

# Generierung und Evaluierung von hochspezifischen Zinkfinger-Nukleasen

Dissertation

zur Erlangung des akademischen Grades

Doctor rerum naturalium

(Dr. rer. nat.)

eingereicht im

Fachbereich Biologie, Chemie, Pharmazie

der Freien Universität Berlin



vorgelegt von

Dipl.-Biol. Felix Müller-Lerch

aus Berlin

November 2009

Die Doktorarbeit wurde in der Zeit von Januar 2006 bis November 2009 am Institut für Virologie der Charité - Universitätsmedizin Berlin (Campus Benjamin Franklin) unter der Leitung von Prof. Dr. Toni Cathomen angefertigt.

### **Erklärung**

Hiermit versichere ich, dass ich die vorliegende Doktorarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Berlin, den 2. November 2009

---

Felix Müller-Lerch

1. Gutachter: Prof. Dr. Toni Cathomen

Institut für Virologie, Charité - Universitätsmedizin Berlin, Campus Benjamin Franklin

2. Gutachter: Prof. Dr. Rupert Mutzel

Institut für Biologie - Mikrobiologie, Freie Universität Berlin

Disputation am: 14.12.2009

# Inhaltsverzeichnis

<b>1</b>	<b>Zusammenfassung/Summary .....</b>	<b>3</b>
1.1	Zusammenfassung.....	3
1.2	Summary.....	4
<b>2</b>	<b>Einleitung.....</b>	<b>5</b>
2.1	Gentherapie und Gentargeting .....	5
2.1.1	Gentherapie .....	5
2.1.2	Gentargeting und homologe Rekombination .....	7
2.1.3	Homologe Rekombination und DNA-Doppelstrangbrüche.....	8
2.2	Zinkfinger.....	9
2.2.1	Zinkfinger-Nukleasen.....	11
2.2.2	Einsatz von Zinkfinger-Nukleasen.....	13
2.2.3	Selektionssysteme für Zinkfinger-Nukleasen .....	15
2.3	Vektoren zum Einbringen von Zinkfinger-Nukleasen.....	18
<b>3</b>	<b>Einzelarbeiten.....</b>	<b>19</b>
3.1	Titel der publizierten Einzelarbeiten.....	19
3.2	Eigenanteil an den publizierten Einzelarbeiten .....	20
3.3	Publizierte Ergebnisse .....	22
3.3.1	<i>Generation and Functional Analysis of Zinc Finger Nucleases</i> .....	22
3.3.2	<i>Unexpected failure rates for modular assembly of engineered zinc fingers</i> .....	23
3.3.3	<i>Rapid ‘‘Open-Source’’ Engineering of Customized Zinc-Finger Nucleases for Highly Efficient Gene Modification</i> .....	25
<b>4</b>	<b>Diskussion .....</b>	<b>28</b>
<b>5</b>	<b>Referenzen .....</b>	<b>40</b>
<b>6</b>	<b>Anhang.....</b>	<b>49</b>
6.1	Danksagung .....	49
6.2	Lebenslauf .....	50
6.3	Publikationen .....	51
6.4	Vorträge.....	51
6.5	Einzelarbeiten .....	52

# 1 Zusammenfassung/Summary

## 1.1 Zusammenfassung

Zinkfinger-Nukleasen (ZFN), welche als chimäre Proteine aus einer hochspezifischen Zinkfinger-DNA-Bindungsdomäne (ZF-DBD) und einer unspezifischen, nur als Dimer aktiven *FokI*-Nukleasedomäne bestehen, stellen ein sehr effizientes Werkzeug für das gezielte Einfügen eines Doppelstrangbruchs (DSB) in jede DNA-Sequenz einer Zelle dar. Die ZF-DBD setzen sich aus einzelnen Zinkfinger (ZF)-Modulen zusammen, wobei jeweils ein Modul an drei Basen innerhalb der gewünschten ZF-Bindungsstelle bindet. So können ZFN sowohl zur gezielten Mutagenese von Genen als auch, in Anwesenheit einer zum geschädigten DNA-Lokus homologen Donor-DNA, zur DSB-induzierten Stimulation des Gentergetings angewandt werden. Das Ziel der Doktorarbeit war der Vergleich zweier bestehender sowie die Entwicklung einer neuen, auf Semliki-Forest-Viren (SFV) basierenden Plattform zur Generierung und Selektion von ZFN. Ausgehend von einer modularen Architektur der ZF-DBD stand am Anfang der Doktorarbeit die Etablierung eines PCR-basierten (*polymerase chain reaction*) Gensynthesystems zum Zusammenfügen einzelner ZF-Module mit bekannten DNA-Bindungseigenschaften (*modular assembly*). Dabei erwies sich das PCR-System hinsichtlich seiner Fehleranfälligkeit als sehr zuverlässig. Allerdings ergaben umfangreiche funktionelle Tests in Bakterien und humanen Zelllinien, dass die Erfolgsrate, mit modular assemblierten, aus drei Modulen bestehenden ZF-DBD funktionelle ZFN-Paare zu generieren, lediglich 3% beträgt. *Modular assembly* scheint deshalb für die Generierung funktioneller ZFN nicht geeignet zu sein, da einzelne Module innerhalb einer ZF-DBD nicht unabhängig voneinander agieren. Zum Vergleich wurden ZF-DBD mit einem kontext-abhängigen Selektionssystem generiert, in dem die ZF-Module im Kontext ihrer benachbarten Module selektioniert wurden. Die Mehrheit der resultierenden ZFN vermochte hocheffizient genetische Modifikationen in einem chromosomal vorliegenden Reportergen sowie in vier endogenen Loci in humanen bzw. pflanzlichen Zellen herbeizuführen, wobei absolute Gentergetingraten von 1-50% erzielt werden konnten. Darüber hinaus wurde mit der Entwicklung des SFV-Systems begonnen, um ZF-Proteine in Zukunft im eukaryotischen Kontext selektionieren zu können. Zusammenfassend lässt sich festhalten, dass im Rahmen dieser Doktorarbeit essentielle Erkenntnisse bezüglich der Generierung und Evaluierung von ZFN gewonnen wurden, die einen wichtigen Beitrag für zukünftige ZFN-vermittelte Genommodifikationen darstellen.



## 1.2 Summary

Zinc finger nucleases (ZFN) are chimeric proteins composed of a highly specific zinc finger DNA-binding domain (ZF-DBD) linked to a non-specific *FokI*-nuclease cleavage domain that is only active as a dimer. These artificial nucleases are efficient tools for the targeted insertion of a double-strand break (DSB) into any chosen DNA sequence of a cell. A ZF-DBD is composed of three single zinc finger (ZF) modules, each one recognizing three nucleotides within the target site. Hence, ZFN can be used for the targeted mutagenesis of genes and, in the presence of a donor DNA that harbors homologous sequences to the damaged locus, for the stimulation of gene targeting. The aim of this doctoral thesis was the comparison of two existing as well as the establishment of a new Semliki-Forest-Virus (SFV) based platform for the generation and selection of ZFN. Based on the modular architecture of ZF-DBD, the first step was to establish a PCR-based (polymerase chain reaction) system for the synthesis of ZF-DBDs composed of three single ZF modules with known DNA-binding properties (modular assembly). The PCR-based gene synthesis system turned out to be very reliable with regard to error proneness. However, large-scale functional tests in bacteria and human cell lines showed that the success rate for generating functional ZFN pairs made by the modular assembly method was only about 3%. Hence, modular assembly was likely to fail because single ZF modules within a ZF-DBD seem not to act independently. For comparison, ZF-DBDs were constructed with a context-sensitive selection system by selecting individual ZF modules in the context to their neighboring modules. The resulting ZFN were able to induce highly efficient genetic modifications in a chromosomally integrated EGFP (enhanced green fluorescent protein) reporter gene as well as in four endogenous loci in human and plant cells, respectively. The absolute frequencies of gene targeting ranged from 1–50%. In addition, the development of the SFV-based selection system that aims at enabling the selection of ZF proteins in a eukaryotic context in the future was initiated. In conclusion, essential results concerning the generation and evaluation of ZFN could be gained in the framework of this doctoral thesis, which will be important for ZFN-mediated genome modifications in the future.

## 2 Einleitung

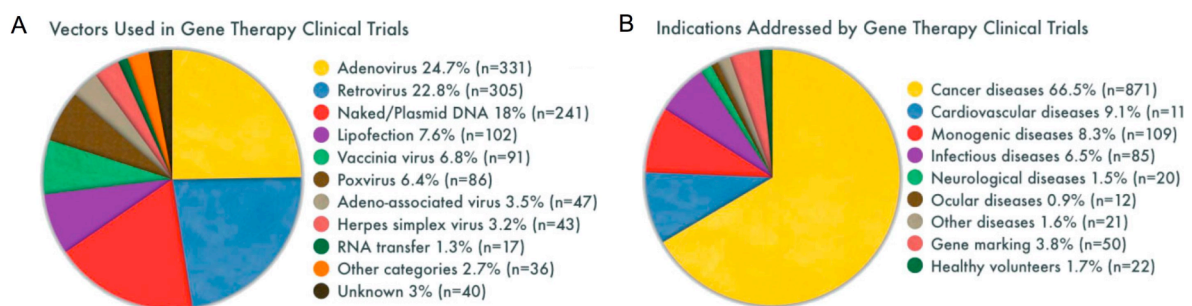
### 2.1 Gentherapie und Gentergeting

Der Bauplan höherer Organismen besteht aus DNA (*deoxyribonucleic acid*, zu deutsch Desoxyribonukleinsäure) und setzt sich im Wesentlichen aus vier verschiedenen Bausteinen, den Nukleobasen Adenin (A), Cytosin (C), Guanin (G) und Thymin (T) zusammen [1]. So wie einzelne Buchstaben aneinander gefügt werden, um Wörter und Sätze zu bilden, bestimmen Anzahl und Abfolge der Nukleobasen den Aufbau der Gene und somit den genetischen Fingerabdruck eines Organismus. Die Gene wiederum kodieren als kleinste funktionelle Einheiten der DNA die mRNA (*messenger ribonucleic acid*, zu deutsch Boten-Ribonukleinsäure) und damit die Proteine, die als Bestandteil von Muskeln, Hormonen, Enzymen und des Immunsystems die essentiellen Funktions- und Strukturkomponenten der Zellen eines Lebewesens sind. Das menschliche Genom, das aus circa 3,4 Milliarden Basenpaaren besteht, beinhaltet ungefähr 25.000 Gene. Trotz dieser enormen Genomgröße kann der Wegfall (Deletion), Austausch (Transversion bzw. Transition) oder Einbau (Insertion) einer einzigen bzw. einiger Basen zu einer fehlerhaften Genexpression, Proteinfehlfaltungen und somit zu Erbkrankheiten wie der Sichelzellanämie [2] oder zu erworbenen Krankheiten wie Krebs [3] führen. An dieser Stelle kann die Gentherapie ansetzen.

#### 2.1.1 Gentherapie

Die Gentherapie ist allgemein definiert als der Transfer genetischen Materials in eine Zelle, ein Gewebe oder ein ganzes Organ mit dem Ziel, eine Krankheit zu heilen, oder zumindest den klinischen Zustand eines Patienten zu verbessern [4]. Das Feld der Gentherapie kann seinerseits in zwei Bereiche, die *Ex-vivo*- und die *In-vivo*-Gentherapie, aufgeteilt werden [5]. Bei dem *Ex-vivo*-Ansatz werden dem Patienten die relevanten Zellen entnommen, im Labor genetisch modifiziert und anschließend wieder reimplantiert [6]. Dieser Ansatz eignet sich am besten für Applikationen, bei denen die Zielzellen - z.B. hämatopoetische Stammzellen - leicht entnommen werden können und, wie im Falle der angeborenen Immunschwächekrankheit X-SCID (*X-linked severe combined immunodeficiency*), nach der erfolgreichen genetischen Modifikation und Reimplantation einen Wachstumsvorteil haben [7]. Bei der *In-vivo*-Methode wird der Vektor mit dem kurativen Genmaterial direkt in den Organismus appliziert, entweder lokal (z.B. intramuskulär) oder systemisch (z.B. intravenös) [8]. Ein gemeinsamer Punkt der beiden genannten Therapieformen ist dabei die Wahl eines geeigneten Gentransfervektors, welcher als Vehikel für den Transport des genetischen

Materials zum Wirkungsort benötigt wird. Die derzeit für die Gentherapie verfügbaren Vektoren lassen sich dabei grundsätzlich in virale und nicht virale Transfervehikel aufteilen (Abb. 1 A), wobei in über zwei Drittel der weltweit bis 2007 durchgeführten Gentherapiestudien virale Vektoren verwendet wurden [9]. Generell sollte der ideale Gentransfervektor unter anderem einfach und effizient produzierbar, nicht immunogen, anhaltend bezüglich seiner Transgenexpression, gewebespezifisch, apathogen bzw. nicht toxisch, ausreichend hinsichtlich seiner Kapazität sowie für eine ortsspezifische Integration und Infektion sich nicht teilender Zellen geeignet sein [10]. Findet die Gentherapie heutzutage auch Anwendung bei erworbenen Krankheiten wie Infektionskrankheiten [11], Herz-Kreislauf-Erkrankungen [12] und vor allem bei Krebs [13] (Abb. 1 B), so wurde sie



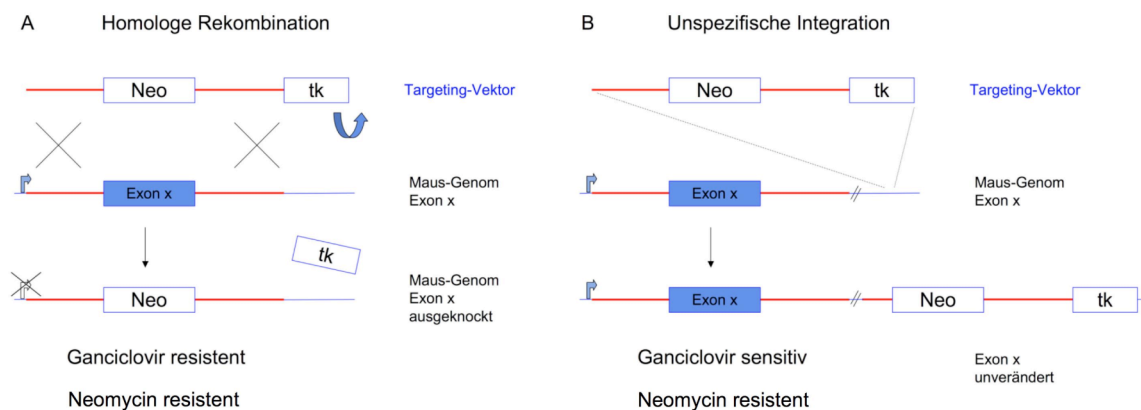
**Abbildung 1 Klinische Gentherapiestudien A)** In klinischen Studien eingesetzte Genterapievektoren. **B)** Anwendungsgebiete der Gentherapie. Abbildungen aus [9]; Stand 2007

ursprünglich zur Behandlung von monogenetischen Erbkrankheiten konzipiert [14]. Die klassische und bis heute erfolgreichste Form der Gentherapie für Erbkrankheiten ist die so genannte Genaddition [15]. Bei dieser wird eine unter der Kontrolle eines exogenen Promotors stehende und mit einem Terminationssignal versehene intakte Wildtypsequenz des mutierten Gens in die Zelle eingeführt. Die Expression des episomal vorliegenden bzw. chromosomal integrierten therapeutischen Transgens [16] sollte dann, zumindest bei rezessiv vererbten Krankheiten mit vorliegenden *loss-of-function*-Mutationen (Funktionsverlustmutationen), zur phänotypischen Korrektur auf zellulärer Ebene führen. Auch wenn mit der Genaddition bereits Erfolge erzielt werden konnten [17-19], birgt diese Variante der Gentherapie auch einige Nachteile. So wurden unerwünschte Nebeneffekte wie die Insertionsmutagenese beschrieben, bei der durch die Insertion eines retroviralen Vektors ein zelluläres Onkogen konstitutiv aktiviert bzw. ein Tumorsuppressorgen inaktiviert werden kann, was dann wiederum unter Umständen zu unkontrolliertem Zellwachstum und Krebs führt [20]. Zudem stellt die vom exogenen Promotor gesteuerte, unregulierbare

Genexpression ein Problem dar. Zusätzlich kann es durch DNA-Methylierung des integrierten Transgens zur Stilllegung (*gene silencing*) der Genexpression kommen [21]. Eine alternative Gentherapiestrategie, welche die genannten Mankos der Genaddition nicht aufweist, basiert auf der Grundlage des so genannten Gentergetings.

### 2.1.2 Gentergeting und homologe Rekombination

Der Begriff Gentergeting wurde vor circa 30 Jahren von Thomas und Capecchi [22] geprägt und bezeichnet die ortsspezifische, gezielte Modifikation eines Genoms, wobei das transferierte Genmaterial nicht mehr in das Genom der Zielzelle eingebaut werden soll, sondern lediglich als exogene Matrize für die homologe Rekombination (HR) mit dem endogenen Ziellokus dient. Ursprünglich wurde das System des Gentergetings in murinen



**Abbildung 2 Selektion von homolog rekombinierten und unspezifisch integrierten Targetingvektoren.**

**A)** Homologe Rekombination. Exon x wird bei der HR mit dem Targetingvektor durch die Neo-Kassette ersetzt, wodurch eine vollständige Transkription des modifizierten Zielgens verhindert wird (durchgestrichener Pfeil). Als nicht homologer Bereich geht die tk-Kassette bei der HR verloren, wodurch die Zellen sowohl Neomycin-resistent als auch unempfindlich gegenüber dem Medikament Ganciclovir sind, welches zum Absterben Thymidin-Kinase-exprimierender Zellen führt. **B)** Unspezifische Integration. Die unspezifische Integration des Targetingvektors inklusive der tk-Kassette bewirkt, dass die betroffenen Zellen zwar Neomycin-resistent, jedoch auch Ganciclovir-sensitiv werden und dementsprechend eine Doppelselektion mit Neomycin und Ganciclovir nicht überleben. Da Exon x unverändert vorliegt, kann die Transkription des Zielgens normal verlaufen (blaue Pfeile); tk: Thymidin-Kinase-Expressionskassette (negativer Selektionsmarker); Neo: Neomycin-Resistenz-Expressionskassette (positiver Selektionsmarker); rote Linie: homologe Bereiche

embryonalen Stammzellen angewandt [23] und zur Herstellung von *Knock-out-* bzw. *Knock-in-*Mäusen konzipiert [24], welche seitdem für die Untersuchung von Genfunktionen und als Modelle für menschliche Krankheiten unentbehrlich geworden sind. Wie in Abbildung 2 dargestellt, liegt dem klassischen Gentergeting mittels linearisierten Plasmidvektoren, welche über Techniken wie Elektroporation oder Mikroinjektion in die Zellen eingebracht werden,

eine Doppelselektion mit zwei verschiedenen Selektionsmarkern zugrunde. Mit dieser kann zwischen HR-basiertem Gentergeting und unspezifischer Integration des Targetingvektors unterschieden werden [25]. Eine häufig eingesetzte Positiv-Negativ-Selektion mit den Wirkstoffen Neomycin und Ganciclovir lässt dabei nur Zellen überleben, deren erworbene Neomycin-Resistenz das Resultat einer HR mit dem Targetingvektor und damit eines erfolgreichen Gentergetings ist. Die niedrige HR-Frequenz in Säugetierzellen (ein Ereignis pro  $10^5$  bis  $10^7$  Zellen) im Verhältnis zum hohen Level an unspezifisch integrierenden Targetingvektoren (ein Ereignis in  $10^2$  bis  $10^4$  Zellen) sowie die Notwendigkeit einer Positiv-Negativ-Doppelselektion hat die Gentergetingstechnik lange Zeit von therapeutischen Anwendungen abgehalten [26]. Die HR ist ein evolutionär hochkonservierter, in Prokaryoten sowie Eukaryoten stattfindender Prozess. Während der Meiose vermittelt die HR die Neukombination von maternalen und paternalen Allelen in Vorläuferzellen der Gameten und trägt damit zur Entstehung genetischer Vielfalt bei. Darüber hinaus dient die HR der Instandsetzung von unvollständigen Telomeren und beschädigten DNA-Replikationsgabeln [27] und ist an der Reparatur von DNA-Doppelstrangbrüchen (DSB) beteiligt. In somatischen Zellen mit einem vorliegenden DSB findet die HR vor der Mitose durch die Rekombination der beschädigten DNA-Sequenz mit dem homologen Bereich des intakten Schwesterchromatids statt und dient damit der Erhaltung der genomischen Integrität der Zelle sowie der Vermeidung von Krebs [28]. Der Grundstein, die Gentergetingstechnik doch eines Tages für therapeutische Anwendungen nutzen zu können, wurde Mitte der achtziger Jahre gelegt. Dabei wurde die Beobachtung gemacht, dass die Anzahl an HR-Ereignissen durch das Einfügen eines DSB gesteigert werden kann.

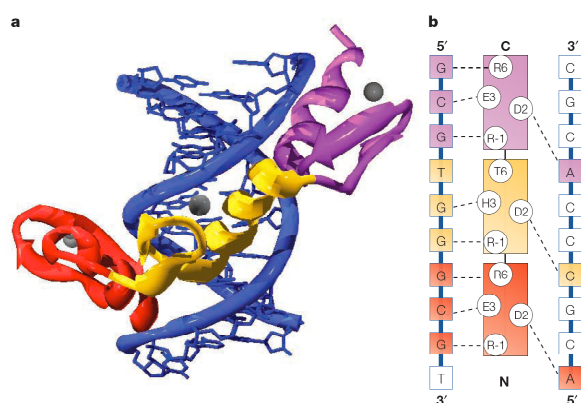
### 2.1.3 Homologe Rekombination und DNA-Doppelstrangbrüche

Erste Untersuchungen in Hefe [29] und Säugetierzellen [30] zeigten, dass der Prozess der HR zwischen einem DNA-Plasmid und einem chromosomalen Locus durch die Induktion eines DSB im Donor-Plasmid stimuliert werden kann. Später konnte darüber hinaus die Stimulation des HR-vermittelten Gentergetings durch das Einfügen eines DSB im Ziellokus demonstriert werden [31]. Zum Einfügen des DSB wurde bei den Initialversuchen die selten schneidende Meganuklease I-*SceI* verwendet, eine Intron-kodierte Homing-Endonuklease aus den Mitochondrien der Bäckerhefe *Saccharomyces cerevisiae* [32]. In seiner natürlichen Funktion katalysiert I-*SceI* seine eigene Duplikation durch das Setzen eines DSB in Empfängerallelen, wodurch der Transfer der entsprechenden I-*SceI*-Gensequenz per HR initiiert wird [33]. Die Erkennungssequenz von I-*SceI* mit einer Länge von 18 Basenpaaren sollte statistisch gesehen

höchsten einmal in  $7 \times 10^{10}$  Basenpaaren ( $4^{18}$ ) existieren und kommt im menschlichen Genom nicht vor [34]. Durch das Einführen der I-SceI-Erkennungssequenz in ein Zielgen konnte das HR-vermittelte Gentergeting in eukaryotischen Zellen nach Expression von I-SceI um mehr als das 1000-fache stimuliert werden [35]. Die I-SceI-vermittelte Stimulation des Gentergetings konnte bereits in verschiedenen Zelllinien und murinen embryonalen Stammzellen gezeigt werden, was darauf hinweist, dass die Induktion der HR durch gezieltes Setzen eines DSB ein universelles zelluläres Phänomen ist [36]. Mittels Reporter-genversuchen wurden unter optimalen Bedingungen Gentergetingraten von ~10% erreicht [37]. Obwohl das I-SceI-System ein großartiges Modell für die Untersuchung der Mechanismen des Gentergetings darstellt, ist die therapeutische Anwendbarkeit von Meganukleasen limitiert, da endogene Gene keine Erkennungssequenzen für I-SceI oder andere Homing-Endonukleasen aufweisen [38]. Um die stimulierende Eigenschaft von DSB im Gentergetingkontext für therapeutische Zwecke nutzen zu können, muss man Werkzeuge zur Hand haben, mit denen man ortsspezifisch DSB in jeder gewünschten chromosomalen Sequenz erzeugen kann. Als dafür geeignet erwiesen haben sich bis jetzt verschiedene Methoden basierend auf Triplex-formenden Oligonukleotiden [39], modifizierten Polyamiden [40], modifizierten Peptid-Nukleinsäuren [41], modifizierten Homing-Endonukleasen [42] und vor allem Zinkfinger-Nukleasen (ZFN).

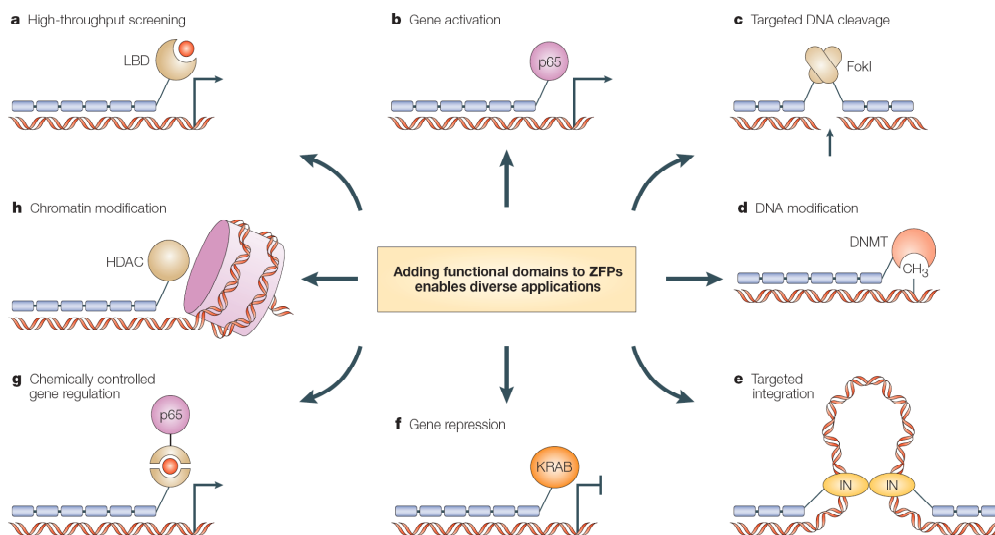
## 2.2 Zinkfinger

Das Motiv des Zinkfingers (ZF) wurde erstmals 1985 von Sir Aaron Klug und Mitarbeitern als repetitiver Bestandteil des Transkriptionsfaktors TFIIIA, welcher aus den Oozyten des Krallenfrosches *Xenopus laevis* isoliert wurde, identifiziert [43]. Bei dem entdeckten Motiv



**Abbildung 3 Charakteristische Bindung von Cys<sub>2</sub>His<sub>2</sub>-ZF an ihre DNA-Zielsequenz.** **a)** Charakteristische ZF-DNA-Bindung. Die drei Cys<sub>2</sub>His<sub>2</sub>-ZF-Module (rot, gelb und lila) einer ZF-DBD binden über ihre  $\alpha$ -Helices an die große Furche der DNA. Zink-Ionen sind als graue Kugeln dargestellt. **b)** Charakteristische ZF-DNA-Interaktion. Die  $\alpha$ -Helices einer ZF-DBD sind als Rechtecke zwischen den beiden DNA-Strängen dargestellt. Über ihre Schlüsselaminosäuren (Kreise) an den Positionen -1, +2, +3 und +6 kontaktiert jede  $\alpha$ -Helix drei Basen (Quadrate) in der großen Furche der DNA, wobei die Aminosäure an Position 2 in der  $\alpha$ -Helix eines Cys<sub>2</sub>His<sub>2</sub>-ZF-Moduls mit einer Base in der Bindungsstelle eines benachbarten ZF-Moduls interagiert. Abbildung aus [56].

handelt es sich um den klassischen Cys<sub>2</sub>His<sub>2</sub>-ZF, welcher aus etwa 30 Aminosäuren besteht. Diese sind in einer ββα-Konfiguration gefaltet, wobei sich die spezifische fingerartige Sekundärstruktur dadurch ausbildet, dass zwei Cysteine im Bereich der N-terminal gelegenen β-Faltblätter und zwei Histidine in der sich am C-Terminus befindenden α-Helix ein als Chelator fungierendes Zink-Ion kontaktieren [44]. Cys<sub>2</sub>His<sub>2</sub>-ZF sind die am häufigsten vorkommenden DNA-Bindungsmotive in eukaryotischen Transkriptionsfaktoren [45]. Neben DNA-bindenden ZF gibt es auch RNA- sowie Protein-bindende ZF [46]. In der Rolle als DNA-Bindungsmotiv bindet jeder Cys<sub>2</sub>His<sub>2</sub>-ZF über seine α-Helix hochspezifisch an drei Basen in der großen Furche der DNA, wobei modular assemblierte Cys<sub>2</sub>His<sub>2</sub>-ZF innerhalb eines Transkriptionsfaktors mehrere Triplets kontaktieren. Dabei koordinieren innerhalb der α-Helix bestimmte Schlüsselaminosäuren an den Positionen -1, +2, +3 und +6 relativ zum Start der Helix den Kontakt zur DNA (Abb. 3) [47]. Natürlich vorkommende DNA-bindende ZF-Proteine bestehen in der Regel aus zwei funktionellen Domänen, der aus mehreren ZF-Modulen bestehenden Zinkfinger-DNA-Bindungsdomäne (ZF-DBD) und einer Effektor-domäne, welche im Fall von Transkriptionsfaktoren eine Aktivierungsdomäne (AD) ist [48]. Dabei erkennt eine aus drei ZF-Modulen bestehende ZF-DBD eine Sequenz von neun Basenpaaren bzw. eine ZF-DBD aus sechs ZF-Modulen (6er-ZF-DBD)



**Abbildung 4 ZF-DBD können mit verschiedenen Effektor-domänen für vielfältige Applikationen kombiniert werden.** Als DBD können ZF-Proteine verbunden werden mit **A)** einer Liganden-Bindungsdomäne (LBD), **B)** einer Aktivierungsdomäne wie der p65-Domäne vom zellulären Transkriptionsfaktor NF-κB, **C)** einer Nukleasedomäne von Typ-II-Restriktionsendonukleasen wie *FokI*, **D)** einer DNA-Methyltransferase, **E)** einer retroviralen Integrase für die Insertion exogener Sequenzen in die DNA, **F)** einer Repressordomäne wie KRAB, **G)** einem Liganden-Aktivator-domänen-Komplex bzw. **H)** einer Domäne zur Modifikation des Chromatinstatus. Dargestellt sind ZF-DBD bestehend aus sechs ZF-Modulen (hellblaue Rechtecke). Abbildung aus [56].



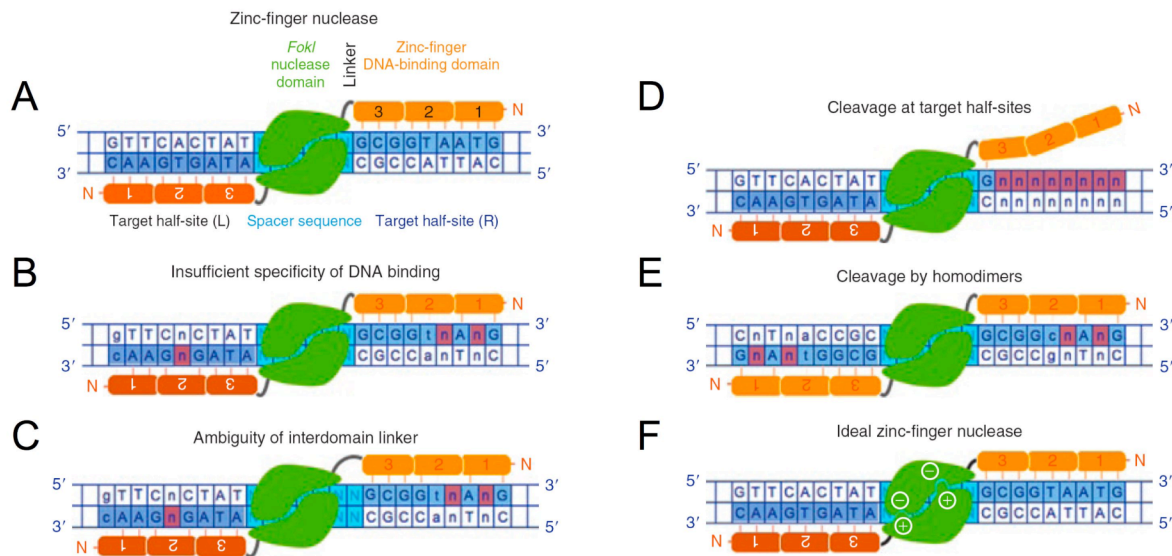
eine 18 Basenpaarsequenz [49]. Durch den gezielten Austausch der Effektor-domäne gegen verschiedene funktionelle Domänen können künstliche DNA-bindende ZF-Proteine für diverse therapeutische und forschungsorientierte Anwendungen hergestellt werden (Abb. 4). So wurden unter anderem bereits DNA-Methyltransferasen [50] zur gezielten Methylierung von DNA-Sequenzen [51], verschiedene Aktivator-domänen wie die VP16-Domäne vom Herpes Simplex Virus [52] oder die p65-Domäne vom zellulären Transkriptionsfaktor NF- $\kappa$ B [53] zur gezielten Transkriptionsaktivierung, Repressor-domänen wie KRAB (*Krüppel associated box*) [54] zur gezielten Transkriptionsstilllegung und Nukleasedomänen von Typ-II-Restriktionsendonukleasen [55] wie *FokI* zum gezielten Einführen von DSB als Effektor-domänen in künstlichen ZF-Proteinen eingesetzt [56].

### 2.2.1 Zinkfinger-Nukleasen

Die ersten ZFN wurden 1996 von Chandrasegaran und Mitarbeitern entwickelt. Dabei wurden Cys<sub>2</sub>His<sub>2</sub>-ZF-Module als spezifische ZF-DBD am N-Terminus der ZFN über eine kurze, *Linker* genannte Proteinsequenz mit der C-terminal gelegenen unspezifischen *FokI*-Nukleasedomäne verbunden [57]. Da das *FokI*-Restriktionsenzym nur als Dimer enzymatisch aktiv ist [58], müssen zwei ZFN (ZFN-Paar) zum Setzen eines DSB auf der DNA-Zielsequenz über ihre *FokI*-C-Termini dimerisieren [59]. Der Bereich zwischen den beiden gegenläufigen ZFN-Bindungsstellen wird dabei als *Spacer* bezeichnet (Abb. 5 A). Wie in chromosomalen Transfektionsansätzen gezeigt wurde, besteht die ideale Bindungsstellen-ZFN-Kombination aus ZFN-Bindungsstellen mit einem dazwischen liegenden sechs Basenpaar langen *Spacer* und ZFN mit vier Aminosäure langen *Linkern* [60]. Die ZF-DBD der meisten bis dato in der Literatur verwendeten ZFN bestehen aus drei bzw. vier ZF-Modulen, welche beliebig tandemartig arrangiert werden können, um so eine Vielfalt an neuen DNA-Zielsequenzen zu kontaktieren [48, 56, 61]. Aufgrund der Tatsache, dass jedes dieser drei bzw. vier ZF-Module drei Basen kontaktiert und ZFN nur als Dimere agieren, besteht eine vollständige ZFN-Paar-Bindungsstelle aus 18 bzw. 24 Basenpaaren, welche statistisch gesehen einzigartig im menschlichen Genom sein sollte. Studien in verschiedenen Zelllinien und Organismen zeigten, dass durch das ortsspezifische, ZFN-vermittelte Setzen eines DSB Gantargetingfrequenzen zwischen 1 und 39% erreicht werden können [62-70]. Trotz dieser Erfolge kam es in menschlichen Zellen zu ZFN-induzierten Toxizitätsreaktionen wie Apoptose und Zelltod, welche wahrscheinlich durch das vermehrte Einführen von DSB an *off-target sites* (Bereiche außerhalb der eigentlichen Bindungsstelle) verursacht wurden [37, 71, 72]. Wie in Abbildung 5 dargestellt, kann das Auftreten der ZFN-vermittelten



Toxizität dabei mehrere Ursachen haben [73]. Eine suboptimale Spezifität der ZF-DBD bewirkt eine unerwünschte Anlagerung der ZFN an eine verwandte Ziellokussequenz und



**Abbildung 5 ZFN-Architektur und mögliche Ursachen für eine ZFN-off-target-Aktivität.** A) Eine ZFN besteht aus einer Zinkfinger-DNA-Bindungsdomäne (ZF-DBD; rot bzw. orange), welche über einen *Linker* mit einer *FokI*-Nukleasedomäne (grün) verbunden ist. Die ZF-DBD setzt sich aus einzelnen ZF-Modulen zusammen, wobei jeweils ein Modul drei Basen in der ZFN-Bindungsstelle kontaktiert. Die beiden invertiert vorliegenden ZFN-Bindungsstellen sind durch einen *Spacer* (türkis) separiert. Mögliche Ursachen für eine ZFN-off-target-Aktivität: B) unspezifische DNA-Bindung, C) die Ambiguität des Aminosäure-Linkers, D) Induktion eines DSB durch Protein-Protein-Interaktion, obwohl nur eine der beiden ZFN an ihre Bindungsstelle gebunden hat, E) Bildung von ZFN-Homodimeren aufgrund von symmetrischen *FokI*-Dimerisierungsbereichen, F) Architektur eines idealen ZFN-Paares durch den Einsatz asymmetrischer und niedrigaffiner *FokI*-Dimerisierungsbereiche, hochaffiner ZF-DBD und einer optimalen *Linker-Spacer*-Kombination. Abbildung aus [73].

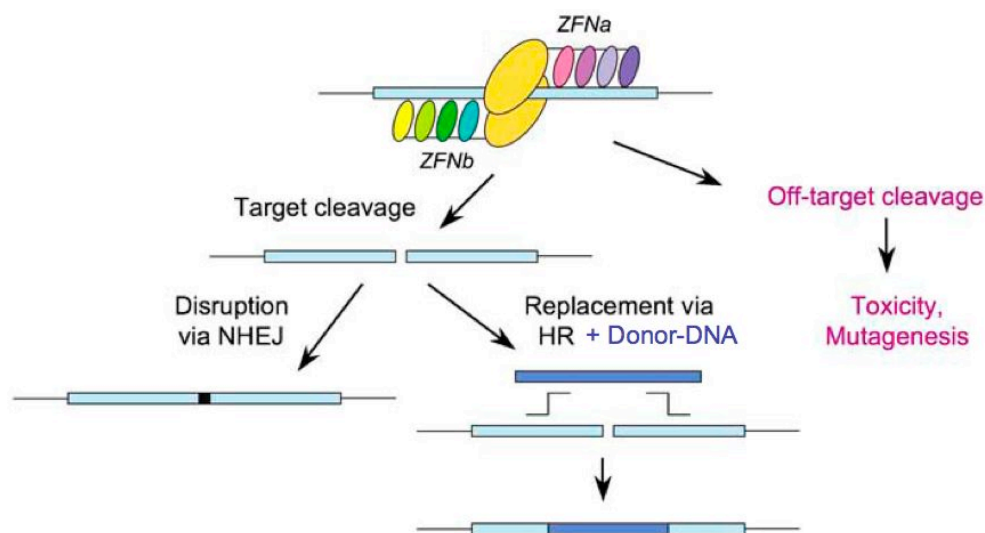
dadurch das Setzen eines unspezifischen DSB (Abb. 5 B) [74]. Ein weiteres Problem könnte sein, dass diese Bindungsstellen für jede einzelne ZFN mehrmals im Genom vorhanden sind, sodass auch suboptimale *Spacer-Linker*-Verhältnisse zu einer Dimerisierung der ZFN führen. Die Ambiguität der *Linker* kann so unter Umständen auch bei räumlich eigentlich zu weit entfernten oder zu nahe liegenden DNA-Bindungsstellen das ungewollte Einfügen eines DSB zur Folge haben (Abb. 5 C) [60, 75]. Zusätzlich kann eine zu hohe Affinität der *FokI*-Domänen zueinander die Induktion eines DSB durch eine Protein-Protein-Interaktion bewirken, obwohl nur eine der beiden ZFN an ihre Bindungsstelle gebunden hat (Abb. 5 D) [76]. Zuletzt kann die Symmetrie der *FokI*-Dimerisierungsbereiche zur Bildung von Homodimeren an zwei ähnlichen Bindungsstellen beitragen, wodurch es zu ungewollten DSB an zusätzlichen DNA-Sequenzen kommen kann (Abb. 5 E) [77]. Zur Reduzierung der genannten ZFN-vermittelten Toxizitätsreaktionen können daher drei verschiedene Strategien angewandt werden. Neben der Erhöhung der DNA-Bindungsspezifität der ZF-DBD [74, 78-

80] tragen der Einsatz einer optimalen *Linker-Spacer*-Kombination [60, 75] und die Verwendung von asymmetrischen, niedrigaffinen *FokI*-Domänen zur Konstruktion eines idealen ZFN-Paares bei (Abb. 5 F). Bezüglich des letzten Punktes wurde zur Abschwächung der hydrophoben Wechselwirkungen zwischen den *FokI*-Dimerisierungsbereichen die Aminosäure Isoleucin (hydrophobe Aminosäure) gegen Alanin bzw. Valin (neutrale Aminosäuren) in den *FokI*-Domänen ausgetauscht [37].

### 2.2.2 Einsatz von Zinkfinger-Nukleasen

Die Reparatur von ZFN-induzierten DSB kann prinzipiell durch einen von zwei potentiellen Reparaturwegen erfolgen; entweder durch das *nonhomologous end-joining* (NHEJ) oder die HR (Abb. 6). Nach der Expression eines ZFN-Paares und der Induktion eines DSB kann es je nach eingeschlagenem Reparaturweg zu verschiedenen Resultaten kommen [81]. So kann unter NHEJ-vermitteltem *gene disruption* und HR-basierter Genkorrektur bzw. gezielter Genaddition unterschieden werden. In Abwesenheit einer homologen Reparaturmatrize für den geschädigten DNA-Lokus wird ein vorliegender DSB durch den Mechanismus des NHEJ repariert, welcher der vorherrschende Reparaturmechanismus für DSB in eukaryotischen Zellen ist. Die durch den DSB entstehenden DNA-Enden werden dabei von Proteinen des NHEJ-Weges erkannt und durch eine Protein-Protein-Interaktion ligiert [82]. Dabei kann die Verknüpfung nicht kompatibler DNA-Enden zu Deletionen bzw. Insertionen und dadurch zur Expression nicht funktioneller Proteine führen, was den NHEJ-Mechanismus zu einem extrem fehlerbehafteten Reparaturmechanismus macht. Forschungsorientierte Anwendungen für das ZFN-stimulierte NHEJ basieren auf dem gezielten *knock-out* gewünschter Gene zwecks *reverse genetics*-Analysen (Gen-Funktionsanalysen) in Pflanzen und Tieren. So wurden bereits gezielt Mutationen in chromosomalen Abschnitten von Säugetierzellen [83] und Organismen wie *Drosophila* [64, 65], *Caenorhabditis elegans* [84], Zebrafischen [85-87] sowie diversen Pflanzen wie *Arabidopsis thaliana* [88] gesetzt. Als therapeutische Applikation würde sich unter anderem der gezielte *knock-out* von dominant-negativ mutierten Allelen, wie sie bei der Huntington-Krankheit [89] vorliegen, oder von Genen zellulärer Rezeptoren, welche von viralen Pathogenen für die Infektion benötigt werden, eignen. So konnte bereits im murinen *In-vivo*-Modell durch ZFN-stimuliertes NHEJ in *ex vivo* infizierten, primären humanen CD4-positiven T-Zellen eine mehr als 50%-ige Zerstörung des Gens des CCR5-Rezeptors (*chemokine receptor 5*), der vom HI-Virus-1 (*human immunodeficiency virus-1*) als Korezeptor für eine erfolgreiche Infektion benötigt wird, gezeigt werden [90]. In Anwesenheit einer zum geschädigten DNA-Lokus homologen

Reparaturmatrize, wobei die HR-Maschinerie an Stelle des Schwesterchromatids auch ein exogen eingeführtes DNA-Fragment akzeptiert, wird ein vorliegender DSB durch den Mechanismus der HR repariert. Durch einen ortsspezifisch gesetzten, ZFN-induzierten DSB kann die HR um mehr als das 1000-fache gesteigert werden [68, 91], wodurch das Gentergeting therapeutisch anwendbar wird. Dabei bestimmt der Aufbau des auch als Donor-DNA bezeichneten exogenen DNA-Fragments, inwiefern sich das HR-vermittelte Gentergeting anwenden lässt. Die Donor-DNA kann dabei so aufgebaut sein, dass entweder eine Mutation eingeführt oder korrigiert wird. Bei der Genkorrektur wird durch die ZFN-stimulierte HR zwischen dem mutierten endogenen Locus und der dazu homologen, exogenen



**Abbildung 6 Reparaturwege von ZFN-induzierten DSB.** Die Reparatur eines ZFN-induzierten DSB kann durch einen von zwei möglichen Reparaturwegen stattfinden. In Abwesenheit einer Donor-DNA wird ein vorliegender DSB durch den Mechanismus des NHEJ repariert. Dadurch kann es zu Veränderungen in der betroffenen Gensequenz und somit zu einer gestörten Genexpression kommen. In Anwesenheit einer Donor-DNA wird ein ZFN-induzierter DSB durch den Mechanismus der HR repariert. Das vermehrte Einführen von DSB durch Bindung der ZFN an *off-target sites* kann zu ungewollten Toxizitätsreaktionen wie der Apoptose bzw. zu einer ZFN-induzierten Mutagenese führen. Jede ZF-DBD des dargestellten ZFN-Paares (ZFNa + ZFNb) besteht jeweils aus vier ZF-Modulen (kleine, bunte Ovale). Die *FokI*-Dimerisierungsbereiche der ZFN sind jeweils als große, gelbe Ovale dargestellt. Die ZF-Bindungsstellen sind als hellblauer, die Donor-DNA als dunkelblauer Balken dargestellt. Modifizierte Abbildung aus [81].

Donor-DNA, welche als intakte Reparaturmatrize fungiert, ein genetischer Defekt direkt *in situ* korrigiert. Initiale Studien zur Therapie der Krankheit X-SCID, hervorgerufen durch Mutationen im Gen des für die T-Zellentwicklung wichtigen Zytokinrezeptors *IL2R $\gamma$*  (*interleukin-2 receptor gamma chain*) [92], zeigten dabei erste deutliche Erfolge der Genkorrektur. Durch das gezielte ZFN-vermittelte Einfügen eines DSB in den *IL2R $\gamma$* -Lokus konnte in bis zu 5% primärer humaner T-Zellen und 18% transient transfizierter K562-Zellen

(humane Erythroleukämiezelllinie [93]) eine Genkorrektur erzielt werden [69], wobei die Anzahl genkorrigierter K562-Zellen durch die Verwendung Integrase-defizienter lentiviraler Vektoren (IDLV) zur Expression der ZFN auf 29% erhöht werden konnte [94].

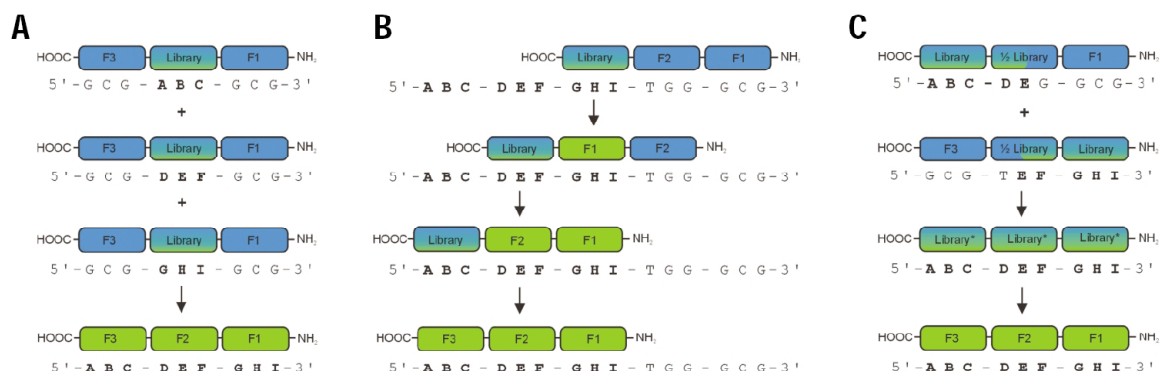
Beim Ansatz der gezielten Genaddition kann die Donor-DNA entweder aus einer Intron-freien cDNA (*complementary DNA*) oder einer vollständigen Expressionskassette mit exogenem Promotor bestehen. Beim cDNA-Ansatz werden die entsprechenden Wildtyp-Exons inklusive des transkriptionellen Terminationssignals vor (*upstream*) die mutierten Exons gezielt in den Ziellokus addiert, wobei so mit nur einer Donor-DNA und einem einzelnen ZFN-Paar mehrere Mutationen auf einmal stromabwärts (*downstream*) des gesetzten DSB „korrigiert“ bzw. phänotypisch behoben werden können. Dabei steht die Genexpression weiter unter der Kontrolle des endogenen Promotors. Die Durchführbarkeit dieses Ansatzes konnte bereits mit der Verwendung Donor-DNA-tragender IDLV und *IL2R $\gamma$* -Lokus-spezifischer ZFN nachgewiesen werden, wobei in 6% humaner lymphoblastoider Zellen eine gezielte Genaddition in den *IL2R $\gamma$* -Lokus erzielt wurde [94]. In der gleichen Publikation wurde auch die gezielte Genaddition einer vollständigen Expressionskassette in eine ungefährliche, als „sicherer Hafen“ (*safe harbor*) bezeichnete DNA-Sequenz untersucht. Ein solch geeigneter DNA-Bereich sollte nach der gezielten Genaddition eine hohe und lang anhaltende Expression des Transgens ohne genotoxische Nebeneffekte ermöglichen. Da Individuen mit einer homozygot vorliegenden CCR5-Gen-Mutation klinisch gesund sind, scheint dieser Locus für eine gezielte Genaddition geeignet zu sein [95, 96]. So konnte bereits in ersten Untersuchungen die gezielte Genaddition einer Expressionskassette für das grün fluoreszierende Reportergen EGFP (*enhanced green fluorescent protein*) nach einem ZFN-induzierten DSB in dem humanen CCR5-Lokus diverser humaner Zelllinien wie Jurkat-Zellen (T-Zell-Lymphomzellen) und embryonaler Stammzellen [97] gezeigt werden [94].

### 2.2.3 Selektionssysteme für Zinkfinger-Nukleasen

Die Generierung funktioneller ZFN, welche jede beliebige Sequenz in den Genomen der Zielorganismen wie Pflanzen und Säugetiere erkennen und schneiden können, ist von der Existenz hochaffiner und hochspezifischer ZF-DBD abhängig. Die Identifizierung und Selektion von ZF-Modulen zur spezifischen Erkennung jedes der 64 möglichen DNA-Triplets ist einer der wichtigsten Bestandteile bei der Konstruktion von künstlichen ZF-DBD [91]. Ein entscheidender Schritt bei der Erschaffung neuer ZF-DBD waren die Röntgenstrukturanalysen des murinen Transkriptionsfaktors Zif268, in denen die Struktur und das charakteristische DNA-Bindungsverhalten von Cys<sub>2</sub>His<sub>2</sub>-ZF entschlüsselt werden konnten

[49]. Durch die Beobachtung, dass bestimmte Schlüsselaminosäuren in der  $\alpha$ -Helix (Positionen -1 bis +6 relativ zum Start der  $\alpha$ -Helix) eines jeden ZF-Moduls für die spezifische Interaktion mit der DNA-Zielsequenz verantwortlich sind [98], wurde der Grundstein zur Konstruktion von künstlichen ZF-DBD gelegt. So basierten erste Versuche der Herstellung künstlicher ZF-DBD auf dem Austausch bzw. der Veränderung der  $\alpha$ -helikalen Schlüsselaminosäuren durch Methoden wie *helix-grafting* [99] bzw. *site-directed mutagenesis* [100]. Ausgehend von der Annahme einer  $\alpha$ -Helix-Konsensussequenz, versuchten einige Arbeitsgruppen, Regeln für ein rationales Design dieser Domäne aufzustellen [101, 102]. Ungefähr zu der gleichen Zeit begannen andere Arbeitsgruppen, erstmals die *phage display*-Methode zur Selektion von ZF-Modulen mit neuen DNA-Bindungseigenschaften durch Randomisierung der  $\alpha$ -Helix einzusetzen [103].

Beim klassischen *phage display* werden 3er-ZF-DBD, welche aus zwei ZF-Modulen mit bekannten DNA-Bindungseigenschaften und einem ZF-Modul mit einer randomisierten  $\alpha$ -Helix bestehen (Abb. 7 A), auf der Oberfläche des filamentösen M13-Bakteriophagen exprimiert. Phagen, die *in vitro* über die ZF-DBD an ihre doppelsträngige DNA-Zielsequenz binden, werden dabei aus einer Phagenbibliothek herausselektioniert. Nach mehreren Amplifikations- und Selektionsrunden wird die DNA der selektionierten Klone sequenziert, wodurch die für die DNA-Bindung verantwortlichen Aminosäuren in der  $\alpha$ -Helix identifiziert werden können [104]. Prinzipiell lassen sich beim *phage display* drei verschiedene Strategien



**Abbildung 7 Strategien für die Produktion von ZF-DBD mit gewünschten DNA-Bindungsspezifitäten.** **A)** Parallele Selektion, **B)** Sequentielle Selektion und **C)** *bipartite*-Selektion. Die parallele Selektion (A) basiert auf der funktionellen Unabhängigkeit der einzelnen ZF-Module. In der sequentiellen Selektionsstrategie (B), werden die in der  $\alpha$ -Helix randomisierten ZF-Module im Kontext zu ihren benachbarten ZF-Modulen Schritt für Schritt selektioniert, wodurch der Fakt der Kreuzverbindungen zwischen nebeneinander liegenden Triplets miteinbezogen wird. Diese Art der Selektion ist sehr zeitintensiv und erfordert die Konstruktion riesiger ZF-Protein-Bibliotheken sowie die Durchführung mehrerer Selektionsrunden. Die *bipartite*-Selektionsstrategie (C) verwendet zwei vorselektionierte ZF-Protein-Bibliotheken von anderthalb-ZF mit randomisierten  $\alpha$ -Helices, welche dann im Kontext einer 3er-ZF-DBD selektioniert werden. Abbildung aus [105].

(Abb. 7) namens parallele Selektion, sequentielle Selektion und *bipartite*-Selektion unterscheiden [105]. Die parallele Selektion (Abb. 7 A) basiert auf der Idee, dass ZF-Module funktionell unabhängig sind und dass individuelle, gegen einzelne Triplets selektionierte ZF-Module modular zusammengefügt (*modular assembly*) werden können, um längere Zielsequenzen mit hoher Spezifität und Affinität zu binden [106]. So werden die DNA-bindenden Aminosäurereste in dem mittleren ZF-Modul (F2) einer 3er-ZF-DBD randomisiert, wodurch eine Bibliothek an Varianten entsteht [107-110]. Die beiden äußeren ZF-Module (F1 und F3) liegen unverändert vor und besitzen dabei eine Ankerfunktion. So können ZF-Modul-Varianten selektiert werden, die das mittlere Triplet der neun Basenpaar langen DNA-Zielsequenz erkennen. Ein Nachteil bei der parallelen Selektion und dem *modular assembly* ist der Fakt, dass die Aminosäure an Position 2 in der  $\alpha$ -Helix eines ZF-Moduls mit einer Base in der Bindungsstelle eines benachbarten ZF-Moduls interagiert (Abb. 3), wodurch Kreuzverbindungen zwischen nebeneinander liegenden Triplets entstehen. Diese Tatsache wurde bei den weiterentwickelten *phage display*-Methoden für die sequentielle Selektion (Abb. 7 B) [111] und *bipartite*-Selektion (Abb. 7 C) [112, 113] berücksichtigt, wobei die ZF-Module mit neuen DNA-Bindungseigenschaften im Kontext der benachbarten ZF-Module selektiert werden (kontext-abhängige Selektion).

Neben dem *in vitro* basierten *phage display* haben sich außerdem das auf Hefen basierende *yeast one-hybrid* System [114, 115] sowie die auf Bakterien basierenden Systeme *bacterial one-hybrid* [79, 116] und vor allem das *bacterial two-hybrid* (B2H) [117] als *In-vivo*-Selektionssysteme für die Selektion neuer ZF-DBD aus randomisierten ZF-Modul-Bibliotheken als geeignet erwiesen. In dem auf *Escherichia coli* basierenden B2H-System wird durch eine erfolgreiche ZF-Bindung die Expression des HIS3-Gens der Hefe ermöglicht, wodurch sich die HIS-auxotrophen *E. coli*-Zellen vermehren können. Dabei wird die 3er-ZF-DBD, welche in einem der drei ZF-Module eine randomisierte Bibliothek trägt [78], mit dem Gal11P-Proteinfragment der Hefe verbunden, während die  $\alpha$ -Untereinheit der bakteriellen RNA-Polymerase mit dem Gal4-Proteinfragment der Hefe fusioniert vorliegt. Nach erfolgreicher Bindung der ZF-DBD an die gewünschte DNA-Bindungsstelle wird über die hochaffine Gal11P/Gal4-Protein-Protein-Interaktion die RNA-Polymerase-Untereinheit zum bakteriellen Promotor rekrutiert, sodass eine HIS3-Expression stattfinden kann. Durch die ZF-DBD-vermittelte Stimulation der HIS3-Transkription können so HIS-auxotrophe *E. coli*-Bakterien auf selektiven Minimalmediumplatten wachsen. Äquivalent zum *phage display*-Ansatz wird die DNA der überlebenden Klone nach der Isolation sequenziert, wodurch die für die DNA-Bindung verantwortlichen  $\alpha$ -helikalen Aminosäuren identifiziert werden können.

Bis dato wurden mit den genannten Selektionsmethoden ZF-Module identifiziert, die alle 16 GXX-Triplets sowie die meisten der CXX-, AXX- und einige TXX-Triplets erkennen können [118].

### **2.3 Vektoren zum Einbringen von Zinkfinger-Nukleasen**

Der gezielte simultane Transfer von zwei ZFN-Expressionskassetten und einer Donor-DNA in die jeweiligen Zielzellen stellt eine der größten Herausforderungen in der ZFN-basierten Gentherapie dar. Bis dato wurden drei verschiedene Vektorsysteme für die transiente Expression von ZFN in menschlichen Zellen zwecks DSB-stimulierter Genommodifikation publiziert. So sind neben den bisher hauptsächlich verwendeten Plasmid-basierten Vektoren [37, 63, 66-68, 72, 74] auch IDLV [94] und modifizierte adenovirale Vektoren [90] zum Einsatz gekommen. Aufgrund ihrer hervorragenden Transduktionseigenschaften stellen IDLV und Vektoren basierend auf den Adenoviren des Typs 5, welche zusätzlich mit Typ 35-Fiberproteinen versehen sind (Ad5/35), viel versprechende Werkzeuge für den Transfer von ZFN-Expressionskassetten in menschliche hämatopoetische Zellen dar. So wurden schon erfolgreich IDLV und Ad5/35-Vektoren zur Expression von ZFN in hämatopoetischen Stammzellen [94] und primären CD4-positiven T-Zellen [90] verwendet. Der Transfer von DNA-Expressionskassetten mit starken Promotoren birgt dabei unabhängig von der Wahl des verwendeten Vektors ein potentielles Risiko für eine Insertionsmutagenese, was sowohl für integrierende als auch für vornehmlich episomal persistierende Vektoren gilt [119-121]. Alternativ dazu könnten ZFN direkt als Proteine transduziert oder über Transfektion einer ZFN-kodierenden mRNA in die Zielzellen eingeschleust werden, um so das Risiko einer Vektorintegration zu reduzieren. So konnte bereits der erfolgreiche Transfer eines künstlichen ZF-Transkriptionsfaktor-Proteins in menschliche Zellen gezeigt werden [122]. Mithilfe der ZFN-mRNA-Transfektion wurden schon erfolgreich DSB in chromosomalen Abschnitten diverser Zellen von Taufliegen [123], Zebrafischen [86] und Ratten [124] gezielt gesetzt. Da für das effiziente Setzen eines DSB allerdings eine hohe intranukleäre Konzentration an ZFN-Proteinen benötigt wird, muss die Frage, ob der Transfer per Proteintransduktion bzw. mRNA-Transfektion für gezielte ZFN-stimulierte Genommodifikationen generell effizient genug ist, noch geklärt werden.

### 3 Einzelarbeiten

#### 3.1 Titel der publizierten Einzelarbeiten

- Arbeit 1:

##### **Generation and Functional Analysis of Zinc Finger Nucleases**

Cathomen T, Segal DJ, Brondani V, **Müller-Lerch F.**

Methods Mol Biol. 2008; 434:277-90.

[http://www.ncbi.nlm.nih.gov/pubmed/18470651?ordinalpos=2&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed\\_ResultsPanel.Pubmed\\_DefaultReportPanel.Pubmed\\_RVDocSum](http://www.ncbi.nlm.nih.gov/pubmed/18470651?ordinalpos=2&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_DefaultReportPanel.Pubmed_RVDocSum)

- Arbeit 2:

##### **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.

Nat Methods. 2008; 5:374-5.

[http://www.ncbi.nlm.nih.gov/pubmed/18446154?ordinalpos=3&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed\\_ResultsPanel.Pubmed\\_DefaultReportPanel.Pubmed\\_RVDocSum](http://www.ncbi.nlm.nih.gov/pubmed/18446154?ordinalpos=3&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_DefaultReportPanel.Pubmed_RVDocSum) [+ Supplementary information]

- Arbeit 3:

##### **Rapid “Open-Source” Engineering of Customized Zinc-Finger Nucleases for Highly Efficient Gene Modification**

Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, Eichinger M, Jiang T, Foley JE, Winfrey RJ, Townsend JA, Unger-Wallace E, Sander JD, **Müller-Lerch F**, Fu F, Pearlberg J, Göbel C, Dassie JP, Pruett-Miller SM, Porteus MH, Sgroi DC, Iafrate AJ, Dobbs D, McCray PB Jr, Cathomen T, Voytas DF, Joung JK.

Mol Cell. 2008; 31:294-301.

[http://www.ncbi.nlm.nih.gov/pubmed/18657511?ordinalpos=1&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed\\_ResultsPanel.Pubmed\\_DefaultReportPanel.Pubmed\\_RVDocSum](http://www.ncbi.nlm.nih.gov/pubmed/18657511?ordinalpos=1&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_DefaultReportPanel.Pubmed_RVDocSum) [+ Supplemental Data]



## 3.2 Eigenanteil an den publizierten Einzelarbeiten

- Arbeit 1:

### **Generation and Functional Analysis of Zinc Finger Nucleases**

An der Konzeption (I), Durchführung (II) und Berichtsabfassung (III) bin ich wie folgt beteiligt gewesen:

**I:** Entwicklung des Systems zur PCR-basierten Konstruktion von ZF-DBD bzw. ZFN auf digitaler und molekularbiologischer Ebene.

Erstellung des Versuchsplans der eukaryotisch basierten ZFN-Experimente.

**II:** Konstruktion der ZFN GZF8-N und GZF9-N sowie der dazugehörigen ZFN-Zielplasmide.

Evaluierung der PCR-generierten ZFN-Expressionskassetten bzw. der ZFN-Expressionsplasmide (GZF8-N und GZF9-N) per Restriktionsverdau und Sequenzierung.

Evaluierung der ZFN GZF8-N und GZF9-N im episomalen HR-System in HEK-293T-Zellen (HEK: *human embryonic kidney*) per Durchflusszytometrie.

Analyse der ZFN-Zielplasmide per Sequenzierung.

Evaluierung der modular assemblierten ZFN GZF1-N im episomalen HR-System in HEK-293T-Zellen per Durchflusszytometrie.

**III:** Erstellung aller ZFN-Manuskript-Absätze.

- Arbeit 2:

### **Unexpected failure rates for modular assembly of engineered zinc fingers**

An der Konzeption (I), Durchführung (II) und Berichtsabfassung (III) bin ich wie folgt beteiligt gewesen:

**I:** Erstellung des Versuchsplans der ZFN-Experimente.

**II:** Konstruktion der ZFN-Zielplasmide sowie der ZFN-Expressionsplasmide, welche für die Durchführung des episomalen HR-Systems zur eukaryotisch basierten Evaluierung der ZFN in HEK-293T-Zellen verwendet wurden.

Überprüfung dieser Expressions- und Zielplasmide hinsichtlich ihrer Richtigkeit mittels Restriktionsverdau und Sequenzierung.

ZFN-Expressionskontrollen mittels Western-Blot-Untersuchungen.

Untersuchung der jeweiligen ZFN-Funktionalität im episomalen HR-System per Durchflusszytometrie.

Beteiligung an der statistischen Auswertung und grafischen Darstellung der erlangten ZFN-Daten.

**III:** Beteiligung an der Überarbeitung des Manuskripts.

- Arbeit 3:

**Rapid “Open-Source” Engineering of Customized Zinc-Finger Nucleases for Highly Efficient Gene Modification**

An der Konzeption (I), Durchführung (II) und Berichtsabfassung (III) bin ich wie folgt beteiligt gewesen:

**I:** Erstellung des Versuchsplans der EGFP-basierten, episomalen ZFN-Experimente.

**II:** EGFP-ZFN-Expressionskontrollen mittels Western-Blot-Untersuchungen.

Untersuchung der jeweiligen ZFN-Funktionalität im episomalen EGFP-NHEJ-Reparatursystem per Durchflusszytometrie.

Beteiligung an der statistischen Auswertung und grafischen Darstellung der erlangten ZFN-Daten.

**III:** Beteiligung an der Überarbeitung des Manuskripts.

### 3.3 Publierte Ergebnisse

#### 3.3.1 *Generation and Functional Analysis of Zinc Finger Nucleases*

Nach der Festlegung einer potentiellen ZFN-Zielsequenz innerhalb eines gewünschten DNA-Abschnittes und dem Erhalt der mittels *phage display*-Parallelselektion identifizierten  $\alpha$ -Helix-Aminosäuresequenzen der einzelnen ZF-Module (<http://www.zincfingertools.org>) benötigt man Techniken für die Generierung und funktionelle Analyse der ZFN. Bezüglich des ersten Punktes wurde im Rahmen dieser Publikation ein PCR-basiertes (*polymerase chain reaction*) System zur Generierung von ZF-DBD entwickelt (Arbeit 1; Fig. 1 A). Dabei wurden in einem ersten Schritt sechs Oligonukleotide (Oligos) über überlappende Bereiche per PCR miteinander verbunden und die entstandenen Lücken zwischen ihnen durch Verlängerung der 3'-Enden aufgefüllt (*assembly PCR*). Ein ebenso funktioneller alternativer Ansatz zur *assembly PCR* war das Verbinden der Oligos durch *In-vitro-annealing*, wobei die Lücken in diesem Fall durch Klenow-*fill-in* aufgefüllt wurden. Die Oligos C1, C2 und C3 bestanden jeweils aus dem konstanten  $\beta$ -Faltblatt-Rückgratbereich eines ZF-Moduls, wohingegen die Oligos V1, V2 und V3 jeweils den variablen  $\alpha$ -Helix-Bereich eines ZF-Moduls aufwiesen. Ein wichtiger Punkt bei diesem System war die Gestaltung der überlappenden Bereiche zwischen den konstanten und variablen (als „n“ dargestellt) Sequenzen (Arbeit 1; Fig. 1 C). Bezüglich dieser konnten aufgrund der Degeneration des genetischen Codes verschiedene Triplets für die identische Abfolge von Aminosäuren gewählt werden, wobei die *annealing*-Temperatur gleich gehalten wurde. So sollte verhindert werden, dass sich z.B. der variable Bereich (V3) des ZF-Moduls ZF3 an den konstanten Bereich (C2) des ZF-Moduls ZF2 anlagert, wodurch eine verkürzte ZF-DBD aus den Sequenzen C1, V1, C2 und V3 entstanden wäre. In einem zweiten Schritt wurde die entstandene, für die drei ZF-Module (ZF1, ZF2 und ZF3) kodierende Sequenz durch die Verwendung von Außenprimern (F und R) mittels PCR vervielfältigt (*amplification PCR*). Anschließend wurde das PCR-Produkt verdaut und in einen prokaryotischen bzw. einen eukaryotischen Expressionsvektor kloniert (Arbeit 1; Fig. 1 B), um die PCR-generierte ZF-DBD am 3'-Ende jeweils mit einer *FokI*-Domäne zu verbinden. Zur funktionellen Analyse auf prokaryotischer Ebene wurden modular assemblierte ZFN (GZF1-N und GZF3-N) in einem *In-vitro*-System evaluiert. Wie im Agarosegel zu erkennen war, führte die Expression funktioneller ZFN zur Spaltung linearer DNA-Substrate, die entweder ZF-Bindungsstellen für GZF1-N (1-1), GZF3-N (3-3) oder GZF3-N und GZF1-N (3-1) enthielten (Arbeit 1; Fig. 2 A+B), in jeweils zwei niedermolekulare DNA-Fragmente. Dabei erkannten die ZFN als Homodimere jeweils ihre entsprechenden Bindungsstellen (1-1 bzw. 3-3), wohingegen sie als

Heterodimere alle Bindungsstellen spalteten (1-1, 3-3 sowie 3-1). Die Verwendung von *EcoRI*, dessen Bindungsstelle in allen verwendeten DNA-Substraten vorhanden war, diente als Positivkontrolle. Im eukaryotischen Kontext wurde die Funktionalität modular assemblierter ZFN exemplarisch durch die GZF1-N-Expression zwecks DSB-induzierter Stimulation der HR zwischen einem Reparaturplasmid (RP), welches ein 5'-deletiertes LacZ-EGFP-Fusionsgen (LacZ: Gen für das Enzym  $\beta$ -Galaktosidase) ohne Promotor trug, und einem Zielplasmid (*target plasmid*, TP) in menschlichen HEK-293T-Zellen untersucht (Arbeit 1; Fig. 2 C). Das TP besaß ein CMV-Promotor gesteuertes (CMV: *cytomegalovirus*) LacZ-Gen, das durch ein Stoppcodon abgeschlossen wurde, sowie ein nicht funktionelles, 5'-deletiertes EGFP-Gen. Die Bindungsstellen für das GZF1-N-Paar sowie für *I-SceI*, welches als Positivkontrolle fungierte, wurden zwischen die beiden offenen Leserahmen platziert. Die Frequenz, der durch episodale Genkorrektur wiederhergestellten EGFP-Expression, wurde anschließend mittels Durchflusszytometrie ermittelt. Wie per Durchflusszytometriemessungen gezeigt wurde, konnte die unstimulierte Genkorrekturfrequenz durch die Expression von GZF1-N von 0,7% auf 15,3% EGFP-positive Zellen erhöht werden (Arbeit 1; Fig. 2 D).

### 3.3.2 *Unexpected failure rates for modular assembly of engineered zinc fingers*

Allerdings wiesen zwei modular assemblierte ZFN (GZF8-N bzw. GZF9-N), welche mittels des beschriebenen PCR-Systems generiert wurden, in dem episodalen HR-System keinerlei stimulierende Eigenschaften auf. Aufgrund dieser Beobachtung und den allgemein widersprüchlichen Ergebnissen hinsichtlich der Erfolgsrate funktionelle ZF-DBD mittels der *modular assembly*-Methode generieren zu können, wurden in einem groß angelegten Experiment 168 modular assemblierte 3er-ZF-DBD, welche für 104 verschiedene, aus drei Triplets bestehende DNA-Zielsequenzen generiert wurden (Arbeit 2; Supplementary Table 2 und Supplementary Methods), hinsichtlich ihrer Funktionalität evaluiert. Die 104 DNA-Zielsequenzen variierten dabei in ihrer GXX-, AXX-, CXX- und TXX-Triplettkomposition, wobei X stellvertretend für jede mögliche Base angegeben war. In einem ersten Schritt wurde die DNA-Bindungsfähigkeit der unter gleichen Bedingungen generierten 168 ZF-DBD im prokaryotischen Kontext mittels des B2H-Systems analysiert. Dabei zeigte sich, dass für 79 der insgesamt 104 getesteten DNA-Zielsequenzen keine einzige ZF-DBD gefunden werden konnte, die positiv im B2H-System bewertet wurde, wobei „positiv“ als >1.57-fache transkriptionelle Aktivierung definiert war. Dies bedeutete insgesamt eine Fehlerrate von ~76% bzw. eine Erfolgsrate von ~24% (Arbeit 2; Figure 1 a). Darüber hinaus war der *modular assembly*-Ansatz wesentlich weniger effektiv für DNA-Bindungsstellen, die nur ein

oder kein GXX-Triplett aufwiesen. Entsprechend der genannten Reihenfolge lagen für jene DNA-Bindungsstellen die im B2H-System gemessenen Fehlerraten bei 88% bzw. 100%. Im Vergleich dazu lagen die Fehlerraten für DNA-Bindungsstellen, die zwei bzw. drei GXX-Triplets aufwiesen, mit 71% bzw. 41% etwas tiefer (Arbeit 2; Figure 1 a). Ausgehend von der Annahme, dass es bezüglich ein und desselben 3er-ZF-DBD-Proteins eine Korrelation zwischen der Fähigkeit zur Transkriptionsaktivierung als ZF-DBD im B2H-System und der Funktionalität als ZFN in menschlichen Zellen gibt, wurde die Aktivität von 23 ZFN in dem in Absatz 3.3.1 (Arbeit 1; Fig. 2 C) beschriebenen, episomalen HR-System untersucht (Arbeit 2; Supplementary Methods). Die DNA-Bindungsstellen für die als Homodimere getesteten ZFN sowie jene für die Positivkontrolle I-SceI lagen dabei jeweils zwischen den offenen Leserahmen des CMV-Promotor gesteuerten LacZ-Gens und des 5'-deletierten EGFP-Gens der einzelnen TP vor. Die modular assemblierten ZF-DBD dieser 23 ZFN aktivierten im B2H-System verschieden stark die Transkription eines LacZ-Reportergens, wobei sie im Vergleich zur Negativkontrolle zu einer 0,85- bis 5,07-fachen Aktivierung führten. Wie der Abbildung *Supplementary Figure 4* (Arbeit 2) zu entnehmen ist, konnten neun von neun der getesteten ZF-DBD, welche im B2H-System eine 1,57-fache bzw. eine geringere Aktivierung bewirkten, im Vergleich zu einer ZFN-Negativkontrolle keine signifikante Aktivität als ZFN in menschlichen Zellen erzielen (rosa Balken; Arbeit 2; Supplementary Figure 4). Somit waren 100% der im B2H-System als negativ bewerteten ZF-DBD auch als ZFN im eukaryotischen Kontext inaktiv. Umgekehrt zeigten neun der übrigen vierzehn getesteten ZF-DBD, die im B2H-System eine mehr als 1,57-fache Aktivierung der Reportergentranskription bewirkten, als ZFN im episomalen HR-System eine signifikant höhere Aktivität (mit Sternchen markierte türkise Balken; Arbeit 2; Supplementary Figure 4). Folglich waren ~64% der im B2H-System als positiv bewerteten ZF-DBD auch als ZFN im eukaryotischen Kontext aktiv. Die ZFN-Aktivität wurde dabei als relative Prozentzahl angegeben, wobei die durch die ZFN-Stimulation hervorgerufene Genkorrekturrate (Anzahl EGFP-positiver Zellen) zu der durch die I-SceI-Positivkontrolle erzielten Genkorrekturrate relativiert wurde. Als Negativkontrolle fungierten dabei modular assemblierte ZFN mit bekannter Inaktivität, wohingegen funktionelle ZFN als Positivkontrolle eingesetzt wurden, deren ZF-Module innerhalb der ZF-DBD im Kontext zu ihren benachbarten ZF-Modulen (kontext-abhängige Selektion) selektioniert wurden (rote bzw. blaue Balken; Arbeit 2; Supplementary Figure 4). Aufgrund der erzielten Ergebnisse scheint das *modular assembly* keine geeignete Methode für eine zuverlässige Generierung von funktionellen ZFN zu sein.

### 3.3.3 Rapid “Open-Source” Engineering of Customized Zinc-Finger Nucleases for Highly Efficient Gene Modification

Obwohl gezeigt werden konnte, dass mittels kontext-abhängigen Selektionsstrategien hochaffine und hochspezifische ZF-DBD mit geringer Toxizität generiert werden können, benötigen diese Strategien eine spezielle Expertise bei der Generierung und Auswertung großer, randomisierter Bibliotheken (typischerweise  $>10^8$ ), was sie extrem arbeitsintensiv und nicht für jedermann zugänglich macht. Aus diesem Grund wurde ein Selektionssystem namens OPEN (*Oligomerized Pool ENgineering*) konstruiert. Dafür wurde ein Archiv von B2H-vorselektionierten ZF-Modul-Bibliotheken generiert, welche jeweils für ein bestimmtes Zieltriplett einen Pool (95 verschiedene) an ZF-Modulen an einer bestimmten Position innerhalb einer 3er-ZF-DBD aufwiesen (Arbeit 3; Figure 1 A). Zum Zeitpunkt der Publikation lagen ZF-Modul-Bibliotheken für alle 48 GXX- und 18 TXX-Triplets vor (Arbeit 3; Figure 1 B). Zur Anwendung der OPEN-Selektion für eine neun Basenpaar lange DNA-Zielsequenz, wurden die Bibliotheken der einzelnen ZF-Module zur Generierung einer 3er-ZF-DBD-Bibliothek ( $95^3 = 8,6 \times 10^5$  Varianten für eine 3er-ZF-DBD) miteinander kombiniert, und mittels des B2H-Systems, in dem durch Bindung einer 3er-ZF-DBD-Variante an die gewünschte Zielsequenz die Expression eines Selektionsmarkers induziert wird, evaluiert (Arbeit 3; Figure 1 C+D). Zur Identifikation von möglichen ZF-DNA-Bindungsstellen, welche durch OPEN-generierte ZFN angesteuert werden können, wurde die Internetsoftware ZiFiT 3.0 (<http://www.zincfingers.org/software-tools.htm>) verwendet. Für den direkten Vergleich von modular assemblierten und OPEN-generierten ZFN, wurden mit beiden Systemen 3er-ZF-DBD für zehn ZF-Bindungsstellen in fünf DNA-Sequenzen im EGFP-Gen konstruiert (Arbeit 3; Figure S2 A; Supplemental Data), und anschließend per B2H-System hinsichtlich ihrer DNA-Bindungseigenschaft bewertet. Dabei wurde eine mehr als dreifache Aktivierung der Transkription des B2H-Reportergens LacZ als Richtwert für hochaffine und hochspezifische ZF-DBD festgelegt. Keine der modular assemblierten ZF-DBD führte zum Erreichen dieses Richtwertes, wohingegen OPEN-selektionierte ZF-DBD an neun der zehn ZF-Bindungsstellen eine über dreifache Transkriptionsaktivierung bewirkten (Arbeit 3; Figure 2 A). Als nächstes wurden modular assemblierte und OPEN-generierte ZFN-Paare in menschlichen Zellen mittels eines NHEJ-Reparatursystems evaluiert, in dem die fehlerhafte NHEJ-Reparatur von ZFN-induzierten DSB im chromosomal stabilisierten EGFP-Lokus zu einer Abnahme der EGFP-Expression führen sollte (Arbeit 3; Figure 2 B). Modular assemblierte ZFN zeigten nur an einer der fünf EGFP-DNA-Sequenzen (EG502) eine signifikante Aktivität. Im Gegensatz dazu waren OPEN-generierte ZFN an vier der fünf

EGFP-Zielsequenzen (EG223, EG292, EG382, und EG502) aktiv, was sich in einer signifikanten Abnahme der EGFP-Expression widerspiegelte (Arbeit 3; Figure 2 C). In einem nächsten Schritt wurden mittels ZiFiT 3.0 und OPEN ZFN-Paare für ZF-DNA-Bindungsstellen in drei humanen (VEGF-A: *vascular endothelial growth factor A*; HoxB13: *homeobox B13*; CFTR: *cystic fibrosis transmembrane conductance regulator*) und einem pflanzlichen endogenen Gen (SuRA: Acetolactatsynthase-Gen) konstruiert (Arbeit 3; Figure S2; Supplemental Data). Mithilfe eines *CEL-I*-Nuklease-basierten NHEJ-Mutation-Detektionssystems (Arbeit 3; Figure 3 A) zum Nachweis von ZFN-induzierten, NHEJ-reparierten DSB, konnte gezeigt werden, dass alle OPEN-generierten ZFN-Paare an zwei der fünf ZF-DNA-Bindungsstellen (VF2468 und VF2471) im VEGF-A-Gen Mutationen verursachten (Arbeit 3; Figure 3 B). Das gleiche galt für alle ZFN-Paare, die für drei der vier ZF-DNA-Bindungsstellen (HX587, HX735 und HX761) im HoxB13-Gen (Arbeit 3; Figure 3 C) mittels OPEN konstruiert wurden. An den restlichen drei ZF-DNA-Bindungsstellen im VEGF-A-Gen (VF3537, VF3542 und VF3552) bzw. der einen im HoxB13-Gen (HX508) induzierten OPEN-generierte ZFN-Paare keine detektierbaren Mutationen (Arbeit 3; Figure 3 B+C). Mittels Sequenzierungsuntersuchungen konnte für den CFTR-Lokus gezeigt werden, dass CFTR-spezifische OPEN-ZFN Insertionen an der CF877-ZF-Bindungsstelle mit einer durchschnittlichen Frequenz von 1,2% induzierten (Arbeit 3; Figure S3 B; Supplemental Data). Bezüglich des pflanzlichen SuRA-Gens ergaben Sequenzierung, dass SuRA-spezifische OPEN-ZFN in drei von 66 Tabakpflanzen Mutationen im SuRA-Gen induzierten, wobei in einer Pflanze in beiden Allelen des SuRA-Gens die gleiche Deletion detektiert wurde (Arbeit 3; Figure S3 C; Supplemental Data).

Um zu testen, ob sich OPEN-selektionierte ZFN für den Einsatz im Gentargetingkontext eignen, wurden funktionelle ZFN-Paare (VF2468/Paar C und VF2471/Paar B; Arbeit 3; Figure 3 B) zur Stimulation der HR zwischen dem VEGF-A-Lokus und der mit einer *SalI*-Schnittstelle versehenen Donor-DNA eingesetzt (Arbeit 3; Figure S4; Supplemental Data). Zum direkten Vergleich dienten *IL2R $\gamma$* -spezifische ZFN mit bekannter Funktionalität [72]. Wie PCR-Restriktionsanalysen in K562-Zellen ergaben, wurden Gentargetingfrequenzen von 7,7%, 4,5% bzw. 4,1% mit den VF2468-, VF2471- bzw. *IL2R $\gamma$* -spezifischen ZFN erreicht (Arbeit 3; Figure 4 A). Southern-Blot-Analysen von Zellen, die mittels VEGF-A-spezifischer ZFN modifiziert wurden, deuteten darauf hin, dass die eigentliche Gentargetingrate sogar höher lag als durch die PCR-Analyse vermutet (Arbeit 3; Figure 4 B). Die Zugabe des Chemotherapeutikums Vinblastin, das zu einer Arretierung des Zellzyklus in der G2-Phase führt, erhöhte dabei die Gentargetingfrequenz von 7,7%, 4,5% bzw. 4,1% auf 54%, 37%,

bzw. 44% (Arbeit 3; Figure 4 C). Eine Analyse zur ZFN-vermittelten Toxizität, die zum Absterben der betroffenen Zellen führen kann, ergab nach Expression der drei einzeln getesteten ZFN eine signifikante Abnahme der durch Kotransfektion vermittelten GFP-Expression (grüne Balken; Arbeit 3; Figure 4 D) sowie der Gentergetingfrequenz (lila Balken; Arbeit 3; Figure 4 D). Daher wurde untersucht, ob die ZFN-vermittelte Toxizität durch die Verwendung von *FokI*-Varianten mit abgeschwächten Dimerisierungsbereichen [72] vermindert werden konnte. Die Varianten der VF2471- bzw. der *IL2R $\gamma$* -ZFN-Paare verursachten keine signifikante Toxizität, wobei die VF2468-ZFN-Paar-Variante eine minimale Toxizität aufwies (Arbeit 3; Figure 4 D). Alle drei ZFN-Paar-Varianten vermochten dabei Gentergeting zu stimulieren, wodurch gezeigt werden konnte, dass die Reduzierung der Toxizität nicht auf eine fehlende ZFN-Aktivität zurückzuführen war (Arbeit 3; Figure 4 E). Das kontext-abhängige OPEN-System stellt ein schnell durchführbares, hocheffektives und für jedermann zugängliches Selektionssystem dar. Innerhalb kürzester Zeit konnten hochspezifische OPEN-ZFN-Paare generiert werden, mit deren Hilfe drei endogene Säugetiergene und ein pflanzliches endogenes Gen erfolgreich modifiziert wurden.



## 4 Diskussion

ZFN stellen ein sehr effizientes Werkzeug für das gezielte Einfügen eines DSB in jede beliebige DNA-Sequenz einer Zelle dar. Abhängig von der Anwesenheit einer Donor-DNA und dem Zellzyklusstadium, in dem sich die behandelte Zelle befindet, können ZFN-induzierte DSB entweder durch den Mechanismus der HR oder des NHEJ repariert werden. So kann die ZFN-Technologie einerseits zur gezielten, NHEJ-vermittelten Mutagenese von Genen genutzt werden, welche auf der Forschungsebene für den gezielten *knock-out* gewünschter Gene zwecks *reverse genetics*-Analysen und auf der therapeutischen Ebene für den gezielten *knock-out* von dominant-negativ mutierten Allelen eingesetzt werden kann. Auf der anderen Seite kann die ZFN-Technologie in Anwesenheit einer zum geschädigten DNA-Lokus homologen Donor-DNA zur DSB-induzierten Stimulation der HR-basierten Genkorrektur bzw. der gezielten Genaddition genutzt werden.

Um diese vielfältig verwendbaren molekularen Werkzeuge zur Verfügung zu haben, benötigt man im Vorfeld effiziente Methoden für deren Konstruktion. Aus diesem Grund stand am Anfang der Doktorarbeit die Entwicklung eines PCR-basierten Systems zum Zusammenfügen einzelner ZF-Module mit bekannten DNA-Bindungseigenschaften für die Konstruktion künstlicher ZF-DBD und somit zur Generierung von ZFN (siehe Absatz 3.3.1). Nach der Festlegung einer gewünschten ZFN-Zielsequenz innerhalb eines bestimmten DNA-Lokus wurden mithilfe der von Carlos Barbas etablierten Internetsoftware „Zinc Finger Tools“ (<http://www.zincfingertools.org>) die  $\alpha$ -Helix-Aminosäuresequenzen (Positionen -1 bis +6 relativ zum Start der Helix) der einzelnen ZF-Module ermittelt [106]. Dabei ist anzumerken, dass das zur Auswahl stehende, mittels *phage display*-Parallelselektion identifizierte ZF-Modul-Set primär aus ZF-Modulen für GXX- (16 von 16), AXX- (15 von 16) und CXX-Tripletts (15 von 16) bestand, wohingegen nur wenige ZF-Module (3 von 16) für TXX-Tripletts zur Verfügung standen (wobei X stellvertretend für jede mögliche Base angegeben ist). Aufgrund der Unvollständigkeit des ZF-Modul-Sets und der Tatsache, dass ZFN nur als Dimere funktionell aktiv sind, wodurch es einen klar definierten Aufbau der beiden invertiert vorliegenden ZFN-Bindungsstellen gibt, wird die Anzahl potentieller ZFN-Bindungsstellen stark reduziert. Neben den erwähnten EGFP-spezifischen ZFN GZF8-N und GZF9-N wurden mittels des etablierten PCR-Systems ZFN für den *cgd*-Lokus (*chronic granulomatous disease*) [19] sowie für den murinen *prkdc*-Lokus (*protein kinase, DNA-activated, catalytic polypeptide*) [125] hergestellt. Obwohl sich das zum Zusammenfügen der einzelnen ZF-Module etablierte PCR-System als sehr zuverlässig hinsichtlich seiner Fehleranfälligkeit erwies, konnte für keine der modular assemblierten ZFN in dem in Absatz 3.3.1 (Arbeit 1;

Fig. 2 C) beschriebenen episomalen HR-System eine stimulierende Eigenschaft nachgewiesen werden. Das für die Generierung der ZF-DBD-Kassetten verwendete PCR-System lässt sich dabei in zwei Teile gliedern. Zuerst wurden in der *assembly* PCR sechs Oligonukleotide über überlappende Bereiche per PCR miteinander verbunden und die entstandenen Lücken zwischen ihnen durch Verlängerung der 3'-Enden geschlossen. Ein ebenso funktioneller alternativer Ansatz zur *assembly* PCR stellte das Verbinden der Oligos durch *In-vitro-annealing* dar, wobei die Lücken in diesem Fall durch Klenow-*fill-in* aufgefüllt wurden. Anschließend wurde in der *amplification* PCR die entstandene, für die drei ZF-Module (ZF1, ZF2 und ZF3) kodierende Sequenz durch die Verwendung von Außenprimern (F und R) mittels PCR vervielfältigt. Jede neu generierte ZF-DBD-Kassette, welche die zusammengefügt ZF-Module enthielt, wurde nach folgendem Schema auf ihre Richtigkeit hin überprüft. Nach sorgfältigen Restriktionsanalysen im Agarosegel, wobei mithilfe des genetischen Codes der variable  $\alpha$ -Helix-Bereich eines ZF-Moduls so gestaltet wurde, dass eine zusätzliche Restriktionsschnittstelle eingefügt werden konnte, wurde die Expression der ZFN mittels Western-Blot-Analysen kontrolliert. An dieser Stelle sei angemerkt, dass durch die Gestaltung der überlappenden Bereiche zwischen den konstanten und variablen Sequenzen (Arbeit 1; Fig. 1 C) in der *assembly* PCR fast ausschließlich vollständige ZF-DBD-Kassetten generiert wurden, die aus drei ZF-Modulen bestanden (3er-ZF-DBD), wobei verkürzte 1er- bzw. 2er-ZF-DBD-Versionen der ZF-DBD-Kassette optisch im Agarosegel nicht detektiert werden konnten. Bezüglich der Gestaltung der überlappenden Bereiche konnten aufgrund der Degeneration des genetischen Codes verschiedene Triplets für die identische Abfolge von Aminosäuren gewählt werden, wobei die *annealing*-Temperatur gleich gehalten wurde. Dadurch konnten Fehlbindungen von variablen Bereichen an nicht dafür vorgesehene konstante Bereiche und die Entstehung verkürzter ZF-DBD-Kassetten verhindert werden. Im Anschluss an die Restriktionsanalyse wurden die positiven Klone sequenziert, um zu bestätigen, dass die ZF-DBD-Kassetten fehlerfrei generiert wurden. Hinsichtlich der Zuverlässigkeit des etablierten PCR-Systems konnte gezeigt werden, dass 90% der anfangs gepickten Klone in den Restriktionsanalysen als korrekt bewertet werden konnten. 75% dieser Klone konnten anschließend in den Expressionskontrollen als positiv eingestuft werden, wobei zwei Drittel dieser positiven Klone laut Sequenzierungsergebnissen die korrekte Sequenz enthielten. Somit kann bezüglich der Fehlerrate des etablierten PCR-Systems allgemein davon ausgegangen werden, dass 45% der anfangs gepickten Klone eine vollständige ZF-DBD-Kassette mit der korrekten Sequenz aufweisen. Die Tatsache, dass die ZFN GZF1-N und GZF3-N als Homodimere jeweils ihre entsprechenden Bindungsstellen (1-

1 bzw. 3-3) und als Heterodimere alle Bindungsstellen spalteten (1-1, 3-3 sowie 3-1), mag mit der Wahl der *FokI*-Dimerisierungsbereiche zusammenhängen (Arbeit 1; Fig. 2 B). Da es sich bei den Nukleasedomänen in den ZFN GZF1-N und GZF3-N um Wildtyp-*FokI*-Domänen handelt, kann die hohe Affinität dieser *FokI*-Domänen zu der Induktion eines DSB durch eine Protein-Protein-Interaktion geführt haben, obwohl nur eine der beiden ZFN an ihre Bindungsstelle gebunden hat [76]. Das ungewollte Setzen eines DSB durch eine Protein-Protein-Interaktion kann durch die Verwendung von asymmetrischen, niedrigaffinen *FokI*-Domänen verhindert werden [37], wodurch ZFN-induzierte Toxizitätsreaktionen reduziert werden können. Trotz der Tatsache, dass die modular assemblierten ZFN GZF1-N und GZF3-N auf prokaryotischer Ebene in dem beschriebenen *In-vitro*-System und darüber hinaus GZF1-N in dem episomalen HR-System funktionell waren, scheint die Erfolgsrate, funktionelle ZF-DBD mittels der *modular assembly*-Methode generieren zu können, generell niedrig zu sein.

Zur Stützung dieser Hypothese und zur Bewertung der Effizienz der *modular assembly*-Methode wurden in einem groß angelegten Experiment 168 modular assemblierte 3er-ZF-DBD, die für 104 verschiedene, neun Basenpaar lange DNA-Zielsequenzen generiert wurden, hinsichtlich ihrer Funktionalität mittels des B2H-Systems evaluiert (siehe Absatz 3.3.2). In dem B2H-System wird die Transkription eines LacZ-Reportergens durch die spezifische Bindung einer ZF-DBD an ihre ZF-Bindungsstelle aktiviert, wobei eine positive Aktivierung als >1.57-fache transkriptionelle Aktivierung im Vergleich zu einer Negativkontrolle definiert wurde. Für 79 der insgesamt 104 getesteten DNA-Zielsequenzen, welche in ihrer GXX-, AXX-, CXX-, und TXX-Triplettkomposition variierten, konnte keine einzige ZF-DBD gefunden werden, welche positiv im B2H-System bewertet wurde. Umgerechnet bedeutete dies eine Fehlerrate von ~76% bzw. eine Erfolgsrate von ~24% (Arbeit 2; Figure 1 a). Bei genauerer Betrachtung der Ergebnisse zeigte sich, dass die Triplettkomposition der ZF-Bindungsstellen einen enormen Einfluss auf die Fehlerrate hat. Die *modular assembly*-Methode war um einiges weniger effektiv für ZF-Bindungsstellen, die ein oder gar kein GXX-Triplett aufwiesen (Fehlerraten von 88% bzw. 100%; Arbeit 2; Figure 1 a). Die Fehlerraten für DNA-Bindungsstellen, die zwei bzw. drei GXX-Triplets aufwiesen, lagen dagegen etwas tiefer (Fehlerraten von 71% bzw. 41%; Arbeit 2; Figure 1 a). Da ZFN nur nach einer Dimerisierung enzymatisch aktiv sind, liegen die theoretischen Fehlerraten, ein funktionelles ZFN-Paar mittels der *modular assembly*-Methode für ZF-Bindungsstellen, die drei, zwei, ein bzw. kein GXX-Triplett beinhalten, zu generieren, bei ~65%, ~92%, ~99%, bzw. 100%. Die Fehlerraten wurden dabei jeweils durch das Multiplizieren der Erfolgsrate

eines ZF-Monomers des ZFN-Paares mit anschließender Subtraktion dieser Prozentzahl von 100% berechnet. Die theoretischen Erfolgsraten mit der *modular assembly*-Methode, funktionelle ZFN-Paare für ZF-Bindungsstellen, die drei, zwei, ein bzw. kein GXX-Triplett aufwiesen, generieren zu können, lagen dementsprechend bei ~35%, ~8%, ~1%, bzw. 0% bzw. für alle getesteten ZF-Bindungsstellen bei ~5%. Aufgrund der Beobachtung, dass nicht alle ZF-DBD, die im B2H-System als funktionell eingestuft wurden, auch als ZFN in einem episomalen HR-System in menschlichen Zellen aktiv waren, scheinen die tatsächlichen Erfolgsraten mittels der *modular assembly*-Methode, funktionelle ZFN-Paare zu generieren, jedoch niedriger zu liegen. So konnte gezeigt werden, dass sich die B2H-Methode dazu eignet, ZF-DBD zu identifizieren, die als ZFN keine Aktivität in menschlichen Zellen besitzen, da 100% (neun von neun ZF-DBD) der im B2H-System als negativ bewerteten ZF-DBD auch als ZFN im eukaryotischen Kontext inaktiv waren (rosa Balken; Arbeit 2; Supplementary Figure 4). Im Gegensatz dazu waren neun von vierzehn ZF-DBD, die im B2H-System durch eine mehr als 1,57-fache Aktivierung der Reportertranskription als positiv bewertet wurden, als ZFN im episomalen HR-System signifikant aktiv (mit Sternchen markierte türkise Balken; Arbeit 2; Supplementary Figure 4). Folglich waren ~64% der im B2H-System als positiv eingestuften ZF-DBD auch als ZFN im eukaryotischen Kontext funktionell. Die theoretischen Erfolgsraten, mit der *modular assembly*-Methode funktionelle ZFN-Paare generieren zu können, müssten daher mit diesen ~64% für die Berechnung der realistischen Erfolgsraten multipliziert werden. Für ZF-Bindungsstellen, die drei, zwei, ein bzw. kein GXX-Triplett aufweisen, betragen dementsprechend die realistischen Erfolgsraten ~22%, ~5%, ~0,6% bzw. 0%, während jene für alle 104 getesteten ZF-Bindungssequenzen ~3% beträgt. In zwei ebenfalls groß angelegten Studien wurde die Behauptung aufgestellt, dass die Erfolgsrate, modular assemblierte, funktionelle 3er-ZF-DBD für neun Basenpaar lange ZF-Bindungsstellen generieren zu können, 100% [126] bzw. 60% [127] beträgt. Ein möglicher Grund für die abweichenden Prozentzahlen hinsichtlich der *modular assembly*-Erfolgsrate ist, dass in den damaligen Studien vornehmlich ZF-Bindungsstellen untersucht wurden, die zwei bzw. drei GXX-Triplets aufwiesen, während in der in Absatz 3.3.2 beschriebenen Studie neun Basenpaar lange ZF-Bindungsstellen mit jeder möglichen GXX-Triplettkomposition analysiert wurden (Arbeit 2; Figure 1 b). Wie jedoch gezeigt wurde, hat die ZF-Bindungsstellen-Komposition einen erheblichen Einfluss auf die ermittelte Fehlerrate (Arbeit 2; Figure 1 a). Die vorliegende Studie stellt dabei eine aussagekräftigere Evaluierung der *modular assembly*-Methode dar, da ZF-Bindungsstellen, welche ein bzw. kein GXX-Triplett aufweisen, die Mehrheit (>75%) der potentiell ansteuerbaren, neun Basenpaar langen

ZF-Bindungsstellen repräsentieren (Arbeit 2; Figure 1 b). Daher muss davon ausgegangen werden, dass sich für diese ZF-Bindungsstellen mittels der *modular assembly*-Methode höchstwahrscheinlich keine funktionellen 3er-ZF-DBD generieren lassen. An dieser Stelle sollte auf eine kürzlich veröffentlichte Studie [128] hingewiesen werden, in der die Vermutung geäußert wurde, dass die Erfolgsrate für die Herstellung von funktionellen, modular assemblierten ZFN von der Art der verwendeten ZF-Module abhängt. Dabei wurden für modular assemblierte 3er-ZFN-Paare Erfolgsraten von ~9%, bzw. für modular assemblierte 4er-ZFN-Paare Erfolgsraten von ~26% beobachtet. Die verwendeten ZFN-Paare bestanden in diesem Fall ausschließlich aus natürlich vorkommenden ZF-Modulen (ToolGen-Module), während in der in Absatz 3.3.2 durchgeführten Studie zusätzlich ZFN-Paare evaluiert wurden, deren ZF-Module mittels *phage display* bzw. mittels ortsspezifischer Mutagenese entwickelt wurden (Barbas- bzw. Sangamo-Module).

Ein fundamentaler Trugschluss bei der *modular assembly*-Methode ist die Vermutung, dass ZF-Module funktionell unabhängig sind. Verschiedene Studien konnten diesbezüglich jedoch zeigen, dass die Aminosäure an Position 2 in der  $\alpha$ -Helix eines ZF-Moduls mit einer Base in der Bindungsstelle eines benachbarten ZF-Moduls interagiert, wodurch es zu Kreuzverbindungen zwischen benachbarten Triplets kommt [47, 129-131]. Diese Fakten werden bei den weiterentwickelten *phage display*-Methoden für die sequentielle Selektion [111] und die *bipartite*-Selektion [112] berücksichtigt, wobei die ZF-Module mit neuen DNA-Bindungseigenschaften im Kontext der benachbarten ZF-Module selektioniert werden. Eine weitere Methode zur kontext-abhängigen Selektion von ZF-Modulen basiert auf dem bakteriellen B2H-System. Dabei werden die ZF-Module zuerst entsprechend ihrer späteren Position im 3er-ZF-DBD-Protein vorselektioniert. Anschließend werden die Pools dieser vorselektionierten ZF-Module miteinander kombiniert und die entstandenen 3er-ZF-DBD als Pool in einer finalen Selektionsrunde hinsichtlich ihrer DNA-Bindungsfähigkeit evaluiert [132]. Ein Nachteil aller erwähnten kontext-abhängigen Selektionsstrategien ist jedoch der Fakt, dass diese sehr arbeitsintensiven Methoden eine spezielle Expertise bei der Generierung und Auswertung großer, randomisierter Bibliotheken (typischerweise  $>10^8$ ) benötigen.

Aus diesem Grund wurde ein schnell durchführbares, hocheffektives und für jedermann zugängliches Selektionssystem namens OPEN (*Oligomerized Pool ENgineering*) entwickelt (siehe Absatz 3.3.3). Mittels OPEN wurden 37 ZFN-Paare generiert, mit denen hocheffizient genetische Modifikationen in vier verschiedenen Sequenzen innerhalb eines chromosomal vorliegenden EGFP-Reportergens, sechs verschiedenen Stellen innerhalb von drei humanen endogenen Genen (VEGF-A, HoxB13 und CFTR) und einem Locus eines pflanzlichen

endogenen Gens (SuRA) herbeigeführt werden konnten. Die absoluten Genmodifikationsraten, welche mittels OPEN-generierter ZFN erzielt werden konnten, lagen dabei bei 1-50%. Wie sich herausstellte, war das OPEN-System für die Generierung funktioneller ZFN um einiges effizienter als die *modular assembly*-Methode. Für einen direkten Vergleich von modular assemblierten und OPEN-generierten ZFN wurden mit beiden Systemen 3er-ZF-DBD für zehn ZF-Bindungsstellen in fünf DNA-Sequenzen im EGFP-Gen hergestellt (Arbeit 3; Figure S2 A; Supplemental Data). Dabei konnte für keine der modular assemblierten ZF-DBD eine Aktivität im B2H-System nachgewiesen werden, wohingegen OPEN-selektionierte ZF-DBD an neun der zehn ZF-Bindungsstellen zu einer mehr als dreifachen Transkriptionsaktivierung führten (Arbeit 3; Figure 2 A). In einem NHEJ-Reparaturexperiment, in dem die fehlerhafte NHEJ-Reparatur von ZFN-induzierten DSB im chromosomal stabilisierten EGFP-Lokus eine Abnahme der EGFP-Expression bewirken sollte, waren nur 18% (2 von 11) der modular assemblierten ZFN-Paare, jedoch 75% (15 von 20) der OPEN-generierten ZFN-Paare in menschlichen Zellen aktiv. Darüber hinaus zeigten die OPEN-konstruierten ZFN-Paare an ZF-Bindungsstellen, für die beide Systeme funktionelle ZFN hervorbringen konnten, eine wesentlich höhere Aktivität als die modular assemblierten ZFN (Arbeit 3; Figure 2 C). Die wesentlich höhere Erfolgsrate des OPEN-Ansatzes ist vornehmlich auf die Tatsache zurückzuführen, dass diese Methode die kontext-abhängigen Effekte bei der DNA-Bindung zwischen den benachbarten ZF-Modulen im Gegensatz zur *modular assembly*-Methode berücksichtigt [47, 130, 131]. Ausgehend von dem Set an ZF-Modulen, welches zum Zeitpunkt der Publikation vorlag (für alle GXX-Triplets und einige TXX-Triplets; Arbeit 3; Figure 1 B), lassen sich mittels des OPEN-Systems 3er-ZFN für  $\sim 4,1\%$  ( $23/64 \times 21/64 \times 22/64$ ) aller möglichen neun Basenpaar langen ZFN-Bindungsstellen bzw. für  $\sim 0,16\%$  ( $4,1\% \times 4,1\%$ ) aller möglichen 18 Basenpaar langen ZFN-Paar-Bindungsstellen generieren. Da ZFN-Paare mit einer definierten *Linker*-Länge auch an Bindungsstellen aktiv sein können [60], in denen die beiden einzelnen, gegenläufigen ZFN-Bindungsstellen durch verschieden lange *Spacer* getrennt vorliegen, erhöht sich entsprechend der Anzahl (N) geeigneter *Spacer*-Längen die Zahl (X) der ansteuerbaren ZFN-Paar-Bindungsstellen. In einem beliebigen DNA-Lokus mit einer Länge von 1000 Basenpaaren lägen dementsprechend „ $X = 0,0016 \times 1000 \times N$ “ ZFN-Paar-Bindungsstellen vor. Folglich würde man für ein OPEN-generiertes ZFN-Paar, das an ZFN-Paar-Bindungsstellen mit fünf, sechs oder sieben Basenpaar langen *Spacern* funktionell ist, ungefähr fünf potentielle ZFN-Paar-Bindungsstellen in einem 1000 Basenpaaren langen DNA-Lokus erwarten. Damit das OPEN-System in näherer Zukunft für weitere ZFN-Paar-Bindungsstellen verwendet werden

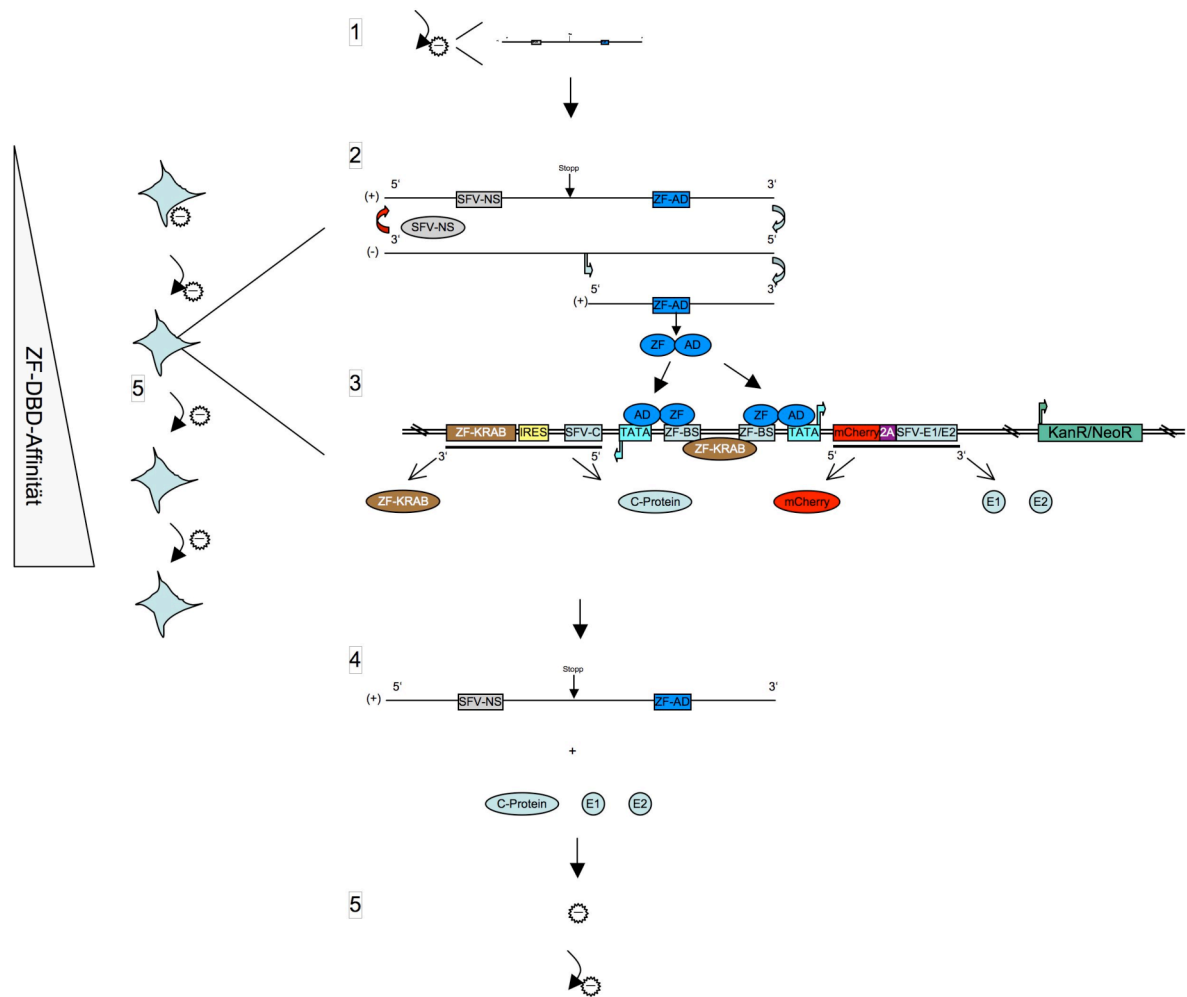
kann, muss darüber hinaus das derzeit verfügbare Set an positionsspezifischen ZF-Modulen (Arbeit 3; Figure 1 B) erweitert werden.

Bei den Experimenten mit den OPEN-generierten ZFN wurden einige interessante Beobachtungen gemacht, die generell bei der Verwendung von ZFN bedacht werden sollten. Entsprechend den in Absatz 3.3.2 erzielten Ergebnissen, waren nicht alle 3er-ZF-DBD, für die im B2H-System eine sequenzspezifische Aktivität nachgewiesen werden konnte, auch als ZFN in menschlichen Zellen funktionell. So waren OPEN-generierte ZFN-Paare nur an vier der fünf ZFN-Paar-Bindungsstellen im chromosomal vorliegenden EGFP-Lokus aktiv (Arbeit 3; Figure 2 C). Am EGFP-Lokus EG568 waren sowohl OPEN-generierte als auch modular assemblierte ZFN-Paare inaktiv, obwohl die für diesen Lokus OPEN-generierten 3er-ZF-DBD auf prokaryotischer Ebene eine mehr als dreifache Transkriptionsaktivierung im B2H-System bewirkten. Auf der anderen Seite waren ZFN am EGFP-Lokus EG502 funktionell, obwohl die 3er-ZF-DBD für die ZFN der linken ZFN-Bindungsstelle (EG502L) im B2H-System inaktiv waren, wohingegen die 3er-ZF-DBD für die ZFN der rechten ZFN-Bindungsstelle (EG502R) im B2H-System zu einer mehr als dreifachen Transkriptionsaktivierung führten. Diese Beobachtungen deuten generell darauf hin, dass man hinsichtlich der Funktionalität von künstlichen ZF-Proteinen keine allgemeingültigen Rückschlüsse aus den Ergebnissen prokaryotisch bzw. eukaryotisch basierter Experimente ziehen kann. Eine weitere interessante Beobachtung bei der Arbeit mit OPEN-generierten ZFN ist, dass die Wirkkraft der ZFN von dem Zelltyp, in dem die ZFN exprimiert werden, abhängig ist. So waren ZFN, welche mittels OPEN für den HoxB13-Lokus generiert wurden, in menschlichen HEK-293-Zellen aktiv, wohingegen sie in humanen K562-Zellen keine Aktivität aufwiesen. Eine mögliche Erklärung hierfür wäre, dass der transkriptionelle Status sowie die Chromatinkonfiguration des angesteuerten Genlokus einen Einfluss darauf haben, ob die ZF-Bindungsstellen innerhalb dieses Lokus für die ZFN-Paare zugänglich sind. Dabei konnte für den HoxB13-Lokus gezeigt werden, dass dieser in den HEK-293-Zellen transkriptionell aktiv ist, während er in den K562-Zellen in einem transkriptionell reprimierten Zustand vorliegt. Aufgrund dieser Erkenntnisse sollte daher vor jedem Einsatz von ZFN die Zugänglichkeit des jeweiligen DNA-Lokus in den Zielzellen überprüft werden. Bezüglich des Einsatzes von ZFN im Genkorrekturkontext sollte folgendes angemerkt werden. Obwohl durch die Zugabe von Vinblastin eine Erhöhung der HR-vermittelten Gentargetingfrequenz bewirkt werden konnte (Arbeit 3; Figure 4 C), deuteten Sequenzierungsuntersuchungen darauf hin, dass einige Allele dennoch NHEJ-vermittelte Deletionen bzw. Insertionen aufwiesen. Darüber hinaus konnte für einzelne Allele gezeigt

werden, dass diese nach einer ZFN-vermittelten Induktion eines DSB durch das HR-vermittelte Gentergeting und zusätzlich durch das fehlerbehaftete NHEJ modifiziert wurden. Aufgrund dieser Erkenntnisse sollten neben Southern-Blot- und PCR-Analysen zusätzlich Sequenzierungsuntersuchungen nach ZFN-Anwendungen durchgeführt werden, in denen eine fehlerfreie Genkorrektur angestrebt wird. Um die enorme Effizienz von OPEN hervorzuheben, sei an dieser Stelle angemerkt, dass mit diesem für jedermann zugänglichen System innerhalb kürzester Zeit die Anzahl an endogenen Säugetiergenen, die mittels ZFN erfolgreich modifiziert wurden, von drei (IL2R $\gamma$ , CCR5 und DHFR) [69, 83, 94] auf sechs verdoppelt werden konnte. Zusätzlich zu diesen drei humanen endogenen Genen (VEGF-A, HoxB13 und CFTR) und dem pflanzlichen endogenen Gen (SuRA) konnten inzwischen mit OPEN-generierten ZFN-Paaren fünf verschiedene Zebrafischgene [133], ein zweites pflanzliches Gen [134] sowie ein weiteres menschliches Gen [135] erfolgreich modifiziert werden.

Eine Alternative zu dem bakteriell basierten OPEN-Selektionssystem könnte ein auf Semliki-Forest-Viren (SFV) [136, 137] basierendes System zur Selektion von hochaffinen ZF-DBD aus einer Library darstellen, dessen Entwicklung im Rahmen dieser Doktorarbeit begonnen wurde (Abb. 8). Dafür werden eukaryotische Zellen mit Plasmiden stabil transfiziert, welche neben einer Antibiotikaresistenzkassette (NeoR: Neomycin-Resistenz) Expressionskassetten für die SFV-Strukturproteine (S-Proteine; SFV-C bzw. SFV-E1/E2) besitzen (Abb. 8; *Flussdiagramm Punkt 3*). Die etablierten Säugetierzellen werden anschließend mit SFV-Vektoren infiziert, die in ihrem Plusstrang-RNA-Genom neben den SFV-Nichtstrukturproteingenen (NS-Proteine; SFV-NS) die Expressionskassette für die mit einer Aktivierungsdomäne (AD) verbundene ZF-DBD aufweisen (Abb. 8; *Flussdiagramm Punkt 1+2*). Das RNA-Genom fungiert gleich als mRNA, wodurch es zur Expression der NS-Proteine kommt. Diese dienen einerseits zur Replikation der Vektorgenome sowie zur Expression der AD-gekoppelten ZF-DBD (ZF-AD) (Abb. 8; *Flussdiagramm Punkt 2*). Die ZF-DBD wiederum erkennt spezifische chromosomal vorliegende Sequenzen innerhalb des bidirektionalen Minimalpromotors (TATA), welcher die Expression der SFV-Strukturproteine kontrolliert. Dadurch sollten die zur SFV-Vektorherstellung benötigten S-Proteine nur exprimiert werden, wenn die ZF-DBD mit hoher Affinität an den Promotor bindet und diesen über die AD aktiviert. Um die Basalaktivität des Minimalpromotors in einem negativen Feedbackzyklus zu unterdrücken, wird bei der Promotoraktivierung ein über eine IRES-Struktur (*internal ribosomal entry site*) an das SFV-Kapsidgen gekoppeltes Repressorprotein (ZF-KRAB) exprimiert. Dieses erkennt seinerseits spezifische Sequenzen innerhalb des





**Abbildung 8 SFV-basiertes Selektionssystem zur Erhöhung der ZF-DBD-Affinität.** Punkt 1+2: Zielzelllinien werden mit SFV-Vektoren, die in ihrem Plusstrang-RNA-Genom neben den SFV-Nichtstrukturproteingenen (NS-Proteine; SFV-NS) die Expressionskassette für die mit einer Aktivierungsdomäne (AD) verbundene ZF-DBD aufweisen, infiziert. Das RNA-Genom fungiert gleich als mRNA, wodurch es zur Expression der NS-Proteine kommt. Diese dienen einerseits zur Replikation der Vektorgenome, sowie zur Expression der AD-gekoppelten ZF-DBD (ZF-AD). Punkt 3: Die ZF-DBD wiederum erkennt spezifische chromosomal vorliegende Sequenzen innerhalb des bidirektionalen Minimalpromotors (TATA), welcher die Expression der SFV-Strukturproteine (S-Proteine; SFV-C bzw. SFV-E1/E2) kontrolliert. Um die Basalaktivität des TATA-Promotors in einem negativen Feedbackzyklus zu unterdrücken, wird bei der Promotoraktivierung ein über eine IRES-Struktur an das SFV-Kapsidgen gekoppeltes Repressorprotein (ZF-KRAB) exprimiert. Dieses erkennt seinerseits spezifische Sequenzen innerhalb des bidirektionalen TATA-Promotors und reprimiert durch die N-terminal-vorliegende KRAB-Domäne nach DNA-Bindung dessen Aktivität. Punkt 4: Am Ende des Replikationszyklus sollten SFV-Vektoren synthetisiert werden, welche die Expressionskassette für die affine ZF-DBD im Genom vorliegen haben. Die Expression des über die 2A-Autoprotease an die E1/E2-Proteine gekoppelten mCherry-Fluoreszenzproteins gibt einem darüber hinaus einen quantitativen Aufschluss über die Expression der S-Proteine. Punkt 5: Durch wiederholte Infektionen der etablierten Zelllinien mit den jeweils neu synthetisierten SFV-Vektoren, sollten dann zellokusspezifische und hochaffine ZF-DBD selektiert werden.

bidirektionalen TATA-Promotors und reprimiert durch die N-terminal-vorliegende KRAB-Domäne nach DNA-Bindung dessen Aktivität (Abb. 8; Flussdiagramm Punkt 3). Am Ende

des Replikationszyklus sollten SFV-Vektoren synthetisiert werden, welche die Expressionskassette für die affine ZF-DBD im Genom vorliegen haben (Abb. 8; *Flussdiagramm Punkt 4*). Die Expression des über die 2A-Autoprotease an die E1/E2-Proteine gekoppelten mCherry-Fluoreszenzproteins gibt einem ferner einen quantitativen Aufschluss über die Expression der S-Proteine und damit darüber, wie spezifisch die ZF-DBD ihre ZF-Bindungsstelle im Promotor erkennt. Durch wiederholte Infektionen der etablierten Zelllinien mit den jeweils neu synthetisierten SFV-Vektoren sollten dann ziellokusspezifische und hochaffine ZF-DBD selektioniert werden (Abb. 8; *Flussdiagramm Punkt 5*). Mittels dieses Systems sollte es möglich sein, für jeden beliebigen Ziellokus hochaffine ZF-DBD im eukaryotischen Kontext generieren zu können. Eine quantitative Beurteilung der Qualität der jeweiligen ZF-DBD hinsichtlich ihrer Fähigkeit zur spezifischen DNA-Bindung kann anschließend mittels Luziferase-Reportergenuntersuchungen vorgenommen werden.

Zusammenfassend lässt sich festhalten, dass im Rahmen dieser Doktorarbeit essentielle Erkenntnisse bezüglich der Generierung und Evaluierung von modular assemblierten ZFN (siehe Absatz 3.3.1) sowie der Erfolgsrate, mit der funktionelle ZFN mittels der *modular assembly*-Methode (siehe Absatz 3.3.2) produziert werden können, gewonnen wurden. Des Weiteren konnte in chromosomalen EGFP-Experimenten gezeigt werden, dass ZFN, die mittels des neu etablierten OPEN-Systems generiert wurden, zu einem wesentlich höheren Prozentsatz funktionell sind als modular assemblierte (siehe Absatz 3.3.3). Darüber hinaus wurde mit der Entwicklung eines SFV-Selektionssystems begonnen, um in Zukunft ZF-DBD im eukaryotischen Kontext selektionieren zu können.

ZFN stellen hocheffektive molekulare Werkzeuge dar, mit deren Hilfe man DSB in jede beliebige DNA-Sequenz eines Zielorganismus einfügen kann. Aufgrund dieser Eigenschaft sind vielfältige Einsatzbereiche denkbar, in denen ZFN in Zukunft zur Geltung kommen können. Zum einen lassen sich ZFN in Anwesenheit einer Donor-DNA zur Stimulation der HR einsetzen [91], wodurch das HR-basierte Gentergeting im Sinne einer Genkorrektur bzw. einer gezielten Genaddition für monogenetische Erbkrankheiten therapeutisch anwendbar wird. So konnten bereits in Studien zur Therapie der Krankheit X-SCID durch das gezielte, ZFN-vermittelte Einfügen eines DSB in den *IL2R $\gamma$* -Lokus erste Erfolge der Genkorrektur und der gezielten Genaddition verbucht werden [69, 94]. Ebenso könnte das ZFN-stimulierte, HR-vermittelte Gentergeting demnächst für *Knock-out*- bzw. *Knock-in*-Ansätze zwecks Erzeugung transgener Tiere und Pflanzen für Forschungs- und Agrarzwecke genutzt werden. Andererseits können ZFN zur gezielten, NHEJ-vermittelten Mutagenese von Genen diverser

Organismen [64, 65, 84-88] genutzt werden. So kann die NHEJ-basierte ZFN-Technologie zu Forschungszwecken für den *knock-out* bestimmter Gene zwecks *reverse genetics*-Analysen und zur Herstellung von transgenen Tiermodellen verwendet werden [86, 124]. Auf therapeutischer Ebene können NHEJ-stimulierende ZFN für den gezielten *knock-out* von dominant-negativ mutierten Allelen oder von Genen zellulärer Rezeptoren, welche von viralen Pathogenen für die Infektion benötigt werden [90], eingesetzt werden. Wie einleitend erwähnt, stellt die Wahl eines geeigneten Gentransfervektors, der als Vehikel für den Transport genetischen Materials zum Wirkungsort benötigt wird, einen essentiellen Bestandteil in der Gentherapie dar. Je nach gewünschter Anwendung und abhängig vom jeweiligen Zielorganismus stehen derzeit verschiedene Vektorsysteme für die Expression von ZFN zur Verfügung. So wurden bis dato vornehmlich Plasmid-basierte Vektoren [37, 63, 66-68, 72, 74] sowie IDLV [94] und modifizierte adenovirale Vektoren [90] für den Transfer von ZFN zwecks einer ortsspezifischen DSB-Induktion verwendet. Aufgrund ihrer hervorragenden Eigenschaften als Gentherapievehikel wurden des Weiteren DNA-Vektoren, welche auf Adeno-assoziierten Viren des Serotyps 2 (AAV-2) basieren [138], als ein potentiell Transportsystem für ZFN evaluiert. Innerhalb dieser Doktorarbeit konnte in Plasmid-basierten Versuchen mit AAV-2-Genomen demonstriert werden, dass ZFN-Paare, die im AAV-2-Kontext exprimiert wurden, ortsspezifisch einen DSB im Genom der verwendeten Zielzellen einfügten (Gellhaus *et al.*; Manuskript in Vorbereitung). Darüber hinaus konnte während der Doktorarbeit gezeigt werden, dass Vektoren, die auf den AAV-Serotypen 2 und 5 basieren, im Mausmodell nach intramuskulärer Injektion retrograd zu den oberen Motoneuronen des zentralen Nervensystems (ZNS) transportiert werden (Müller-Lerch *et al.*; Manuskript in Vorbereitung) [139, 140]. So könnten ZFN-exprimierende AAV-Vektoren zur Therapie von neurodegenerativen Erkrankungen für den gezielten *knock-out* von dominant-negativ mutierten Allelen, wie sie beispielsweise bei der Huntington-Krankheit [89] vorliegen, eingesetzt werden. In einer kürzlich veröffentlichten Studie konnten mittels ZFN gezielt DSB in CAG-Triplett-Wiederholungen eingefügt werden, welche als genetische Ursache für derartige Erkrankungen gelten [141]. Alternativ zu den genannten Vektorsystemen, in denen die ZFN-Expressionskassetten als DNA vorliegen, wodurch eine Vektorintegration zu einer Insertionsmutagenese führen kann, könnten ZFN direkt als Proteine transduziert oder durch eine Transfektion der ZFN-kodierenden mRNA in die Zielzellen eingebracht werden. Bezüglich der ZF-Proteintransduktion konnte bereits der Transfer eines künstlichen ZF-Transkriptionsfaktorproteins in menschliche Zellen demonstriert werden [122]. Mithilfe der ZFN-mRNA-Transfektion wurden darüber hinaus schon erfolgreich DSB in chromosomalen

Sequenzen diverser Zellen von Taufliegen [123], Zebrafischen [86] und Ratten [124] ortsspezifisch gesetzt. Eine weitere Möglichkeit, die Gefahr einer Insertionsmutagenese durch eine Vektorintegration zu umgehen, wäre die Verwendung von SFV-Vektoren, in deren Vektorgenomen die ZFN-Expressionskassetten als RNA vorliegen würden.

Die in dieser Arbeit gewonnen Erkenntnisse hinsichtlich der Generierung und Evaluierung von ZFN stellen dabei einen wichtigen Beitrag für die zukünftigen ZFN-vermittelten Genommodifikationen dar.

## 5 Referenzen

1. Szathmary, E (2003). Why are there four letters in the genetic alphabet? *Nat Rev Genet* **4**: 995-1001.
2. Platt, OS, *et al.* (1994). Mortality in sickle cell disease. Life expectancy and risk factors for early death. *N Engl J Med* **330**: 1639-1644.
3. Hahn, WC, and Weinberg, RA (2002). Modelling the molecular circuitry of cancer. *Nat Rev Cancer* **2**: 331-341.
4. Verma, IM, and Weitzman, MD (2005). Gene therapy: twenty-first century medicine. *Annu Rev Biochem* **74**: 711-738.
5. Vassaux, G, and Groot-Wassink, T (2003). In Vivo Noninvasive Imaging for Gene Therapy. *J Biomed Biotechnol* **2003**: 92-101.
6. Kaji, EH, and Leiden, JM (2001). Gene and stem cell therapies. *JAMA* **285**: 545-550.
7. Hacein-Bey-Abina, S, *et al.* (2002). Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy. *N Engl J Med* **346**: 1185-1193.
8. van de Loo, FA, Smeets, RL, and van den Berg, WB (2004). Gene therapy in animal models of rheumatoid arthritis: are we ready for the patients? *Arthritis Res Ther* **6**: 183-196.
9. Edelstein, ML, Abedi, MR, and Wixon, J (2007). Gene therapy clinical trials worldwide to 2007--an update. *J Gene Med* **9**: 833-842.
10. Somia, N, and Verma, IM (2000). Gene therapy: trials and tribulations. *Nat Rev Genet* **1**: 91-99.
11. Bunnell, BA, and Morgan, RA (1998). Gene therapy for infectious diseases. *Clin Microbiol Rev* **11**: 42-56.
12. Isner, JM (2002). Myocardial gene therapy. *Nature* **415**: 234-239.
13. Lowenstein, PR (1997). Why are we doing so much cancer gene therapy? Disentangling the scientific basis from the origins of gene therapy. *Gene Ther* **4**: 755-756.
14. Kay, MA, and Woo, SL (1994). Gene therapy for metabolic disorders. *Trends Genet* **10**: 253-257.
15. Bohne, J, and Cathomen, T (2008). Genotoxicity in gene therapy: an account of vector integration and designer nucleases. *Curr Opin Mol Ther* **10**: 214-223.
16. Hirata, RK, and Russell, DW (2000). Design and packaging of adeno-associated virus gene targeting vectors. *J Virol* **74**: 4612-4620.
17. Aiuti, A, *et al.* (2009). Gene therapy for immunodeficiency due to adenosine deaminase deficiency. *N Engl J Med* **360**: 447-458.

18. Cavazzana-Calvo, M, *et al.* (2000). Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science* **288**: 669-672.
19. Ott, MG, *et al.* (2006). Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nat Med* **12**: 401-409.
20. Baum, C, *et al.* (2004). Chance or necessity? Insertional mutagenesis in gene therapy and its consequences. *Mol Ther* **9**: 5-13.
21. Seger, RA (2008). Modern management of chronic granulomatous disease. *Br J Haematol* **140**: 255-266.
22. Thomas, KR, and Capecchi, MR (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**: 503-512.
23. Capecchi, MR (1989). The new mouse genetics: altering the genome by gene targeting. *Trends Genet* **5**: 70-76.
24. Manis, JP (2007). Knock out, knock in, knock down--genetically manipulated mice and the Nobel Prize. *N Engl J Med* **357**: 2426-2429.
25. Capecchi, MR (2005). Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat Rev Genet* **6**: 507-512.
26. Cathomen, T (2004). AAV vectors for gene correction. *Curr Opin Mol Ther* **6**: 360-366.
27. Sung, P, and Klein, H (2006). Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nat Rev Mol Cell Biol* **7**: 739-750.
28. San Filippo, J, Sung, P, and Klein, H (2008). Mechanism of eukaryotic homologous recombination. *Annu Rev Biochem* **77**: 229-257.
29. Orr-Weaver, TL, Szostak, JW, and Rothstein, RJ (1981). Yeast transformation: a model system for the study of recombination. *Proc Natl Acad Sci U S A* **78**: 6354-6358.
30. Kucherlapati, RS, Eves, EM, Song, KY, Morse, BS, and Smithies, O (1984). Homologous recombination between plasmids in mammalian cells can be enhanced by treatment of input DNA. *Proc Natl Acad Sci U S A* **81**: 3153-3157.
31. Johnson, RD, and Jasin, M (2001). Double-strand-break-induced homologous recombination in mammalian cells. *Biochem Soc Trans* **29**: 196-201.
32. Jasin, M (1996). Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet* **12**: 224-228.
33. Belfort, M, and Roberts, RJ (1997). Homing endonucleases: keeping the house in order. *Nucleic Acids Res* **25**: 3379-3388.
34. de Piedoue, G, *et al.* (2005). Improving gene replacement by intracellular formation of linear homologous DNA. *J Gene Med* **7**: 649-656.

35. Rouet, P, Smih, F, and Jasin, M (1994). Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc Natl Acad Sci U S A* **91**: 6064-6068.
36. Donoho, G, Jasin, M, and 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**: 4070-4078.
37. Szczepek, M, Brondani, V, Buchel, J, Serrano, L, Segal, DJ, and Cathomen, T (2007). Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat Biotechnol* **25**: 786-793.
38. Porteus, MH, Connelly, JP, and Pruett, SM (2006). A look to future directions in gene therapy research for monogenic diseases. *PLoS Genet* **2**: e133.
39. Knauert, MP, Kalish, JM, Hegan, DC, and Glazer, PM (2006). Triplex-stimulated intermolecular recombination at a single-copy genomic target. *Mol Ther* **14**: 392-400.
40. Dervan, PB, and Edelson, BS (2003). Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr Opin Struct Biol* **13**: 284-299.
41. Kaihatsu, K, Janowski, BA, and Corey, DR (2004). Recognition of chromosomal DNA by PNAs. *Chem Biol* **11**: 749-758.
42. Epinat, JC, *et al.* (2003). A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res* **31**: 2952-2962.
43. Miller, J, McLachlan, AD, and Klug, A (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J* **4**: 1609-1614.
44. Wolfe, SA, Nekludova, L, and Pabo, CO (2000). DNA recognition by Cys2His2 zinc finger proteins. *Annu Rev Biophys Biomol Struct* **29**: 183-212.
45. Tupler, R, Perini, G, and Green, MR (2001). Expressing the human genome. *Nature* **409**: 832-833.
46. Mackay, JP, and Crossley, M (1998). Zinc fingers are sticking together. *Trends Biochem Sci* **23**: 1-4.
47. Elrod-Erickson, M, Rould, MA, Nekludova, L, and Pabo, CO (1996). Zif268 protein-DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions. *Structure* **4**: 1171-1180.
48. Blancafort, P, Segal, DJ, and Barbas, CF, 3rd (2004). Designing transcription factor architectures for drug discovery. *Mol Pharmacol* **66**: 1361-1371.
49. Pavletich, NP, and Pabo, CO (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252**: 809-817.
50. Xu, GL, and Bestor, TH (1997). Cytosine methylation targeted to pre-determined sequences. *Nat Genet* **17**: 376-378.

51. Carvin, CD, Parr, RD, and Kladde, MP (2003). Site-selective in vivo targeting of cytosine-5 DNA methylation by zinc-finger proteins. *Nucleic Acids Res* **31**: 6493-6501.
52. Beerli, RR, Segal, DJ, Dreier, B, and Barbas, CF, 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**: 14628-14633.
53. Yaghmai, R, and Cutting, GR (2002). Optimized regulation of gene expression using artificial transcription factors. *Mol Ther* **5**: 685-694.
54. Thiesen, HJ, Bellefroid, E, Revelant, O, and Martial, JA (1991). Conserved KRAB protein domain identified upstream from the zinc finger region of Kox 8. *Nucleic Acids Res* **19**: 3996.
55. Carroll, D (2004). Using nucleases to stimulate homologous recombination. *Methods Mol Biol* **262**: 195-207.
56. Jamieson, AC, Miller, JC, and Pabo, CO (2003). Drug discovery with engineered zinc-finger proteins. *Nat Rev Drug Discov* **2**: 361-368.
57. Kim, YG, Cha, J, and Chandrasegaran, S (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A* **93**: 1156-1160.
58. Bitinaite, J, Wah, DA, Aggarwal, AK, and Schildkraut, I (1998). FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci U S A* **95**: 10570-10575.
59. Mani, M, Smith, J, Kandavelou, K, Berg, JM, and 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**: 1191-1197.
60. Handel, EM, Alwin, S, and 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**: 104-111.
61. Klug, A (2005). Towards therapeutic applications of engineered zinc finger proteins. *FEBS Lett* **579**: 892-894.
62. Miller, DG, Petek, LM, and Russell, DW (2004). Adeno-associated virus vectors integrate at chromosome breakage sites. *Nat Genet* **36**: 767-773.
63. Alwin, S, *et al.* (2005). Custom zinc-finger nucleases for use in human cells. *Mol Ther* **12**: 610-617.
64. Beumer, K, Bhattacharyya, G, Bibikova, M, Trautman, JK, and Carroll, D (2006). Efficient gene targeting in *Drosophila* with zinc-finger nucleases. *Genetics* **172**: 2391-2403.
65. Bibikova, M, Beumer, K, Trautman, JK, and Carroll, D (2003). Enhancing gene targeting with designed zinc finger nucleases. *Science* **300**: 764.



66. Moehle, EA, *et al.* (2007). Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc Natl Acad Sci U S A* **104**: 3055-3060.
67. Porteus, MH (2006). Mammalian gene targeting with designed zinc finger nucleases. *Mol Ther* **13**: 438-446.
68. Porteus, MH, and Baltimore, D (2003). Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**: 763.
69. Urnov, FD, *et al.* (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**: 646-651.
70. Wright, DA, *et al.* (2005). High-frequency homologous recombination in plants mediated by zinc-finger nucleases. *Plant J* **44**: 693-705.
71. Porteus, MH, Cathomen, T, Weitzman, MD, and Baltimore, D (2003). Efficient gene targeting mediated by adeno-associated virus and DNA double-strand breaks. *Mol Cell Biol* **23**: 3558-3565.
72. Miller, JC, *et al.* (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol* **25**: 778-785.
73. Cathomen, T, and Joung, JK (2008). Zinc-finger nucleases: the next generation emerges. *Mol Ther* **16**: 1200-1207.
74. Cornu, TI, *et al.* (2008). DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Mol Ther* **16**: 352-358.
75. Bibikova, M, *et al.* (2001). Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol* **21**: 289-297.
76. Smith, J, Bibikova, M, Whitby, FG, Reddy, AR, Chandrasegaran, S, and Carroll, D (2000). Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res* **28**: 3361-3369.
77. Camenisch, TD, Brilliant, MH, and Segal, DJ (2008). Critical parameters for genome editing using zinc finger nucleases. *Mini Rev Med Chem* **8**: 669-676.
78. Joung, JK, Ramm, EI, and Pabo, CO (2000). A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc Natl Acad Sci U S A* **97**: 7382-7387.
79. Durai, S, Bosley, A, Abulencia, AB, Chandrasegaran, S, and Ostermeier, M (2006). A bacterial one-hybrid selection system for interrogating zinc finger-DNA interactions. *Comb Chem High Throughput Screen* **9**: 301-311.
80. Meng, X, Thibodeau-Beganny, S, Jiang, T, Joung, JK, and Wolfe, SA (2007). Profiling the DNA-binding specificities of engineered Cys2His2 zinc finger domains using a rapid cell-based method. *Nucleic Acids Res* **35**: e81.
81. Carroll, D (2008). Progress and prospects: zinc-finger nucleases as gene therapy agents. *Gene Ther* **15**: 1463-1468.

82. Lieber, MR, Ma, Y, Pannicke, U, and Schwarz, K (2003). Mechanism and regulation of human non-homologous DNA end-joining. *Nat Rev Mol Cell Biol* **4**: 712-720.
83. Santiago, Y, *et al.* (2008). Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc Natl Acad Sci U S A* **105**: 5809-5814.
84. Morton, J, Davis, MW, Jorgensen, EM, and Carroll, D (2006). Induction and repair of zinc-finger nuclease-targeted double-strand breaks in *Caenorhabditis elegans* somatic cells. *Proc Natl Acad Sci U S A* **103**: 16370-16375.
85. Ekker, SC (2008). Zinc finger-based knockout punches for zebrafish genes. *Zebrafish* **5**: 121-123.
86. Doyon, Y, *et al.* (2008). Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat Biotechnol* **26**: 702-708.
87. Meng, X, Noyes, MB, Zhu, LJ, Lawson, ND, and Wolfe, SA (2008). Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol* **26**: 695-701.
88. Lloyd, A, Plaisier, CL, Carroll, D, and Drews, GN (2005). Targeted mutagenesis using zinc-finger nucleases in *Arabidopsis*. *Proc Natl Acad Sci U S A* **102**: 2232-2237.
89. Imarisio, S, *et al.* (2008). Huntington's disease: from pathology and genetics to potential therapies. *Biochem J* **412**: 191-209.
90. Perez, EE, *et al.* (2008). Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol* **26**: 808-816.
91. Durai, S, Mani, M, Kandavelou, K, Wu, J, Porteus, MH, and Chandrasegaran, S (2005). Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Res* **33**: 5978-5990.
92. Notarangelo, LD, *et al.* (2000). Of genes and phenotypes: the immunological and molecular spectrum of combined immune deficiency. Defects of the gamma(c)-JAK3 signaling pathway as a model. *Immunol Rev* **178**: 39-48.
93. Lozzio, CB, and Lozzio, BB (1975). Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. *Blood* **45**: 321-334.
94. Lombardo, A, *et al.* (2007). Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol* **25**: 1298-1306.
95. Lusso, P (2006). HIV and the chemokine system: 10 years later. *EMBO J* **25**: 447-456.
96. Lim, JK, Glass, WG, McDermott, DH, and Murphy, PM (2006). CCR5: no longer a "good for nothing" gene--chemokine control of West Nile virus infection. *Trends Immunol* **27**: 308-312.
97. Cowan, CA, *et al.* (2004). Derivation of embryonic stem-cell lines from human blastocysts. *N Engl J Med* **350**: 1353-1356.

98. Elrod-Erickson, M, and Pabo, CO (1999). Binding studies with mutants of Zif268. Contribution of individual side chains to binding affinity and specificity in the Zif268 zinc finger-DNA complex. *J Biol Chem* **274**: 19281-19285.
99. Shi, Y, and Berg, JM (1995). A direct comparison of the properties of natural and designed zinc-finger proteins. *Chem Biol* **2**: 83-89.
100. Thukral, SK, Morrison, ML, and Young, ET (1992). Mutations in the zinc fingers of ADR1 that change the specificity of DNA binding and transactivation. *Mol Cell Biol* **12**: 2784-2792.
101. Desjarlais, JR, and Berg, JM (1992). Toward rules relating zinc finger protein sequences and DNA binding site preferences. *Proc Natl Acad Sci U S A* **89**: 7345-7349.
102. Choo, Y, and Klug, A (1994). Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. *Proc Natl Acad Sci U S A* **91**: 11163-11167.
103. Rebar, EJ, and Pabo, CO (1994). Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* **263**: 671-673.
104. Choo, Y, and Klug, A (1995). Designing DNA-binding proteins on the surface of filamentous phage. *Curr Opin Biotechnol* **6**: 431-436.
105. Beerli, RR, and Barbas, CF, 3rd (2002). Engineering polydactyl zinc-finger transcription factors. *Nat Biotechnol* **20**: 135-141.
106. Mandell, JG, and Barbas, CF, 3rd (2006). Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res* **34**: W516-523.
107. Segal, DJ, Dreier, B, Beerli, RR, and Barbas, CF, 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**: 2758-2763.
108. Wu, H, Yang, WP, and Barbas, CF, 3rd (1995). Building zinc fingers by selection: toward a therapeutic application. *Proc Natl Acad Sci U S A* **92**: 344-348.
109. Dreier, B, Segal, DJ, and Barbas, CF, 3rd (2000). Insights into the molecular recognition of the 5'-GNN-3' family of DNA sequences by zinc finger domains. *J Mol Biol* **303**: 489-502.
110. Dreier, B, Beerli, RR, Segal, DJ, Flippin, JD, and Barbas, CF, 3rd (2001). Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem* **276**: 29466-29478.
111. Greisman, HA, and Pabo, CO (1997). A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* **275**: 657-661.
112. Isalan, M, Klug, A, and Choo, Y (2001). A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nat Biotechnol* **19**: 656-660.

113. Isalan, M, and Choo, Y (2001). Rapid, high-throughput engineering of sequence-specific zinc finger DNA-binding proteins. *Methods Enzymol* **340**: 593-609.
114. Cheng, X, Boyer, JL, and Juliano, RL (1997). Selection of peptides that functionally replace a zinc finger in the Sp1 transcription factor by using a yeast combinatorial library. *Proc Natl Acad Sci U S A* **94**: 14120-14125.
115. Bartsevich, VV, and Juliano, RL (2000). Regulation of the MDR1 gene by transcriptional repressors selected using peptide combinatorial libraries. *Mol Pharmacol* **58**: 1-10.
116. Meng, X, Brodsky, MH, and Wolfe, SA (2005). A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. *Nat Biotechnol* **23**: 988-994.
117. Joung, JK (2001). Identifying and modifying protein-DNA and protein-protein interactions using a bacterial two-hybrid selection system. *J Cell Biochem Suppl* **Suppl 37**: 53-57.
118. Fu, F, *et al.* (2009). Zinc Finger Database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays. *Nucleic Acids Res* **37**: D279-283.
119. Hacein-Bey-Abina, S, *et al.* (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**: 415-419.
120. Hacein-Bey-Abina, S, *et al.* (2003). A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N Engl J Med* **348**: 255-256.
121. Donsante, A, *et al.* (2007). AAV vector integration sites in mouse hepatocellular carcinoma. *Science* **317**: 477.
122. Tachikawa, K, Schroder, O, Frey, G, Briggs, SP, and Sera, T (2004). Regulation of the endogenous VEGF-A gene by exogenous designed regulatory proteins. *Proc Natl Acad Sci U S A* **101**: 15225-15230.
123. Beumer, KJ, *et al.* (2008). Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. *Proc Natl Acad Sci U S A* **105**: 19821-19826.
124. Geurts, AM, *et al.* (2009). Knockout rats via embryo microinjection of zinc-finger nucleases. *Science* **325**: 433.
125. Araki, R, *et al.* (1997). Nonsense mutation at Tyr-4046 in the DNA-dependent protein kinase catalytic subunit of severe combined immune deficiency mice. *Proc Natl Acad Sci U S A* **94**: 2438-2443.
126. Segal, DJ, *et al.* (2003). Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry* **42**: 2137-2148.
127. Bae, KH, *et al.* (2003). Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat Biotechnol* **21**: 275-280.

128. Kim, HJ, Lee, HJ, Kim, H, Cho, SW, and Kim, JS (2009). Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res* **19**: 1279-1288.
129. Isalan, M, Klug, A, and Choo, Y (1998). Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers. *Biochemistry* **37**: 12026-12033.
130. Isalan, M, Choo, Y, and Klug, A (1997). Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proc Natl Acad Sci U S A* **94**: 5617-5621.
131. Wolfe, SA, Grant, RA, Elrod-Erickson, M, and Pabo, CO (2001). Beyond the "recognition code": structures of two Cys2His2 zinc finger/TATA box complexes. *Structure* **9**: 717-723.
132. Hurt, JA, Thibodeau, SA, Hirsh, AS, Pabo, CO, and Joung, JK (2003). Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc Natl Acad Sci U S A* **100**: 12271-12276.
133. Foley, JE, *et al.* (2009). Rapid mutation of endogenous zebrafish genes using zinc finger nucleases made by Oligomerized Pool ENgineering (OPEN). *PLoS One* **4**: e4348.
134. Townsend, JA, *et al.* (2009). High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature* **459**: 442-445.
135. Zou, J, *et al.* (2009). Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell* **5**: 97-110.
136. Smerdou, C, and Liljestrom, P (1999). Two-helper RNA system for production of recombinant Semliki forest virus particles. *J Virol* **73**: 1092-1098.
137. Lundstrom, K (2005). Biology and application of alphaviruses in gene therapy. *Gene Ther* **12 Suppl 1**: S92-97.
138. Grimm, D, and Kay, MA (2003). From virus evolution to vector revolution: use of naturally occurring serotypes of adeno-associated virus (AAV) as novel vectors for human gene therapy. *Curr Gene Ther* **3**: 281-304.
139. Ermolayev, V, *et al.* (2009). Impaired axonal transport in motor neurons correlates with clinical prion disease. *PLoS Pathog* **5**: e1000558.
140. Ermolayev, V, *et al.* (2009). Ultramicroscopy reveals axonal transport impairments in cortical motor neurons at prion disease. *Biophys J* **96**: 3390-3398.
141. Mittelman, D, *et al.* (2009). Zinc-finger directed double-strand breaks within CAG repeat tracts promote repeat instability in human cells. *Proc Natl Acad Sci U S A* **106**: 9607-9612.

## **6 Anhang**

### **6.1 Danksagung**

Mein Dank gilt vor allem Herrn Prof. Dr. Toni Cathomen für die Möglichkeit, an diesem sehr interessanten Thema in seiner Arbeitsgruppe forschen zu können. Ebenso möchte ich mich für seine freundliche Betreuung und Unterstützung bedanken.

In diesem Zusammenhang möchte ich mich auch bei der gesamten AG Cathomen für wertvolle Tipps und ein produktives Arbeitsklima bedanken.

Ein weiteres großes Dankeschön geht an Herrn Prof. Dr. Rupert Mutzel für seine Bereiterklärung, diese externe Doktorarbeit zu begutachten.

Nicht zuletzt möchte ich mich bei meiner wunderbaren Frau Maren und meiner kleinen Tochter Junia, meiner Familie und meinen Freunden für jegliche Unterstützung (nicht nur während der Doktorarbeitszeit) bedanken.

## **6.2 Lebenslauf**

Der Lebenslauf wurde aus datenschutzrechtlichen Gründen entfernt.

### 6.3 Publikationen

1. Cathomen T, Segal DJ, Brondani V, Müller-Lerch F.: **Generation and Functional Analysis of Zinc Finger Nucleases**; Methods Mol Biol. 2008; 434:277-90.
2. 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.: **Unexpected failure rates for modular assembly of engineered zinc fingers**; Nat Methods. 2008; 5:374-5.
3. Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, Eichtinger M, Jiang T, Foley JE, Winfrey RJ, Townsend JA, Unger-Wallace E, Sander JD, Müller-Lerch F., Fu F, Pearlberg J, Göbel C, Dassie JP, Pruett-Miller SM, Porteus MH, Sgroi DC, Iafrate AJ, Dobbs D, McCray PB Jr, Cathomen T, Voytas DF, Joung JK.: **Rapid “Open-Source” Engineering of Customized Zinc-Finger Nucleases for Highly Efficient Gene Modification**; Mol Cell. 2008; 31:294-301.
4. Gellhaus, K., Müller-Lerch F., Heilbronn, R., Cathomen, T.: **Zinc-finger nuclease induced gene targeting using AAV vectors**; (in Vorbereitung)
5. Müller-Lerch, F., Ermolayev, V., Guhl, E., Hochgräfe, K., Klein, M., Flechsig, E., Cathomen, T.: **AAV vectors endorse efficient retrograde transduction of upper motor neurons in the CNS**; (in Vorbereitung)

### 6.4 Vorträge

1. **Generation of artificial Zinc Fingers**; *nuclear fate*-Workshop im Rahmen des DFG-Schwerpunktprogramms 1230; München, März 2008



2. **AAV vectors endorse efficient retrograde transduction of upper motor neurons in the CNS**; Jahrestagung der Deutschen Gesellschaft für Gentherapie (DG-GT); Berlin, Oktober 2008

## **6.5 Einzelarbeiten**

## Generation and Functional Analysis of Zinc Finger Nucleases

Toni Cathomen, David J. Segal, Vincent Brondani, and Felix Müller-Lerch

### Summary

The recent development of artificial endonucleases with tailored specificities has opened the door for a wide range of new applications, including the correction of mutated genes directly in the chromosome. This kind of gene therapy is based on homologous recombination, which can be stimulated by the creation of a targeted DNA double-strand break (DSB) near the site of the desired recombination event. Artificial nucleases containing zinc finger DNA-binding domains have provided important proofs of concept, showing that inserting a DSB in the target locus leads to gene correction frequencies of 1–18% in human cells. In this paper, we describe how zinc finger nucleases are assembled by polymerase chain reaction (PCR) and present two methods to assess these custom nucleases quickly in vitro and in a cell-based recombination assay.

**Key Words:** ZFN; artificial nuclease; custom nuclease; homologous recombination; HR; DNA double strand break; DSB.

### 1. Introduction

The targeted modification of complex genomes by homologous recombination (HR) is a powerful tool for genetic studies, biotechnology and gene therapy. Because HR is a rare event in mammalian cells (*I*), the development of HR-based therapies has only become a possibility after the observation that the HR frequency can be dramatically enhanced by the creation of a targeted DNA double-strand break (DSB) near the site of the desired recombination

event. The recent development of artificial endonucleases with tailored specificities has thus opened the door for a wide range of new applications, including the correction of mutated genes directly in the chromosome. Because of their modularity in structure and function, zinc finger-based DNA-binding domains are highly suitable to direct an endonuclease domain to a specific genomic target site. Artificial nucleases containing a zinc finger DNA-binding domain have provided important proofs of concept. Several recent studies have reported gene correction frequencies of 1–18% in human cells when the targeted DSB was introduced by so-called zinc finger nucleases (ZFNs) (2–5). ZFNs have also been successfully used to stimulate HR in whole organisms such as drosophila (6) and plants (7).

ZFNs are composed of an engineered zinc finger DNA-binding domain tethered to the catalytic domain of the type II restriction enzyme *FokI* (8,9). A single zinc finger motif consists of about 30 amino acids and includes an  $\alpha$ -helix that generally contacts three bases in the major groove of DNA (10). Through a combination of rational design and selection by phage display, a database of zinc fingers motifs has been generated that includes DNA-binding domains able to recognize most of the 64 possible triples (11–19). Through modular assembly of such zinc finger motifs into tandem arrays, artificial DNA-binding domains can be engineered that recognize a wide variety of novel DNA target sites. Because the active ZFN is a dimer, two subunits are typically designed to recognize the target sequence in a tail-to-tail conformation (20,21). The C-terminal *FokI* domains are positioned in proximity on the same face of the DNA helix by separating the two zinc finger-binding sites by a 6-bp spacer (22). The required combination of two ZFN subunits, each consisting of three fingers, thus provides enough specificity in any mammalian genome by defining a unique 18-bp site.

In this chapter, we describe how ZFNs are generated by using PCR and present two methods to assess them quickly in vitro or in a cell-based recombination assay.

## 2. Materials

### 2.1. Generation of ZFNs

1. Oligonucleotides for assembly and amplification PCR (see Fig. 1A): The oligonucleotides were ordered at MWG Biotech (High Purity Salt Free Purification; see Note 1):

C1: 5'-cggggagaaaccctataagtgtccggagtgtggcaagtcgttctc

C2: 5'-gcgtaccacacggcgcaaaagccgtacaatgccagaatgcggtaaatccttcagc

C3: 5'-tcaacggacgcatacaggagagaagccatacaaatgtcccgaatgtgggaagagtttag



V1: 5'-gcccgtgtgggtacgctgatgnnnnnnnnnnnnnnnnnnncgagaacgactgccacac  
 V2: 5'-cctgatgctccggtgatgnnnnnnnnnnnnnnnnnnngctgaaggattaccgcat  
 V3: 5'-gtgtgagtgcgctggtggnnnnnnnnnnnnnnnnnnactaaaactcttcccattcg  
 ZF-F: 5'-caggacctcgagcccggggagaaacctataag  
 ZF-R: 5'-caggacaccggtgtgagtgcgctggtg

2. PCR: Vent DNA polymerase (New England BioLabs, Frankfurt, Germany) supplied with 10× ThermoPol-reaction buffer; dNTPs (10 mM, Roche, Mannheim, Germany).
3. TAE running buffer: 40 mM Tris–HCl, pH 8.0, 20 mM acetic acid, 1 mM ethylenediaminetetraacetic acid (EDTA); 10 mg/ml ethidium bromide; 6× loading dye [20% (w/v) Ficoll 400, 0.25% (w/v) bromophenol blue and 0.25 % (w/v) xylene cyanol FF, dissolved in water].
4. Agarose gel electrophoresis apparatus.
5. DNA extraction kit, such as Invisorb Spin DNA Extraction Kit (Invitex, Berlin, Germany).
6. T4 DNA ligase (400,000 U/ml) supplied with 10× ligase buffer and restriction endonucleases from New England BioLabs.
7. XL1-Blue cells (Stratagene, La Jolla, CA, USA).

## 2.2. *In Vitro* Nuclease Assay

1. The plasmid pMAL-c2X (New England BioLabs) is a prokaryotic expression vector that appends the maltose-binding protein (MBP) to the N-terminus of the expressed protein. The MBP fusion greatly improves the solubility of the ZFNs and does not appear to interfere with activity. The MBP fusion also enables purification to >90% in a single step over amylose resin (*see Note 2*).
2. Plasmid pMAL.ZFN (*see Fig. 1B*), a modified pMAL vector that places the N-terminal MBP in-frame with an engineered zinc finger peptide and a C-terminal *FokI* cleavage domain, can be obtained from the authors.
3. DNA cleavage substrates: There are many ways to design target DNA for these experiments. The binding sites can be easily incorporated into a plasmid by constructing complementary oligonucleotides that contain the desired target site flanked by appropriate restriction sites (*see Fig. 2A–C*). We typically use the same target plasmids (TPs) that are used in the episomal recombination assay (*see Subheading 3.3.*). For the *in vitro* cleavage assay, the plasmid should be linearized in such a manner that cleavage by the ZFN will produce two fragments of different sizes. ZFN cleavage at the appropriate site will therefore indicate activity. Specificity of ZFNs can be assessed by using DNA substrates that include altered or non-target sites (*see Fig. 2B*).



Fig. 1. (Continued) Overlapping sequences between the oligonucleotides are underlined (dashed for the outside primers) and relevant cloning sites are indicated.

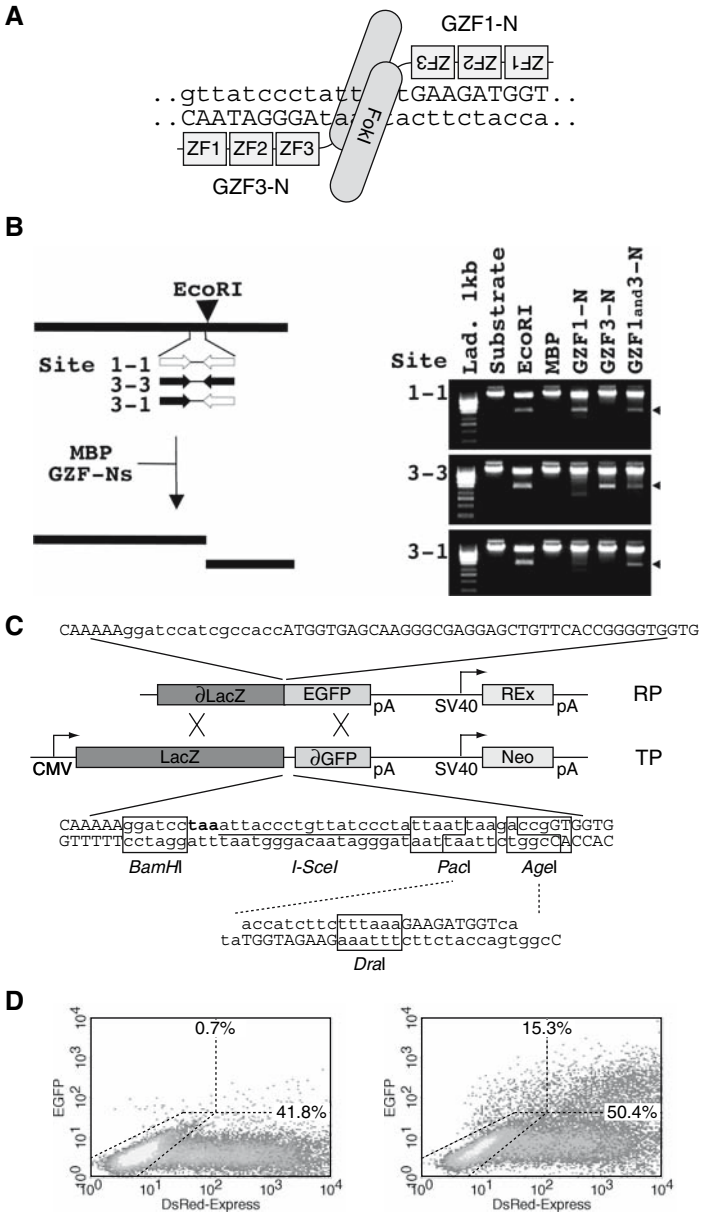


Fig. 2. Evaluation of zinc finger nucleases (ZFNs). (A) Schematic of a pair of ZFNs (GZF1-N and GZF3-N) bound to DNA. Individual ZF motifs (ZF1–ZF3) are shown as boxes. Each motif contacts three nucleotides of the target sequence (nucleotides contacted by the ZFs in capitals). The zinc fingers are fused to the catalytic domain of the *FokI* endonuclease (*FokI*), which must dimerize in order to become active.

4. Cells: *Escherichia coli* ER2508 (New England BioLabs). These cells provide tight repression of protein expression before induction and high expression levels after induction.
5. Luria Broth (LB) powder (Sigma). Suspend 20 g in 1 l of distilled water; Autoclave. Carbenicillin (Omega Scientific, Tarzana, CA, USA), 100 mg/ml in water; Filter sterilize, and store at  $-20^{\circ}\text{C}$ . The antibiotic (a more stable analog of ampicillin) is added to LB media at 50  $\mu\text{g}/\text{ml}$  just before use.
6. Isopropyl-beta-D-thiogalactopyranoside (IPTG) (Sigma), 0.5 M in water. Filter sterilize and store at  $-20^{\circ}\text{C}$ .
7.  $\text{ZnCl}_2$ , 100 mM in water.
8. Column buffer: 10 mM Tris-HCl (pH 8.0), 90 mM KCl, 0.1 mM  $\text{ZnCl}_2$ , 5 mM DTT, 0.2% Triton-X-100, 1 $\times$  Complete protease inhibitor without EDTA (Roche). Prepare 50 ml for each ZFN to be purified.
9. Elution buffer: Dissolve 36 mg of maltose in 10 ml of column buffer (10 mM final).
10. Amylose resin (New England BioLabs).
11. Econo-Pac columns with 2-way stopcocks (Bio-Rad, Hercules, CA, USA), ring stand, clamps.
12. Sonicator, such as a Fisher Scientific Sonic Dismembrator, Model 100.



Fig. 2. (*Continued*) Consequently, the full target sequence is made up of two 9-bp half sites, which are separated by a 6-bp spacer. **(B)** In vitro cleavage assay. A schematic of the in vitro cleavage assay is shown on the left. A linear DNA substrate containing binding sites for GZF1-N (1-1), GZF3-N (3-3), or GZF3-N and GZF1-N (3-1) is cleaved in different size products. The right panel shows analysis of the cleavage reaction. Purified ZFNs were incubated with the linear DNA substrate and the extent of cleavage analyzed by agarose gel electrophoresis. Control reactions (substrate alone and cleavage with *EcoRI*) and protein control (MBP) are indicated. An arrow points to the cleavage products. **(C)** Experimental set up of cell-based recombination assay. Repair plasmid (RP) and target plasmid (TP) are shown schematically. The relevant sequences are given above or below, including LacZ- and EGFP-coding sequences (capital letters), the recognition site for *I-SceI* (underlined), and the relevant cloning sites (boxed). LacZ-EGFP expression from TP is prevented by a stop codon (bold), which terminates translation of LacZ, and a 33-bp truncation at the 5'-end of the EGFP open-reading frame. RP contains a 5'-deleted LacZ gene fused in-frame with a downstream EGFP open-reading frame. Expression of a LacZ-EGFP fusion protein is prevented by the 1197-bp deletion at the 5'-end and by omitting a promoter. To identify transfected cells, an expression cassette for DsRed-Express (Clontech, Mountain View, CA, USA) is located further downstream. Inserting a target site between the *PacI* and *AgeI* sites generates a customized TP for assessing novel ZFNs. An example is given for GZF1-N. **(D)** Analysis of gene repair by using flow cytometry. HEK293T cells transfected with RP, TP and a ZFN expression vector are analyzed by flow cytometry after 2 days. The right panel displays the extent of gene repair of a corresponding TP stimulated by the expression of GZF1-N, while the left panel depicts unstimulated recombination.

13. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE): 10–20% gradient Tris–HCl Ready-Gels, 50  $\mu$ l, 10 wells (Bio-Rad), Broad Range SDS–PAGE Standards (Bio-Rad), and appropriate apparatus, buffers and stains.
14. BSA standards prepared to 50, 100, 200 and 400  $\mu$ g/ml.
15. 10 $\times$  Reaction buffer: 100 mM Tris–HCl (pH 8.0), 900 mM KCl, 1 mM ZnCl<sub>2</sub>, 50 mM DTT. Prepare 1 ml.
16. MgCl<sub>2</sub>, 200 mM in water.
17. 5 $\times$  Loading Buffer: 4 mM Orange-G (Sigma), 30% glycerol.

### 2.3. Cell-Based Recombination Assay

1. Plasmids: Eukaryotic expression plasmid pPGK.ZFN, TP and repair plasmid (RP) are described in (5). The plasmids and the corresponding plasmid maps can be obtained upon request.
2. Generation of TP: Oligonucleotides ZFBS-F: 5'-nnnnnnnnntttaannnnnnnnnca and ZFBS-R: 5'-ccggtgnnnnnnnnntttaannnnnnnnnat (*see Note 3*); T4 DNA ligase (400,000 U/ml) supplied with 10 $\times$  ligase buffer and restriction endonucleases from New England BioLabs; XL1-Blue cells (Stratagene).
3. HEK293T cells. These cells are a highly transfectable derivative of the 293 cell line into which the temperature sensitive gene for SV40 T-antigen was inserted (ATCC). The cells are maintained in Dulbecco's modified Eagle Medium (DMEM with high glucose; Gibco/Invitrogen, Karlsruhe, Germany) supplemented with 10% fetal bovine serum (FBS; Gibco/Invitrogen) and penicillin/streptomycin (100 $\times$  solution; Gibco/Invitrogen).
4. Trypsin solution (0.05% Trypsin–EDTA; Gibco/Invitrogen).
5. CO<sub>2</sub> incubator with 5% CO<sub>2</sub> atmosphere, such as Heraeus HeraCell240.
6. 2 $\times$  BES-buffered saline (BBS) for transfection: dissolve 4.28 g of BES (Sigma), 6.4 g of NaCl and 0.108 g of Na<sub>2</sub>HPO<sub>4</sub> (or 0.161 g of Na<sub>2</sub>HPO<sub>4</sub>•7H<sub>2</sub>O) in 360 ml of ddH<sub>2</sub>O. Adjust pH to 6.96 with HCl at room temperature (RT) before adjusting volume to 400 ml with ddH<sub>2</sub>O. Passage solution through a 0.2- $\mu$ m filter and prepare 5 ml aliquots stored at –20°C.
7. 2.5 M CaCl<sub>2</sub>: dissolve 13.5 g of CaCl<sub>2</sub> • 6H<sub>2</sub>O (or 7.35 g of CaCl<sub>2</sub> • 2H<sub>2</sub>O) in 20 ml of ddH<sub>2</sub>O. Filter sterilize (0.2  $\mu$ m) and prepare 1 ml aliquots stored at –20°C.
8. Flow cytometer, such as FacsCalibur (BD Bioscience, Heidelberg, Germany), and software for analysis, such as CellQuest (BD Bioscience).

## 3. Methods

### 3.1. Generation of ZFNs

1. Design zinc finger sequence and the corresponding oligonucleotides: To determine the sequence that encodes a specific zinc finger DNA-binding domain recognizing your chosen target sequence, visit either one of two websites: <http://www.zincfingertools.org> [(23) Carlos Barbas' lab at The Scripps Research



Institute in La Jolla, CA, USA) or <http://bindr.gdcb.iastate.edu/ZiFiT/> (Dobbs & Voytas laboratories at Iowa State University, Ames, IA, USA) (see **Note 4**).

2. Generate the sequence coding for the three zinc finger modules by assembly PCR, in which the overlapping oligonucleotides are annealed and the gaps between them closed (see **Fig. 1A**). Set up the PCR by mixing the following ingredients (keep all solutions on ice): 1  $\mu\text{l}$  of each oligo (C1, C2, C3, V1, V2, and V3 at 25 pmol/ $\mu\text{l}$ ); for sequence information see **Subheading 2.1.**; see **Note 5**), 0.5  $\mu\text{l}$  of dNTPs (10 mM), 2.5  $\mu\text{l}$  10 $\times$  ThermoPol-buffer, 0.125  $\mu\text{l}$  of Vent polymerase (2 U/ $\mu\text{l}$ ), and 15.9  $\mu\text{l}$  of ddH<sub>2</sub>O (end volume is 25  $\mu\text{l}$ ). Run PCR using the following cycling conditions: 1 cycle at 95°C for 2 min; 12 cycles at 95°C/56°C/72°C, for 30 s each segment; 1 cycle at 72°C for 5 min; keep at 4°C until further use.
3. Set up the PCR for the second engineering step, in which the product of **step 1** is used as a template for the amplification PCR (see **Fig. 1A**): 1  $\mu\text{l}$  of the assembly PCR product, 0.75  $\mu\text{l}$  of each primer (ZF-F and ZF-R at 10 pmol/ $\mu\text{l}$ ); for sequence information see **Subheading 2.1.**; see **Note 6**), 0.5  $\mu\text{l}$  of dNTPs (10 mM), 2.5  $\mu\text{l}$  of 10 $\times$  ThermoPol-buffer, 0.125  $\mu\text{l}$  of Vent polymerase (2 U/ $\mu\text{l}$ ) and 19.4  $\mu\text{l}$  of ddH<sub>2</sub>O (end volume is 25  $\mu\text{l}$ ). Run PCR using the following cycling conditions: 1 cycle at 95°C for 2 min; 25 cycles at 95°C/56°C/72°C, for 30 s each segment; 1 cycle at 72°C for 5 min; last cycle at 4°C until further use.
4. Pour a TAE-buffered 1.5% agarose gel containing 1  $\mu\text{g}/\text{ml}$  ethidium bromide. Add 5  $\mu\text{l}$  of 6 $\times$  loading dye to the PCR product and load on gel. Run gel at 80 V for 40 min and isolate the 280-bp fragment from the gel. Extract the DNA using a DNA extraction kit and elute DNA in 30  $\mu\text{l}$  of 10 mM Tris-HCl pH 8.0.
5. Digest 20  $\mu\text{l}$  of the eluted DNA and a respective expression vector (pMAL.ZFN or pPGK.ZFN) with *AgeI* and *XhoI* (see **Fig. 2B and 2C**) according standard procedures. Ligate the digested insert and backbone and use the ligation product to transform XL1-Blue cells (see **Note 7**).

### 3.2. In Vitro Nuclease Assay

1. Clone the assembled zinc fingers from **Subheading 3.1** into the pMAL.ZFN prokaryotic expression vector and place the zinc fingers in-frame between an N-terminal MBP and the C-terminal *FokI* cleavage domain (see **Fig. 1B**). This plasmid is used to transform *E. coli* ER2508 cells.
2. Inoculate a 5-ml culture in LB/carbenicillin (50  $\mu\text{g}/\text{ml}$ ) and incubate overnight in a 37°C shaker.
3. Back dilute 2.5 ml of the overnight culture into 50 ml LB/carbenicillin (50  $\mu\text{g}/\text{ml}$ ) in a 250-ml flask. Incubate in a 37°C shaker until  $A_{600} \approx 0.7$  OD ( $\sim 2$  h).
4. Induce ZFN expression by adding 30  $\mu\text{l}$  of 0.5 M IPTG (0.3 mM final) and 50  $\mu\text{l}$  of 100 mM ZnCl<sub>2</sub> (100  $\mu\text{M}$  final). Continue incubation in a 37°C shaker for 2 h.
5. Transfer the culture to a 50-ml conical tube. Pellet the cells in a bench-top centrifuge at 1900  $\times g$  (Beckman CS-6R) for 20 min.

6. Discard the supernatant as media waste. If desired, the pellets can be frozen and stored at  $-80^{\circ}\text{C}$  for several days. Thaw at RT.
7. Resuspend the pellet in 5 ml Column Buffer (*see Note 8*). The samples should be kept on ice during all subsequent steps.
8. Sonicate on ice 2 min at 50% duty cycle (10 s pulse, 10 s pause).
9. Pellet insoluble debris in high-speed centrifuge at  $44,000 \times g$  (Beckman J2-21M in JA20 rotor) for 30 min at  $4^{\circ}\text{C}$ . Alternatively, split sample into five 1.5-ml microcentrifuge tubes and centrifuge at  $21,000 \times g$  (Eppendorf Refrigerated Microcentrifuge 5417R) for 15 min at  $4^{\circ}\text{C}$ .
10. Filter the supernatant with a 0.2- $\mu\text{m}$  syringe filter.
11. Prepare a chromatography column with 3 ml of amylose resin. This should produce a packed resin bed of approximately 1.5 ml. Do not let the column run dry at any step. Use of a flow cap or valve is recommended for this purpose. We typically perform **steps 11–15** at RT for convenience, but a cold room can be used if desired.
12. Equilibrate the resin with 10 vol (15 ml) of Column Buffer.
13. Apply the filtered supernatant from **step 10** to the column.
14. Wash with 10 vol (15 ml) of Column Buffer.
15. Elute in  $10 \times 0.5$  ml of Column Buffer/10 mM maltose. Protein should elute in fractions 3–10. Add 0.25 ml glycerol (30% final) and store at  $-20^{\circ}\text{C}$ .
16. Analyze 10  $\mu\text{l}$  of the fractions by using SDS–PAGE. Samples should be compared to BSA standards of 0.5, 1, 2 and 4  $\mu\text{g}$  (i.e., 10  $\mu\text{l}$  of stocks) (*see Note 9*).
17. In a 30- $\mu\text{l}$  reaction volume, prepare  $1 \times$  Reaction Buffer, 1  $\mu\text{g}$  ( $\approx 200$  fmol) of linear DNA substrate and approximately 50 ng ( $\approx 500$  fmol) of full-length (80 kDa) purified protein. It is often useful to examine a range of protein concentrations.
18. Incubate at RT ( $25^{\circ}\text{C}$ ) for 15 min.
19. Add 1.5  $\mu\text{l}$  of 200 mM  $\text{MgCl}_2$  (final = 10 mM) to initiate cleavage.
20. Incubate at  $37^{\circ}\text{C}$  for 30 min.
21. Add 5  $\mu\text{l}$  of  $5 \times$  Loading Buffer and analyze the reactions on a 1% agarose gel containing 0.5  $\mu\text{g}/\text{ml}$  ethidium bromide (*see Fig. 2B*).

### 3.3. Cell-Based Recombination Assay

1. Clone the PCR product from **Subheading 3.1** into plasmid pPGK.ZFN, placing the zinc finger DNA-binding domain in-frame between an N-terminal HA tag and the C-terminal *FokI* cleavage domain (*see Fig. 1B*).
2. Clone two antiparallel oligonucleotides encoding the zinc finger-binding site into the TP. To this end, digest 100 ng TP (*see Fig. 2C*) with 1 unit each of *PacI* and *AgeI* in a total volume of 20  $\mu\text{l}$  for 2 h at  $37^{\circ}\text{C}$  (*see Note 10*). Then, heat inactivate the restriction enzymes for 20 min at  $80^{\circ}\text{C}$ . In the meantime, mix the two oligonucleotides: 1  $\mu\text{l}$  each of ZFBS-F and ZFBS-R (100 pmol/ $\mu\text{l}$  each) and 98  $\mu\text{l}$  of 150 mM NaCl. Heat the mix to  $95^{\circ}\text{C}$  for 5 min, then allow the heat block to slowly cool to RT (about 60 min), which permits efficient annealing of the two oligonucleotides.

3. Set up ligation by mixing 5  $\mu$ l of the digested plasmid with 1  $\mu$ l of double-stranded oligos, 4  $\mu$ l ddH<sub>2</sub>O, 1.2  $\mu$ l 10 $\times$  ligase buffer and 0.8  $\mu$ l T4 DNA ligase (400 U/ $\mu$ l). Incubate for 15 min at RT and use 5  $\mu$ l to transform 45  $\mu$ l of chemically competent XL1-Blue cells. Successful cloning should be confirmed by *Dra*I digest (see **Note 3**), followed by sequencing.
4. For the cell-based recombination assay, seed 75,000 HEK293T cells in 0.5-ml culture medium into each well of a 24-well plate and incubate cells for 24 h.
5. Replace culture medium carefully with 0.5 ml fresh medium 1 h before transfection and prepare transfection mixture in Eppendorf tube. Mix 1.5  $\mu$ g plasmid DNA (20 ng TP; 1.0  $\mu$ g RP; 0.5  $\mu$ g pPGK.ZFN; see **Note 11**) with 22.5  $\mu$ l of H<sub>2</sub>O, 2.5  $\mu$ l of 2.5 M CaCl<sub>2</sub> and 25  $\mu$ l of 2 $\times$  BBS buffer. Mix transfection solution gently by pipetting up and down and incubate for 10 min at RT. Then, overlay cells carefully drop by drop with the mixture and incubate the culture overnight in the CO<sub>2</sub> incubator.
6. The next day, aspirate the transfection medium carefully from cells and overlay cells with 0.5 ml of fresh tissue culture medium. Place the cells in the CO<sub>2</sub> incubator for another 24 h.
7. The next day, aspirate medium and wash cells carefully with 1 ml PBS. Remove all PBS, add 80  $\mu$ l of trypsin solution and leave in CO<sub>2</sub> incubator until cells detach. Add 400  $\mu$ l of PBS supplemented with 20% FBS, resuspend the cells thoroughly by pipetting up and down several times, transfer cell suspension to a FACS tube (BD Bioscience) and keep tubes on ice until analysis.
8. Set up your flow cytometer (see **Note 12**).
9. Count 50,000 cells and evaluate the number of EGFP- and DsRedExpress-positive cells as shown in **Fig. 2D**. The fraction of transfected cells in which an HR event took place is calculated by dividing the number of EGFP-positive cells by the number of DsRedExpress-positive cells. An example for the calculation is given for the right panel: 15.3% (green cells)/50.4% (red cells)  $\times$  100 = 30.3% (corrected cells).

#### 4. Notes

1. A relatively low degree of purification [high purity salt free (HPSF)] is sufficient to create zinc finger domains by the method described. We were very cautious when designing the overlapping sequences between the constant and the variable regions and between the different zinc finger modules in order to make sure that the PCR assembly strategy is as robust and as error-free as possible.
2. This detailed protocol is based on the general protocol provided in the New England BioLabs Protein Expression and Purification Kit manual.
3. Subcloning of the two annealed oligonucleotides into TP generates an additional *Dra*I site (see **Fig. 2C**), which can be used for screening mini-prep DNA.
4. The assumption that zinc finger motifs are functionally independent subunits is somewhat of an oversimplification. Cooperativity in zinc finger–DNA interaction and recognition of a fourth base in the target sequence (target-site-overlap)

can lead to the generation of DNA-binding domains with both low affinity and low specificity. Such issues can be addressed by elaborate selection strategies (24–28), which are, however, beyond the scope of this chapter. Generally, zinc finger DNA-binding motifs recognizing 5'-GNNGNNGNN target sites tend to perform with sufficient affinity and specificity (3,5). Further information on the zinc finger technology in general can be found on our website (<http://www.charite.de/cathomen/>) or the website of The Zinc Finger Consortium (<http://www.zincfingers.org/>), which has been established to promote the development of the engineered zinc fingers.

5. If possible, the variable region encoding the recognition helix (shown as “n” in **Subheading 2.1**) should be designed such that an analytic recognition site for a restriction endonuclease is introduced. This greatly facilitates restriction analysis and later discrimination between different zinc finger domains.
6. To recombine the PCR-amplified zinc finger cassette with a desired expression plasmid, primer ZF-F contains an *XhoI* recognition site whereas primer ZF-R comprises an *AgeI* site.
7. Every newly engineered zinc finger cassette was scrutinized in the following order: After a thorough restriction analysis, expression of the ZFN was assessed by western blot analysis. Positive clones were then verified by sequencing. Using the described method, about 90% of mini-preps tested positive in the restriction analysis. About 75% of these minis were subsequently positively evaluated for ZFN expression by western blot analysis using an anti-HA antibody (Novus Biologicals, Littleton, CO, USA No. NB600-363) and about 2/3 of clones sequenced contained the correct sequence.
8.  $Mg^{2+}$  is required by the *FokI* cleavage domain for activity. The purification is performed without  $Mg^{2+}$  so that the protein can be stored without loss of activity.  $Mg^{2+}$  is added in **step 19** before the cleavage reaction.
9. Often, the proteins are only partially purified by this protocol. However, this level of purification is typically sufficient to assess the binding and cleavage activity.
10. Alternatively, the oligonucleotides encoding the ZFN target site can be designed to be inserted into TP through *BamHI* and *AgeI*. This, however, removes the binding site for *I-SceI*, which usually serves as a positive control in the cell-based recombination assay.
11. Plasmid pRK5.LHA-Sce1 (29), which codes for the yeast homing endonuclease *I-SceI*, can be used instead of pPGK.ZFN as a positive control, whereas pCMV.Luc (5) encoding luciferase can be used as a negative control (*see Fig. 2C–D*).
12. For a FacsCalibur (BD Bioscience), we typically use the following settings for Voltage and AmpGain: FSC (E-1/4.40), SSC (270/6.23), FL1 (500/1.00), and FL2 (445/1.00). Compensations between the green and red channels are FL1–1.2% FL2 and FL2–34.5% FL1 (FSC, Forward Scatter; SSC, Side Scatter; FL1, green channel; FL2, red channel). You may have to adjust these numbers to some extent, depending on the age of your laser and/or the batch of HEK293T cells you are using in your laboratory.

## Acknowledgments

The authors thank Shamim H. Rahman for discussions and careful reading of the manuscript. This chapter is based on work supported by grants CA311/1 from the German Research Foundation (T.C.) and CA103651 from the National Cancer Institute, NIH (D.J.S.).

## References

1. Vasquez, K. M., Marburger, K., Intody, Z., and Wilson, J. H. (2001) Manipulating the mammalian genome by homologous recombination. *Proc Natl Acad Sci USA* **98**:8403–8410.
2. Urnov, F. D., Miller, J. C., Lee, Y. L., Beausejour, C. M., Rock, J. M., Augustus, S., Jamieson, A. C., Porteus, M. H., Gregory, P. D., and Holmes, M. C. (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**:646–651.
3. Porteus, M. H. (2006) Mammalian gene targeting with designed zinc finger nucleases. *Mol Ther* **13**:438–446.
4. Porteus, M. H. and Baltimore, D. (2003) Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**:763.
5. Alwin, S., Gere, M. B., Guhl, E., Effertz, K., Barbas, C. F., 3rd, Segal, D. J., Weitzman, M. D., and Cathomen, T. (2005) Custom zinc-finger nucleases for use in human cells. *Mol Ther* **12**:610–617.
6. Beumer, K., Bhattacharyya, G., Bibikova, M., Trautman, J. K., and Carroll, D. (2006) Efficient gene targeting in *Drosophila* with zinc-finger nucleases. *Genetics* **172**:2391–2403.
7. Wright, D. A., Townsend, J. A., Winfrey, R. J., Jr., Irwin, P. A., Rajagopal, J., Lonosky, P. M., Hall, B. D., Jondle, M. D., and Voytas, D. F. (2005) High-frequency homologous recombination in plants mediated by zinc-finger nucleases. *Plant J* **44**:693–705.
8. Kim, Y. G., Cha, J., and Chandrasegaran, S. (1996) Hybrid restriction enzymes: zinc finger fusions to FokI cleavage domain. *Proc Natl Acad Sci USA* **93**:1156–1160.
9. Durai, S., Mani, M., Kandavelou, K., Wu, J., Porteus, M. H., and Chandrasegaran, S. (2005) Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Res* **33**:5978–5990.
10. Pavletich, N. P. and Pabo, C. O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252**:809–817.
11. Rebar, E. J. and Pabo, C. O. (1994) Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* **263**:671–673.
12. Choo, Y. and Klug, A. (1994) Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. *Proc Natl Acad Sci USA* **91**:11168–11172.

13. Jamieson, A. C., Wang, H., and Kim, S. H. (1996) A zinc finger directory for high-affinity DNA recognition. *Proc Natl Acad Sci USA* **93**:12834–12839.
14. Wu, H., Yang, W. P., and Barbas, C. F., 3rd. (1995) Building zinc fingers by selection: toward a therapeutic application. *Proc Natl Acad Sci USA* **92**: 344–348.
15. Liu, Q., Xia, Z., Zhong, X., and Case, C. C. (2002) Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *J Biol Chem* **277**: 3850–3856.
16. Segal, D. J., Dreier, B., Beerli, R. R., and 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 USA* **96**:2758–2763.
17. Dreier, B., Beerli, R. R., Segal, D. J., Flippin, J. D., and Barbas, C. F., 3rd. (2001) Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem* **276**:29466–29478.
18. Dreier, B., Fuller, R. P., Segal, D. J., Lund, C. V., Blancafort, P., Huber, A., Koksich, B., and Barbas, C. F., 3rd. (2005) Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem* **280**:35588–35597.
19. Blancafort, P., Magnenat, L., and Barbas, C. F., 3rd. (2003) Scanning the human genome with combinatorial transcription factor libraries. *Nat Biotechnol* **21**: 269–274.
20. Smith, J., Bibikova, M., Whitby, F. G., Reddy, A. R., Chandrasegaran, S., and Carroll, D. (2000) Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res* **28**: 3361–3369.
21. Bitinaite, J., Wah, D. A., Aggarwal, A. K., and Schildkraut, I. (1998) FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci USA* **95**:10570–10575.
22. Bibikova, M., Carroll, D., Segal, D. J., Trautman, J. K., Smith, J., Kim, Y. G., and Chandrasegaran, S. (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol* **21**:289–297.
23. Mandell, J. G. and Barbas, C. F., 3rd. (2006) Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res* **34**:W516–523.
24. Greisman, H. A. and Pabo, C. O. (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* **275**:657–661.
25. Isalan, M., Klug, A., and Choo, Y. (2001) A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nat Biotechnol* **19**:656–660.
26. Cheng, X., Boyer, J. L., and Juliano, R. L. (1997) Selection of peptides that functionally replace a zinc finger in the Sp1 transcription factor by using a yeast combinatorial library. *Proc Natl Acad Sci USA* **94**:14120–14125.
27. Hughes, M. D., Zhang, Z. R., Sutherland, A. J., Santos, A. F., and Hine, A. V. (2005) Discovery of active proteins directly from combinatorial randomized protein

- libraries without display, purification or sequencing: identification of novel zinc finger proteins. *Nucleic Acids Res* **33**:e32.
28. Hurt, J. A., Thibodeau, S. A., Hirsh, A. S., Pabo, C. O., and Joung, J. K. (2003) Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc Natl Acad Sci USA* **100**:12271–12276.
  29. Radecke, F., Peter, I., Radecke, S., Gellhaus, K., Schwarz, K., and 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.



<sup>1</sup>Laboratory for Cell Function and Dynamics, Brain Science Institute, RIKEN, 2-1 Hirosawa, Wako-city, Saitama 351-0198, Japan. <sup>2</sup>Life Function and Dynamics, Exploratory Research for Advanced Technology, Japan Science and Technology Agency, 2-1 Hirosawa, Wako-city, Saitama 351-0198, Japan. e-mail: matsushi@brain.riken.jp

1. Denk, W. *et al. Science* **248**, 73–76 (1990).
2. Zipfel, W.R. *et al. Nat. Biotechnol.* **21**, 1369–1377 (2003).
3. Shaner, N.C. *et al. Nat. Methods* **2**, 905–909 (2005).
4. Kogure, T. *et al. Nat. Biotechnol.* **24**, 577–581 (2006).

## Unexpected failure rates for modular assembly of engineered zinc fingers

**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<sup>1</sup> (**Supplementary Fig. 1** online). We found 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 Fig. 2** online). Two large-scale surveys have suggested that modular assembly is highly effective (100% (ref. 2) and 60% (ref. 3)

success rates) for making three-finger arrays designed to bind 9-bp target sites. The Zinc Finger Consortium recently assembled an archive of 141 previously published finger modules<sup>3–5</sup> encoded on a standardized platform<sup>6</sup>. Our initial experiences using these reagents suggested that modular assembly was inefficient (**Supplementary Discussion** and **Supplementary Table 1** online).

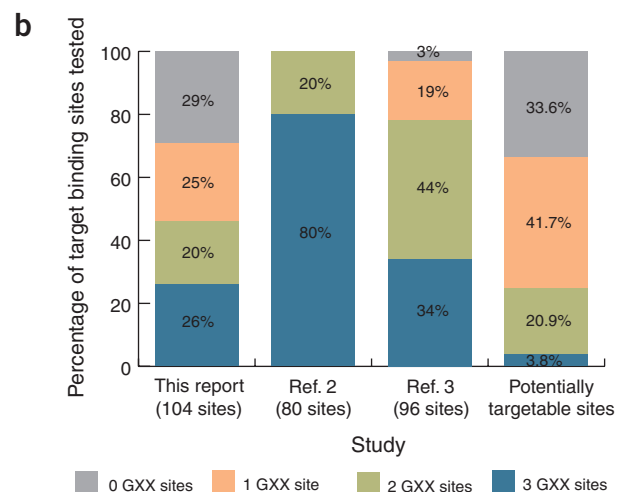
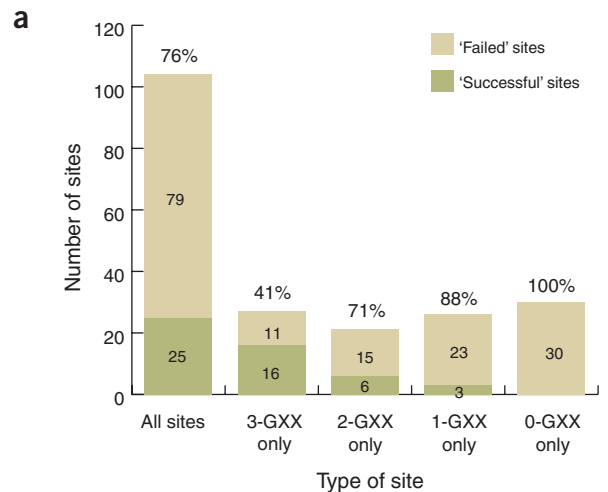
To perform a larger-scale test, we assembled 168 zinc-finger arrays designed for 104 diverse target DNA sites (**Supplementary Table 2** and **Supplementary Methods** online). We tested these domains for DNA binding using a bacterial two-hybrid (B2H) assay<sup>6</sup>, which accurately identifies arrays that lack activity as ZFNs in human cells (**Supplementary Discussion** and **Supplementary Figs. 3** and **4** online). For 79 of the 104 target sites, we did not obtain a single three-finger array that scored positively in the B2H assay (overall failure rate of ~76%; **Fig. 1a** and **Supplementary Table 2**). Notably, modular assembly 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 (**Fig. 1a**). Additionally, because ZFNs function as dimers, we would expect failure rates for making a functional ZFN pair to be even higher (**Supplementary Discussion**). Notably, these values are all likely underestimates of actual failure rates because not all zinc-finger arrays that are positive in the B2H assay will be active as ZFNs in human cells (**Supplementary Fig. 4**).



**Figure 1** | Large-scale evaluation of the modular assembly method for engineering zinc-finger arrays. **(a)** Data for success and failure of modular assembly as judged by the B2H assay for all 104 target DNA sites (see **Supplementary Discussion** for definition of success threshold) and for subsets of target sites containing three, two, one or no GXX subsites. 'Successful' sites are those for which at least one functional zinc-finger array was identified; 'failed' sites are those for which we failed to obtain a single successful array (see **Supplementary Table 2** for details). Predicted failure rates for each set of target sites are indicated above the bars. **(b)** Distributions of target DNA sites used in different modular assembly evaluation studies according to the number of GXX subsites present. 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 (ref. 6) is also shown (far right bar).

One reason for the apparent discrepancy between previously published studies<sup>2,3</sup> and our results is that the former primarily used 9-bp sites composed of two or three GXX subsites, whereas we used sites with a more varied number of GXX subsites (**Fig. 1b**); this difference will critically affect the observed failure rate (**Fig. 1a**). Our study represents a more meaningful evaluation because target sites containing one or no GXX subsites (under-represented in previous studies) encompass the majority (>75%) of the 107,011 potential 9-bp sites that can be targeted with existing modules (**Fig. 1b**).

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



Note: Supplementary information is available on the Nature Methods website.

#### ACKNOWLEDGMENTS

We thank D. Dobbs and P. Zaback for helpful discussions and E. Unger-Wallace (Iowa State University) for contributions to zinc-finger array construction. J.K.J. is supported by the US National Institutes of Health (R21 RR024189), the Cystic Fibrosis Foundation (MCCRAY07G0) and the Massachusetts General Hospital Department of Pathology; D.F.V. is supported by a US National Science Foundation 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.

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

<sup>1</sup>Molecular Pathology Unit, Center for Cancer Research, and Center for Computational and Integrative Biology, 149 13<sup>th</sup> Street, Room 7132, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA.

<sup>2</sup>Biological and Biomedical Sciences Program, Division of Medical Sciences, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>3</sup>Department of Genetics, Development and Cell Biology, Iowa State University, 1035A Roy J. Carver Co-Lab, Ames, Iowa 50011, USA. <sup>4</sup>Charité Medical School, Institute of Virology, Hindenburgdamm 27, D-12203 Berlin, Germany. <sup>5</sup>Interdepartmental Graduate Program in Bioinformatics and Computational Biology, 2114 Molecular Biology Building, Iowa State University, Ames, Iowa 50011, USA. <sup>6</sup>Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>7</sup>These authors contributed equally to this work.

e-mail: jjoung@partners.org

- Porteus, M.H. & Carroll, D. *Nat. Biotechnol.* **23**, 967–973 (2005).
- Segal, D.J. *et al. Biochemistry* **42**, 2137–2148 (2003).
- Bae, K.H. *et al. Nat. Biotechnol.* **21**, 275–280 (2003).
- Mandell, J.G. & Barbas, C.F. *Nucleic Acids Res.* **34**, W516–W523 (2006).
- Liu, Q., Xia, Z., Zhong, X. & Case, C.C. *J. Biol. Chem.* **277**, 3850–3856 (2002).
- Wright, D.A. *et al. Nat. Protoc.* **1**, 1637–1652 (2006).

Correspondence

## **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, Jeffrey D Sander, Fengli Fu, Jeffrey A Townsend, Toni Cathomen, Daniel F Voytas & J Keith Joung

Supplementary figures and text:

**Supplementary Figure 1** ZFNs induce highly efficient gene targeting events.

**Supplementary Figure 2** Schematic illustrating the “modular assembly” method of engineering multi-finger domains.

**Supplementary Figure 3** Schematic of the bacterial two-hybrid (B2H) reporter used to assess DNA-binding activities of zinc finger arrays.

**Supplementary Figure 4** The B2H assay identifies zinc-finger arrays that fail to show significant activity as ZFNs in human cells.

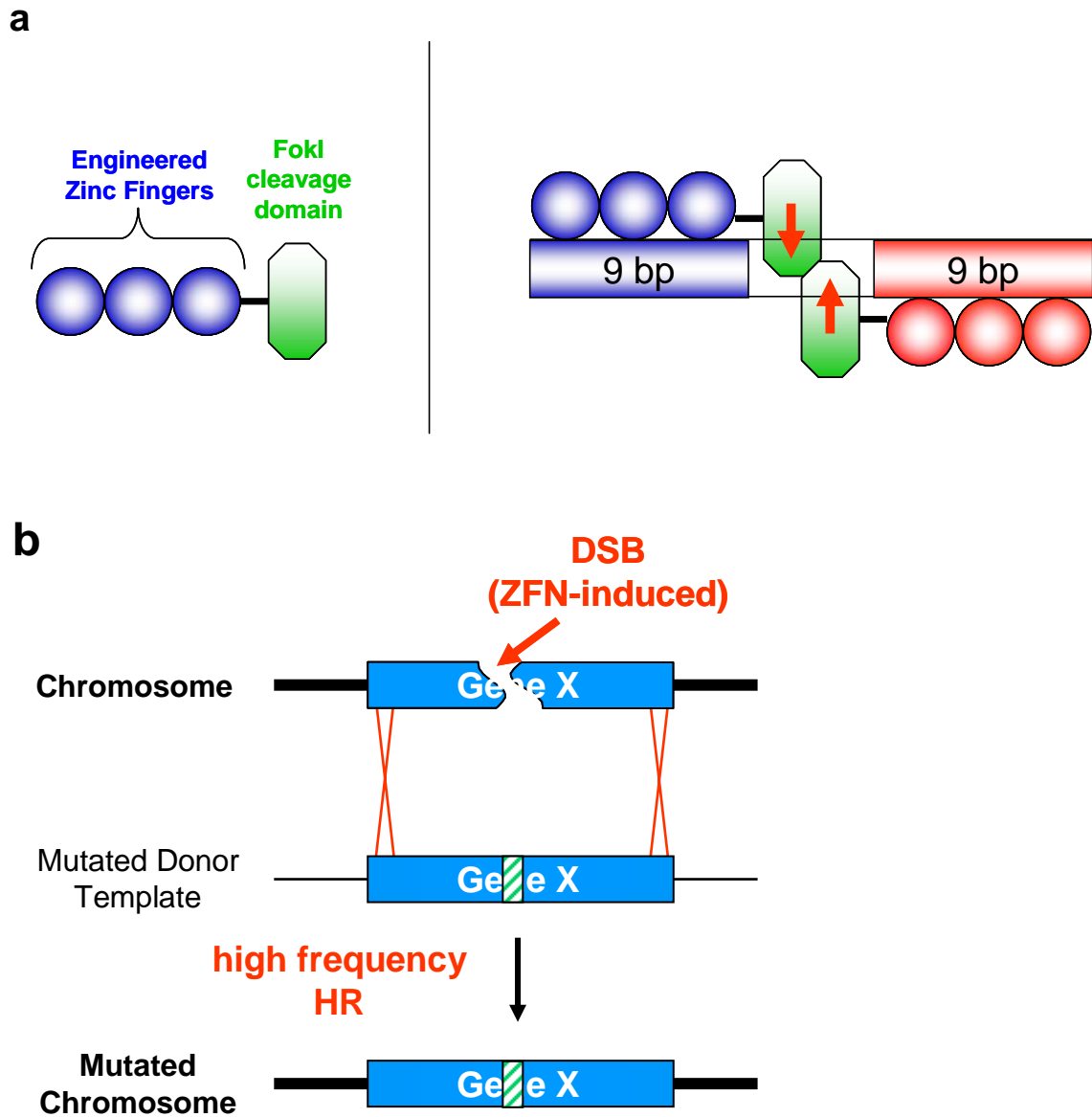
**Supplementary Table 1** Small-scale tests of modular assembly using various activity assays.

**Supplementary Table 2** Modularly assembled zinc finger arrays, cognate target binding sites, and their activities in the B2H assay.

**Supplementary Discussion**

**Supplementary Methods**

**Supplementary Figure 1** ZFNs induce highly efficient gene targeting events



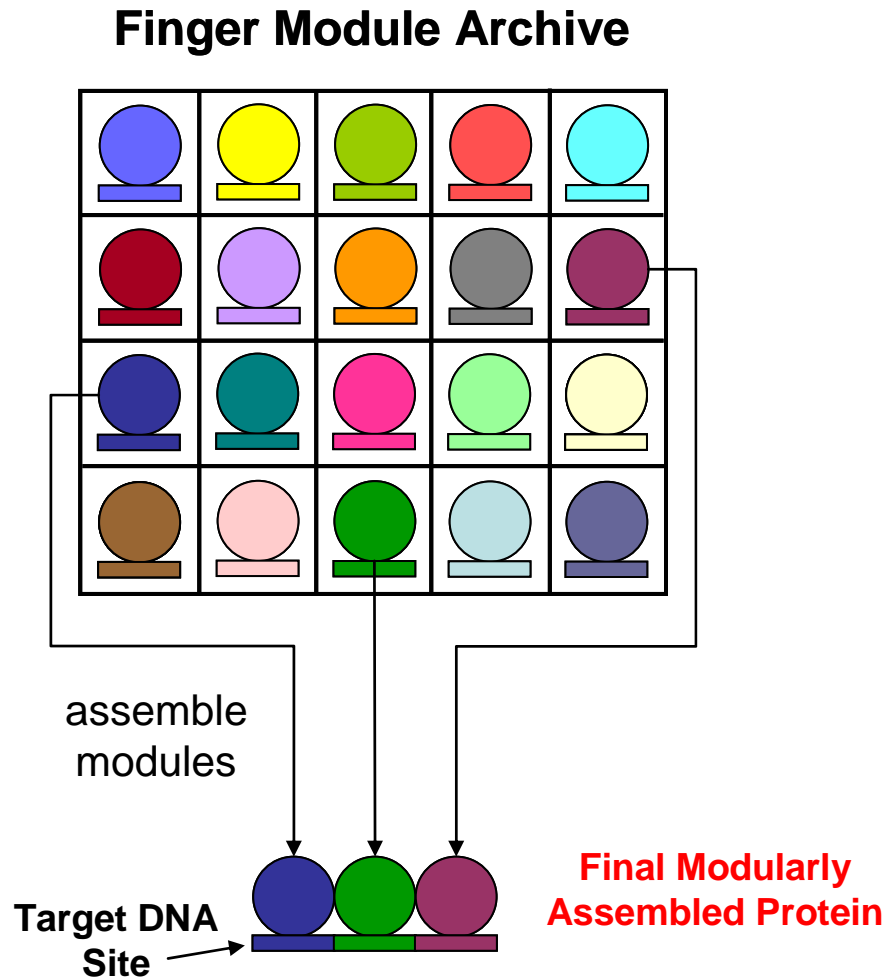
(a) Schematic of a ZFN monomer (left) and a pair of ZFNs cleaving DNA as a dimer (right).

(b) Schematic representing ZFN-enhanced recombination-based genome manipulation.

ZFNs introduce site-specific double-strand DNA breaks (DSBs) that can be harnessed to

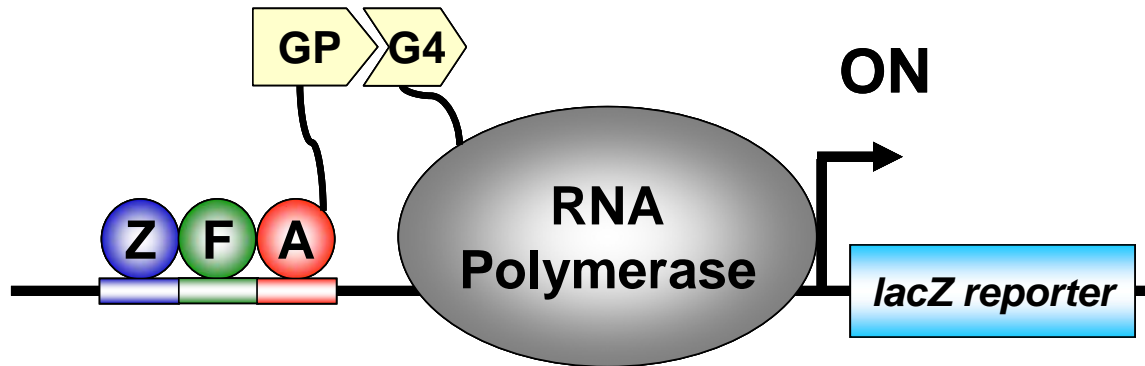
mediate gene targeting via homologous recombination with an exogenously introduced “donor template.”

**Supplementary Figure 2** Schematic illustrating the “modular assembly” method of engineering multi-finger domains



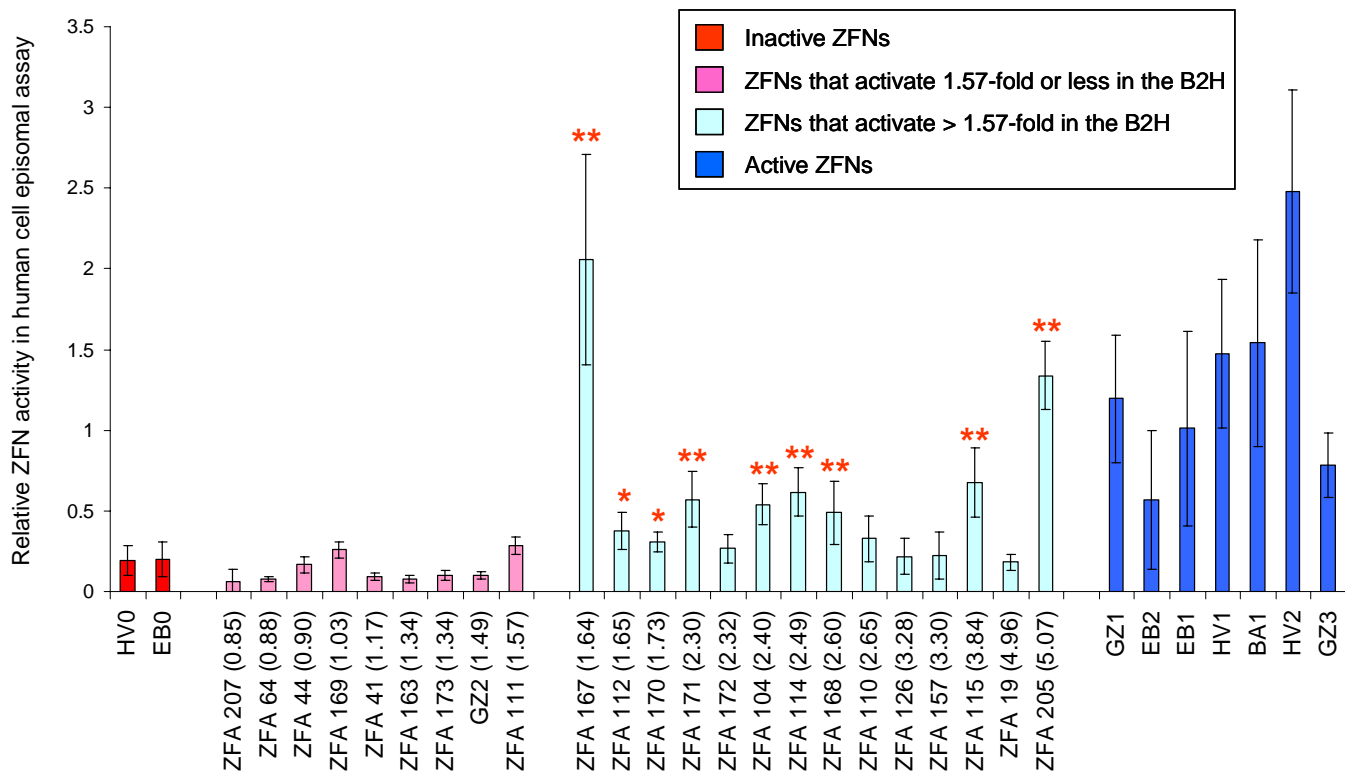
Fingers from archives of pre-characterized modules are joined together to create three-finger domains capable of recognizing a 9 bp target DNA site. Single fingers (colored spheres) recognize their cognate 3 bp “subsites” (colored rectangles).

**Supplementary Figure 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  $\beta$ -galactosidase.<sup>1-3</sup> 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 Gal11 IP protein (GP) and the other a hybrid of the RNA polymerase  $\alpha$ -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 R $\alpha$  hybrid proteins are recruited to the promoter and expression of the downstream *lacZ* reporter is activated.

**Supplementary Figure 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.<sup>1</sup> 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 1 and 2** online. ZFN activity is reported as a percentage of EGFP-rescue relative to an *I-SceI* 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<sup>1</sup> 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.<sup>1</sup> 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 Figure 4 Reference:

1. Cornu, T. I. et al. DNA-binding Specificity Is a Major Determinant of the Activity and Toxicity of Zinc-finger Nucleases. *Mol Ther* **16**, 352-8 (2008).



**Supplementary Table 1** Small-scale tests of modular assembly using various activity assays

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	+

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,<sup>1</sup> 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-stranded 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 **Supplementary Methods** online) 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 **Supplementary Methods** online). A plus sign indicates that ZFN activity in stimulating HR was >40% as compared to the activity of a control I-SceI 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-*SceI* on the same target template. For further information on the activity assays, see **Supplementary Methods** online.

\*\*Finger arrays with backbone sequences that differ from the ZF finger archive.

\*\*\*Finger arrays previously described.<sup>2</sup>

Reference for Supplementary Table 1:

1. Wright, D. A. et al. Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat Protoc* **1**, 1637-52 (2006).
2. Alwin, S. et al. Custom zinc-finger nucleases for use in human cells. *Mol Ther* **12**, 610-7 (2005).

**Supplementary Table 2** Modularly assembled zinc finger arrays, cognate target binding sites, and their activities in the B2H assay

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

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

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

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

See legend to **Supplementary Table 1** online 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,<sup>1</sup> S=Sangamo Biosciences,<sup>2,3</sup> and T=Toolgen, Inc<sup>4</sup>.) is indicated. Because the Zinc Finger Consortium Modular Assembly Kit 1.0 includes more than one module for certain subsites,<sup>3</sup> 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 **Supplementary Methods** online. Fold-activation of transcription in the B2H assay was determined for each zinc finger array using  $\beta$ -galactosidase assays as previously described.<sup>5</sup>  $\beta$ -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).

References for Supplementary Table 2:

1. Mandell, J.G. & Barbas, C.F., 3rd Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res* **34**, W516-523 (2006).

2. Liu, Q., Xia, Z., Zhong, X. & Case, C.C. Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *J Biol Chem* **277**, 3850-3856 (2002).
3. Wright, D.A. et al. Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat Protoc* **1**, 1637-1652 (2006).
4. Bae, K.H. et al. Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat Biotechnol* **21**, 275-280 (2003).
5. Cornu, T.I. et al. DNA-binding Specificity Is a Major Determinant of the Activity and Toxicity of Zinc-finger Nucleases. *Mol Ther* (2007).



### **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 1** online. 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 1** online), 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.<sup>1</sup> To further assess this correlation, we tested the activities of 23 ZFNs using a human cell-based episomal homologous recombination (HR) assay (**Supplementary Methods** online). 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 Fig. 4** online, 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 Fig.**

4 online). 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 Fig. 4** online). 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; **Fig. 1a**) compared with those that contain three GXX subsites (41% failure rate; **Fig. 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.<sup>2-5</sup> 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.<sup>6</sup> Isalan, Klug, and Choo described a “bi-partite” 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.<sup>7</sup> 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.<sup>8</sup> 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.

#### References for Supplementary Discussion:

1. Cornu, T. I. et al. DNA-binding Specificity Is a Major Determinant of the Activity and Toxicity of Zinc-finger Nucleases. *Mol Ther* (2007).
2. Isalan, M., Choo, Y. & Klug, A. Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proc Natl Acad Sci U S A* **94**, 5617-21 (1997).
3. Isalan, M., Klug, A. & Choo, Y. Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers. *Biochemistry* **37**, 12026-33 (1998).
4. Elrod-Erickson, M., Rould, M. A., Nekludova, L. & Pabo, C. O. Zif268 protein-DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions. *Structure* **4**, 1171-80 (1996).

5. Wolfe, S. A., Grant, R. A., Elrod-Erickson, M. & Pabo, C. O. Beyond the "recognition code": structures of two Cys2His2 zinc finger/TATA box complexes. *Structure (Camb)* **9**, 717-23 (2001).
6. Greisman, H. A. & Pabo, C. O. A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* **275**, 657-61 (1997).
7. Isalan, M., Klug, A. & Choo, Y. A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nat Biotechnol* **19**, 656-60 (2001).
8. Hurt, J. A., Thibodeau, S. A., Hirsh, A. S., Pabo, C. O. & Joung, J. K. Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc Natl Acad Sci U S A* **100**, 12271-6 (2003).

## **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.<sup>1</sup> B2H expression vectors encoding different three-finger arrays were constructed essentially as previously described<sup>1</sup> using the Zinc Finger Modular Assembly Kit 1.0 which includes 141 modules made by the Barbas group, Sangamo Biosciences, and Toolgen, Inc.<sup>1</sup> 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,<sup>2</sup> (2) the Toolgen group discovered that their human zinc fingers worked best with one another and not as well with other engineered modules,<sup>3</sup> 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.<sup>4</sup> 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-Gal1 1P expression vector and a compatible plasmid encoding an *E. coli* RNA polymerase alpha-Gal4 hybrid protein.<sup>1</sup>

Fold-activation was calculated as the ratio of the level of  $\beta$ -galactosidase (*lacZ*) reporter activity in the presence and absence of the zinc finger array as previously described.<sup>5</sup> All  $\beta$ -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  $\beta$ -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  $\beta$ -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  $\beta$ -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  $\mu$ g 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-SceI.

*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.<sup>6</sup> 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.<sup>1</sup> 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  $0.59 \times 0.59 = 0.35$  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.

#### References for Supplementary Methods:

1. Wright, D. A. et al. Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat Protoc* **1**, 1637-52 (2006).
2. Mandell, J. G. & Barbas, C. F., 3rd. Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res* **34**, W516-23 (2006).
3. Bae, K. H. et al. Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat Biotechnol* **21**, 275-80 (2003).
4. Liu, Q., Xia, Z., Zhong, X. & Case, C. C. Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *J Biol Chem* **277**, 3850-6 (2002).
5. Thibodeau, S. A., Fang, R. & Joung, J. K. High-throughput beta-galactosidase assay for bacterial cell-based reporter systems. *Biotechniques* **36**, 410-5 (2004).
6. Sander, J. D., Zaback, P., Joung, J. K., Voytas, D. F. & Dobbs, D. Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucleic Acids Res* **35**, W599-605 (2007).

# Rapid “Open-Source” Engineering of Customized Zinc-Finger Nucleases for Highly Efficient Gene Modification

Morgan L. Maeder,<sup>1,2</sup> Stacey Thibodeau-Beganny,<sup>1,2</sup> Anna Osiak,<sup>4</sup> David A. Wright,<sup>5</sup> Reshma M. Anthony,<sup>7</sup> Magdalena Eichtinger,<sup>1,2,8</sup> Tao Jiang,<sup>1,2,8</sup> Jonathan E. Foley,<sup>1,2</sup> Ronnie J. Winfrey,<sup>5</sup> Jeffrey A. Townsend,<sup>5</sup> Erica Unger-Wallace,<sup>5</sup> Jeffry D. Sander,<sup>5,6</sup> Felix Müller-Lerch,<sup>4</sup> Fengli Fu,<sup>5,6</sup> Joseph Pearlberg,<sup>9</sup> Carl Göbel,<sup>1,2</sup> Justin P. Dassisti,<sup>7</sup> Shondra M. Pruett-Miller,<sup>10</sup> Matthew H. Porteus,<sup>10</sup> Dennis C. Sgroi,<sup>1,8</sup> A. John Iafrate,<sup>3,8</sup> Drena Dobbs,<sup>5</sup> Paul B. McCray, Jr.,<sup>7</sup> Toni Cathomen,<sup>4</sup> Daniel F. Voytas,<sup>5,11</sup> and J. Keith Joung<sup>1,2,8,\*</sup>

<sup>1</sup>Molecular Pathology Unit and Center for Cancer Research, Massachusetts General Hospital, Charlestown, MA 02129, USA

<sup>2</sup>Center for Computational and Integrative Biology

<sup>3</sup>Pathology Service

Massachusetts General Hospital, Boston, MA 02114, USA

<sup>4</sup>Institute of Virology (CBF), Charité Medical School, D-12203 Berlin, Germany

<sup>5</sup>Department of Genetics, Development and Cell Biology

<sup>6</sup>Interdepartmental Graduate Program in Bioinformatics and Computational Biology

Iowa State University, Ames, IA 50011, USA

<sup>7</sup>Department of Pediatrics, Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA

<sup>8</sup>Department of Pathology

<sup>9</sup>Department of Biological Chemistry and Molecular Pharmacology

Harvard Medical School, Boston, MA 02115, USA

<sup>10</sup>Department of Pediatrics, UT Southwestern Medical Center, Dallas, TX 75390, USA

<sup>11</sup>Present address: Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, MN 55455, USA

\*Correspondence: [jjoung@partners.org](mailto:jjoung@partners.org)

DOI 10.1016/j.molcel.2008.06.016

## SUMMARY

Custom-made zinc-finger nucleases (ZFNs) can induce targeted genome modifications with high efficiency in cell types including *Drosophila*, *C. elegans*, plants, and humans. A bottleneck in the application of ZFN technology has been the generation of highly specific engineered zinc-finger arrays. Here we describe OPEN (Oligomerized Pool ENGINEERING), a rapid, publicly available strategy for constructing multifinger arrays, which we show is more effective than the previously published modular assembly method. We used OPEN to construct 37 highly active ZFN pairs which induced targeted alterations with high efficiencies (1%–50%) at 11 different target sites located within three endogenous human genes (*VEGF-A*, *HoxB13*, and *CFTR*), an endogenous plant gene (tobacco *SuRA*), and a chromosomally integrated *EGFP* reporter gene. In summary, OPEN provides an “open-source” method for rapidly engineering highly active zinc-finger arrays, thereby enabling broader practice, development, and application of ZFN technology for biological research and gene therapy.

## INTRODUCTION

A reliable, publicly available method for creating targeted genome modifications with high efficiency would be broadly useful for bio-

logical research and gene therapy (Cathomen and Joung, 2008). Although strategies have been developed using viral vectors or short double- or single-stranded DNA molecules to direct homologous recombination (HR), these methods generally have low efficiencies even under optimal cell culture conditions (Cathomen, 2004). An alternative and more efficient strategy involves introducing a double-stranded break (DSB) into a target locus of interest. Repair of this DSB by cellular mechanisms greatly increases rates of HR-mediated gene targeting with an exogenous DNA molecule (Jasin, 1996) or of gene mutation by error-prone nonhomologous end joining (NHEJ) (Bibikova et al., 2002). However, this approach depends on the capability to create “designer” nucleases targeted to specific genomic sequences of interest.

Engineered ZFNs are customized endonucleases that induce site-specific DSBs and genome modifications (as high as 50%) in *Drosophila*, somatic *C. elegans*, plant, and human cells (Alwin et al., 2005; Beumer et al., 2006; Bibikova et al., 2002, 2003; Lloyd et al., 2005; Lombardo et al., 2007; Miller et al., 2007; Moehle et al., 2007; Morton et al., 2006; Porteus and Baltimore, 2003; Szczeppek et al., 2007; Urnov et al., 2005; Wright et al., 2005). ZFNs function as dimers with each monomer composed of a nonspecific cleavage domain from the *FokI* endonuclease fused to a zinc-finger array engineered to bind a target DNA sequence of interest (Cathomen and Joung, 2008; Durai et al., 2005; Porteus and Carroll, 2005). Because a single zinc-finger domain binds a 3 bp subsite, a ZFN dimer can recognize 18 or 24 bp target sites, depending on the number of zinc-fingers in each ZFN monomer.

Publicly available zinc-finger engineering methods described in the literature can be grouped into two general categories.

“Modular assembly” involves joining together single fingers with precharacterized specificities (Bae et al., 2003; Beerli and Barbas, 2002; Liu et al., 2002; Mandell and Barbas, 2006; Segal et al., 2003). Although easy to perform, modular assembly has an efficacy rate for making functional ZFN pairs that is less than 6% (Ramirez et al., 2008) and can yield ZFNs with low activities and/or high toxicities (Cornu et al., 2008; Pruett-Miller et al., 2008). Alternative approaches involve combinatorial selection-based methods that yield multifinger domains possessing high DNA-binding affinities and specificities (Greisman and Pabo, 1997; Hurt et al., 2003; Isalan et al., 2001) and high activities and low toxicities when expressed as ZFNs in human cells (Cornu et al., 2008; Pruett-Miller et al., 2008). However, selection-based methods require construction and interrogation of large randomized libraries (typically  $>10^8$  in size) and, therefore, remain intractable for all but a few labs that possess the required expertise.

Here we describe the development and validation of OPEN (Oligomerized Pool ENgineering), a facile, robust, and publicly available platform for engineering zinc-finger arrays. OPEN is enabled by an archive of zinc-finger pools constructed by the Zinc Finger Consortium, a group of academic laboratories committed to developing engineered zinc-finger technology (<http://www.zincfingers.org>). The Consortium used OPEN to rapidly engineer 37 ZFN pairs that induce modifications at 11 different sites located within three endogenous human genes (*VEGF-A*, *HoxB13*, and *CFTR*), an endogenous plant gene (tobacco *SuRA*), and the *EGFP* reporter gene with efficiencies ranging from 1%–50%. The publicly available OPEN platform will enable routine practice and further development of ZFN technology.

## RESULTS

### OPEN: A Rapid and Robust Strategy for Engineering Zinc-Finger Arrays

OPEN selections require an archive of preselected zinc-finger pools, each containing a mixture of fingers targeted to a different three base pair subsite at a defined position within a three-finger protein (Figure 1A and Experimental Procedures). Fully enabling OPEN will require 192 finger pools (64 potential three bp target subsites for each position in a three-finger protein). In this initial report, we created pools (each containing a maximum of 95 different fingers) targeted to 66 subsites (48 GXX subsites and 18 TXX subsites; Figure 1B). To perform an OPEN selection for a target site, appropriate finger pools from the archive are recombined to create a small library of variants ( $95^3 = 8.6 \times 10^5$  members for a three-finger domain) that is interrogated using a bacterial two-hybrid (B2H) selection system in which binding of a zinc-finger domain to its cognate site activates expression of selectable marker genes (Figures 1C and 1D) (Hurt et al., 2003; Jung et al., 2000). To simplify the identification of potential three-finger ZFN sites that can be targeted by OPEN, we created a new version of our web-based ZiFIT software (Sander et al., 2007), ZiFIT 3.0 (Figure S1) (available at <http://www.zincfingers.org/software-tools.htm>).

### Comparing ZFNs Made by Modular Assembly and OPEN

To compare the efficacy of OPEN with modular assembly, we used both strategies to construct multifinger arrays for five sites

in *EGFP* (ten “half-sites”; Figure S2 and Tables S1 and S2 available online) and assessed the DNA-binding activities of these proteins using a quantitative B2H assay (Supplemental Experimental Procedures). Zinc-finger arrays with high affinities and specificities activate transcription by more than 3-fold in the B2H system (Hurt et al., 2003). None of the modularly assembled arrays we tested activated transcription by more than 3-fold (Figure 2A), even though western blots showed that all proteins were expressed (data not shown). By contrast, OPEN selections yielded at least one—in most cases many—zinc-finger protein that activated transcription by more than 3-fold for 9 of the 10 target half-sites (Figure 2A and Table S1).

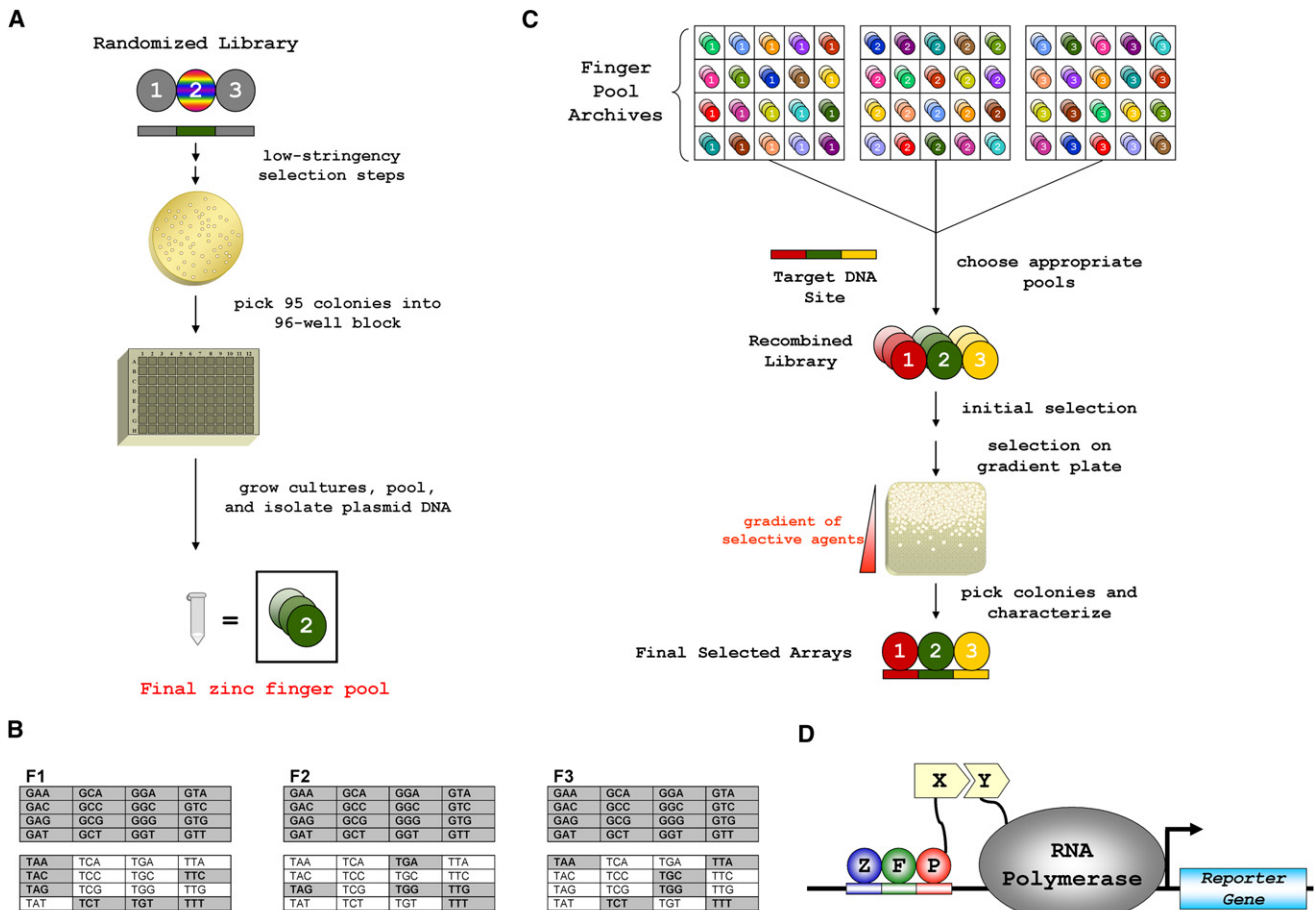
For each of the five full *EGFP* target sites, various pairs of modularly assembled and OPEN-selected ZFNs (Table S2) were tested in human cells using an assay in which repair of ZFN-induced DSBs by error-prone NHEJ leads to insertions and deletions in a chromosomally integrated *EGFP* gene (Figure 2B). Modular assembly yielded ZFN pairs with activities above background for only one of the five sites (EG502) (Figure 2C). By contrast, OPEN yielded ZFN pairs that were active for four of the five full ZFN target sites (EG223, EG292, EG382, and EG502). Although both methods produced active ZFN pairs for the EG502 site, the pairs made by OPEN were more active than those made by modular assembly (Figure 2C). Western blots verified the expression of all ZFNs tested (data not shown).

### OPEN Selection of Zinc-Finger Arrays that Bind to Sequences in Endogenous Human and Plant Genes

We used ZiFIT 3.0 to identify 14 potential ZFN target sites in three endogenous human genes (*VEGF-A*, *HoxB13*, and *CFTR*) and one endogenous plant gene (tobacco *SuRA*) (Table S2 and Figure S2). OPEN selections were performed for the 28 “half sites” within these 14 full ZFN target sites (Tables S1 and S2), and 24 of 28 were deemed successful (Table S1). Using a subset of these finger arrays (Table S2), we constructed ZFN pairs for five sites in *VEGF-A*, four sites in *HoxB13*, one site in *CFTR*, and one site in the *SuRA* gene.

### OPEN ZFNs Induce Highly Efficient Mutation of Endogenous Human and Plant Genes

To test whether OPEN ZFNs induce mutations by NHEJ at endogenous human genes, we employed a CEL I nuclease-based mutation detection assay (Figure 3A) (Lombardo et al., 2007; Miller et al., 2007). We found that all four VF2468 and all four VF2471 ZFN pairs we tested generated detectable mutations in the endogenous *VEGF-A* gene in human 293 cells (Figure 3B). In addition, mutations could be detected at the endogenous *HoxB13* gene in 293 cells for all ZFN pairs tested at the HX587, HX735, and HX761 sites (Figure 3C). DNA sequencing of *HoxB13* alleles amplified from cells modified by HX587 ZFN pair B (Supplemental Experimental Procedures) revealed mutations at the expected location with an average frequency of 9.6% (Figure S3A). ZFN pairs targeted to three sites in the *VEGF-A* gene (VF3537, VF3542, and VF3552) and to one site in the *HoxB13* gene (HX508) failed to induce detectable levels of mutation (Figures 3B and 3C). Due to polymorphisms at the CF877 site in *CFTR* alleles of 293 cells, we could not assess



**Figure 1. OPEN Method for Engineering Zinc-Finger Arrays**

(A) OPEN zinc-finger pool construction. Zinc-finger domains are shown as spheres, and associated 3 bp subsites are shown as rectangles. Randomized finger in the library is rainbow colored. Note that the figure illustrates how finger pools for the middle position in a three-finger domain were made, but that pools for amino- or carboxy-terminal fingers were also obtained by building libraries in which finger 1 or finger 3 were randomized, respectively (Experimental Procedures).

(B) GXX and TXX target subsites for which finger pools have been constructed (highlighted in gray).

(C) Schematic overview of OPEN selection for a target DNA site. Zinc-fingers and associated subsites represented as in (A). Details are in the Supplemental Experimental Procedures.

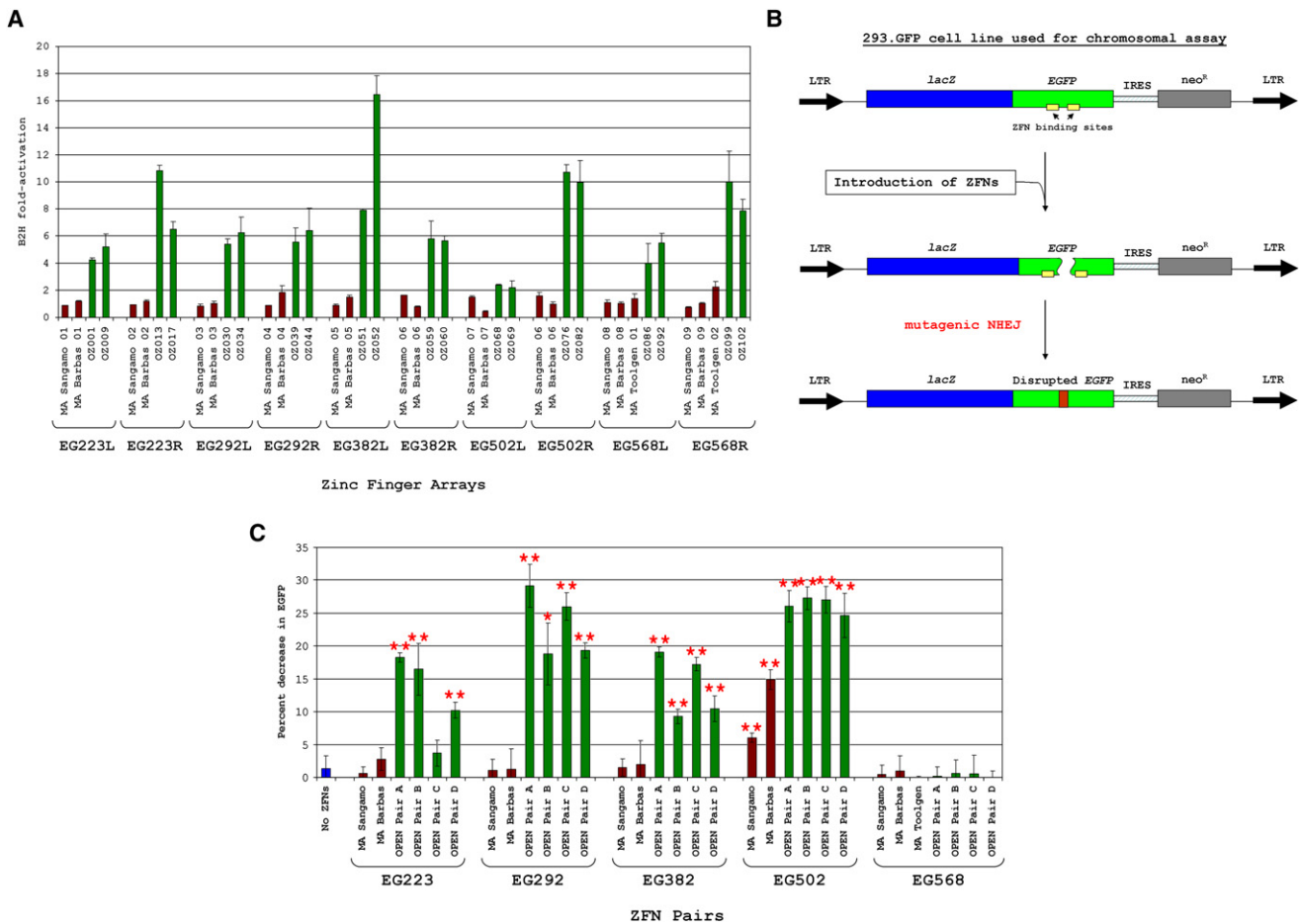
(D) Schematic of the bacterial two-hybrid (B2H) system. ZFP, zinc-finger protein; X and Y, arbitrary interacting proteins.

the CF877 ZFN pairs using the CEL I assay (data not shown). However, DNA sequencing revealed ZFN-induced insertions at the CF877 cleavage site with an average frequency of 1.2% (Figure S3B).

We also tested whether OPEN ZFNs induced mutations in endogenous plant genes. Tobacco protoplasts transformed with a construct encoding the SR2163 ZFN pair were regenerated into individual plants (Experimental Procedures). The SR2163 site is present in both *SuRA* and *SuRB*, and each plant was examined for evidence of cleavage at these loci using DNA sequencing. Among 66 transgenic plants surveyed, three had mutations in *SuRA*, all of which were deletions of a single base (Figure S3C). In one plant, both alleles of *SuRA* had the same deletion. No mutated alleles of *SuRB* were detected. This frequency of mutagenesis by NHEJ (~2% of potential target alleles) is comparable to what we observed for other OPEN ZFNs in human cells.

### OPEN ZFNs Induce Highly Efficient Gene Targeting at an Endogenous Human Gene

We next tested whether our OPEN ZFNs could induce high efficiency gene targeting at the endogenous human *VEGF-A* gene. For these experiments, we used VF2468 pair C and VF2471 pair B (Table S2) and directly compared their activities with a previously published ZFN pair (Miller et al., 2007), which cleaves the human *IL2R $\gamma$*  gene. Gene-targeting frequencies using *VEGF-A*- or *IL2R $\gamma$* -specific ZFNs and matched donor templates (Figure S4) were measured in human K562 cells using a limited-cycle PCR/restriction digest assay (Experimental Procedures) (Urnov et al., 2005): mean targeting efficiencies of 7.7%, 4.5%, and 4.1% were observed with the VF2468, VF2471, and *IL2R $\gamma$*  ZFNs, respectively (Figure 4A). Efficient gene targeting required both the donor construct and ZFN expression vectors (Figure 4A). Control Southern blot assays of cells modified by *VEGF-A*-specific ZFNs confirm that, if anything, the PCR-based



**Figure 2. OPEN ZFNs Engineered to Cleave EGFP Gene Sequences**

(A) Quantitative B2H assay of modular assembly (MA; red bars) and OPEN (green bars) zinc-finger arrays. Mean fold-activation values (colored bars) and standard deviations (error bars) from three independent assays are shown.

(B) EGFP disruption assay for testing ZFN activities in human cells.

(C) Modularly assembled and OPEN ZFNs assessed using the EGFP-disruption assay. Error bars represent standard deviations. \* $p < 0.05$ ; \*\* $p < 0.01$ .

assay tended to underestimate gene targeting rates (Figure 4B and Figure S5).

Consistent with previous studies (Urnov et al., 2005), we observed higher rates of gene targeting in K562 cells that had been transiently arrested in G2 with vinblastine: 54%, 37%, and 44% mean efficiencies with VF2468, VF2471, and *IL2R $\gamma$*  ZFNs, respectively (Figure 4C). Notably, vinblastine treatment greatly reduced the number of viable cells (data not shown). Sequencing of *VEGF-A* and *IL2R $\gamma$*  alleles revealed gene targeting events at the expected locations and with the anticipated frequencies (Figure S6), but for all three target sites, we also observed high frequencies of insertion and deletion events at the ZFN cleavage sites (presumably caused by error-prone NHEJ). Unexpectedly, we found that 1.8% and 9.0% of alleles from cells treated with VF2468 and *IL2R $\gamma$*  ZFNs, respectively, contained evidence of both HR-mediated gene targeting and NHEJ-mediated insertion events at a single allele. Interestingly, analysis of single cell clones from vinblastine-treated populations of *VEGF-A* ZFN-modified cells indicates that gene targeting events are stably

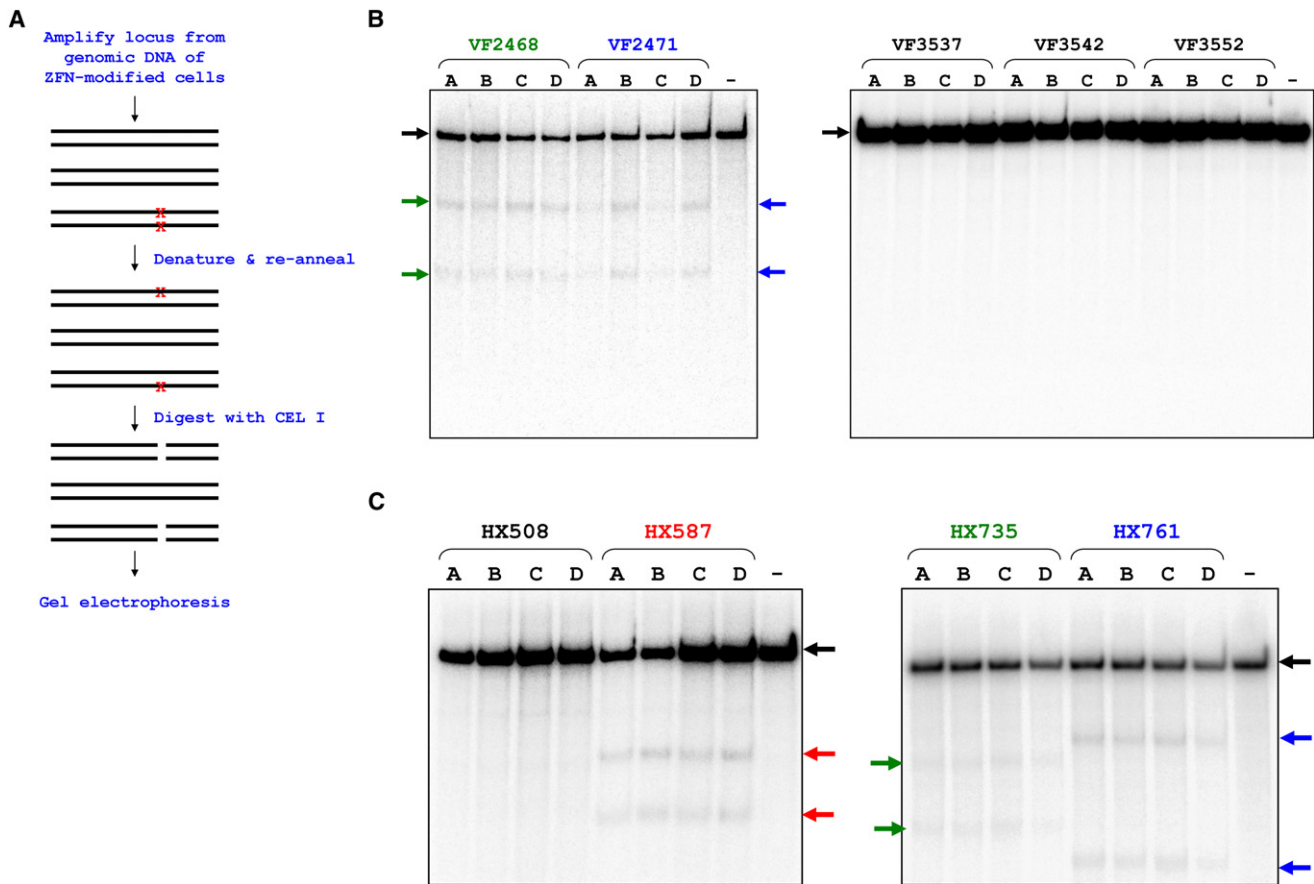
maintained (even after 35 days) and, strikingly, can be induced in as many as four copies of *VEGF-A* in polyploid cells (Figure S7).

#### Toxicity Profiles of OPEN ZFNs

We compared the relative toxicities of two pairs of three-finger OPEN *VEGF-A* ZFNs and of a pair of four-finger *IL2R $\gamma$*  ZFNs using cell survival assays in which K562 cells were transfected with ZFN expression vectors, a donor plasmid, and a GFP-expression plasmid. Previous studies demonstrated that toxic ZFNs reduce both the percentage of GFP-positive cells (Cornu et al., 2008; Pruett-Miller et al., 2008) and cells that have undergone gene targeting (Porteus and Baltimore, 2003) over time. All three ZFN pairs we tested showed significant reductions in the relative number of GFP-positive cells by posttransfection day 7 (Figure 4D, green bars) and analogous decreases in the percentage of gene targeting events (Figure 4D, purple bars).

We examined whether toxicity profiles of the *IL2R $\gamma$*  and our OPEN VF2468 and VF2471 ZFNs could be improved by using





**Figure 3. Highly Efficient Mutagenesis of Endogenous Human Genes by OPEN ZFNs**

(A) Schematic of CEL I assay for assaying ZFN-induced mutations.

(B and C) Mutation of the endogenous human *VEGF-A* gene (B) and *HoxB13* gene (C) by OPEN ZFNs. Colored arrows indicate expected CEL I digestion products. Images shown are from representative experiments.

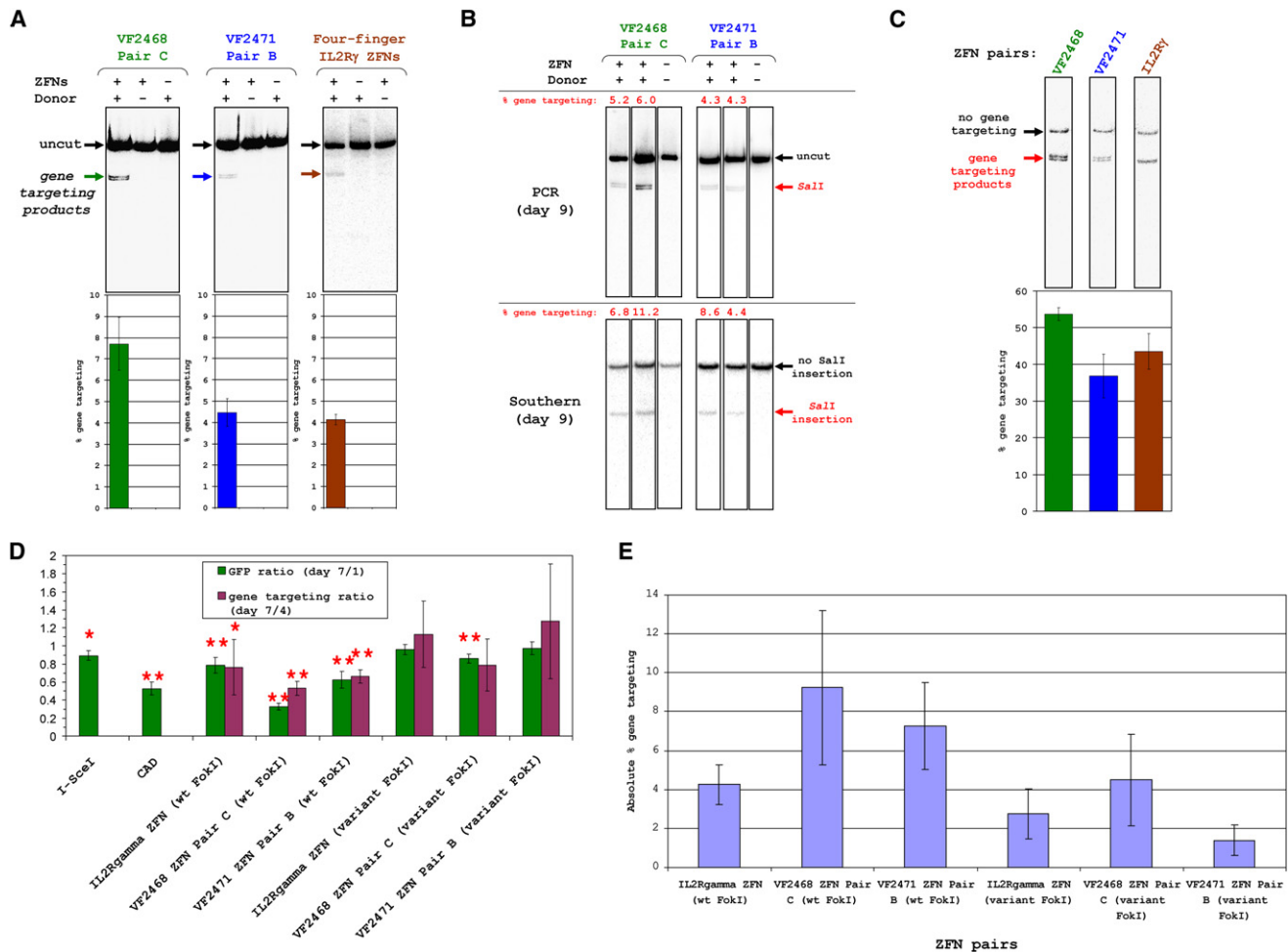
obligate heterodimeric *FokI* nuclease domain variants which significantly reduce ZFN-associated toxicity (Miller et al., 2007; Szczepek et al., 2007). Both the variant *IL2R $\gamma$*  ZFN pair and the variant VF2471 ZFN pair showed no significant toxicity as judged by the GFP toxicity assay (Figure 4D, green bars). In addition, the variant VF2468 ZFN pair revealed minimal toxicity (Figure 4D) similar to that of *I-SceI*, a highly specific meganuclease used previously as a control (Porteus and Baltimore, 2003). Comparable effects were also observed in the relative percentage of gene-targeting events (Figure 4D, purple bars). We note that all three variant ZFN pairs induce efficient gene targeting (Figure 4E), demonstrating that the absence of observable toxicity is not due to lack of ZFN activity.

## DISCUSSION

The OPEN reagents and methods we describe in this report provide a rapid, highly effective, and publicly available platform for engineering zinc-finger arrays. With OPEN, we created 37 ZFN pairs that mediate highly efficient modification at four different sites within a chromosomally integrated *EGFP* reporter gene,

six different sites within three endogenous human genes (*VEGF-A*, *HoxB13*, and *CFTR*), and one site within an endogenous plant gene (*SuRA*). The absolute rates of gene modification induced by our OPEN ZFNs ranged from 1%–50%, and we were able to alter as many as two alleles in a single plant cell and four alleles in a single polyploid human cell.

We found OPEN to be more effective than previously described modular assembly approaches for making zinc-finger arrays. In our studies targeting the *EGFP* gene, only two of 11 (~18%) modularly assembled ZFN pairs tested showed activity in human cells, in contrast to 15 of 20 (75%) OPEN ZFN pairs. In addition, at the one target site for which both methods were successful, OPEN ZFNs showed significantly more activity than modularly assembled ZFNs. The low efficacy rate observed with modular assembly is consistent with the results of a large-scale assessment of this method recently conducted by our groups (Ramirez et al., 2008). The higher success rate of OPEN is likely attributable to its consideration of context-dependent effects on DNA-binding among neighboring zinc-fingers (Elrod-Erickson et al., 1996; Isalan et al., 1997; Wolfe et al., 1999, 2001), which are largely ignored by modular assembly.



**Figure 4. Highly Efficient Gene Targeting of Endogenous Human Loci by OPEN ZFNs**

(A) OPEN *VEGF-A* ZFNs and previously described *IL2R $\gamma$*  ZFNs induce efficient gene targeting at endogenous genes in human K562 cells. Top part shows representative gel images from limited-cycle PCR/restriction digest assays and bottom part shows gene targeting frequency means (colored bars) and standard errors (error bars) from multiple experiments.

(B) Gene targeting efficiencies of OPEN *VEGF-A* ZFNs assessed 9 days posttransfection by limited-cycle PCR/restriction digest and Southern blot assays.

(C) Vinblastine enhances gene targeting by OPEN *VEGF-A* and four-finger *IL2R $\gamma$*  ZFNs. Assays performed 4 days posttransfection. Data presented as in (A).

(D) Toxicities of OPEN *VEGF-A* and four-finger *IL2R $\gamma$*  ZFNs in human K562 cells. Means of GFP (green bars) and gene targeting ratios (purple bars) are shown. Error bars represent standard deviations. \* $p < 0.05$ ; \*\* $p < 0.01$ .

(E) Gene targeting efficiencies of OPEN *VEGF-A* and four-finger *IL2R $\gamma$*  ZFNs in toxicity experiments of (D). Means and standard deviations (error bars) are shown of PCR-based assays performed 4 days posttransfection.

With the set of zinc-finger pools described in this report, we estimate that OPEN can be used to engineer three-finger proteins for  $\sim 4.1\%$  ( $[23 \times 21 \times 22]/[64 \times 64 \times 64]$ ) of all possible 9 bp target sites or  $\sim 0.16\%$  ( $4.1\% \times 4.1\%$ ) of all possible 18 bp ZFN sites. Because ZFNs can bind to sites in which two "half sites" are separated by "spacer sequences" of five, six, or seven base pairs (Bibikova et al., 2001; Porteus and Baltimore, 2003) (K. Wilson and M.H.P., unpublished data), one should be able to find approximately five full ZFN sites in any given kb of random sequence ( $0.0016 \times 1000 \times 3$ ). Thus, with the pools described in this report, one should be able to target multiple ZFN sites within a typical size gene. Important goals for future work will be to generate additional finger pools that expand the targeting range of

OPEN and to test whether the approach can also be used to generate arrays composed of more than three fingers.

While using ZFNs to modify human genes, we observed two limitations that have not been emphasized in previous reports. First, not all zinc-finger arrays that possess sequence-specific DNA-binding activities (as measured in the well-established B2H method) will function as ZFNs in human cells. ZFNs for one of the four sites targeted in the *HoxB13* locus, for three of the five sites targeted in the *VEGF-A* locus, and for one of the five sites targeted in an integrated *EGFP* reporter gene failed to induce mutagenic NHEJ repair in human cells. In addition, some ZFNs we made to the *HoxB13* gene were active in 293 cells (Figure 3C), but not in K562 cells (data not shown). We speculate that the

transcriptional status and chromatin configuration of a target site may influence ZFN access to target sites: *HoxB13* is transcriptionally active in 293 cells (data not shown) but bears chromatin marks consistent with a repressive state in K562 cells (B. Bernstein, personal communication). Additional studies will be needed to determine whether lack of ZFN activity results from chromatin effects on DNA accessibility or other reasons such as ZFN expression/stability or target site methylation. Second, although the use of vinblastine increased the frequency of gene targeting, DNA sequencing reveals that many alleles still underwent insertions or deletions caused by error-prone NHEJ and that some alleles underwent both a gene targeting event and an insertion. These findings demonstrate limitations in relying solely on PCR- or Southern blot-based assays and suggest that DNA sequencing should always be performed to verify ZFN-induced gene targeting events.

Because OPEN is rapid, reliable, and publicly available, it will foster wider usage and larger-scale applications of engineered zinc-finger technology. OPEN selections are performed in *E. coli* and do not require specialized equipment. Although our ZFN validation experiments were performed in plant and human cells, OPEN should also be useful for generating ZFNs that function well in other organisms such as zebrafish, mosquitos, *Drosophila*, and *C. elegans*. The rapidity and effectiveness of OPEN should enable genome-scale ZFN projects (e.g., developing ZFNs for all human kinase genes or for every zebrafish gene). We note that this report doubles the number of endogenous mammalian genes described in the published literature—from three (*IL2R $\gamma$* , *CCR5*, and *DHFR*) (Lombardo et al., 2007; Santiago et al., 2008; Urnov et al., 2005) to six—that have been successfully modified using ZFNs. In conclusion, our publicly available OPEN platform will enable scientists to perform the research and development required to move ZFN technology forward for applications in biological research and gene therapy.

## EXPERIMENTAL PROCEDURES

Additional details for all methods are provided in the [Supplemental Experimental Procedures](#).

### Construction of Zinc-Finger Pools

Randomized zinc-finger libraries constructed by cassette mutagenesis were introduced into "B2H selection strains" harboring a 9 bp target site as described (Thibodeau-Beganny and Jung, 2007) and plated on histidine-deficient selective media (NM media) containing 3-AT, a competitive inhibitor of the HIS3 enzyme. Zinc-finger-encoding phagemids were rescued from surviving colonies, reintroduced into fresh B2H selection cells, and plated on NM media containing 3-AT and streptomycin. 95 surviving colonies were inoculated into a 96-well block for growth, and plasmids were isolated to obtain the final finger pools.

### OPEN Selections

Each OPEN selection was performed in two steps: first, B2H selection strain cells were infected with randomized zinc-finger phage libraries consisting of three recombinant finger pools and then plated on NM media containing 3-AT and streptomycin. Second, zinc-finger-encoding phagemids were rescued from surviving colonies, reintroduced into fresh B2H selection strain cells, and plated on NM media containing a gradient of 3-AT and streptomycin. For a small number of the OPEN selections, we performed selections in a single step (see the [Supplemental Experimental Procedures](#)).

### CEL I Nuclease Assay for NHEJ-Mediated Mutation

Fip-In T-REX 293 cells (Invitrogen) were transfected with ZFN expression plasmids, and genomic DNA was isolated 3 days posttransfection. Limited-cycle PCR was performed with radiolabeled nucleotides and *VEGF-A*- or *HoxB13*-specific primers. PCR products were treated with CEL I nuclease and then separated on 10% polyacrylamide gels and visualized using a phosphorimaging screen.

### Gene Targeting Assays

Human K562 cells were transfected with ZFN expression plasmids and donor constructs, and genomic DNA was harvested 4 days posttransfection. Limited-cycle PCR was performed with radiolabeled nucleotides and *VEGF-A*- or *IL2R $\gamma$* -specific primers. PCR products digested with *Sall* (for *VEGF-A*) or *BsrBI* (for *IL2R $\gamma$* ) were separated on 10% polyacrylamide gels and visualized using a phosphorimaging screen. Additional details and Southern blot assays are described in the [Supplemental Experimental Procedures](#).

### Tobacco Transformation and Assay for Mutations

The transformation of tobacco protoplasts by electroporation, selection for kanamycin resistance, and regeneration into plantlets was carried out as previously described (Wright et al., 2005). DNA was prepared from tissue harvested from individual plantlets, and *SuRA* and *SuRB* alleles were amplified by PCR, gel purified, and sequenced to identify mutations.

### ZFN Toxicity Assays

ZFN expression vectors, donor templates, and plasmid pmaxGFP (Amaya) were transfected into K562 cells. Cells were assayed for GFP expression at days 1 and 7 posttransfection using a FACScan cytometer and for gene targeting efficiencies at days 4 and 7 posttransfection using the limited-cycle PCR/restriction digest assay.

## SUPPLEMENTAL DATA

The Supplemental Data include Supplemental Experimental Procedures, seven Figures, and three tables and can be found with this article online at <http://www.molecule.org/cgi/content/full/31/2/294/DC1/>.

## ACKNOWLEDGMENTS

We thank Chong Jin Park for help with statistical calculations and Andrew Hirsh for helpful discussions and comments. J.K.J. is supported by the NIH (R01GM069906, R24GM078369, and R21RR024189), the Cystic Fibrosis Foundation (CFF) (MCCRAY07G0), and the MGH Pathology Service. D.F.V. is supported by the NSF (DBI 0501678). T.C. is supported by the European Commission's 6th Framework Programme (037783 ZNIP). P.B.M. is supported by the CFF (MCCRAY07G0) and the Roy J. Carver Charitable Trust. A.J.I. is supported by the MGH Pathology Service. D.C.S. is supported by the NIH (R01CA112021 and NCI SPORE in Breast Cancer at MGH). M.H.P. is supported by the NIH (R01 HL079295). J.K.J. dedicates this paper to the memory of Robert L. Burghoff and his belief in the power of methods.

Received: March 24, 2008

Revised: May 16, 2008

Accepted: June 4, 2008

Published: July 24, 2008

## REFERENCES

- Alwin, S., Gere, M.B., Guhl, E., Effertz, K., Barbas, C.F., III, Segal, D.J., Weitzman, M.D., and Cathomen, T. (2005). Custom zinc-finger nucleases for use in human cells. *Mol. Ther.* 12, 610–617.
- Bae, K.H., Do Kwon, Y., Shin, H.C., Hwang, M.S., Ryu, E.H., Park, K.S., Yang, H.Y., Lee, D.K., Lee, Y., Park, J., et al. (2003). Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat. Biotechnol.* 21, 275–280.



- Beerli, R.R., and Barbas, C.F., III. (2002). Engineering polydactyl zinc-finger transcription factors. *Nat. Biotechnol.* **20**, 135–141.
- Beumer, K., Bhattacharyya, G., Bibikova, M., Trautman, J.K., and Carroll, D. (2006). Efficient gene targeting in *Drosophila* with zinc-finger nucleases. *Genetics* **172**, 2391–2403.
- Bibikova, M., Carroll, D., Segal, D.J., Trautman, J.K., Smith, J., Kim, Y.G., and Chandrasegaran, S. (2001). Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol. Cell. Biol.* **21**, 289–297.
- Bibikova, M., Golic, M., Golic, K.G., and Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics* **161**, 1169–1175.
- Bibikova, M., Beumer, K., Trautman, J.K., and Carroll, D. (2003). Enhancing gene targeting with designed zinc finger nucleases. *Science* **300**, 764.
- Cathomen, T. (2004). AAV vectors for gene correction. *Curr. Opin. Mol. Ther.* **6**, 360–366.
- Cathomen, T., and Joung, J.K. (2008). Zinc Finger Nucleases - The Next Generation Emerges. *Mol. Ther.* **16**, 1200–1207.
- Cornu, T.I., Thibodeau-Beganny, S., Guhl, E., Alwin, S., Eichinger, M., Joung, J., and Cathomen, T. (2008). DNA-binding Specificity Is a Major Determinant of the Activity and Toxicity of Zinc-finger Nucleases. *Mol. Ther.* **16**, 352–358.
- Durai, S., Mani, M., Kandavelou, K., Wu, J., Porteus, M.H., and Chandrasegaran, S. (2005). Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Res.* **33**, 5978–5990.
- Elrod-Erickson, M., Rould, M.A., Nekudova, L., and Pabo, C.O. (1996). Zif268 protein-DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions. *Structure* **4**, 1171–1180.
- Greisman, H.A., and Pabo, C.O. (1997). A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* **275**, 657–661.
- Hurt, J.A., Thibodeau, S.A., Hirsh, A.S., Pabo, C.O., and Joung, J.K. (2003). Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc. Natl. Acad. Sci. USA* **100**, 12271–12276.
- Isalan, M., Choo, Y., and Klug, A. (1997). Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proc. Natl. Acad. Sci. USA* **94**, 5617–5621.
- Isalan, M., Klug, A., and Choo, Y. (2001). A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nat. Biotechnol.* **19**, 656–660.
- Jasin, M. (1996). Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet.* **12**, 224–228.
- Joung, J.K., Ramm, E.I., and Pabo, C.O. (2000). A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc. Natl. Acad. Sci. USA* **97**, 7382–7387.
- Liu, Q., Xia, Z., Zhong, X., and Case, C.C. (2002). Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *J. Biol. Chem.* **277**, 3850–3856.
- Lloyd, A., Plaisier, C.L., Carroll, D., and Drews, G.N. (2005). Targeted mutagenesis using zinc-finger nucleases in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **102**, 2232–2237.
- Lombardo, A., Genovese, P., Beausejour, C.M., Colleoni, S., Lee, Y.L., Kim, K.A., Ando, D., Urnov, F.D., Galli, C., Gregory, P.D., et al. (2007). Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat. Biotechnol.* **25**, 1298–1306.
- Mandell, J.G., and Barbas, C.F., III. (2006). Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res.* **34**, W516–W523.
- Miller, J.C., Holmes, M.C., Wang, J., Guschin, D.Y., Lee, Y.L., Rupniewski, I., Beausejour, C.M., Waite, A.J., Wang, N.S., Kim, K.A., et al. (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.* **25**, 778–785.
- Moehle, E.A., Rock, J.M., Lee, Y.L., Jouvenot, Y., Dekelver, R.C., Gregory, P.D., Urnov, F.D., and Holmes, M.C. (2007). Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc. Natl. Acad. Sci. USA* **104**, 3055–3060.
- Morton, J., Davis, M.W., Jorgensen, E.M., and Carroll, D. (2006). Induction and repair of zinc-finger nuclease-targeted double-strand breaks in *Caenorhabditis elegans* somatic cells. *Proc. Natl. Acad. Sci. USA* **103**, 16370–16375.
- Porteus, M.H., and Baltimore, D. (2003). Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**, 763.
- Porteus, M.H., and Carroll, D. (2005). Gene targeting using zinc finger nucleases. *Nat. Biotechnol.* **23**, 967–973.
- Pruett-Miller, S.M., Connelly, J.P., Maeder, M.L., Joung, J.K., and Porteus, M.H. (2008). Comparison of Zinc Finger Nucleases for Use in Gene Targeting in Mammalian Cells. *Mol. Ther.* **10**, 1038/mt.2008.1020.
- 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., et al. (2008). Unexpected failure rates for modular assembly of engineered zinc-fingers. *Nat. Methods* **5**, 374–375.
- Sander, J.D., Zaback, P., Joung, J.K., Voytas, D.F., and Dobbs, D. (2007). Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucleic Acids Res.* **35**, W599–W605.
- Santiago, Y., Chan, E., Liu, P.Q., Orlando, S., Zhang, L., Urnov, F.D., Holmes, M.C., Guschin, D., Waite, A., Miller, J.C., et al. (2008). Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc. Natl. Acad. Sci. USA* **105**, 5809–5814.
- Segal, D.J., Beerli, R.R., Blancafort, P., Dreier, B., Effertz, K., Huber, A., Koksche, B., Lund, C.V., Magnenat, L., Valente, D., and Barbas, C.F., III. (2003). Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry* **42**, 2137–2148.
- Szcepek, M., Brondani, V., Buchel, J., Serrano, L., Segal, D.J., and Cathomen, T. (2007). Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat. Biotechnol.* **25**, 786–793.
- Thibodeau-Beganny, S., and Joung, J.K. (2007). Engineering Cys2His2 Zinc Finger Domains Using a Bacterial Cell-Based Two-Hybrid Selection System. *Methods Mol. Biol.* **408**, 317–334.
- Urnov, F.D., Miller, J.C., Lee, Y.L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D., and Holmes, M.C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646–651.
- Wolfe, S.A., Greisman, H.A., Ramm, E.I., and Pabo, C.O. (1999). Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code. *J. Mol. Biol.* **285**, 1917–1934.
- Wolfe, S.A., Grant, R.A., Elrod-Erickson, M., and Pabo, C.O. (2001). Beyond the "recognition code": structures of two Cys2His2 zinc finger/TATA box complexes. *Structure* **9**, 717–723.
- Wright, D.A., Townsend, J.A., Winfrey, R.J., Jr., Irwin, P.A., Rajagopal, J., Lonosky, P.M., Hall, B.D., Jondle, M.D., and Voytas, D.F. (2005). High-frequency homologous recombination in plants mediated by zinc-finger nucleases. *Plant J.* **44**, 693–705.

## **Supplemental Experimental Procedures**

### **B2H selection media**

NM medium has been previously described (Thibodeau-Beganny and Joung, 2007).

NM/CCK medium plates contain 100 µg/mL carbenicillin, 30 µg/mL chloramphenicol, 30 µg/mL kanamycin, and 1.5% Bacto-agar.

### **Construction of zinc finger pools**

All master randomized zinc finger libraries were constructed in a standard framework consisting of three tandem repeats of the middle finger of the murine transcription factor Zif268 in which the recognition helix residues have been altered. For each library, recognition helix residues -1, 1, 2, 3, 5, and 6 were randomized using 24 codons (degenerate sequence 5'VNS3') encoding 16 amino acids (excluding cysteine and the aromatics). The theoretical complexity of each library is therefore  $24^6 = \sim 2 \times 10^8$  members. Each library was converted into infectious M13 phage particles as previously described (Joung et al., 2000).

B2H selection strains each harbor: (1) a single copy episome bearing a target site of interest positioned upstream of a promoter which drives co-cistronic expression of two selectable markers (the yeast *HIS3* gene and the bacterial *aadA* gene) and (2) a low copy number plasmid expressing the RNA polymerase  $\alpha$ -subunit/yeast Gal4 hybrid protein ( $\alpha$ -Gal4). Strains were constructed as previously described (Thibodeau-Beganny and Joung, 2007). The target binding site of each strain was verified by DNA sequencing.

Zinc finger pools were obtained using two selection steps: In a first step,  $10^9$  ampicillin-transducing-units (ATU) of randomized zinc finger phage library were introduced into  $>3 \times 10^9$  B2H selection strain cells harboring a target subsite of interest. Transformed cells were plated on histidine-deficient NM/CCK medium plates containing 50  $\mu$ M isopropyl  $\beta$ -d-thiogalactoside (IPTG) and 10 mM 3-aminotriazole (3-AT), a competitive inhibitor of the HIS3 enzyme. After incubation for 24 hours at 37°C followed by 18 hours at room temperature, surviving colonies were scraped from the plates and infected with M13K07 helper phage to rescue the zinc finger-encoding phagemids as infectious phage. In a second step, this enriched phage library was then used to re-infect fresh B2H selection strain cells and the resulting transformants plated on NM/CCK medium plates containing 50  $\mu$ M IPTG, 10 mM 3-AT, and 20  $\mu$ g/ml streptomycin. After incubation at 37°C for 48 hours, we inoculated 95 surviving colonies of various sizes into individual wells of a 96-well block containing 1ml of Terrific Broth and 50  $\mu$ g/ml carbenicillin. These cultures were grown overnight at 37°C and then a 96-pin replicator was used to inoculate a second block of identical cultures. Glycerol was added to a final concentration of 15% to all wells in the first block and this was stored at -80°C. The second block was grown overnight at 37°C and then the 95 cultures were pooled together. Plasmid DNA encoding the finger pools was isolated from 10 ml of the pooled culture using a QIAgen miniprep kit and for many pools a small number of random clones were sequenced with primer OK61. For finger pools against nine subsites, sequencing revealed no strong consensus (i.e.—the sequences were diverse and did not resemble one another), suggesting that selective pressure for those sites was relatively weak under the initial selection conditions. For each of these nine subsites, the second step of selection was repeated and

plated on higher stringency plates to obtain sequences that more closely resembled one another. The higher stringency plates used were NM/CCK medium plates containing 50  $\mu$ M IPTG, 20 mM 3-AT, and 30  $\mu$ g/ml streptomycin (for seven subsites: F1 GAT, F2 GAC, F2 GAG, F2 GCG, F2 TGA, F2 TAG, and F2 GTT) or 50  $\mu$ M IPTG, 25 mM 3-AT, and 40  $\mu$ g/ml streptomycin (for two subsites: F2 GAA, F2 TGG).

### **OPEN selections**

To create libraries for use in OPEN selections, finger pools were amplified by PCR. For each amplification, first five and then 20 cycles of PCR were performed using the following primers and annealing temperatures (primer names; initial annealing temp; final annealing temp): for finger 1: OK1424 and OK1425; 55°C; 59°C, for finger 2: OK1426 and OK1427; 52°C; 57°C, for finger 3: OK1428 and OK1429; 41°C; 56°C. Amplified individual finger pools were isolated from 10% polyacrylamide gels and fused together by PCR. To do this, equal concentrations of the three finger pool fragments were fused using the following PCR conditions: 94°C, 5 minutes; 10 cycles of 94°C, 30 sec; 50°C, 30 sec; 72°C, 2 min; final extension 72°C, 7 min. This fusion product was purified using a QIAgen PCR purification kit and then amplified by PCR using primers OK1430 and OK1432 with 10 initial cycles of 94°C, 30 sec; 56°C, 30 sec; 72°C, 1 min and 20 additional cycles of 94°C, 30 sec; 64°C, 30 sec; 72°C, 1 min; final extension 72°C, 7 min. The final PCR product (encoding a library of three-finger arrays) was isolated on a 5% polyacrylamide gel and treated with *Pfu* polymerase and T4 polynucleotide kinase to create overhangs. This fragment was then ligated with *Bbs*I-digested pBR-UV5-GP-FD2 vector backbone which results in a plasmid that expresses

the zinc finger array as a FLAG-tagged Gal11P fusion in the B2H system. This ligation was then introduced into *E. coli* XL-1 Blue cells by electroporation and each library was constructed from  $>3 \times 10^6$  independent transformants, ensuring at least three-fold oversampling of the theoretical library complexity of  $\sim 8.6 \times 10^5$  ( $95^3$ ). Libraries were then converted into infectious M13 phage as previously described (Thibodeau-Beganny and Joung, 2007).

All but six (see below) of the OPEN selections were performed in two stages. In the first stage, an OPEN three-finger library was introduced by infection into a B2H selection strain harboring the full target DNA sequence of interest.  $>2.2 \times 10^6$  ATU of OPEN phage library (a number which ensures adequate oversampling of the maximal theoretical library diversity) were used to infect  $>2 \times 10^8$  B2H selection strain cells and the resulting transformants were plated on two different NM/CCK medium plates containing 50  $\mu$ M IPTG, 10 mM 3AT, and 20  $\mu$ g/mL streptomycin or 50  $\mu$ M IPTG, 25 mM 3AT, and 40  $\mu$ g/mL streptomycin. After 36-48 hours of incubation, colonies were harvested from the highest stringency plate yielding at least 1000 colonies. These cells were infected with M13K07 helper phage to rescue zinc finger-encoding phagemids, thereby creating an enriched library. In the second stage,  $2.5 \times 10^6$  ATU of this enriched library were used to infect  $>2 \times 10^8$  fresh B2H selection strain cells and the resulting transformants plated on a 245x245mm NM/CCK medium plate containing parallel gradients of 3-AT and streptomycin ranging from 0 mM to 80 mM and 0  $\mu$ g/mL to 100  $\mu$ g/mL, respectively. Gradient plates were poured using the method of Szybalski (Bryson and Szybalski, 1952). After incubation at 37°C for a minimum of 48 and a maximum of 96 hours, 12

surviving colonies were picked from the highest stringency edge of the plate and ZFP-encoding plasmids were isolated by plasmid miniprep and sequenced with primer OK61.

For six of the sites we targeted using OPEN, selections were (for historical reasons) performed in a single step instead of two steps. For five sites (EG223L, EG223R, EG292L, EG292R, and EG382L),  $>3.1 \times 10^6$  ATU of OPEN phage library were used to infect  $>4 \times 10^8$  B2H selection strain cells and the resulting transformants were plated on a series of NM/CCK medium plates containing 50  $\mu$ M IPTG, 40 mM 3AT, and 60  $\mu$ g/mL streptomycin, 50  $\mu$ M IPTG, 60 mM 3AT, and 80  $\mu$ g/mL streptomycin, 0  $\mu$ M IPTG, 40 mM 3AT, and 60  $\mu$ g/mL streptomycin or 0  $\mu$ M IPTG, 60 mM 3AT, and 80  $\mu$ g/mL streptomycin. For each target site, we picked colonies from the highest stringency plates that yielded colonies. For one site (EG382R),  $4.25 \times 10^8$  ATU of OPEN phage library were used to infect  $>10^9$  B2H selection strain cells and the resulting transformants were plated on a 245x245mm NM/CCK medium plate containing parallel gradients of 3-AT and streptomycin ranging from 0 mM to 80 mM and 0  $\mu$ g/mL to 100  $\mu$ g/mL, respectively. Colonies were picked from the highest stringency edge of the plate.

### **Construction of modularly assembled zinc finger arrays**

Modularly assembled zinc finger arrays were assembled using the Zinc Finger Consortium Modular Assembly Kit v1.0 as previously described (Wright et al., 2006). For each target half-site in the *EGFP* reporter gene, we assembled three-finger arrays using modules from the Barbas, Sangamo, and Toolgen archives but we did not mix and match modules across platforms because: (1) the Barbas group does not suggest 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 but not as well with other engineered modules (Bae et al., 2003), and (3) the Sangamo modules were designed to be finger-position-specific and have non-canonical linkers joining them that differ from the TGEKP linker used by the Barbas and Toolgen modules (Liu et al., 2002).

### **Quantitative bacterial two-hybrid (B2H) assays**

Zinc-finger-encoding plasmids identified from OPEN selections were co-transformed with an  $\alpha$ -Gal4 expression plasmid into a “B2H reporter strain” harboring a single copy bacterial plasmid with a target binding site positioned upstream of a weak promoter driving *lacZ* expression. B2H reporter strains were constructed as described (Wright et al., 2006).  $\beta$ -galactosidase assays were performed in triplicate as described (Thibodeau et al., 2004).

### **ZFN expression vectors**

All zinc finger arrays were expressed as ZFNs using the Zinc Finger Consortium mammalian expression vector pST1374 (Wright et al., 2006). Zinc finger arrays were excised directly from B2H expression vectors on an *XbaI/BamHI* fragment and cloned into pST1374. In this configuration, zinc finger arrays are joined to the *FokI* nuclease domain by a four amino acid linker of sequence LRGS. We constructed expression vectors encoding the four-finger *IL2R $\gamma$*  ZFNs using amino acid sequences obtained from a previously published study (Miller et al., 2007).

### **Human cell-based EGFP-disruption assay**

Human 293.EGFP cells harbor an integrated retroviral construct which constitutively expresses a  $\beta$ -galactosidase-EGFP fusion protein (Figure 2B). 293.EGFP cells were transfected in triplicate in 24-well plates using calcium phosphate precipitation as previously described (Cathomen et al., 2001). Transfection cocktails included 300 ng each of a CMV promoter-controlled zinc finger nuclease expression vector, 100 ng pDS.RedExpress (Clontech, Mountain view, CA) and pUC118 to 1.5  $\mu$ g. 600 ng of pRK5.SceI plasmid (Alwin et al., 2005), which expresses the meganuclease I-SceI, was used in place of GFP-ZFN-encoding plasmids for negative controls. 50,000 cells were analyzed by flow cytometry two and five days post-transfection to determine the percentage of EGFP-negative cells. The number of REx-positive cells at day 2 was used to normalize for transfection efficiency. Statistical significance was determined using a two-sided student's t-test with unequal variance.

### **CEL I nuclease assay for NHEJ-mediated mutation**

In this assay, limited-cycle PCR is used to amplify a locus of interest from the genomic DNA of a population of human cells transfected with ZFN expression vectors (Figure 3A). The resulting PCR product is denatured and re-annealed and heteroduplex DNA will form if mutated alleles are present in the population. These DNA fragments can be cleaved at the site of mismatch by the CEL I enzyme into smaller products of predictable size. Specifically,  $2 \times 10^6$  human Flp-In T-REx 293 cells (Invitrogen) were transfected with pairs of ZFN-encoding plasmids (100 or 250 ng of each ZFN-encoding plasmid) using Lipofectamine 2000 (Invitrogen). Genomic DNA was isolated from nuclease-



treated cells at 3 days post-transfection using the QIAGEN Blood Mini kit. Limited-cycle PCR (24 cycles) was performed using Platinum PCR SuperMix Hi-Fidelity (Invitrogen) or its equivalent constituent components (Invitrogen) with 50 ng of genomic DNA as template, 8  $\mu$ Ci of each [ $\alpha$ - $^{32}$ P]-dATP and dCTP, 1  $\mu$ M each of gene-specific primers (primers OK1681 and OK1682 for *VEGF-A* sites VF2468/VF2471, primers OK1706 and OK1718 for *VEGF-A* sites VF3537/VF3542/VF3552, primers OK1733 and OK1734 for *HoxB13* sites HX508/HX587 or primers OK1736 and OK1738 for *HoxB13* sites HX735/HX761) and 1.25  $\mu$ l DMSO in a 25  $\mu$ l reaction volume. PCR products were cleaned up using Sephadex G-50 columns (Roche) and then melted/re-annealed using the following conditions: 95°C for 10 min; 95°C to 85°C cooling at a rate of -2°C/sec; 85°C to 25°C cooling at a rate of -0.1°C/sec; rapid cool to 4°C. Re-annealed PCR products were diluted 1:3.75 in a buffer of 20 mM Tris-HCl (pH 8.8), 2 mM MgSO<sub>4</sub>, 60 mM KCl, 0.1% Triton X-100 and treated with 1  $\mu$ l CEL I enzyme (Surveyor nuclease S; Transgenomic) and 1  $\mu$ l Surveyor Enhancer S (Transgenomic) in a 15  $\mu$ l reaction incubated at 42°C for 20 min. Products were visualized by electrophoresis on a 0.8mm thick, 10% 1X TBE polyacrylamide gel which was dried down and exposed overnight to a phosphorimaging screen. All experiments were performed a minimum of two times.

### **Gene targeting assays**

2 x 10<sup>6</sup> Flp-In T-REx 293 cells were transfected with pairs of plasmids expressing ZFNs (7.5  $\mu$ g of each ZFN-encoding plasmid) and 50  $\mu$ g of donor plasmid using nucleofection with solution V and program Q001 (Amaxa). 2 x 10<sup>6</sup> K562 cells were transfected with ZFN expression plasmid pairs (5 or 7.5  $\mu$ g of each) and matched donor construct (25 or

50 µg donor plasmid) using nucleofection with solution V and program T-16. Genomic DNA was harvested 3 or 4 days post-transfection for 293 or K562 cells, respectively, using a QIAgen Blood Mini kit. Transfection efficiencies were monitored by including a GFP-encoding plasmid in each transfection and determining the percentage of GFP-positive cells by flow cytometry one day post-transfection. For experiments in which cells were arrested in G2 phase, 0.2 µM vinblastine was added 24 hours post-transfection and then removed by washing three times with phosphate buffered saline 14-18 hours later. Limited-cycle PCR assays (24 cycles) were performed using Platinum PCR SuperMix Hi-Fidelity (Invitrogen) or its equivalent constituent components (Invitrogen) with 4 ng genomic DNA, 8 µCi of each [ $\alpha$ - $^{32}$ P]-dATP and dCTP, 1 µM each of gene-specific primers (OK1776 and OK1777 for *VEGF-A* or OK1845 and OK1846 for *IL2R $\gamma$* ) and 1.25 µl DMSO in a 25 µl reaction. Purified PCR product was digested with 25 units *SalI* or 10 units *BsrBI* restriction enzyme for 2 hours and the resulting products were visualized by electrophoresis on a 10% 1X TBE polyacrylamide gel. This gel was dried down and exposed overnight to a phosphorimaging screen. Quantification of bands was performed using Quantity One software (Bio-Rad).

For Southern blots, 15 µg of genomic DNA (15 µg) was digested with *MscI* and *SalI* restriction enzymes for 20 hrs, electrophoresed in 0.8% tris-acetate agarose gels (100 mM Tris-HCl, 10 mM EDTA, pH 8.0 with acetic acid), and transferred to Zeta-probe nylon membrane (BioRad) using 25 mM sodium phosphate (pH 6.5) according to the procedure of Southern (Southern, 1975) as modified for use with the Turbo-Blot downward transfer apparatus (Schleicher & Schuell). The *VEGF-A* DNA probe was generated by PCR

amplification of a cloned human *VEGF-A* DNA template using primers OK1823 and OK1824 which was subsequently labeled (25 ng) with [ $\alpha$ - $^{32}$ P]-dCTP) using Rediprime II random priming reagents (Amersham). Following hybridization (20 hrs) at 65°C in 5 mL ExpressHyb solution (Clontech), the filters were washed with 0.1X SSC/0.1% SDS at 65°C (2 hrs), blotted dry and exposed to a phosphorimager screen and/or film. The filters were scanned in the Typhoon 8600 phosphorimager and relative band intensities were quantified by volume analysis using ImageQuant software (GE Healthcare/Amersham).

### **Sequencing of modified genomic alleles**

The region encompassing each potential ZFN cleavage site was amplified from genomic DNA isolated from populations of human Flp-In T-REx 293 or K562 cells that had been transfected with ZFN expression plasmids alone or with ZFN expression and donor plasmids. PCR conditions for these amplifications were the same as those used for the CEL I (for assessing NHEJ events) or limited-cycle PCR/restriction digest (for assessing gene targeting events) assays but with all components doubled to a final volume of 50 $\mu$ l. CF877 was amplified using primers OK1711 and 1713. PCR reactions were purified using the QIAgen Minelute PCR Purification kit and eluted with 15  $\mu$ l 0.1X EB buffer (QIAgen). PCR fragments were cloned into the pCR4Blunt-TOPO plasmid using the Zero Blunt TOPO PCR Cloning Kit for Sequencing (Invitrogen). The TOPO cloning reaction used 4  $\mu$ l purified PCR product, 1  $\mu$ l salt solution and 1  $\mu$ l TOPO vector. 2  $\mu$ l of TOPO reaction were transformed into One Shot Mach1-T1 chemically competent cells (Invitrogen) or chemically competent Top10 cells (Invitrogen) and plated on LB plates

containing 50 µg/ml kanamycin. Plasmid DNAs from transformants were sequenced with a primer designed to bind internal to the PCR product (OK1773 or T3 (Invitrogen) for *VEGF-A*, OS216 for *HoxB13*, M13 primer (Invitrogen) for *CFTR* or OK1838 for *IL2Rγ*).

### **Tobacco transformation and assay for mutations**

The transformation of tobacco protoplasts by electroporation was carried out as previously described (Wright et al., 2005). Plasmids introduced into protoplasts (10 µg each) included those expressing ZFNs that recognize the left (pRW242) and right (pRW246) half sites of target 2163 in *SuRA*. Note that these constructs do not express the heterodimeric variants of *FokI* endonuclease. Also transformed into protoplasts was a plasmid expressing neomycin phosphotransferase (NPTII) (pDW998). The CaMV 35S promoter was used to drive expression of both the ZFNs and NPTII. Plasmid DNAs were linearized with *BglIII* prior to transformation. Protoplasts were allowed to recover and then selected for kanamycin resistance and regenerated into plantlets as previously described (Wright et al., 2005).

DNA was prepared from tissue harvested from individual plantlets using the Epicentre MasterPure Plant Leaf DNA Purification Kit following the manufacturer's directions. An initial PCR screen for mutations at the site of ZFN cleavage was performed using primers DVO4461 and DVO4462 to amplify a 445 bp fragment from both the *SuRA* and *SuRB* loci. PCR was performed with 100 ng of genomic DNA and the following PCR conditions: 94°C 2 min, followed by 34 cycles of 94°C 15 sec, 61°C 15 sec, 72°C 30 sec,

and then 72°C for 5 min. The reactions were run out on a 0.8% agarose gel, purified using a QIAGEN QIAquick Gel Extraction kit, and sequenced with DVO4462. The resulting sequences were examined for double peaks, which not only indicate sequence differences between *SuRA* and *SuRB*, but also identify potential insertion/deletion events at the ZFN cleavage site in either or both loci.

DNA from candidate mutants was then PCR amplified using a set of nested, allele-specific primers in two consecutive PCR reactions to confirm the mutation and determine if it occurred in *SuRA* or *SuRB*. The primary reaction amplified a 2.15 kb fragment using approximately 100 ng of genomic DNA as template and primers DVO4565 and DVO4461 for *SuRA* and DVO4429 and DVO4461 for *SuRB*. The second PCR reaction amplified a 2 kb fragment using 1 µl of the primary PCR reaction as template and primers DVO4444 and DVO4461 for *SuRA* and DVO4445 and DVO4461 for *SuRB*. All PCR reactions were performed using a Clontech Advantage cDNA Polymerase kit and the following PCR conditions: 94°C 1 min, followed by 34 cycles of 94°C 30 sec, 66°C 30 sec, 68°C 3 min, and then 68°C for 5 min. The reactions were run out on a 0.8% agarose gel, purified with a QIAGEN QIAquick Gel Extraction kit and sequenced with primer DVO4462.

### **ZFN toxicity assays**

ZFN expression plasmids and donor templates were transfected into K562 cells using nucleofection as described above. For these experiments, 5 µg of each ZFN expression plasmid, 25 µg of donor, and 15 ng of pmaxGFP (encoding a GFP variant; Amaxa) were included in each transfection. For controls, we transfected: (1) 10 µg of plasmid

encoding I-SceI meganuclease (Porteus and Baltimore, 2003) with 25 µg pUC118 and 15 ng of pmaxGFP, (2) 10 µg of plasmid encoding CAD (caspase-activated DNase) protein (Pruett-Miller et al., 2008) with 25 µg pUC118 and 15 ng of pmaxGFP, or (3) 35 µg of pUC118 and 15 ng of pmaxGFP. Cells were assayed for GFP expression at post-transfection days 1 and 7 with a FACScan cytometer. GFP ratios shown in Figure 4D (green bars) were calculated using the formula:

$$\frac{(\%GFP+ \text{ in ZFN-transfected cells on day 7} / \%GFP+ \text{ in ZFN-transfected cells on day 1})}{(\%GFP+ \text{ in pUC-transfected cells on day 7} / \%GFP+ \text{ in pUC-transfected cells on day 1})}$$

In addition, genomic DNA was harvested using a QIAgen Blood Mini Prep kit on post-transfection days 4 and 7 and assayed for gene targeting using the limited-cycle PCR/restriction digest assay as described above. Gene targeting ratios shown in Figure 4D (purple bars) were calculated by dividing the gene targeting rate on day 7 by the gene targeting rate on day 4. All assays (both GFP and gene targeting) were performed on at least three-independent samples and t-tests of significance were performed by comparing experimentally determined ratios to a fixed ratio value of 1 (i.e.--no change in value). Variant heterodimer *FokI* domains were constructed by introducing the “+” and “-“ mutations (previously described in Miller et al., 2007) into the wild-type *FokI* nuclease domain.

### **Fluorescence in situ hybridization (FISH)**

Two-color fluorescence in situ hybridization (FISH) was performed on 3:1 methanol-acetic acid fixed cell lines using bacterial artificial chromosome clones RP11-710L16 (6p21.1; *VEGF-A*) labeled in Spectrum Orange (Abbott-Vysis, Downer's Grove, Il.), or

RP11-142P4 (14q31.1; copy number control) labeled in Spectrum Green using standard protocols. Images were captured using an Olympus BX61 fluorescent microscope equipped with a CCD camera, and analysis was performed with Cytovision software (Applied Imaging, San Jose, CA).

## **Supplemental Figure and Table Legends:**

### **Figure S1 ZiFiT 3.0 -- Web-based software for identifying potential sites targetable by OPEN selection**

(A) Screen shot of ZiFiT 3.0 input page. We previously described ZiFiT, a web-based software program that enables users to rapidly identify potential zinc finger nuclease target sites within genes of interest (Sander et al., 2007; Wright et al., 2006). The original implementations of ZiFiT (versions 1.0 and 2.0) were both geared toward making zinc finger nucleases by modular assembly, and the output cross-referenced zinc finger modules available in the Zinc Finger Consortium Modular Assembly Kit 1.0 (Wright et al., 2006). With the implementation of OPEN, we have released ZiFiT Version 3.0 (<http://www.zincfingers.org/software-tools.htm>), which gives users the option of searching for target sites for which OPEN pools are available. The user pastes a DNA sequence (limited to 10,000 characters) into the text box at the top of the main window. Underneath the main window are three boxes corresponding to finger positions in the zinc finger array. Each box contains a list of the 64 possible nucleotide triplets. By default, triplets are checked for which OPEN pools are currently available from the Zinc Finger Consortium; however, users can manually select or deselect triplet pools to customize their search. Target site composition can be further restricted by options available under the "Advanced/Basic" toggle near the bottom of the screen. By selecting the "Submit" button, the sequence is searched for all target sites using the specified pools and an output is returned.

(B) Screen shot of ZiFiT 3.0 output page. In addition to identifying zinc finger nuclease



target sites in the input sequence, it may be desirable to identify matching or highly similar sites within a genome of interest. ZiFiT Version 3.0 enables the user to BLAST the genome sequence of a desired organism to identify duplicate or related target sites. The BLAST search is implemented by a button link and an organism selection menu provided for each target site returned in the output on the results page. Requests are submitted to the NCBI BLAST servers to provide maximum reliability and speed, and results are typically returned to ZiFiT within a minute. Note that BLAST scoring methods do not identify all matches near the end of a hit if they are separated by mismatches that lower the score of the local alignment. As a result, hits to highly similar target sites lacking information at the ends may need to be analyzed manually.

**Figure S2 Schematic of target sites in the *EGFP*, *VEGF-A*, *HoxB13*, and *CFTR* genes**

Using ZFNs made by OPEN selections, we targeted (A) five full ZFN sites in the *EGFP* reporter gene (EG223, EG292, EG382, EG502, and EG568), (B) six sites in open chromatin regions of the human *VEGF-A* gene (VF2468, VF2471, VF3537, VF3540, VF3542, and VF3552), (C) five sites in the first coding exon of the human *HoxB13* gene (HX500, HX508, HX587, HX735, and HX761) and one site in the last coding exon of the human *HoxB13* gene (HX2119, not shown), and (D) one site in exon 10 of the human *CFTR* gene (CF877) positioned within 100 base pairs of the  $\Delta$ F508 deletion.

**Figure S3 DNA sequence analysis of endogenous human and plant genes  
mutagenized by OPEN ZFNs**

(A-C) Sequences of (A) *HoxB13* alleles from human 293 cells transfected with HX587 ZFN pair B, (B) *CFTR* alleles from human K562 cells transfected with CF877 ZFNs, and (C) Sequences of *SuRA* alleles from tobacco plants transfected with SR2163 ZFNs (matching sequence from the *SuRB* gene is also shown). Numbers of each allele identified are shown in parentheses. ZFN recognition sites are in bold orange print.

**Figure S4 Schematics of donor templates and primers used for gene targeting  
experiments at the human *VEGF-A* and *IL2R $\gamma$*  genes.**

Donor templates were constructed for OPEN ZFN pairs that cleave at the VF2468 and VF2471 sites and for a previously described four-finger ZFN pair that cleaves in the *IL2R $\gamma$*  gene (Urnov et al., 2005). Each donor template consists of 1.5 kb of genomic DNA sequence centered on the cleavage site and introduces either an 11 bp insertion encoding a *SalI* restriction site at the center of the cleavage site (for *VEGF-A* targets) or a translationally silent point mutation which creates a *BsrBI* restriction site adjacent to the ZFN cleavage site (for *IL2R $\gamma$* ) (Urnov et al., 2005). Arrows indicate primers used for limited-cycle PCR/restriction digest assay described in the text.

**Figure S5 Comparison of PCR-based and Southern blot methods for assaying gene targeting efficiencies.**

Limited-cycle PCR and Southern blot were used to assess gene targeting efficiencies in cells treated with different ZFNs under various conditions. Dotted red line represents where data points would fall if the two methods were perfectly concordant.

**Figure S6 DNA sequence analysis of endogenous human alleles that have undergone gene targeting induced by ZFNs**

(A-C) Sequences of alleles sequenced from human K562 cells transfected with (A) VF2468 ZFNs and donor, (B) VF2471 ZFNs and donor, and (C) *IL2R $\gamma$*  ZFNs and donor. Data are presented as in Figure S3.

**Figure S7 OPEN ZFNs induce stable multi-allelic alterations of an endogenous human gene**

FISH analysis indicated that K562 cells harbor four alleles of the *VEGF-A* gene (A, upper left panel) and we therefore wished to determine how many of these alleles could be altered in a single cell. Limiting dilution cloning was used to isolate single cell clones from two vinblastine-treated K562 cell populations which had undergone high frequencies of gene targeting: 45% and 26% with the VF2468 and VF2471 ZFN pairs, respectively (B, upper panel). (The VF2468 ZFN-treated cells also showed evidence of a 625 bp deletion (blue asterisks), a finding confirmed by sequencing of alleles from these cells (data not shown).) Individual clones from VF2468 ZFN-treated cells harbored no, one, three, or four alleles that had undergone a gene targeting event with some of the

clones also containing the 625 bp deletion (B, lower left panel). Genotype analysis of individual cell clones from the VF2471 ZFN-treated cells revealed clones in which no, one, two, or three *VEGF* alleles had undergone a gene targeting event (B, lower right panel). All individual cell clones were genotyped more than thirty-five days post-transfection, demonstrating that gene targeting events we observed are stably maintained. FISH analysis of three cell clones in which all four *VEGF-A* alleles had undergone the VF2468 ZFN-induced gene targeting event confirmed the continued presence of four copies of *VEGF-A* per cell (A, upper right, and lower left and right panels). FISH was performed with a probe for *VEGF-A* (red) and a control probe for 14q which is present in two copies per cell (green). Taken together, these results demonstrate that OPEN ZFNs can be used to induce permanent alterations in as many as four alleles in a single human cell.

**Table S1: Sequences and B2H assays of zinc finger arrays selected by OPEN**

The subset of zinc finger arrays we converted to ZFNs for testing in human or plant cells are highlighted in green. As noted in the text, a small number of selections were performed in a single step on plates with fixed concentrations of IPTG, 3-AT, and streptomycin. For these selections, IPTG concentration shown is in  $\mu\text{M}$ , 3-AT concentration shown is in mM, and streptomycin concentration shown is in  $\mu\text{g/ml}$ . Fold-activation of transcription in the B2H system was determined by comparing *lacZ* expression from B2H reporter strains harboring a zinc finger array (fused to a fragment of the yeast Gal11P protein) to matched strains harboring a control that does not express a zinc finger array. All determinations were performed in triplicate and the means and

standard deviations of these means are shown. Certain target sites cause a higher basal level of *lacZ* expression in the quantitative B2H assay and the finger arrays tested on these sites may exhibit lower apparent fold-activations due to this effect (affected values are highlighted in yellow). Certain zinc finger arrays (marked by a “+” in the last column) harbor additional mutations outside of the three randomized recognition helices that were most likely introduced during library construction. The identities of these various mutations are available upon request.

Although half-site EG502L did not yield OPEN zinc-finger arrays which activated by more than three-fold, we note that the basal level of transcription from the B2H reporter bearing this site was high, a situation which can artifactually lower the apparent fold-activation observed.

25 of the 28 OPEN selections we performed for the human *VEGF-A*, human *HoxB13*, human *CFTR*, and tobacco *SuRA* genes yielded zinc-finger arrays whose sequences closely resembled one another. However, one of these 25 selections (for half-site VF3540R) yielded fingers with sequences that appeared to bind to an alternative site (discussed below). For 22 of the other 24 selections, we obtained at least one zinc-finger array which activated *lacZ* expression by more than three-fold in the quantitative B2H assay. The two remaining selections were also deemed to be successful because their reporters possessed a high basal level of transcription which can artifactually mask a higher true fold-activation. For one selection (site VF3540R), examination of the recognition helix sequences of the arrays obtained from OPEN selection suggests that

these finger arrays do not bind to the intended nine base pair site but rather to the site that is shifted 3' in register by 1 bp (i.e.--these arrays bind to the site: 5'-GCG GCG GAC-3' instead of 5'-GGC GGC GGA-3'). For example, for finger 1, the Asp and Glu selected at position -1 and the Asn selected at position 3 of the recognition helix would be expected to specify cytosine and adenine, respectively, and not the adenine and guanine actually present in the subsite. Similarly, for finger 3, the Arg selected at position -1 and the Ser and Asp selected at position 3 of the recognition helix would be expected to specify guanine and cytosine, respectively, and not the cytosine and guanine actually present. Thus, this selection was deemed to be unsuccessful.

**Table S2 ZFN target sites and associated zinc-finger arrays and ZFNs**

All target sites and half-sites shown are written 5' to 3'. Names of zinc-finger arrays are the same as in Table S1.

**Table S3: Primers used in this study**

### **Supplemental References:**

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

Bae, K. H., Do Kwon, Y., Shin, H. C., Hwang, M. S., Ryu, E. H., Park, K. S., Yang, H. Y., Lee, D. K., Lee, Y., Park, J., *et al.* (2003). Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat Biotechnol* 21, 275-280.

Bryson, V., and Szybalski, W. (1952). Microbial Selection. *Science* 116, 45-51.

Cathomen, T., Stracker, T. H., Gilbert, L. B., and Weitzman, M. D. (2001). A genetic screen identifies a cellular regulator of adeno-associated virus. *Proc Natl Acad Sci U S A* 98, 14991-14996.

Joung, J. K., Ramm, E. I., and 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, 7382-7387.

Liu, Q., Xia, Z., Zhong, X., and Case, C. C. (2002). Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *J Biol Chem* 277, 3850-3856.

Mandell, J. G., and Barbas, C. F., 3rd (2006). Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res* 34, W516-523.

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

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

Pruett-Miller, S. M., Connelly, J. P., Maeder, M. L., Joung, J. K., and Porteus, M. H. (2008). Comparison of Zinc Finger Nucleases for Use in Gene Targeting in Mammalian Cells. *Molecular Therapy*, doi: 10.1038/mt.2008.1020.

Sander, J. D., Zaback, P., Joung, J. K., Voytas, D. F., and Dobbs, D. (2007). Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucleic Acids Res* 35, W599-605.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98, 503-517.

Thibodeau, S. A., Fang, R., and Joung, J. K. (2004). High-throughput beta-galactosidase assay for bacterial cell-based reporter systems. *Biotechniques* 36, 410-415.

Thibodeau-Beganny, S., and Joung, J. K. (2007). Engineering Cys2His2 Zinc Finger Domains Using a Bacterial Cell-Based Two-Hybrid Selection System. *Methods in Molecular Biology* 408, 317-334.

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

Wright, D. A., Thibodeau-Beganny, S., Sander, J. D., Winfrey, R. J., Hirsh, A. S., Eichinger, M., Fu, F., Porteus, M. H., Dobbs, D., Voytas, D. F., and Joung, J. K. (2006). Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat Protoc* 1, 1637-1652.



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

**Figure S1 ZiFiT 3.0 -- Web-based software for identifying potential sites targetable by OPEN selection**

**A.**

IOWA STATE UNIVERSITY  
Dobbs & Voytas Laboratories  
**ZiFiT Version 3.0**

**ZiFiT: Zinc Finger Targeter (V3.0)**

Sequence:

Left Array  Spacer  Right Array

Position 1      Position 2      Position 3

GGG  GGA  GGT  GGC  
 GAG  GAA  GAT  GAC  
 GTA  GTT  GTC  
 GCA  GCT  GCC  
 AGG  AGA  AGT  AGC  
 AAG  AAA  AAT  AAC  
 ATA  ATT  ATC  
 ACA  ACT  ACC

GGG  GGA  GGT  GGC  
 GAG  GAA  GAT  GAC  
 GTA  GTT  GTC  
 GCA  GCT  GCC  
 AGG  AGA  AGT  AGC  
 AAG  AAA  AAT  AAC  
 ATA  ATT  ATC  
 ACA  ACT  ACC

GGG  GGA  GGT  GGC  
 GAG  GAA  GAT  GAC  
 GTA  GTT  GTC  
 GCA  GCT  GCC  
 AGG  AGA  AGT  AGC  
 AAG  AAA  AAT  AAC  
 ATA  ATT  ATC  
 ACA  ACT  ACC

IOWA STATE UNIVERSITY  
**Becoming the best.**  
Ames, Iowa 50011, (515) 284-4111. Published by: University Relations, [online@iastate.edu](mailto:online@iastate.edu)  
Copyright © 1995-2005, Iowa State University of Science and Technology. All rights reserved.  
[Non-Discrimination Statement and Information Disclosure](#)

**B.**

IOWA STATE UNIVERSITY  
Dobbs & Voytas Laboratories  
**ZiFiT Version 3.0**

293 gCGCGCCTCGGCCTCGGCGGCGCC 318  
293 cCGCGGAGCGGCGGAGCGGCGGCGGg 318

Finger	POOL	TRIPLET	REFERENCE NUMBER	MODULE SOURCE
Left1	1	GCG	P1-9	Joung
Left2	2	GCG	P2-9	Joung
Left3	3	GAG	P3-5	Joung

Finger	POOL	TRIPLET	REFERENCE NUMBER	MODULE SOURCE
Right1	1	GCC	P1-11	Joung
Right2	2	GCG	P2-3	Joung
Right3	3	GCC	P3-3	Joung

Homo sapiens (human) Build 36

323 cCGCCTCGCGGACCGAGCGGAGg 348  
323 gCGGAGCGGCGGCGGAGCGGAGg 348

Finger	POOL	TRIPLET	REFERENCE NUMBER	MODULE SOURCE
Left1	1	GCG	P1-9	Joung
Left2	2	GAG	P2-5	Joung
Left3	3	GCC	P3-3	Joung

Finger	POOL	TRIPLET	REFERENCE NUMBER	MODULE SOURCE
Right1	1	GAG	P1-5	Joung

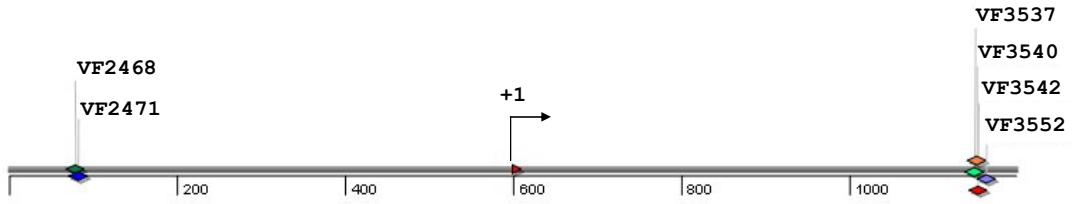
IOWA STATE UNIVERSITY  
**Becoming the best.**  
Ames, Iowa 50011, (515) 284-4111. Published by: University Relations, [online@iastate.edu](mailto:online@iastate.edu)  
Copyright © 1995-2005, Iowa State University of Science and Technology. All rights reserved.  
[Non-Discrimination Statement and Information Disclosure](#)

**Figure S2 Schematic of target sites in the *EGFP*, *VEGF-A*, *HoxB13*, and *CFTR* genes**

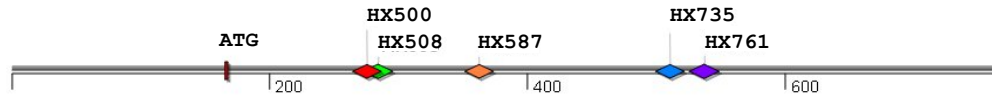
**A.**



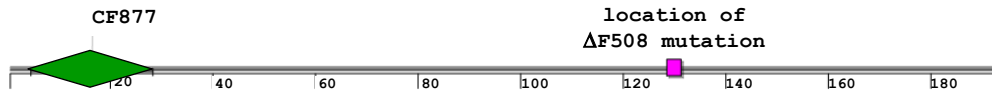
**B.**



**C.**



**D.**



## Figure S3 DNA sequence analysis of endogenous human and plant genes mutagenized by OPEN ZFNs

### A.

		% mutants
<b><u>HX587, Experiment #1</u></b>		
wild-type:	TCGGCGGAGCCGCCAAAGCAA <b>TGCCACCCA</b> <b>tgcccTGGGGTGCC</b> CCAGGGGACGTCCCCA (94x)	
	HX587 ZFN site	
insertions:	 CTG CCG CCTG CCCG CCTT	(1x) (2x) (14x) (1x) (1x)
		<b>16.8%</b>
<b><u>HX587, Experiment #2</u></b>		
wild-type:	TCGGCGGAGCCGCCAAAGCAA <b>TGCCACCCA</b> <b>tgcccTGGGGTGCC</b> CCAGGGGACGTCCCCA (100x)	
	HX587 ZFN site	
insertion:	 CCTG	(8x)
		<b>7.4%</b>
<b><u>HX587, Experiment #3</u></b>		
wild-type:	TCGGCGGAGCCGCCAAAGCAA <b>TGCCACCCA</b> <b>tgcccTGGGGTGCC</b> CCAGGGGACGTCCCCA (102x)	
	HX587 ZFN site	
insertions:	 CCTG CCG	(2x) (2x)
deletion:	█	(1x)
		<b>4.7%</b>

### B.

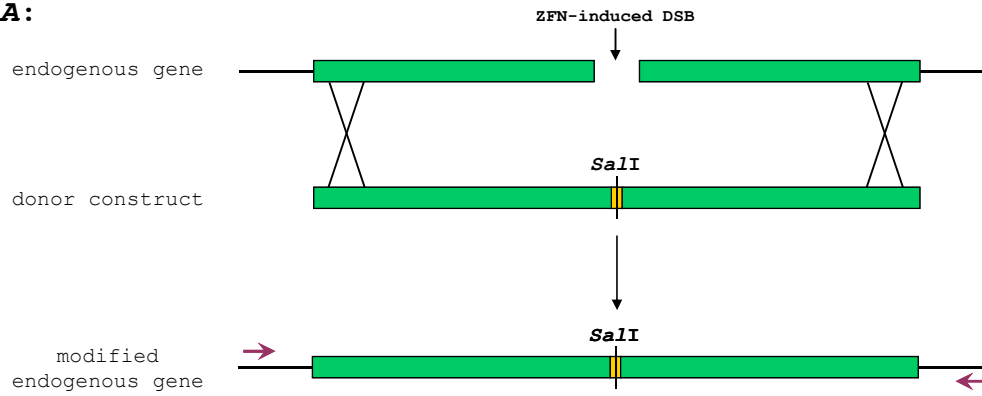
		% mutants
<b><u>CF877, Experiment #1</u></b>		
wild-type:	CCAGACTT <b>CACTTCTAAtggtgATTATGGGAGAACTGGAG</b> (188x)	
	CF877 ZFN site	
insertions:	 CCCA TGACA TGTC TCTCA TGTGA	(1x) (1x) (1x) (1x) (1x)
		<b>2.6%</b>
<b><u>CF877, Experiment #2</u></b>		
wild-type:	CCAGACTT <b>CACTTCTAAtggtgATTATGGGAGAACTGGAG</b> (198x)	
	CF877 ZFN site	
insertions:	 TGTCA TGTGA	(1x) (1x)
		<b>1.0%</b>
<b><u>CF877, Experiment #3</u></b>		
wild-type:	CCAGACTT <b>CACTTCTAAtggtgATTATGGGAGAACTGGAG</b> (200x)	
	CF877 ZFN site	
<b>no insertions or deletions</b>		
		<b>0%</b>

### C.

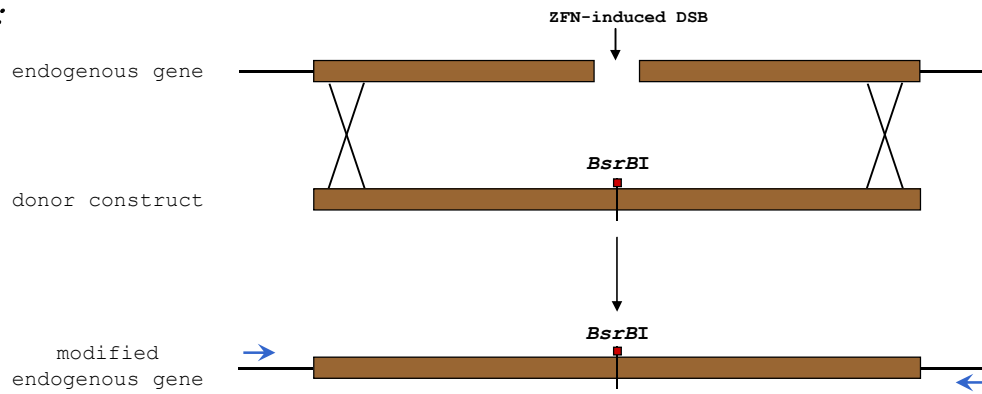
	SR2163 ZFN site
wild-type <i>SuRA</i> :	GCACACACATACCTGGG <b>GAATCCTT</b> Ctaatga <b>GCGGAGAT</b> CCTTCCTAATATGTTG
wild-type <i>SuRB</i> :	GCACACACATACCTGGG <b>GAATCCTT</b> Ctaatga <b>GCGGAGAT</b> CCTTCCTAATATGCTG
deletion in <i>SuRA</i> :	GCACACACATACCTGGG <b>GAATCCTT</b> █TGAGCGGAGATCCTTCCTAATATGCTG

**Figure S4 Schematics of donor templates and primers used for gene targeting experiments at the human *VEGF-A* and *IL2R $\gamma$*  genes.**

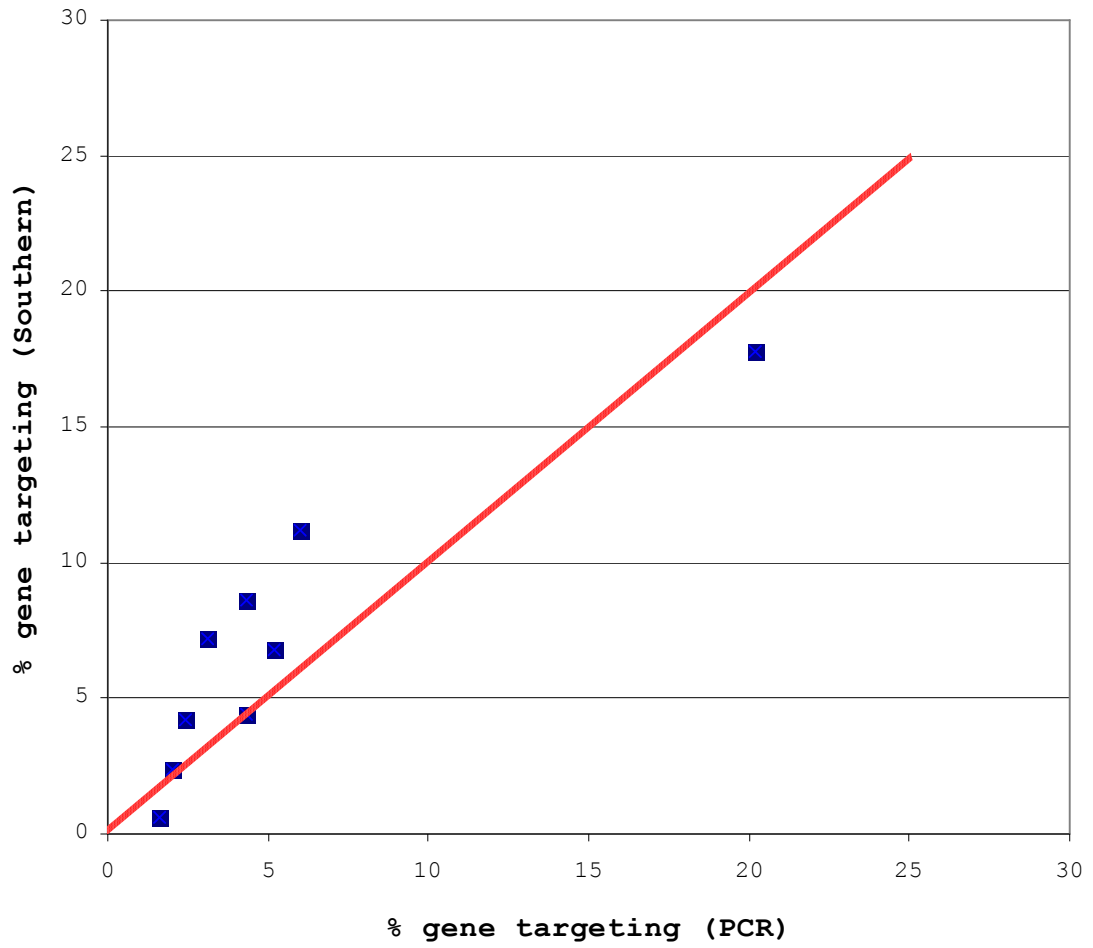
***VEGF-A*:**



***IL2R $\gamma$* :**



**Figure S5 Comparison of PCR-based and Southern blot methods for assaying gene targeting efficiencies**



**Figure S6 DNA sequence analysis of endogenous human alleles that have undergone gene targeting induced by ZFNs**

**A.**

VF2468 ZFN Pair C:		% of total
	VF2468 ZFN site	
wild-type:	AAGCTGGGTGAATGGAGCGAGCAGCGTCTtcgagAGTGAGGACGTGTGTCTGTGTGGGT (18x)	15.9%
GT events:	           TGGGGTCGACC	(56x) 49.6%
GT+insertion:	TGGGGTCGACC	(2x) 1.8%
insertions:	GATTC           GATC           GTTC           GA           GT           GTCGACC	(7x) 16.8% (3x) (5x) (1x) (2x) (1x)
deletions:		(1x) 15.9% (3x) (1x) (2x) (1x) (1x) (3x) (2x) (1x) (1x) (1x) (1x)

**B.**

VF2471 ZFN Pair B:		% of total
	VF2471 ZFN site	
wild-type:	GCTGGGTGAATGGAGCGAGCAGCGTCTTCgagagtGAGGACGTGTGTGTCTGTGTGGGTGAGT (17x)	14.9%
GT events:	           TGGGGTCGACC	(54x) 47.4%
insertions:	AGAG           AG           AGG	(14x) 21.9% (9x) (2x)
deletions:		(1x) 15.8% (1x) (1x) (3x) (2x) (2x) (4x) (2x) (2x)

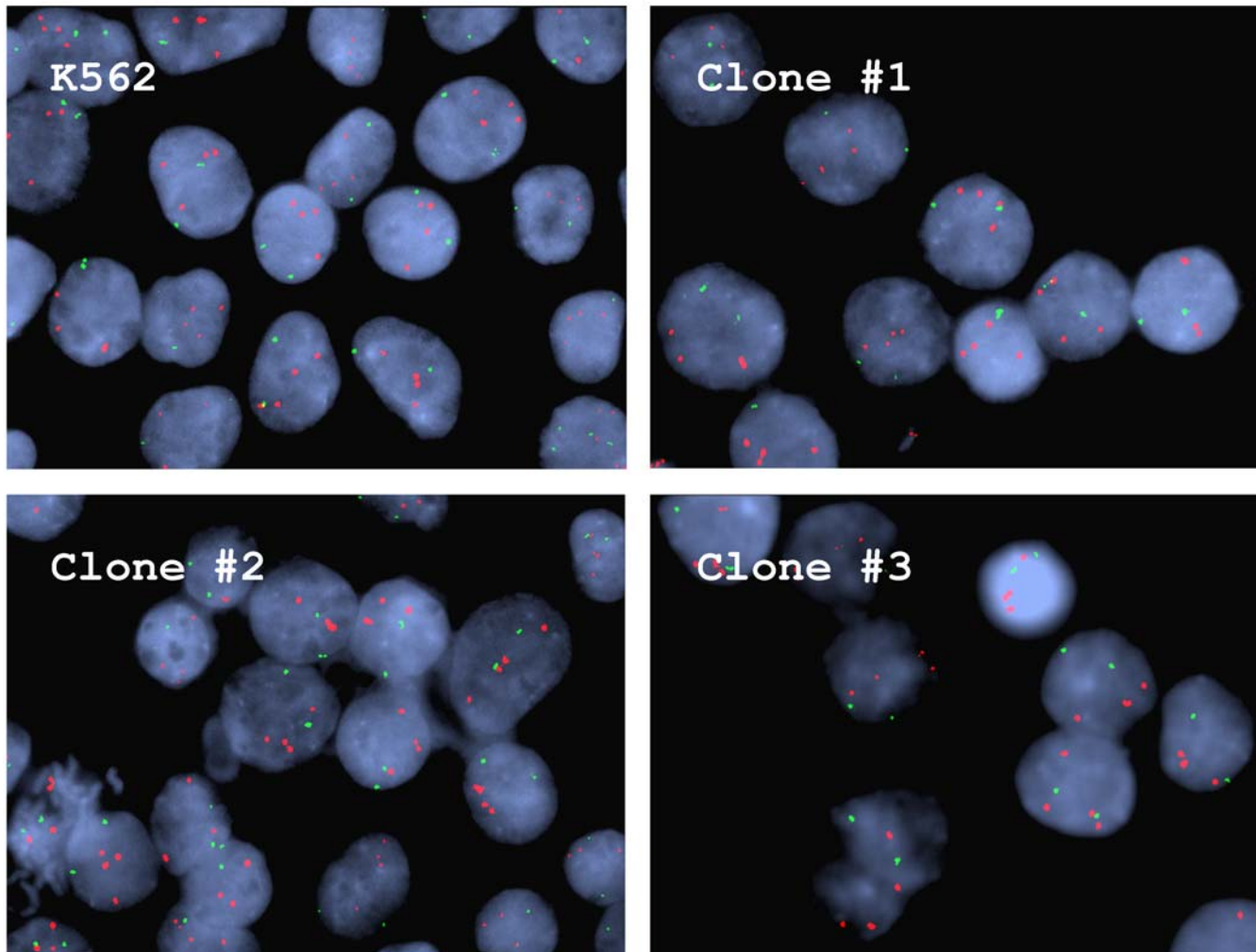
### C.

<i>IL2Ry</i> ZFN pair:		% of total
wild-type:	<div style="text-align: center;"> <p><i>IL2Ry</i> ZFN site</p> <p>CACGTTTCGTGTT<b>CGGAGCCGCTTTaacc</b>ACTCTGTGGAAGTGCTCAGCATTGGAGTGAATG</p> </div>	(39x) <b>32.0%</b>
GT events:	<div style="text-align: center;"> </div>	(31x) <b>25.4%</b>
GT+insertion:	<div style="text-align: center;"> </div>	(11x) <b>9.0%</b>
<i>IL2Ry</i> ZFN site		
wild-type:	CACGTTTCGTGTT <b>CGGAGCCGCTTTaacc</b> ACTCTGTGGAAGTGCTCAGCATTGGAGTGAATG	
insertions:	<div style="text-align: center;"> </div>	(1x) (26x) <b>24.6%</b> (3x)
deletions:	<div style="text-align: center;"> </div>	(1x) <b>9.0%</b> (1x) (2x) (2x) (1x) (1x) (2x) (1x)



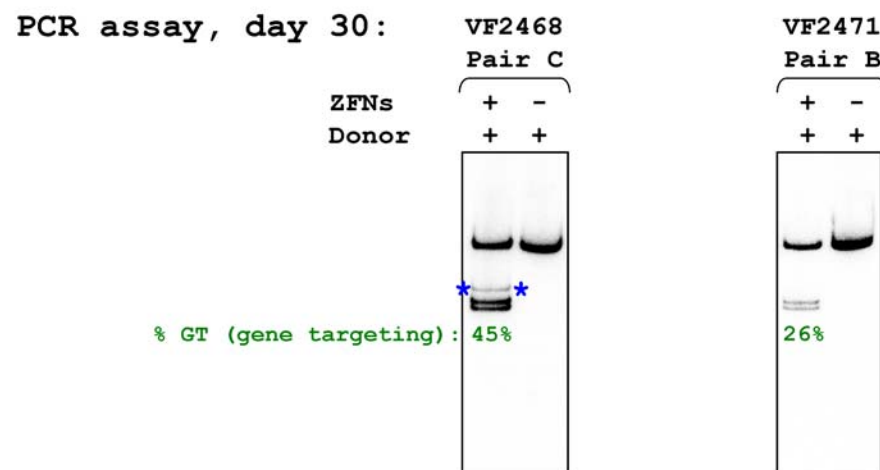
**Figure S7 OPEN ZFNs induce stable multi-allelic alterations of an endogenous human gene**

**A.**

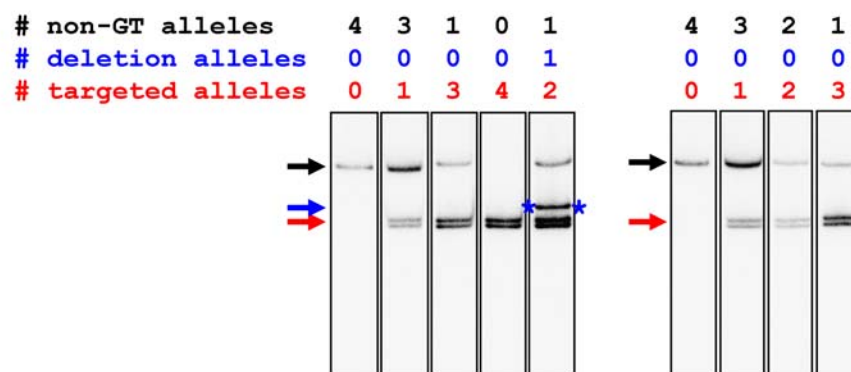


**Figure S7 OPEN ZFNs induce stable multi-allelic alterations of an endogenous human gene**

**B.**



Single cell clone genotyping:



**Table S2**

Full ZFN site name	Full ZFN site sequence	Half-site name	Half-site sequence	# of GXX triplets	Successful B2H selection?	Arrays converted to ZFNs	ZFN Pair A	ZFN Pair B	ZFN Pair C	ZFN Pair D
EG223	cTACCCCGACcacatGAAGCAGCAc	EG223L	gGTCGGGGTAg	3	YES	OZ001	X	X		
						OZ009			X	X
		EG223R	tGAAGCAGCAc	3	YES	OZ013	X		X	
						OZ017		X		X
EG292	cACCATCTTcttcaagGACGACGGCa	EG292L	aGAAGATGGTg	3	YES	OZ030	X	X		
						OZ034			X	X
		EG292R	gGACGACGGCa	3	YES	OZ039	X		X	
						OZ044		X		X
EG382	gGGCATCGACTtcaagGAGGACGGCa	EG382L	aGTCGATGCCc	3	YES	OZ051	X	X		
						OZ052			X	X
		EG382R	gGAGGACGGCa	3	YES	OZ059	X		X	
						OZ060		X		X
EG502	gATCCGCCACaacatcGAGGACGGCa	EG502L	tGTGGCGGATa	3	YES**	OZ068	X	X		
						OZ069			X	X
		EG502R	cGAGGACGGCa	3	YES	OZ076	X		X	
						OZ082		X		X
EG568	cGGCGACGGCcccgtGCTGCTGCCc	EG568L	gGCCGTCGCCg	3	YES	OZ086	X	X		
						OZ092			X	X
		EG568R	tGCTGCTGCCc	3	YES	OZ099	X		X	
						OZ102		X		X
CF877	tCACTTCTAAtggtgaTTATGGGAGa	CF877L	aTTAGAAGTGa	2	YES	OZ118	X			
		CF877R	aTTATGGGAGa	1	YES	OZ134	X			
VF2471	cAGCGTCTTCgagagtGAGGACGTGt	VF2471L	cGAAGACGCTg	3	YES	OZ146	X	X		
						OZ154			X	X
		VF2471R	tGAGGACGTGt	3	YES	OZ162	X		X	
						OZ166***		X		X
VF3537	tACCACCTCctccccgGCCGGCGGCg	VF3537L	aGGAGGTGGTa	3	YES	OZ172	X	X		
						OZ177			X	X
		VF3537R	gGCCGGCGGCg	3	YES	OZ182	X		X	
						OZ185		X		X
VF3540	cACCTCCTCCcgggccGGCGGCGGAc	VF3540L	gGGAGGAGGTg	3	YES	---				
		VF3540R	cGGCGGCGGAc		NO	---				

**Table S2**

VF2468	gAGCAGCGTCttcgaGAGTGAGGAc	VF2468L	aGACGCTGCTc	3	YES	OZ211	X	X			
						OZ217			X	X	
		VF2468R	aGAGTGAGGAc	<b>2</b>	YES	OZ227	X		X		
						OZ230		X		X	
VF3542	cCTCCTCCCCggccgGCGGCGGACa	VF3542L	cGGGGAGGAGg	3	YES	OZ233	X	X			
						OZ234			X	X	
		VF3542R	gGCGGCGGACa	3	YES	OZ244	X		X		
						OZ246		X		X	
VF3552	gGCCGGCGGCggacaGTGGACGCGg	VF3552L	cGCCGCCGGCc	3	YES	OZ253	X	X			
						OZ254			X	X	
		VF3552R	aGTGGACGCGg	3	YES	OZ264	X		X		
						OZ272		X		X	
HX508	aCCCAGCGGCgcctacGCTGATGCct	HX508L	cGCCGCTGGGt	3	YES	OZ276	X	X			
						OZ277			X	X	
		HX508R	cGCTGATGCct	3	YES	OZ284	X		X		
						OZ285		X		X	
HX761	cCCCAGCCGCcccactGAGTTTGCct	HX761L	gGCGGCTGGGg	3	YES	OZ297	X	X			
						OZ300			X	X	
		HX761R	tGAGTTTGCct	<b>2</b>	YES	OZ303	X		X		
						OZ304		X		X	
HX500	gACCAGCCACccagcGGCGCCTACg	HX500L	gGTGGCTGGTc	3	YES	---					
		HX500R	cGGCGCCTACg	<b>2</b>	NO	---					
HX587	aTGCCACCCAtgcccTGGGGTGCCc	HX587L	aTGGGTGGCAat	<b>2</b>	YES	OZ338	X	X			
						OZ340			X	X	
		HX587R	cTGGGGTGCCc	<b>2</b>	YES	OZ345	X		X		
						OZ349		X		X	
HX735	gAGACTCCCAcggccGGGGAAGAGt	HX735L	gTGGGAGTCTc	<b>1</b>	YES	OZ357	X	X			
						OZ363			X	X	
		HX735R	cGGGGAAGAGt	3	YES	OZ370	X		X		
						OZ371		X		X	
HX2119	gAGCGCCAGAttaccaTCTGGTTTCa	HX2119L	aTCTGGCGCTc	<b>2</b>	NO	---					
		HX2119R	aTCTGGTTTCa	<b>1</b>	NO	---					
SR2163	gGAATCCTTctaataGAGCGGAGATc	SR2163L	aGAAGGATTc	<b>2</b>	YES	OZ391	X				
		SR2163R	aGGCGGAGATc*	3	YES**	OZ396	X				

## Table S2

### Notes:

\*This site contained a *dam* methylation site (GATC) and therefore the 3' most C was mutated in the B2H selection strain and B2H reporter strain.

\*\*Although these arrays failed to activate transcription by more than three-fold in the quantitative B2H assay, they were deemed to be successful because their reporters possessed a high basal level of transcription which can often mask a higher true fold-activation.

\*\*\*A PCR-induced mutation was introduced into this array when it was transferred to the human ZFN expression vector (a Thr→Ala mutation at position 8 of its F1 recognition helix).

**Table S3**

Primer Name	Sequence
OK61	5'-GGGTAGTACGATGACGGAACCTGTC-3'
OK1424	5'-GAGCGCCCTTCCAGTGTCGC-3'
OK1425	5'-CGCATAACAGATCCGACACTGAAACGG-3'
OK1426	5'-GTGTCGGATCTGTATGCGAAATTTCTCC-3'
OK1427	5'-TCGGCATTGGAATGGCTTCTCG-3'
OK1428	5'-GCCATTCCAATGCCGAATATGCA-3'
OK1429	5'-CCCTCAGGTGGGTTTTTAGGTG-3'
OK1430	5'-GGGGAGCGCCCTTCCAGTGTCGC-3'
OK1432	5'-GTGCAGAGGATCCCCTCAGGTGGGTTTTTAGGTG-3'
OK1681	5'-GGCTCTCTGTACATGAAGCAACTCC-3'
OK1682	5'-TCACAGCCTGAAAATTACCCATCC-3'
OK1706	5'-CCCTCTCTCTCTCCTTCTTCTT-3'
OK1711	5'-GTGCCCCTTCTCTGTCAACCTCTAT-3'
OK1713	5'-GCATGCTTTGATGACGCTTCTGTA-3'
OK1718	5'-GATTGCTCTACTTCCCCAAATCACTG-3'
OK1733	5'-GGATGGAGCCAAGGATATCGAAG-3'
OK1734	5'-GTAAGTGGCCATAGGCTGGTAGGTT-3'
OK1736	5'-TACGCTGATGCCTGCTGTCAACTAT-3'
OK1738	5'-ACTGTCCACAGGCAACAGGGAGT-3'
OK1773	5'-GAGAGCCGTTCCCTCTTTG-3'
OK1776	5'-GGAAGATGTGGAGAGTTGGAGGAAA-3'
OK1777	5'-CGACTTCTCTCTGGAGCTCTTGCTA-3'
OK1823	5'-AAGTCAAGTTGTTTCAGGGGGCTAAG-3'
OK1824	5'-AAGCCACTAACTTTCCCTTGTGC-3'
OK1838	5'-AGGGGAGGCATTAGATTCAAGTCAG-3'
OK1845	5'-GATTCAACCAGACAGATAGAAGG-3'
OK1846	5'-TTACTGTCTCATCCTTTACTCC-3'
OS216	5'-GAAGGCAAACCTCAGTGGGGCGGC-3'
DVO4429	5'-TTTCAGTTTCGTGGCGCTCTCGTC-3'
DVO4444	5'-CCATCTCCCTCTTCTCCGC-3'
DVO4445	5'-GCGGCGGCTCCATCTCCCTC-3'
DVO4461	5'-GATAATACAAGGCCTAGAACTAGCTCTGTAGC-3'
DVO4462	5'-GTGATGGCAGTTTCATCATGAATG-3'
DVO4565	5'-CCCCTTTTAGCTTAGTAGTGTCTCATG-3'
T3 (Invitrogen)	5'-ATTAACCCTCACTAAAGGGA-3'
M13 (Invitrogen)	5'-CAGGAAACAGCTATGAC-3'