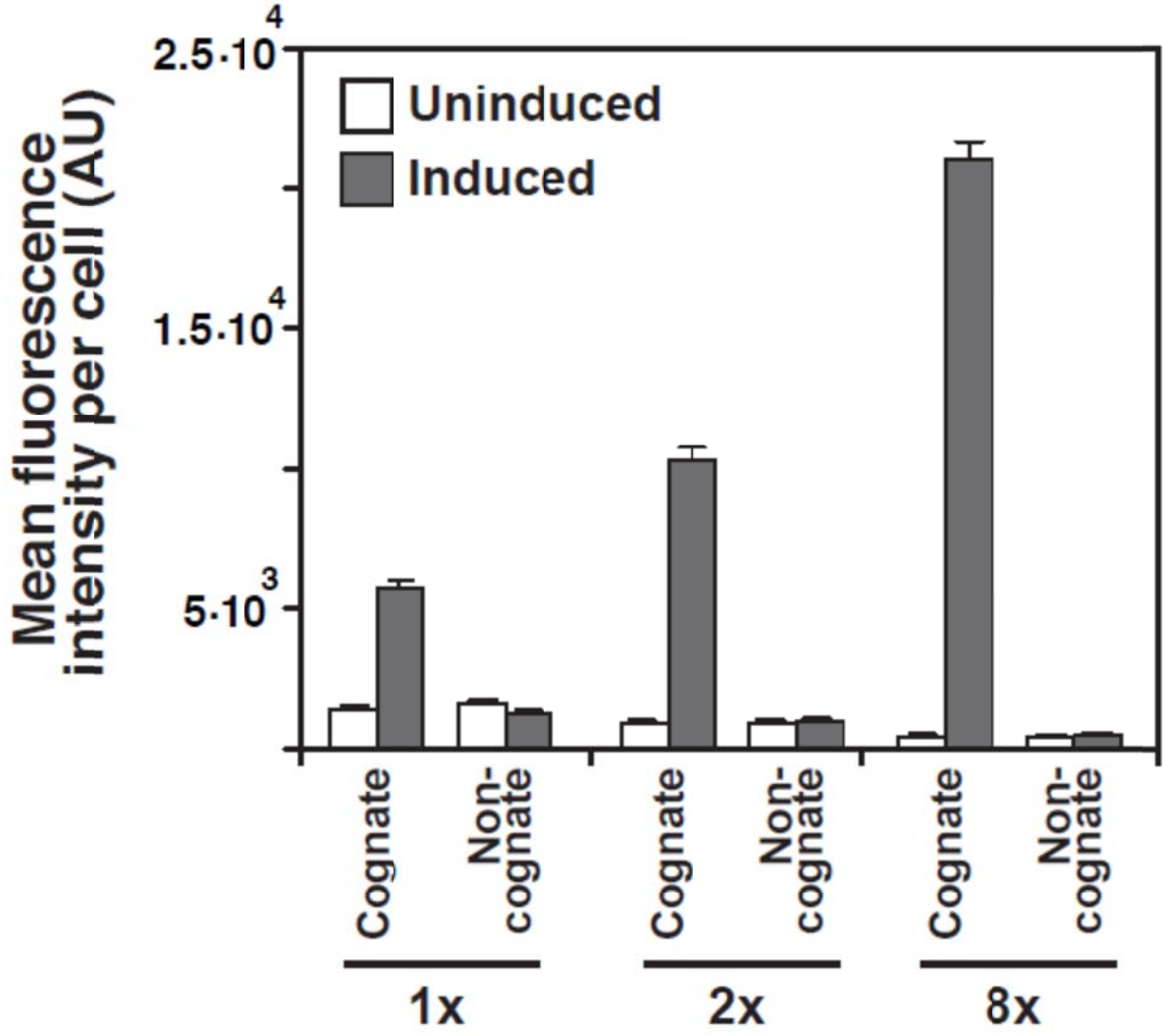
Supplementary Figure A1.1: Demonstration of yEGFP fluorescence measurements as a reliable indicator of steady-state transcriptional circuit output
(a) yEGFP fluorescence values were measured by FACS for the circuit comprised of $sTF_{43\text{-}8}$ for a time course from 9.5 – 18.5 h after circuit induction. The circuit exhibited a linear rate of yEGFP accumulation for the entirety of the time course at three different induction levels: 500 ng/mL ATc (black), 175 ng/mL ATc (red), 0 ng/mL ATc (grey).
(b) In order to demonstrate that the observed linear rate of yEGFP accumulation corresponded to a steady state in transcriptional circuit output, RT-PCR measurements were made of relative mRNA abundance of yEGFP transcript. Between 9.5 and 18.5 h, transcript levels did not vary substantially, and proportional steady-state values of transcript were observed for the three induction conditions. Points represent mean values for three experiments ± s.e.m. Taken together, these results indicate that, for the selected time course, yEGFP accumulates in a linear manner in proportion to a steady-state level of transcript. This argues that FACS-based yEGFP endpoint measurements (14 – 16 h post-induction) that are used throughout the study are a reliable indicator of circuit transcriptional output.



Supplementary Figure A1.2: Flow cytometry data of transcriptional activation by a library of synthetic transcription factor (sTF)-promoter pairs
yEGFP expression for cells harboring only synthetic promoters, uninduced (light red) and induced (red), and for cells harboring full sTF circuits, uninduced (light blue) and induced (blue).

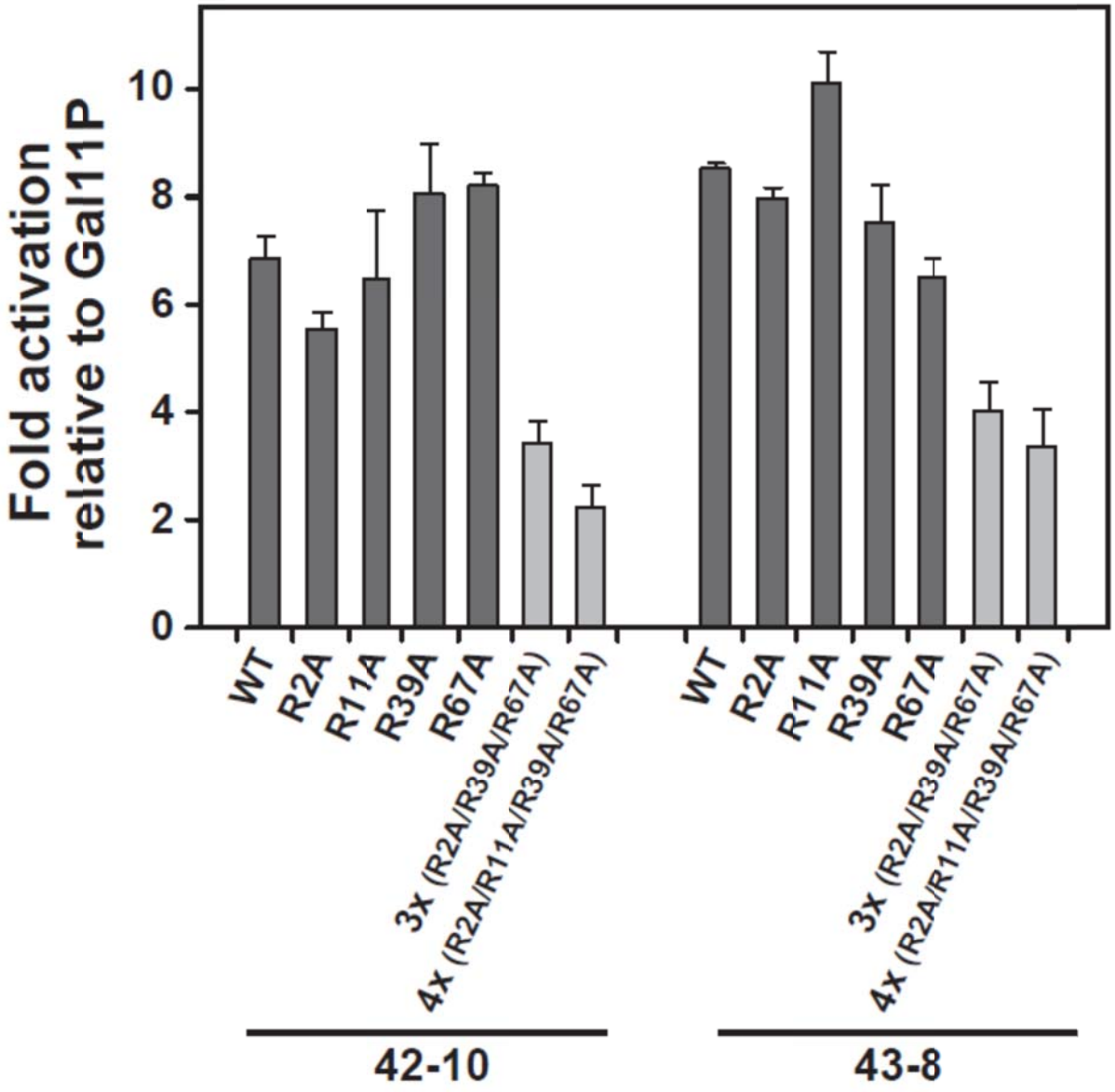


Supplementary Figure A1.3. Growth assays for a representative set of sTF circuit strains Longer time-course growth assays for sTF circuits based on: (a) No ZF control; (b) ZF 42-10, showing the propagation of a fitness cost; (**c**) ZF 42-10-4x, showing that the fitness cost associated with ZF 42-10 is rescued by the phosphate backbone mutations; (d) ZF 13-6, a representative strain exhibiting a severe fitness cost; (e) ZF 129-3, a representative strain exhibiting little fitness cost (induced (red) and uninduced (black))**.** Points represent mean values for three experiments ± standard deviation.



Supplementary Figure A1.4. sTFs constructed from OPEN ZFs operate orthogonally to one another

Transcriptional activation by $sTF_{43\text{-}8}$ on cognate synthetic promoters with one, two, and eight tandem operators, and on orthogonal synthetic promoters with one, two, and eight tandem operators for ZF 42-10.



Supplementary Figure A1.5. DNA-binding activities of engineered ZF proteins possessing various combinations of non-specific affinity mutations

Alanine residue substitution mutations were made to four arginine residues in the three-finger array that are known to mediate non-specific interactions with the DNA phosphate backbone. The affinity mutants were assayed with a previously described bacterial-two-hybrid (B2H) system (see Methods).

Appendix 1 Supplementary Methods:

Quantitative real-time PCR

Transcriptional outputs of our circuits were analyzed by quantitative real-time PCR (RT-PCR). Total RNA was collected from yeast at different induction levels and time points (Supplementary Figure A1.1). RNA extraction was performed using RNeasy Mini Kit (Qiagen), and DNA contamination was eliminated using TURBO DNA-Free (Ambion) according to the manufacturers' protocols. Concentration of RNA in each sample was estimated using an ND-1000 NanoDrop spectrophotometer. Template cDNA was synthesized from RNA using the Superscript III First Strand Synthesis kit (Invitrogen) and stored at -20°C. RT-PCR reactions were prepared using LightCycler 480 SYBR Green I Master Kit (Roche Applied Science) according to the manufacturer's instructions, and RT-PCR was performed using a LightCycler 480 (Roche). Relative mRNA expression level of *yEGFP* in a given sample was determined by normalization to the transcript level of the reference gene *TAF10* (Teste et al., 2009).

Growth assays

Longer time-course growth assays were performed for a representative set of sTF circuit strains (Supplementary Figure A1.3). Growth assays were performed as described previously (see Methods), except that OD_{600} measurements were taken at 8 h time points for 64 h.

Bibliography

AJO-FRANKLIN, C. M., DRUBIN, D. A., ESKIN, J. A., GEE, E. P., LANDGRAF, D., PHILLIPS, I. & SILVER, P. A. 2007. Rational design of memory in eukaryotic cells. *Genes & development,* 21**,** 2271-6.

ALWIN, S., GERE, M. B., GUHL, E., EFFERTZ, K., BARBAS, C. F., 3RD, SEGAL, D. J., WEITZMAN, M. D. & CATHOMEN, T. 2005. Custom zinc-finger nucleases for use in human cells. *Mol Ther,* 12**,** 610-7.

ANDRIANANTOANDRO, E., BASU, S., KARIG, D. K. & WEISS, R. 2006. Synthetic biology: new engineering rules for an emerging discipline. *Mol Syst Biol,* 2**,** 2006-2006.

ARNOULD, S., DELENDA, C., GRIZOT, S., DESSEAUX, C., PAQUES, F., SILVA, G. H. & SMITH, J. 2011. The I-CreI meganuclease and its engineered derivatives: applications from cell modification to gene therapy. *Protein engineering, design & selection : PEDS,* 24**,** 27-31.

BAE, K. H., KWON, Y. D., SHIN, H. C., HWANG, M. S., RYU, E. H., PARK, K. S., YANG, H. Y., LEE, D. K., LEE, Y., PARK, J., KWON, H. S., KIM, H. W., YEH, B. I., LEE, H. W., SOHN, S. H., YOON, J., SEOL, W. & KIM, J. S. 2003. Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat Biotechnol,* 21**,** 275-80.

BASHOR, C. J., HELMAN, N. C., YAN, S. & LIM, W. A. 2008. Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics. *Science,* 319**,** 1539-1543.

BASHOR, C. J., HORWITZ, A. A., PEISAJOVICH, S. G. & LIM, W. A. 2010. Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living systems. *Annu Rev Biophys,* 39**,** 515-537.

BECSKEI, A. & SERRANO, L. 2000. Engineering stability in gene networks by autoregulation. *Nature,* 405**,** 590-593.

BEERLI, R. R. & BARBAS, C. F. 2002. Engineering polydactyl zinc-finger transcription factors. *Nat Biotechnol,* 20**,** 135-141.

BEERLI, R. R., SEGAL, D. J., DREIER, B. & BARBAS, C. F., 3RD 1998. Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc Natl Acad Sci U S A,* 95**,** 14628-33.

BENABDALLAH, B. F., ALLARD, E., YAO, S., FRIEDMAN, G., GREGORY, P. D., ELIOPOULOS, N., FRADETTE, J., SPEES, J. L., HADDAD, E., HOLMES, M. C. & BEAUSEJOUR, C. M. 2010. Targeted gene addition to human mesenchymal stromal cells as a cell-based plasma-soluble protein delivery platform. *Cytotherapy,* 12**,** 394-9.

BETTINI, M., XI, H., MILBRANDT, J. & KERSH, G. J. 2002. Thymocyte development in early growth response gene 1-deficient mice. *J Immunol,* 169**,** 1713-20.

BEUMER, K., BHATTACHARYYA, G., BIBIKOVA, M., TRAUTMAN, J. K. & CARROLL, D. 2006. Efficient gene targeting in Drosophila with zinc-finger nucleases. *Genetics,* 172**,** 2391-403.

BEUMER, K. J., TRAUTMAN, J. K., BOZAS, A., LIU, J. L., RUTTER, J., GALL, J. G. & CARROLL, D. 2008. Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. *Proc Natl Acad Sci U S A,* 105**,** 19821-6.

BHAKTA, M. S. & SEGAL, D. J. 2010. The generation of zinc finger proteins by modular assembly. *Methods Mol Biol,* 649**,** 3-30.

BIBIKOVA, M., BEUMER, K., TRAUTMAN, J. K. & CARROLL, D. 2003. Enhancing gene targeting with designed zinc finger nucleases. *Science,* 300**,** 764.

BIBIKOVA, M., CARROLL, D., SEGAL, D. J., TRAUTMAN, J. K., SMITH, J., KIM, Y. G. & CHANDRASEGARAN, S. 2001. Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol,* 21**,** 289-97.

BIBIKOVA, M., GOLIC, M., GOLIC, K. G. & CARROLL, D. 2002. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics,* 161**,** 1169-75.

BITINAITE, J., WAH, D. A., AGGARWAL, A. K. & SCHILDKRAUT, I. 1998. FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci U S A,* 95**,** 10570-5.

BLANCAFORT, P., CHEN, E. I., GONZALEZ, B., BERGQUIST, S., ZIJLSTRA, A., GUTHY, D., BRACHAT, A., BRAKENHOFF, R. H., QUIGLEY, J. P., ERDMANN, D. & BARBAS, C. F. 2005. Genetic reprogramming of tumor cells by zinc finger transcription factors. *Proc Natl Acad Sci U S A,* 102**,** 11716-11721.

BOCH, J., SCHOLZE, H., SCHORNACK, S., LANDGRAF, A., HAHN, S., KAY, S., LAHAYE, T., NICKSTADT, A. & BONAS, U. 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science,* 326**,** 1509-12.

BOGDANOVE, A. J. & VOYTAS, D. F. 2011. TAL effectors: customizable proteins for DNA targeting. *Science,* 333**,** 1843-6.

BOUCHARD, V. J., ROULEAU, M. & POIRIER, G. G. 2003. PARP-1, a determinant of cell survival in response to DNA damage. *Experimental hematology,* 31**,** 446-54.

BOZAS, A., BEUMER, K. J., TRAUTMAN, J. K. & CARROLL, D. 2009. Genetic analysis of zinc-finger nuclease-induced gene targeting in Drosophila. *Genetics,* 182**,** 641-51.

BRAYER, K. J., KULSHRESHTHA, S. & SEGAL, D. J. 2008. The protein-binding potential of C2H2 zinc finger domains. *Cell Biochem Biophys,* 51**,** 9-19.

BRUNET, E., SIMSEK, D., TOMISHIMA, M., DEKELVER, R., CHOI, V. M., GREGORY, P., URNOV, F., WEINSTOCK, D. M. & JASIN, M. 2009. Chromosomal translocations induced at specified loci in human stem cells. *Proc Natl Acad Sci U S A,* 106**,** 10620-5.

BRYANT, G. O. & PTASHNE, M. 2003. Independent recruitment in vivo by Gal4 of two complexes required for transcription. *Mol Cell,* 11**,** 1301-1309.

BULTMANN, S., MORBITZER, R., SCHMIDT, C. S., THANISCH, K., SPADA, F., ELSAESSER, J., LAHAYE, T. & LEONHARDT, H. 2012. Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic acids research*.

BULYK, M. L., HUANG, X., CHOO, Y. & CHURCH, G. M. 2001. Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. *Proc Natl Acad Sci U S A,* 98**,** 7158-63.

CARBERY, I. D., JI, D., HARRINGTON, A., BROWN, V., WEINSTEIN, E. J., LIAW, L. & CUI, X. 2010. Targeted genome modification in mice using zinc-finger nucleases. *Genetics,* 186**,** 451-9.

CATHOMEN, T. & SCHAMBACH, A. 2009. Zinc-finger nucleases meet iPS cells: Zinc positive: tailored genome engineering meets reprogramming. *Gene Ther,* 17**,** 1-3.

CATTO, L. E., BELLAMY, S. R., RETTER, S. E. & HALFORD, S. E. 2008. Dynamics and consequences of DNA looping by the FokI restriction endonuclease. *Nucleic Acids Res,* 36**,** 2073-81.

CERMAK, T., DOYLE, E. L., CHRISTIAN, M., WANG, L., ZHANG, Y., SCHMIDT, C., BALLER, J. A., SOMIA, N. V., BOGDANOVE, A. J. & VOYTAS, D. F. 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic acids research,* 39**,** e82.

CERTO, M. T., RYU, B. Y., ANNIS, J. E., GARIBOV, M., JARJOUR, J., RAWLINGS, D. J. & SCHARENBERG, A. M. 2011. Tracking genome engineering outcome at individual DNA breakpoints. *Nat Methods,* 8**,** 671-6.

CHAN, S. H., STODDARD, B. L. & XU, S. Y. 2011. Natural and engineered nicking endonucleases--from cleavage mechanism to engineering of strand-specificity. *Nucleic Acids Res,* 39**,** 1-18.

CHANDARLAPATY, S. & ERREDE, B. 1998. Ash1, a daughter cell-specific protein, is required for pseudohyphal growth of Saccharomyces cerevisiae. *Mol Cell Biol,* 18**,** 2884-2891.

CHEN, F., PRUETT-MILLER, S. M., HUANG, Y., GJOKA, M., DUDA, K., TAUNTON, J., COLLINGWOOD, T. N., FRODIN, M. & DAVIS, G. D. 2011. High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat Methods*.

CHEN, J., XU, J., YING, K., CAO, G., HU, G., WANG, L., LUO, C., LOU, M., MAO, Y., XIE, Y. & LU, Y. 2004. Molecular cloning and characterization of a novel human BTB domain-containing gene, BTBD10, which is down-regulated in glioma. *Gene,* 340**,** 61-9.

CHENG, L., BLAZAR, B., HIGH, K. & PORTEUS, M. 2011. Zinc fingers hit off target. *Nat Med,* 17**,** 1192-3.

CHOO, Y. & ISALAN, M. 2000. Advances in zinc finger engineering. *Curr Opin Struct Biol,* 10**,** 411-6.

CHOO, Y. & KLUG, A. 1994. Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. *Proc Natl Acad Sci U S A,* 91**,** 11168-72.

CHOO, Y., SANCHEZ-GARCIA, I. & KLUG, A. 1994. In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature,* 372**,** 642-5.

CLARK, K. J., VOYTAS, D. F. & EKKER, S. C. 2011. A TALE of two nucleases: gene targeting for the masses? *Zebrafish,* 8**,** 147-9.

CORMACK, B. P., BERTRAM, G., EGERTON, M., GOW, N. A., FALKOW, S. & BROWN, A. J. 1997. Yeast-enhanced green fluorescent protein (yEGFP): a reporter of gene expression in Candida albicans. *Microbiology,* 143 (Pt 2)**,** 303-311.

CORNU, T. I. & CATHOMEN, T. 2007. Targeted genome modifications using integrase-deficient lentiviral vectors. *Mol Ther,* 15**,** 2107-13.

CORNU, T. I., THIBODEAU-BEGANNY, S., GUHL, E., ALWIN, S., EICHTINGER, M., JOUNG, J. K. & CATHOMEN, T. 2008. DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Mol Ther,* 16**,** 352-8.

COST, G. J., FREYVERT, Y., VAFIADIS, A., SANTIAGO, Y., MILLER, J. C., REBAR, E., COLLINGWOOD, T. N., SNOWDEN, A. & GREGORY, P. D. 2009. BAK and BAX deletion using zinc-finger nucleases yields apoptosis-resistant CHO cells. *Biotechnol Bioeng,* 105**,** 330-40.

CRADICK, T. J., AMBROSINI, G., ISELI, C., BUCHER, P. & MCCAFFREY, A. P. 2011. ZFN-Site searches genomes for zinc finger nuclease target sites and off-target sites. *BMC Bioinformatics,* 12**,** 152.

CRADICK, T. J., KECK, K., BRADSHAW, S., JAMIESON, A. C. & MCCAFFREY, A. P. 2010. Zinc-finger nucleases as a novel therapeutic strategy for targeting hepatitis B virus DNAs. *Mol Ther,* 18**,** 947-54.

CRAVEN, S. E. & BREDT, D. S. 1998. PDZ proteins organize synaptic signaling pathways. *Cell,* 93**,** 495-498.

CUI, X., JI, D., FISHER, D. A., WU, Y., BRINER, D. M. & WEINSTEIN, E. J. 2011. Targeted integration in rat and mouse embryos with zinc-finger nucleases. *Nat Biotechnol,* 29**,** 64-7.

CURTIN, S. J., ZHANG, F., SANDER, J. D., HAUN, W. J., STARKER, C., BALTES, N. J., REYON, D., DAHLBORG, E. J., GOODWIN, M. J., COFFMAN, A. P., DOBBS, D., JOUNG, J. K., VOYTAS, D. F. & STUPAR, R. M. 2011. Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. *Plant physiology,* 156**,** 466-73.

DAVIS, L. & MAIZELS, N. 2011. DNA nicks promote efficient and safe targeted gene correction. *PLoS One,* 6**,** e23981.

DENG, D., YAN, C., PAN, X., MAHFOUZ, M., WANG, J., ZHU, J. K., SHI, Y. & YAN, N. 2012. Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science,* 335**,** 720-3.

DONOHO, G., JASIN, M. & BERG, P. 1998. Analysis of gene targeting and intrachromosomal homologous recombination stimulated by genomic double-strand breaks in mouse embryonic stem cells. *Mol Cell Biol,* 18**,** 4070-8.

DOYON, Y., MCCAMMON, J. M., MILLER, J. C., FARAJI, F., NGO, C., KATIBAH, G. E., AMORA, R., HOCKING, T. D., ZHANG, L., REBAR, E. J., GREGORY, P. D., URNOV, F. D. & AMACHER, S. L. 2008. Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat Biotechnol,* 26**,** 702-8.

DOYON, Y., VO, T. D., MENDEL, M. C., GREENBERG, S. G., WANG, J., XIA, D. F., MILLER, J. C., URNOV, F. D., GREGORY, P. D. & HOLMES, M. C. 2011. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat Methods,* 8**,** 74-9.

ELLIOTT, B., RICHARDSON, C., WINDERBAUM, J., NICKOLOFF, J. A. & JASIN, M. 1998. Gene conversion tracts from double-strand break repair in mammalian cells. *Mol Cell Biol,* 18**,** 93-101.

ELLIS, T., WANG, X. & COLLINS, J. J. 2009. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat Biotechnol,* 27**,** 465-471.

ELOWITZ, M. & LIM, W. A. 2010. Build life to understand it. *Nature,* 468**,** 889-890.

ELOWITZ, M. B. & LEIBLER, S. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature,* 403**,** 335-338.

ELROD-ERICKSON, M., BENSON, T. E. & PABO, C. O. 1998. High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure,* 6**,** 451-64.

ELROD-ERICKSON, M., ROULD, M. A., NEKLUDOVA, L. & PABO, C. O. 1996. Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions. *Structure,* 4**,** 1171-80.

FISCHER, J. A., GINIGER, E., MANIATIS, T. & PTASHNE, M. 1988. GAL4 activates transcription in Drosophila. *Nature,* 332**,** 853-856.

FLISIKOWSKA, T., THOREY, I. S., OFFNER, S., ROS, F., LIFKE, V., ZEITLER, B., ROTTMANN, O., VINCENT, A., ZHANG, L., JENKINS, S., NIERSBACH, H., KIND, A. J., GREGORY, P. D., SCHNIEKE, A. E. & PLATZER, J. 2011. Efficient immunoglobulin gene disruption and targeted replacement in rabbit using zinc finger nucleases. *PLoS One,* 6**,** e21045.

FOLEY, J. E., MAEDER, M. L., PEARLBERG, J., JOUNG, J. K., PETERSON, R. T. & YEH, J. R. 2009a. Targeted mutagenesis in zebrafish using customized zinc-finger nucleases. *Nat Protoc,* 4**,** 1855-67.

FOLEY, J. E., YEH, J. R., MAEDER, M. L., REYON, D., SANDER, J. D., PETERSON, R. T. & JOUNG, J. K. 2009b. Rapid mutation of endogenous zebrafish genes using zinc finger nucleases made by Oligomerized Pool ENgineering (OPEN). *PLoS One,* 4**,** e4348.

FU, F., SANDER, J. D., MAEDER, M., THIBODEAU-BEGANNY, S., JOUNG, J. K., DOBBS, D., MILLER, L. & VOYTAS, D. F. 2009. Zinc Finger Database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays. *Nucleic Acids Res,* 37**,** D279-83.

GABRIEL, R., LOMBARDO, A., ARENS, A., MILLER, J. C., GENOVESE, P., KAEPPEL, C., NOWROUZI, A., BARTHOLOMAE, C. C., WANG, J., FRIEDMAN, G., HOLMES, M. C., GREGORY, P. D., GLIMM, H., SCHMIDT, M., NALDINI, L. & VON KALLE, C. 2011. An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol,* 29**,** 816-23.

GARDNER, T. S., CANTOR, C. R. & COLLINS, J. J. 2000. Construction of a genetic toggle switch in Escherichia coli. *Nature,* 403**,** 339-342.

GELLHAUS, K., CORNU, T. I., HEILBRONN, R. & CATHOMEN, T. 2010. Fate of recombinant adeno-associated viral vector genomes during DNA double-strand break-induced gene targeting in human cells. *Hum Gene Ther,* 21**,** 543-53.

GEURTS, A. M., COST, G. J., FREYVERT, Y., ZEITLER, B., MILLER, J. C., CHOI, V. M., JENKINS, S. S., WOOD, A., CUI, X., MENG, X., VINCENT, A., LAM, S., MICHALKIEWICZ, M., SCHILLING, R., FOECKLER, J., KALLOWAY, S., WEILER, H., MENORET, S., ANEGON, I., DAVIS, G. D., ZHANG, L., REBAR, E. J., GREGORY, P. D., URNOV, F. D., JACOB, H. J. & BUELOW, R. 2009a. Knockout rats via embryo microinjection of zinc-finger nucleases. *Science,* 325**,** 433.

GEURTS, A. M., COST, G. J., REMY, S., CUI, X., TESSON, L., USAL, C., MENORET, S., JACOB, H. J., ANEGON, I. & BUELOW, R. 2009b. Generation of gene-specific mutated rats using zinc-finger nucleases. *Methods Mol Biol,* 597**,** 211-25.

GOLDBERG, A. D., BANASZYNSKI, L. A., NOH, K. M., LEWIS, P. W., ELSAESSER, S. J., STADLER, S., DEWELL, S., LAW, M., GUO, X., LI, X., WEN, D., CHAPGIER, A., DEKELVER, R. C., MILLER, J. C., LEE, Y. L., BOYDSTON, E. A., HOLMES, M. C., GREGORY, P. D., GREALLY, J. M., RAFII, S., YANG, C., SCAMBLER, P. J., GARRICK, D., GIBBONS, R. J., HIGGS, D. R., CRISTEA, I. M., URNOV, F. D., ZHENG, D. & ALLIS, C. D. 2010. Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell,* 140**,** 678-91.

GONCALVES, M. A., VAN NIEROP, G. P., HOLKERS, M. & DE VRIES, A. A. 2011. Concerted nicking of donor and chromosomal acceptor DNA promotes homology-directed gene targeting in human cells. *Nucleic acids research*.

GREISMAN, H. A. & PABO, C. O. 1997. A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science,* 275**,** 657-61.

GUET, C. C., ELOWITZ, M. B., HSING, W. & LEIBLER, S. 2002. Combinatorial synthesis of genetic networks. *Science,* 296**,** 1466-1470.

GUO, J., GAJ, T. & BARBAS, C. F., 3RD 2010. Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. *J Mol Biol,* 400**,** 96-107.

GUPTA, A., MENG, X., ZHU, L. J., LAWSON, N. D. & WOLFE, S. A. 2011. Zinc finger protein-dependent and -independent contributions to the in vivo off-target activity of zinc finger nucleases. *Nucleic Acids Res,* 39**,** 381-92.

GUSCHIN, D. Y., WAITE, A. J., KATIBAH, G. E., MILLER, J. C., HOLMES, M. C. & REBAR, E. J. 2010. A rapid and general assay for monitoring endogenous gene modification. *Methods Mol Biol,* 649**,** 247-56.

HAHN, S. & YOUNG, E. T. 2011. Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics,* 189**,** 705-736.

HALFORD, S. E., CATTO, L. E., PERNSTICH, C., RUSLING, D. A. & SANDERS, K. L. 2011. The reaction mechanism of FokI excludes the possibility of targeting zinc finger nucleases to unique DNA sites. *Biochem Soc Trans,* 39**,** 584-8.

HANDEL, E. M., ALWIN, S. & CATHOMEN, T. 2009. Expanding or restricting the target site repertoire of zinc-finger nucleases: the inter-domain linker as a major determinant of target site selectivity. *Mol Ther,* 17**,** 104-11.

HANDEL, E. M. & CATHOMEN, T. 2011. Zinc-finger nuclease based genome surgery: it's all about specificity. *Curr Gene Ther,* 11**,** 28-37.

HARRIS, B. Z., HILLIER, B. J. & LIM, W. A. 2001. Energetic determinants of internal motif recognition by PDZ domains. *Biochemistry,* 40**,** 5921-5930.

HARRIS, B. Z. & LIM, W. A. 2001. Mechanism and role of PDZ domains in signaling complex assembly. *J Cell Sci,* 114**,** 3219-3231.

HARTLERODE, A., ODATE, S., SHIM, I., BROWN, J. & SCULLY, R. 2011. Cell cycle-dependent induction of homologous recombination by a tightly regulated I-SceI fusion protein. *PLoS One,* 6**,** e16501.

HARTLERODE, A. J. & SCULLY, R. 2009. Mechanisms of double-strand break repair in somatic mammalian cells. *Biochem J,* 423**,** 157-68.

HASTY, J., MCMILLEN, D. & COLLINS, J. J. 2002. Engineered gene circuits. *Nature,* 420**,** 224-230.

HOCKEMEYER, D. & JAENISCH, R. 2011. Gene targeting in human pluripotent cells. *Cold Spring Harb Symp Quant Biol,* 75**,** 201-9.

HOCKEMEYER, D., SOLDNER, F., BEARD, C., GAO, Q., MITALIPOVA, M., DEKELVER, R. C., KATIBAH, G. E., AMORA, R., BOYDSTON, E. A., ZEITLER, B., MENG, X., MILLER, J. C., ZHANG, L., REBAR, E. J., GREGORY, P. D., URNOV, F. D. & JAENISCH, R. 2009. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol,* 27**,** 851-7.

HOCKEMEYER, D., WANG, H., KIANI, S., LAI, C. S., GAO, Q., CASSADY, J. P., COST, G. J., ZHANG, L., SANTIAGO, Y., MILLER, J. C., ZEITLER, B., CHERONE, J. M., MENG, X., HINKLEY, S. J., REBAR, E. J., GREGORY, P. D., URNOV, F. D. & JAENISCH, R. 2011. Genetic engineering of human pluripotent cells using TALE nucleases. *Nature biotechnology,* 29**,** 731-4.

HOLLIDAY, R. 2007. A mechanism for gene conversion in fungi. *Genet Res,* 89**,** 285-307.

HOLMQUIST, G. P. 1998. Endogenous lesions, S-phase-independent spontaneous mutations, and evolutionary strategies for base excision repair. *Mutat Res,* 400**,** 59-68.

HOLT, N., WANG, J., KIM, K., FRIEDMAN, G., WANG, X., TAUPIN, V., CROOKS, G. M., KOHN, D. B., GREGORY, P. D., HOLMES, M. C. & CANNON, P. M. 2010. Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. *Nat Biotechnol,* 28**,** 839-47.

HOOVER, D. M. & LUBKOWSKI, J. 2002. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res,* 30**,** e43.

HUANG, P., XIAO, A., ZHOU, M., ZHU, Z., LIN, S. & ZHANG, B. 2011. Heritable gene targeting in zebrafish using customized TALENs. *Nature biotechnology,* 29**,** 699-700.

HURT, J. A., THIBODEAU, S. A., HIRSH, A. S., PABO, C. O. & JOUNG, J. K. 2003. Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc Natl Acad Sci U S A,* 100**,** 12271-6.

HUTTER, G., NOWAK, D., MOSSNER, M., GANEPOLA, S., MUSSIG, A., ALLERS, K., SCHNEIDER, T., HOFMANN, J., KUCHERER, C., BLAU, O., BLAU, I. W., HOFMANN, W. K. & THIEL, E. 2009. Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *The New England journal of medicine,* 360**,** 692-8.

ISALAN, M. & CHOO, Y. 2001. Rapid, high-throughput engineering of sequence-specific zinc finger DNA-binding proteins. *Methods Enzymol,* 340**,** 593-609.

ISALAN, M., CHOO, Y. & KLUG, A. 1997. Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proceedings of the National Academy of Sciences of the United States of America,* 94**,** 5617-21.

ISALAN, M., KLUG, A. & CHOO, Y. 1998. Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers. *Biochemistry,* 37**,** 12026-33.

ISALAN, M., KLUG, A. & CHOO, Y. 2001. A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nat Biotechnol,* 19**,** 656-60.

JAMIESON, A. C., KIM, S. H. & WELLS, J. A. 1994. In vitro selection of zinc fingers with altered DNA-binding specificity. *Biochemistry,* 33**,** 5689-95.

JAMIESON, A. C., MILLER, J. C. & PABO, C. O. 2003. Drug discovery with engineered zinc-finger proteins. *Nat Rev Drug Discov,* 2**,** 361-8.

JASIN, M. 1996. Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet,* 12**,** 224-8.

JENSEN, N. M., DALSGAARD, T., JAKOBSEN, M., NIELSEN, R. R., SORENSEN, C. B., BOLUND, L. & JENSEN, T. G. 2011. An update on targeted gene repair in mammalian cells: methods and mechanisms. *J Biomed Sci,* 18**,** 10.

JOUNG, J. K., KOEPP, D. M. & HOCHSCHILD, A. 1994. Synergistic activation of transcription by bacteriophage lambda cI protein and E. coli cAMP receptor protein. *Science,* 265**,** 1863-1866.

JOUNG, J. K., LE, L. U. & HOCHSCHILD, A. 1993. Synergistic activation of transcription by Escherichia coli cAMP receptor protein. *Proc Natl Acad Sci U S A,* 90**,** 3083-3087.

JOUNG, J. K., RAMM, E. I. & PABO, C. O. 2000. A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc Natl Acad Sci U S A,* 97**,** 7382-7.

JOUNG, J. K., VOYTAS, D. F. & CATHOMEN, T. 2010. Reply to "Genome editing with modularly assembled zinc-finger nucleases". *Nat Methods,* 7**,** 91-92.

KASSIR, Y., ADIR, N., BOGER-NADJAR, E., RAVIV, N. G., RUBIN-BEJERANO, I., SAGEE, S. & SHENHAR, G. 2003. Transcriptional regulation of meiosis in budding yeast. *Int Rev Cytol,* 224**,** 111-171.

KHALIL, A. S. & COLLINS, J. J. 2010. Synthetic biology: applications come of age. *Nat Rev Genet,* 11**,** 367-379.

KIM, E., KIM, S., DUK HYOUNG, K., CHOI, B. S., CHOI, I. Y. & KIM, J. S. 2012. Precision genome engineering with programmable DNA-nicking enzymes. *Genome research*.

KIM, H. J., LEE, H. J., KIM, H., CHO, S. W. & KIM, J. S. 2009. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res,* 19**,** 1279-88.

KIM, J. S., LEE, H. J. & CARROLL, D. 2010. Genome editing with modularly assembled zinc-finger nucleases. *Nat Methods,* 7**,** 91; author reply 91-2.

KIM, Y. G., CHA, J. & CHANDRASEGARAN, S. 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A,* 93**,** 1156-60.

KRAMER, B. P., FISCHER, C. & FUSSENEGGER, M. 2004a. BioLogic gates enable logical transcription control in mammalian cells. *Biotechnol Bioeng,* 87**,** 478-484.

KRAMER, B. P., VIRETTA, A. U., DAOUD-EL-BABA, M., AUBEL, D., WEBER, W. & FUSSENEGGER, M. 2004b. An engineered epigenetic transgene switch in mammalian cells. *Nat Biotechnol,* 22**,** 867-870.

LAM, K. N., VAN BAKEL, H., COTE, A. G., VAN DER VEN, A. & HUGHES, T. R. 2011. Sequence specificity is obtained from the majority of modular C2H2 zinc-finger arrays. *Nucleic acids research,* 39**,** 4680-90.

LEE, G. S., NEIDITCH, M. B., SALUS, S. S. & ROTH, D. B. 2004. RAG proteins shepherd double-strand breaks to a specific pathway, suppressing error-prone repair, but RAG nicking initiates homologous recombination. *Cell,* 117**,** 171-84.

LEE, H. J., KIM, E. & KIM, J. S. 2009. Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res,* 20**,** 81-9.

LEE, M. S., GIPPERT, G. P., SOMAN, K. V., CASE, D. A. & WRIGHT, P. E. 1989. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science,* 245**,** 635-7.

LEVSKAYA, A., CHEVALIER, A. A., TABOR, J. J., SIMPSON, Z. B., LAVERY, L. A., LEVY, M., DAVIDSON, E. A., SCOURAS, A., ELLINGTON, A. D., MARCOTTE, E. M. & VOIGT, C. A. 2005. Synthetic biology: engineering Escherichia coli to see light. *Nature,* 438**,** 441-442.

LI, H., HAURIGOT, V., DOYON, Y., LI, T., WONG, S. Y., BHAGWAT, A. S., MALANI, N., ANGUELA, X. M., SHARMA, R., IVANCIU, L., MURPHY, S. L., FINN, J. D., KHAZI, F. R., ZHOU, S., PASCHON, D. E., REBAR, E. J., BUSHMAN, F. D., GREGORY, P. D., HOLMES, M. C. & HIGH, K. A. 2011. In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature,* 475**,** 217-21.

LI, L., WU, L. P. & CHANDRASEGARAN, S. 1992. Functional domains in Fok I restriction endonuclease. *Proc Natl Acad Sci U S A,* 89**,** 4275-9.

LIM, W. A. 2010. Designing customized cell signalling circuits. *Nature reviews. Molecular cell biology,* 11**,** 393-403.

LIU, J. & STORMO, G. D. 2005. Combining SELEX with quantitative assays to rapidly obtain accurate models of protein-DNA interactions. *Nucleic Acids Res,* 33**,** e141.

LIU, P. Q., CHAN, E. M., COST, G. J., ZHANG, L., WANG, J., MILLER, J. C., GUSCHIN, D. Y., REIK, A., HOLMES, M. C., MOTT, J. E., COLLINGWOOD, T. N. & GREGORY, P. D. 2010. Generation of a triple-gene knockout mammalian cell line using engineered zinc-finger nucleases. *Biotechnol Bioeng,* 106**,** 97-105.

LIU, P. Q., REBAR, E. J., ZHANG, L., LIU, Q., JAMIESON, A. C., LIANG, Y., QI, H., LI, P. X., CHEN, B., MENDEL, M. C., ZHONG, X., LEE, Y. L., EISENBERG, S. P., SPRATT, S. K., CASE, C. C. & WOLFFE, A. P. 2001. Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. *J Biol Chem,* 276**,** 11323-34.

LIU, Q., XIA, Z., ZHONG, X. & CASE, C. C. 2002. Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *J Biol Chem,* 277**,** 3850-6.

LLOYD, A., PLAISIER, C. L., CARROLL, D. & DREWS, G. N. 2005. Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. *Proc Natl Acad Sci U S A,* 102**,** 2232-7.

LOMBARDO, A., CESANA, D., GENOVESE, P., DI STEFANO, B., PROVASI, E., COLOMBO, D. F., NERI, M., MAGNANI, Z., CANTORE, A., LO RISO, P., DAMO, M., PELLO, O. M., HOLMES, M. C., GREGORY, P. D., GRITTI, A., BROCCOLI, V., BONINI, C. & NALDINI, L. 2011. Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat Methods*.

LOMBARDO, A., GENOVESE, P., BEAUSEJOUR, C. M., COLLEONI, S., LEE, Y. L., KIM, K. A., ANDO, D., URNOV, F. D., GALLI, C., GREGORY, P. D., HOLMES, M. C. & NALDINI, L. 2007. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol,* 25**,** 1298-306.

LU, T. K., KHALIL, A. S. & COLLINS, J. J. 2009. Next-generation synthetic gene networks. *Nat Biotechnol,* 27**,** 1139-1150.

MA, J., PRZIBILLA, E., HU, J., BOGORAD, L. & PTASHNE, M. 1988. Yeast activators stimulate plant gene expression. *Nature,* 334**,** 631-633.

MA, J. & PTASHNE, M. 1988. Converting a eukaryotic transcriptional inhibitor into an activator. *Cell,* 55**,** 443-446.

MAEDER, M. L., THIBODEAU-BEGANNY, S., OSIAK, A., WRIGHT, D. A., ANTHONY, R. M., EICHTINGER, M., JIANG, T., FOLEY, J. E., WINFREY, R. J., TOWNSEND, J. A., UNGER-WALLACE, E., SANDER, J. D., MULLER-LERCH, F., FU, F., PEARLBERG, J., GOBEL, C., DASSIE, J. P., PRUETT-MILLER, S. M., PORTEUS, M. H., SGROI, D. C., IAFRATE, A. J., DOBBS, D., MCCRAY, P. B., JR., CATHOMEN, T., VOYTAS, D. F. & JOUNG, J. K. 2008. Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell,* 31**,** 294-301.

MAEDER, M. L., THIBODEAU-BEGANNY, S., SANDER, J. D., VOYTAS, D. F. & JOUNG, J. K. 2009. Oligomerized pool engineering (OPEN): an 'open-source' protocol for making customized zinc-finger arrays. *Nature Protocols,* 4**,** 1471-501.

MAK, A. N., BRADLEY, P., CERNADAS, R. A., BOGDANOVE, A. J. & STODDARD, B. L. 2012. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science,* 335**,** 716-9.

MANDELL, J. G. & BARBAS, C. F., 3RD 2006. Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic acids research,* 34**,** W516-23.

MANI, M., SMITH, J., KANDAVELOU, K., BERG, J. M. & CHANDRASEGARAN, S. 2005. Binding of two zinc finger nuclease monomers to two specific sites is required for effective double-strand DNA cleavage. *Biochem Biophys Res Commun,* 334**,** 1191-7.

MAO, Z., BOZZELLA, M., SELUANOV, A. & GORBUNOVA, V. 2008. DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle,* 7**,** 2902-6.

MASHIMO, T., TAKIZAWA, A., VOIGT, B., YOSHIMI, K., HIAI, H., KURAMOTO, T. & SERIKAWA, T. 2010. Generation of knockout rats with X-linked severe combined immunodeficiency (X-SCID) using zinc-finger nucleases. *PLoS One,* 5**,** e8870.

MAXON, M. E. & HERSKOWITZ, I. 2001. Ash1p is a site-specific DNA-binding protein that actively represses transcription. *Proc Natl Acad Sci U S A,* 98**,** 1495-1500.

MCCONNELL SMITH, A., TAKEUCHI, R., PELLENZ, S., DAVIS, L., MAIZELS, N., MONNAT, R. J., JR. & STODDARD, B. L. 2009. Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. *Proc Natl Acad Sci U S A,* 106**,** 5099-104.

MENG, X., NOYES, M. B., ZHU, L. J., LAWSON, N. D. & WOLFE, S. A. 2008. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol,* 26**,** 695-701.

MENG, X., THIBODEAU-BEGANNY, S., JIANG, T., JOUNG, J. K. & WOLFE, S. A. 2007. Profiling the DNA-binding specificities of engineered Cys2His2 zinc finger domains using a rapid cell-based method. *Nucleic Acids Res,* 35**,** e81.

MESELSON, M. S. & RADDING, C. M. 1975. A general model for genetic recombination. *Proc Natl Acad Sci U S A,* 72**,** 358-61.

METZGER, M. J., MCCONNELL-SMITH, A., STODDARD, B. L. & MILLER, A. D. 2011. Single-strand nicks induce homologous recombination with less toxicity than double-strand breaks using an AAV vector template. *Nucleic Acids Res,* 39**,** 926-35.

MEYER, M., DE ANGELIS, M. H., WURST, W. & KUHN, R. 2010. Gene targeting by homologous recombination in mouse zygotes mediated by zinc-finger nucleases. *Proc Natl Acad Sci U S A,* 107**,** 15022-6.

MILLER, J., MCLACHLAN, A. D. & KLUG, A. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *Embo J,* 4**,** 1609-14.

MILLER, J. C., HOLMES, M. C., WANG, J., GUSCHIN, D. Y., LEE, Y. L., RUPNIEWSKI, I., BEAUSEJOUR, C. M., WAITE, A. J., WANG, N. S., KIM, K. A., GREGORY, P. D., PABO, C. O. & REBAR, E. J. 2007. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol,* 25**,** 778-85.

MILLER, J. C., TAN, S., QIAO, G., BARLOW, K. A., WANG, J., XIA, D. F., MENG, X., PASCHON, D. E., LEUNG, E., HINKLEY, S. J., DULAY, G. P., HUA, K. L., ANKOUDINOVA, I., COST, G. J., URNOV, F. D., ZHANG, H. S., HOLMES, M. C., ZHANG, L., GREGORY, P. D. & REBAR, E. J. 2011. A TALE nuclease architecture for efficient genome editing. *Nature biotechnology,* 29**,** 143-8.

MOEHLE, E. A., ROCK, J. M., LEE, Y. L., JOUVENOT, Y., DEKELVER, R. C., GREGORY, P. D., URNOV, F. D. & HOLMES, M. C. 2007. Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc Natl Acad Sci U S A,* 104**,** 3055-60.

MOORE, M., KLUG, A. & CHOO, Y. 2001. Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. *Proc Natl Acad Sci U S A,* 98**,** 1437-41.

MORENO, C., HOFFMAN, M., STODOLA, T. J., DIDIER, D. N., LAZAR, J., GEURTS, A. M., NORTH, P. E., JACOB, H. J. & GREENE, A. S. 2011. Creation and characterization of a renin knockout rat. *Hypertension,* 57**,** 614-9.

MORTON, J., DAVIS, M. W., JORGENSEN, E. M. & CARROLL, D. 2006. Induction and repair of zinc-finger nuclease-targeted double-strand breaks in Caenorhabditis elegans somatic cells. *Proc Natl Acad Sci U S A,* 103**,** 16370-5.

MOSCOU, M. J. & BOGDANOVE, A. J. 2009. A simple cipher governs DNA recognition by TAL effectors. *Science,* 326**,** 1501.

MUKHERJI, S. & VAN OUDENAARDEN, A. 2009. Synthetic biology: understanding biological design from synthetic circuits. *Nat Rev Genet,* 10**,** 859-871.

MURPHY, K. F., BALAZSI, G. & COLLINS, J. J. 2007. Combinatorial promoter design for engineering noisy gene expression. *Proc Natl Acad Sci U S A,* 104**,** 12726-12731.

MUSSOLINO, C., MORBITZER, R., LUTGE, F., DANNEMANN, N., LAHAYE, T. & CATHOMEN, T. 2011. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res,* 39**,** 9283-9293.

NAWA, M., KANEKURA, K., HASHIMOTO, Y., AISO, S. & MATSUOKA, M. 2008. A novel Akt/PKB-interacting protein promotes cell adhesion and inhibits familial amyotrophic lateral sclerosis-linked mutant SOD1-induced neuronal death via inhibition of PP2A-mediated dephosphorylation of Akt/PKB. *Cell Signal,* 20**,** 493-505.

NISSIM, L. & BAR-ZIV, R. H. 2010. A tunable dual-promoter integrator for targeting of cancer cells. *Mol Syst Biol,* 6**,** 444-444.

OCHIAI, H., FUJITA, K., SUZUKI, K., NISHIKAWA, M., SHIBATA, T., SAKAMOTO, N. & YAMAMOTO, T. 2010. Targeted mutagenesis in the sea urchin embryo using zinc-finger nucleases. *Genes Cells,* 15**,** 875-85.

OLEYKOWSKI, C. A., BRONSON MULLINS, C. R., GODWIN, A. K. & YEUNG, A. T. 1998. Mutation detection using a novel plant endonuclease. *Nucleic Acids Res,* 26**,** 4597-602.

OLSEN, P. A., GELAZAUSKAITE, M., RANDOL, M. & KRAUSS, S. 2010. Analysis of illegitimate genomic integration mediated by zinc-finger nucleases: implications for specificity of targeted gene correction. *BMC Mol Biol,* 11**,** 35.

OLSEN, P. A., SOLHAUG, A., BOOTH, J. A., GELAZAUSKAITE, M. & KRAUSS, S. 2009. Cellular responses to targeted genomic sequence modification using single-stranded oligonucleotides and zinc-finger nucleases. *DNA Repair (Amst),* 8**,** 298-308.

ORLANDO, S. J., SANTIAGO, Y., DEKELVER, R. C., FREYVERT, Y., BOYDSTON, E. A., MOEHLE, E. A., CHOI, V. M., GOPALAN, S. M., LOU, J. F., LI, J., MILLER, J. C., HOLMES, M. C., GREGORY, P. D., URNOV, F. D. & COST, G. J. 2010. Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. *Nucleic Acids Res,* 38**,** e152.

OSBORN, M. J., DEFEO, A. P., BLAZAR, B. & TOLAR, J. 2011. Synthetic Zinc Finger Nuclease Design and Rapid Assembly. *Hum Gene Ther*.

PABO, C. O., PEISACH, E. & GRANT, R. A. 2001. Design and selection of novel Cys2His2 zinc finger proteins. *Annu Rev Biochem,* 70**,** 313-340.

PAQUES, F. & DUCHATEAU, P. 2007. Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Curr Gene Ther,* 7**,** 49-66.

PARRAGA, G., HORVATH, S. J., EISEN, A., TAYLOR, W. E., HOOD, L., YOUNG, E. T. & KLEVIT, R. E. 1988. Zinc-dependent structure of a single-finger domain of yeast ADR1. *Science,* 241**,** 1489-92.

PATTANAYAK, V., RAMIREZ, C. L., JOUNG, J. K. & LIU, D. R. 2011. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Methods,* 8**,** 765-70.

PAVLETICH, N. P. & PABO, C. O. 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science,* 252**,** 809-17.

PEARSON, H. 2008. Protein engineering: The fate of fingers. *Nature,* 455**,** 160-4.

PEDRAZA, J. M. & VAN OUDENAARDEN, A. 2005. Noise propagation in gene networks. *Science,* 307**,** 1965-1969.

PEREZ, E. E., WANG, J., MILLER, J. C., JOUVENOT, Y., KIM, K. A., LIU, O., WANG, N., LEE, G., BARTSEVICH, V. V., LEE, Y. L., GUSCHIN, D. Y., RUPNIEWSKI, I., WAITE, A. J., CARPENITO, C., CARROLL, R. G., ORANGE, J. S., URNOV, F. D., REBAR, E. J., ANDO, D., GREGORY, P. D., RILEY, J. L., HOLMES, M. C. & JUNE, C. H. 2008. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol,* 26**,** 808-16.

PETEK, L. M., RUSSELL, D. W. & MILLER, D. G. 2010. Frequent endonuclease cleavage at off-target locations in vivo. *Mol Ther,* 18**,** 983-6.

PINGOUD, A., FUXREITER, M., PINGOUD, V. & WENDE, W. 2005. Type II restriction endonucleases: structure and mechanism. *Cell Mol Life Sci,* 62**,** 685-707.

PINGOUD, A. & SILVA, G. H. 2007. Precision genome surgery. *Nat Biotechnol,* 25**,** 743-4.

POMERANTZ, J. L., WOLFE, S. A. & PABO, C. O. 1998. Structure-based design of a dimeric zinc finger protein. *Biochemistry,* 37**,** 965-970.

PORTEUS, M. H. & BALTIMORE, D. 2003. Chimeric nucleases stimulate gene targeting in human cells. *Science,* 300**,** 763.

PORTEUS, M. H. & CARROLL, D. 2005. Gene targeting using zinc finger nucleases. *Nature biotechnology,* 23**,** 967-73.

PORTEUS, M. H., CATHOMEN, T., WEITZMAN, M. D. & BALTIMORE, D. 2003. Efficient gene targeting mediated by adeno-associated virus and DNA double-strand breaks. *Mol Cell Biol,* 23**,** 3558-65.

PRUETT-MILLER, S. M., CONNELLY, J. P., MAEDER, M. L., JOUNG, J. K. & PORTEUS, M. H. 2008. Comparison of zinc finger nucleases for use in gene targeting in mammalian cells. *Mol Ther,* 16**,** 707-17.

PTASHNE, M. 1986. Gene regulation by proteins acting nearby and at a distance. *Nature,* 322**,** 697-701.

PTASHNE, M. 1988. How eukaryotic transcriptional activators work. *Nature,* 335**,** 683-9.

PTASHNE, M. & GANN, A. 1997. Transcriptional activation by recruitment. *Nature,* 386**,** 569-577.

PTASHNE, M. & GANN, A. 2002. *Genes & signals* Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press.

RADECKE, F., PETER, I., RADECKE, S., GELLHAUS, K., SCHWARZ, K. & CATHOMEN, T. 2006. Targeted chromosomal gene modification in human cells by single-stranded oligodeoxynucleotides in the presence of a DNA double-strand break. *Mol Ther,* 14**,** 798-808.

RADECKE, S., RADECKE, F., CATHOMEN, T. & SCHWARZ, K. 2010. Zinc-finger nuclease-induced gene repair with oligodeoxynucleotides: wanted and unwanted target locus modifications. *Mol Ther,* 18**,** 743-53.

RAHMAN, S. H., MAEDER, M. L., JOUNG, J. K. & CATHOMEN, T. 2011. Zinc-finger nucleases for somatic gene therapy: the next frontier. *Hum Gene Ther*.

RAMALINGAM, S., KANDAVELOU, K., RAJENDERAN, R. & CHANDRASEGARAN, S. 2010. Creating designed zinc-finger nucleases with minimal cytotoxicity. *J Mol Biol,* 405**,** 630-41.

RAMIREZ, C. L., CERTO, M. T., MUSSOLINO, C., GOODWIN, M. J., CRADICK, T. J., MCCAFFREY, A. P., CATHOMEN, T., SCHARENBERG, A. M. & JOUNG, J. K. 2012. Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic acids research*.

RAMIREZ, C. L., FOLEY, J. E., WRIGHT, D. A., MULLER-LERCH, F., RAHMAN, S. H., CORNU, T. I., WINFREY, R. J., SANDER, J. D., FU, F., TOWNSEND, J. A., CATHOMEN, T., VOYTAS, D. F. & JOUNG, J. K. 2008. Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat Methods,* 5**,** 374-5.

REBAR, E. J. & PABO, C. O. 1994. Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science,* 263**,** 671-3.

REYON, D., KIRKPATRICK, J. R., SANDER, J. D., ZHANG, F., VOYTAS, D. F., JOUNG, J. K., DOBBS, D. & COFFMAN, C. R. 2011. ZFNGenome: a comprehensive resource for locating zinc finger nuclease target sites in model organisms. *BMC Genomics,* 12**,** 83.

REYON, D., TSAI, S. Q., KHAYTER, C., FODEN, J. A., SANDER, J. D. & JOUNG, J. K. 2012. FLASH assembly of TALENs for high-throughput genome editing. *Nature biotechnology*.

ROSENFELD, N., YOUNG, J. W., ALON, U., SWAIN, P. S. & ELOWITZ, M. B. 2005. Gene regulation at the single-cell level. *Science,* 307**,** 1962-5.

ROUET, P., SMIH, F. & JASIN, M. 1994. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc Natl Acad Sci U S A,* 91**,** 6064-8.

ROZEN, S. & SKALETSKY, H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol,* 132**,** 365-86.

RUBIN-BEJERANO, I., MANDEL, S., ROBZYK, K. & KASSIR, Y. 1996. Induction of meiosis in Saccharomyces cerevisiae depends on conversion of the transcriptional

represssor Ume6 to a positive regulator by its regulated association with the transcriptional activator Ime1. *Mol Cell Biol,* 16**,** 2518-2526.

SALEH-GOHARI, N., BRYANT, H. E., SCHULTZ, N., PARKER, K. M., CASSEL, T. N. & HELLEDAY, T. 2005. Spontaneous homologous recombination is induced by collapsed replication forks that are caused by endogenous DNA single-strand breaks. *Mol Cell Biol,* 25**,** 7158-69.

SANDER, J. D., CADE, L., KHAYTER, C., REYON, D., PETERSON, R. T., JOUNG, J. K. & YEH, J. R. 2011a. Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nature biotechnology,* 29**,** 697-8.

SANDER, J. D., DAHLBORG, E. J., GOODWIN, M. J., CADE, L., ZHANG, F., CIFUENTES, D., CURTIN, S. J., BLACKBURN, J. S., THIBODEAU-BEGANNY, S., QI, Y., PIERICK, C. J., HOFFMAN, E., MAEDER, M. L., KHAYTER, C., REYON, D., DOBBS, D., LANGENAU, D. M., STUPAR, R. M., GIRALDEZ, A. J., VOYTAS, D. F., PETERSON, R. T., YEH, J. R. & JOUNG, J. K. 2011b. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat Methods,* 8**,** 67-9.

SANDER, J. D., MAEDER, M. L., REYON, D., VOYTAS, D. F., JOUNG, J. K. & DOBBS, D. 2010a. ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Res,* 38**,** W462-8.

SANDER, J. D., REYON, D., MAEDER, M. L., FOLEY, J. E., THIBODEAU-BEGANNY, S., LI, X., REGAN, M. R., DAHLBORG, E. J., GOODWIN, M. J., FU, F., VOYTAS, D. F., JOUNG, J. K. & DOBBS, D. 2010b. Predicting success of oligomerized pool engineering (OPEN) for zinc finger target site sequences. *BMC Bioinformatics,* 11**,** 543.

SANDER, J. D., YEH, J. R., PETERSON, R. T. & JOUNG, J. K. 2011c. Engineering zinc finger nucleases for targeted mutagenesis of zebrafish. *Methods in cell biology,* 104**,** 51-8.

SANDER, J. D., ZABACK, P., JOUNG, J. K., VOYTAS, D. F. & DOBBS, D. 2007. Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucleic Acids Res,* 35**,** W599-605.

SANDER, J. D., ZABACK, P., JOUNG, J. K., VOYTAS, D. F. & DOBBS, D. 2009. An affinity-based scoring scheme for predicting DNA-binding activities of modularly assembled zinc-finger proteins. *Nucleic Acids Res,* 37**,** 506-15.

SANDERS, K. L., CATTO, L. E., BELLAMY, S. R. & HALFORD, S. E. 2009. Targeting individual subunits of the FokI restriction endonuclease to specific DNA strands. *Nucleic Acids Res,* 37**,** 2105-15.

SANTIAGO, Y., CHAN, E., LIU, P. Q., ORLANDO, S., ZHANG, L., URNOV, F. D., HOLMES, M. C., GUSCHIN, D., WAITE, A., MILLER, J. C., REBAR, E. J., GREGORY, P. D., KLUG, A. & COLLINGWOOD, T. N. 2008. Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc Natl Acad Sci U S A,* 105**,** 5809-14.

SEBASTIANO, V., MAEDER, M. L., ANGSTMAN, J. F., HADDAD, B., KHAYTER, C., YEO, D. T., GOODWIN, M. J., HAWKINS, J. S., RAMIREZ, C. L., BATISTA, L. F., ARTANDI, S. E., WERNIG, M. & JOUNG, J. K. 2011. In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. *Stem cells,* 29**,** 1717-26.

SEGAL, D. J. 2011. Zinc-finger nucleases transition to the CoDA. *Nature methods,* 8**,** 53-5.

SEGAL, D. J., BEERLI, R. R., BLANCAFORT, P., DREIER, B., EFFERTZ, K., HUBER, A., KOKSCH, B., LUND, C. V., MAGNENAT, L., VALENTE, D. & BARBAS, C. F., 3RD 2003. Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry,* 42**,** 2137-48.

SEGAL, D. J., DREIER, B., BEERLI, R. R. & BARBAS, C. F., 3RD 1999. Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc Natl Acad Sci U S A,* 96**,** 2758-63.

SESSA, C. 2011. Update on PARP1 inhibitors in ovarian cancer. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO,* 22 Suppl 8**,** viii72-viii76.

SHIMIZU, Y., SOLLU, C., MECKLER, J. F., ADRIAENSSENS, A., ZYKOVICH, A., CATHOMEN, T. & SEGAL, D. J. 2011. Adding Fingers to an Engineered Zinc Finger Nuclease Can Reduce Activity. *Biochemistry,* 50**,** 5033-5041.

SHUKLA, V. K., DOYON, Y., MILLER, J. C., DEKELVER, R. C., MOEHLE, E. A., WORDEN, S. E., MITCHELL, J. C., ARNOLD, N. L., GOPALAN, S., MENG, X., CHOI, V. M., ROCK, J. M., WU, Y. Y., KATIBAH, G. E., ZHIFANG, G., MCCASKILL, D., SIMPSON, M. A., BLAKESLEE, B., GREENWALT, S. A., BUTLER, H. J., HINKLEY, S. J., ZHANG, L., REBAR, E. J., GREGORY, P. D. & URNOV, F. D. 2009. Precise genome modification in the crop species Zea mays using zinc-finger nucleases. *Nature,* 459**,** 437-41.

SIMSEK, D., BRUNET, E., WONG, S. Y., KATYAL, S., GAO, Y., MCKINNON, P. J., LOU, J., ZHANG, L., LI, J., REBAR, E. J., GREGORY, P. D., HOLMES, M. C. & JASIN, M. 2011. DNA Ligase III Promotes Alternative Nonhomologous End-Joining during Chromosomal Translocation Formation. *PLoS Genet,* 7**,** e1002080.

SOLDNER, F., LAGANIERE, J., CHENG, A. W., HOCKEMEYER, D., GAO, Q., ALAGAPPAN, R., KHURANA, V., GOLBE, L. I., MYERS, R. H., LINDQUIST, S., ZHANG, L., GUSCHIN, D., FONG, L. K., VU, B. J., MENG, X., URNOV, F. D., REBAR, E. J., GREGORY, P. D., ZHANG, H. S. & JAENISCH, R. 2011. Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell,* 146**,** 318-31.

SOLLU, C., PARS, K., CORNU, T. I., THIBODEAU-BEGANNY, S., MAEDER, M. L., JOUNG, J. K., HEILBRONN, R. & CATHOMEN, T. 2010. Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. *Nucleic Acids Res,* 38**,** 8269-76.

STRICKER, J., COOKSON, S., BENNETT, M. R., MATHER, W. H., TSIMRING, L. S. & HASTY, J. 2008. A fast, robust and tunable synthetic gene oscillator. *Nature,* 456**,** 516-519.

SZCZEPEK, M., BRONDANI, V., BUCHEL, J., SERRANO, L., SEGAL, D. J. & CATHOMEN, T. 2007. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat Biotechnol,* 25**,** 786-93.

TAKASU, Y., KOBAYASHI, I., BEUMER, K., UCHINO, K., SEZUTSU, H., SAJWAN, S., CARROLL, D., TAMURA, T. & ZUROVEC, M. 2010. Targeted mutagenesis in the silkworm Bombyx mori using zinc finger nuclease mRNA injection. *Insect Biochem Mol Biol,* 40**,** 759-65.

TAKATA, M., SASAKI, M. S., SONODA, E., MORRISON, C., HASHIMOTO, M., UTSUMI, H., YAMAGUCHI-IWAI, Y., SHINOHARA, A. & TAKEDA, S. 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 J,* 17**,** 5497-508.

TAMSIR, A., TABOR, J. J. & VOIGT, C. A. 2011. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature,* 469**,** 212-215.

TESTE, M. A., DUQUENNE, M., FRANCOIS, J. M. & PARROU, J. L. 2009. Validation of reference genes for quantitative expression analysis by real-time RT-PCR in Saccharomyces cerevisiae. *BMC molecular biology,* 10**,** 99.

THIBODEAU-BEGANNY, S. & JOUNG, J. K. 2007. Engineering Cys2His2 zinc finger domains using a bacterial cell-based two-hybrid selection system. *Methods Mol Biol,* 408**,** 317-334.

THIBODEAU, S. A., FANG, R. & JOUNG, J. K. 2004. High-throughput beta-galactosidase assay for bacterial cell-based reporter systems. *Biotechniques,* 36**,** 410-5.

TIGGES, M., MARQUEZ-LAGO, T. T., STELLING, J. & FUSSENEGGER, M. 2009. A tunable synthetic mammalian oscillator. *Nature,* 457**,** 309-312.

TOWNSEND, J. A., WRIGHT, D. A., WINFREY, R. J., FU, F., MAEDER, M. L., JOUNG, J. K. & VOYTAS, D. F. 2009. High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature,* 459**,** 442-5.

TUPLER, R., PERINI, G. & GREEN, M. R. 2001. Expressing the human genome. *Nature,* 409**,** 832-3.

URNOV, F. D. 2002. A feel for the template: zinc finger protein transcription factors and chromatin. *Biochem Cell Biol,* 80**,** 321-33.

URNOV, F. D., MILLER, J. C., LEE, Y. L., BEAUSEJOUR, C. M., ROCK, J. M., AUGUSTUS, S., JAMIESON, A. C., PORTEUS, M. H., GREGORY, P. D. & HOLMES, M. C. 2005. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature,* 435**,** 646-51.

URNOV, F. D., REBAR, E. J., HOLMES, M. C., ZHANG, H. S. & GREGORY, P. D. 2010. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet,* 11**,** 636-46.

VAN GENT, D. C., HOEIJMAKERS, J. H. & KANAAR, R. 2001. Chromosomal stability and the DNA double-stranded break connection. *Nat Rev Genet,* 2**,** 196-206.

VAN NIEROP, G. P., DE VRIES, A. A., HOLKERS, M., VRIJSEN, K. R. & GONCALVES, M. A. 2009. Stimulation of homology-directed gene targeting at an endogenous human locus by a nicking endonuclease. *Nucleic Acids Res,* 37**,** 5725-36.

VANAMEE, E. S., SANTAGATA, S. & AGGARWAL, A. K. 2001. FokI requires two specific DNA sites for cleavage. *J Mol Biol,* 309**,** 69-78.

VASQUEZ, K. M., MARBURGER, K., INTODY, Z. & WILSON, J. H. 2001. Manipulating the mammalian genome by homologous recombination. *Proc Natl Acad Sci U S A,* 98**,** 8403-10.

WANG, B. S., GRANT, R. A. & PABO, C. O. 2001. Selected peptide extension contacts hydrophobic patch on neighboring zinc finger and mediates dimerization on DNA. *Nat Struct Biol,* 8**,** 589-593.

WANG, B. S. & PABO, C. O. 1999. Dimerization of zinc fingers mediated by peptides evolved in vitro from random sequences. *Proc Natl Acad Sci U S A,* 96**,** 9568-9573.

WANG, J., FRIEDMAN, G., DOYON, Y., WANG, N. S., LI, C. J., MILLER, J. C., HUA, K. L., YAN, J. J., BABIARZ, J. E., GREGORY, P. D. & HOLMES, M. C. 2012. Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. *Genome research*.

WANG, X., LIU, Y., YANG, Z., ZHANG, Z., ZHOU, W., YE, Z., ZHANG, W., ZHANG, S., FENG, X., CHEN, F. & HU, R. 2011. Glucose metabolism-related protein 1 (GMRP1) regulates pancreatic beta cell proliferation and apoptosis via activation of Akt signalling pathway in rats and mice. *Diabetologia,* 54**,** 852-63.

WATANABE, M., UMEYAMA, K., MATSUNARI, H., TAKAYANAGI, S., HARUYAMA, E., NAKANO, K., FUJIWARA, T., IKEZAWA, Y., NAKAUCHI, H. & NAGASHIMA, H. 2010. Knockout of exogenous EGFP gene in porcine somatic cells using zinc-finger nucleases. *Biochem Biophys Res Commun,* 402**,** 14-8.

WAUGH, D. S. & SAUER, R. T. 1993. Single amino acid substitutions uncouple the DNA binding and strand scission activities of Fok I endonuclease. *Proc Natl Acad Sci U S A,* 90**,** 9596-600.

WEBER, W. & FUSSENEGGER, M. 2009. Engineering of synthetic mammalian gene networks. *Chem Biol,* 16**,** 287-297.

WEBSTER, N., JIN, J. R., GREEN, S., HOLLIS, M. & CHAMBON, P. 1988. The yeast UASG is a transcriptional enhancer in human HeLa cells in the presence of the GAL4 trans-activator. *Cell,* 52**,** 169-178.

WEINSTOCK, D. M. & JASIN, M. 2006. Alternative pathways for the repair of RAG-induced DNA breaks. *Mol Cell Biol,* 26**,** 131-9.

WHYTE, J. J., ZHAO, J., WELLS, K. D., SAMUEL, M. S., WHITWORTH, K. M., WALTERS, E. M., LAUGHLIN, M. H. & PRATHER, R. S. 2011. Gene targeting with zinc finger nucleases to produce cloned eGFP knockout pigs. *Mol Reprod Dev,* 78**,** 2.

WIGLER, M., SWEET, R., SIM, G. K., WOLD, B., PELLICER, A., LACY, E., MANIATIS, T., SILVERSTEIN, S. & AXEL, R. 1979. Transformation of mammalian cells with genes from procaryotes and eucaryotes. *Cell,* 16**,** 777-85.

WILEN, C. B., WANG, J., TILTON, J. C., MILLER, J. C., KIM, K. A., REBAR, E. J., SHERRILL-MIX, S. A., PATRO, S. C., SECRETO, A. J., JORDAN, A. P., LEE, G., KAHN, J., AYE, P. P., BUNNELL, B. A., LACKNER, A. A., HOXIE, J. A., DANET-DESNOYERS, G. A., BUSHMAN, F. D., RILEY, J. L., GREGORY, P. D., JUNE, C. H., HOLMES, M. C. & DOMS, R. W. 2011. Engineering HIV-resistant human CD4+ T cells with CXCR4-specific zinc-finger nucleases. *PLoS Pathog,* 7**,** e1002020.

WIN, M. N. & SMOLKE, C. D. 2008. Higher-order cellular information processing with synthetic RNA devices. *Science,* 322**,** 456-460.

WOLFE, S. A., GRANT, R. A., ELROD-ERICKSON, M. & PABO, C. O. 2001. Beyond the "recognition code": structures of two Cys2His2 zinc finger/TATA box complexes. *Structure,* 9**,** 717-23.

WOLFE, S. A., GREISMAN, H. A., RAMM, E. I. & PABO, C. O. 1999. Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code. *J Mol Biol,* 285**,** 1917-34.

WOLFE, S. A., RAMM, E. I. & PABO, C. O. 2000. Combining structure-based design with phage display to create new Cys(2)His(2) zinc finger dimers. *Structure,* 8**,** 739-50.

WOOD, A. J., LO, T. W., ZEITLER, B., PICKLE, C. S., RALSTON, E. J., LEE, A. H., AMORA, R., MILLER, J. C., LEUNG, E., MENG, X., ZHANG, L., REBAR, E. J., GREGORY, P. D., URNOV, F. D. & MEYER, B. J. 2011. Targeted genome editing across species using ZFNs and TALENs. *Science,* 333**,** 307.

WRIGHT, D. A., THIBODEAU-BEGANNY, S., SANDER, J. D., WINFREY, R. J., HIRSH, A. S., EICHTINGER, M., FU, F., PORTEUS, M. H., DOBBS, D., VOYTAS, D. F. & JOUNG, J. K. 2006. Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat Protoc,* 1**,** 1637-52.

WRIGHT, D. A., TOWNSEND, J. A., WINFREY, R. J., JR., IRWIN, P. A., RAJAGOPAL, J., LONOSKY, P. M., HALL, B. D., JONDLE, M. D. & VOYTAS, D. F. 2005. High-frequency homologous recombination in plants mediated by zinc-finger nucleases. *Plant J,* 44**,** 693-705.

WU, H., YANG, W. P. & BARBAS, C. F., 3RD 1995. Building zinc fingers by selection: toward a therapeutic application. *Proc Natl Acad Sci U S A,* 92**,** 344-8.

XIE, Z., WROBLEWSKA, L., PROCHAZKA, L., WEISS, R. & BENENSON, Y. 2011. Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science,* 333**,** 1307-1311.

YANG, D., YANG, H., LI, W., ZHAO, B., OUYANG, Z., LIU, Z., ZHAO, Y., FAN, N., SONG, J., TIAN, J., LI, F., ZHANG, J., CHANG, L., PEI, D., CHEN, Y. E. & LAI, L. 2011. Generation of PPARgamma mono-allelic knockout pigs via zinc-finger nucleases and nuclear transfer cloning. *Cell Res,* 21**,** 979-82.

YANOVER, C. & BRADLEY, P. 2011. Extensive protein and DNA backbone sampling improves structure-based specificity prediction for C2H2 zinc fingers. *Nucleic Acids Res*.

YOKOBAYASHI, Y., WEISS, R. & ARNOLD, F. H. 2002. Directed evolution of a genetic circuit. *Proc Natl Acad Sci U S A,* 99**,** 16587-16591.

YOUNG, J. J., CHERONE, J. M., DOYON, Y., ANKOUDINOVA, I., FARAJI, F. M., LEE, A. H., NGO, C., GUSCHIN, D. Y., PASCHON, D. E., MILLER, J. C., ZHANG, L., REBAR, E. J., GREGORY, P. D., URNOV, F. D., HARLAND, R. M. & ZEITLER, B. 2011. Efficient targeted gene disruption in the soma and germ line of the frog Xenopus tropicalis using engineered zinc-finger nucleases. *Proc Natl Acad Sci U S A,* 108**,** 7052-7.

ZHANG, F., MAEDER, M. L., UNGER-WALLACE, E., HOSHAW, J. P., REYON, D., CHRISTIAN, M., LI, X., PIERICK, C. J., DOBBS, D., PETERSON, T., JOUNG, J. K. & VOYTAS, D. F. 2010. High frequency targeted mutagenesis in Arabidopsis thaliana using zinc finger nucleases. *Proc Natl Acad Sci U S A,* 107**,** 12028-33.

ZOU, J., COCHRAN, R. & CHENG, L. 2010. Double knockouts in human embryonic stem cells. *Cell Res,* 20**,** 250-2.

ZOU, J., MAEDER, M. L., MALI, P., PRUETT-MILLER, S. M., THIBODEAU-BEGANNY, S., CHOU, B. K., CHEN, G., YE, Z., PARK, I. H., DALEY, G. Q., PORTEUS, M. H.,

JOUNG, J. K. & CHENG, L. 2009. Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell,* 5**,** 97-110.

ZOU, J., SWEENEY, C. L., CHOU, B. K., CHOI, U., PAN, J., WANG, H., DOWEY, S. N., CHENG, L. & MALECH, H. L. 2011. Oxidase-deficient neutrophils from X-linked chronic granulomatous disease iPS cells: functional correction by zinc finger nuclease-mediated safe harbor targeting. *Blood,* 117**,** 5561-72.

ZYKOVICH, A., KORF, I. & SEGAL, D. J. 2009. Bind-n-Seq: high-throughput analysis of in vitro protein-DNA interactions using massively parallel sequencing. *Nucleic Acids Res,* 37**,** e151.