megaTALs: a novel rare-cleaving nuclease platform for therapeutic genome engineering

Sandrine Boissel

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:

David Baker, Chair

Andrew Scharenberg

Barry Stoddard

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

megaTALs: a novel rare-cleaving nuclease platform for therapeutic genome engineering

Sandrine Boissel

Chair of the Supervisory Committee:

David Baker

Biochemistry

Rare-cleaving endonucleases have emerged as important tools for making targeted genome modifications. While multiple platforms are now available to generate reagents for research applications, each existing platform has significant limitations in one or more of three key properties necessary for therapeutic application: efficiency of cleavage at the desired target site, specificity of cleavage (i.e. rate of cleavage at "off-target" sites), and efficient/facile means for delivery to desired target cells. Here, we describe the development of a single-chain rare-cleaving nuclease architecture, which we designate "megaTAL", in which a TAL effector nuclease is used to "address" a site specific meganuclease adjacent to a single desired target site. This architecture allows the generation of extremely active and hyperspecific compact nucleases that are compatible with all current viral and non-viral cell delivery methods.

Table of contents

List of Figures

List of Tables

Chapter 1

Introduction: A primer on genome engineering

1.1 Gene manipulation through the years

A brief history of genetic engineering

The ability to manipulate genes is a powerful tool that was first employed circa 10,000 BC, when farmers began selecting and crossing agricultural crops and livestock for desired traits. These early geneticists had little idea of the effect they were having, doing nothing more than choosing the sturdiest, most flavorful plants and animals to raise. It was not until Mendel's famous discovery in 1866, on the inheritance pattern of traits in pea plants, that the concept of hardcoded elements, later called genes, was first introduced.[1] However this idea was largely ignored until several decades later, when the molecular unit of inheritance, DNA, was first identified in 1944 and its structure subsequently identified by Watson and Crick.[2,3]

Over the course of the next twenty years, a number of discoveries, including the existence of DNA ligase, restriction enzymes and plasmids, enabled scientists to precisely and predictably manipulate the genetic code for the first time.[4–6] This led to the generation of the first recombinant DNA molecule by Paul Berg and the first genetically modified organism by Boyer and Cohen in 1972.[7,8] Since this time, advances reducing both the difficulty and cost of copying, synthesizing and sequencing DNA have allowed recombinant DNA technology to become widespread. Furthermore, methods for introducing exogenous DNA have been developed for a number of organisms, including many plants and animals.

A brief history of gene therapy

With the advent of recombinant DNA technology, gene therapy emerged as a possible answer to treating genetic disorders. First proposed in the early 1970s as a way to replace defective genes, Friedmann and Roblin warned readers of the potential dangers of *in vivo* gene addition.[9] Nonetheless, the promise of eradicating seriously debilitating and life-threatening disorders was great and so the first clinical trial for gene therapy was performed in 1990 to treat 4-yr old Ashanti DeSilva, suffering from adenosine deaminase deficient severe combined immunodeficiency (ADA-SCID).[10] This initial trial found moderate success, having no apparent adverse effects and resulting in long-term persistence of modified cells,[11] encouraging a number of further trials soon thereafter.

The most notable of these was a clinical trial performed to eradicate X-linked SCID in ten children that resulted in initial excitement over the effective correction of the immunodeficiency in nine of these patients.[12–14] This success was later overshadowed by the discovery that five of the treated patients developed leukemia as a result of random integration of the transgene into regions of the genome neighboring proto-oncogenes.[15,16] Subsequent studies revealed that many viral vectors preferentially integrate transgenes either within or immediately 5' of gene coding regions which can lead to inappropriate gene expression. Furthermore, while only an extremely small subset of treated cells may initially carry transgene integrations that turn on proto-oncogenes, the highly proliferative nature of these cells lead to their clonal dominance of the population.[17–20] These devastating findings revealed the need for more stringent and directed delivery methods to ensure safe use of therapeutic transgene delivery.

1.2 Targeted genome modification

The obvious solution to the issues arising from random transgene integration was to develop methods by which a DNA modification can be introduced at a pre-determined locus within the genome. Targeted genome modification can refer to transgene delivery to a safe harbor locus, as well as direct *in cellulo* manipulation of the therapeutic gene of interest.

Safe harbor transgene delivery

Safe harbor transgene delivery makes use of known genomic loci, such as the AAVS1 locus, at which exogenous material can be introduced without resulting pathologic effects.[21,22] Recombinases and transposases are two classes of DNA-modifying enzymes that can be used for safe harbor delivery due to their ability to catalyze the transfer of genetic material from a donor to a recipient. A major advantage of these reagents is that they are autonomous in that they do not rely on the cell's DNA repair machinery, making them able to elicit the transfer of DNA regardless of the state of the recipient cell.

The major drawback of using recombinases for targeted engineering is that they require an initial genome modification step to introduce their natural recognition sequences at the desired locus. Some success has been found in re-engineering the native specificity of recombinases, however, this has proven to be a non-trivial endeavor. Transposases, on the other hand, exhibit no natural sequence specificity and while some work has shown that these enzymes can be fused to DNA-binding domains to bring about gene targeting, the likelihood of sequence-independent gene transfer remains high.[23,24] Furthermore, while safe harbor transgene delivery would provide a much safer alternative to semi-random integration, it remains to be seen whether a definitively safe locus can be identified and off-target integration can be truly ruled out.

Endogenous genetic manipulation

Targeting the root of the cause, the endogenous DNA from which the genetic disorder arises, provides a number of significant advantages over transgene delivery that simply supplements patients with a corrected copy of the gene. Modifying the gene in its native environment ensures normal regulation and expression of its protein product, avoids any potential issues arising from copy number variation and can be used to treat disorders arising from dominant negative mutations. Additionally, targeted gene knockout may be achieved by altering a DNA sequence to encode a non-functional or truncated protein or by removing a gene entirely. This use opens up the potential for a wider range of gene therapy applications, for example, to eliminate virus in infected patients by knocking out a critical viral protein.[25] Lastly, altering endogenous genomic DNA would facilitate the alteration of both coding and non-coding sequences, thus enabling manipulation of such non-coding elements as microRNAs, gene regulatory elements and repetitive DNA for either therapeutic or research purposes.

Classical gene targeting has been used to alter endogenous genetic material by providing a donor DNA fragment with homology arms matching the DNA flanking the site at which alteration is desired. The two DNA molecules can undergo homologous recombination (HR), resulting in copying of the donor sequence to the targeted genomic site. While this strategy works powerfully in mice and is commonly used to generate transgenic mouse models, a much lower efficiency is observed in human cells, diminishing its utility for therapeutic use. Triplex forming oligonucleotides (TFOs), short single-stranded DNA fragments that can anneal to complementary genomic sequences, can be fused to a double-stranded DNA donor to increase levels of homologous recombination 100-fold over classical HR methods. However, TFOs exhibit a number of disadvantages that do not make them an ideal choice for targeted genome

modification. Namely, they have thus far been only used to correct point mutations, typically require purine-rich TFO recognition sites and are difficult to deliver into cells.[23]

The formation of a double-strand break at the desired gene targeting locus was found to further increase the efficiency of gene transfer, when made in the presence of a substrate for homologous recombination.[26] These breaks may be precisely administered using nucleases that exhibit DNA specificity and result in a 1000-fold increase in levels of HR over classical gene targeting.[23] Moreover, repeated DSB break formation can lead to the introduction of indels at the site of the DNA break due to mutagenic non-homologous end joining (mutNHEJ). These indels can create nonsense and missense mutations in genes to achieve targeted gene knockout.[24] While nucleases offer perhaps the safest and most effective way to administer genomic modifications via DSBs, issues such as the engineerability, specificity and deliverability of these enzymes still remain before they can be considered ideal therapeutic reagents.

1.3 Essential properties of nuclease platforms for targeted gene therapy

Engineerability

Manipulation of endogenous genomic DNA relies on the ability to generate reagents capable of affecting a target locus. While natural and hybrid nucleases with pre-encoded sequence-specificity exist, these represent a finite set and do not cover the entire space of potential sequences needed to target any locus of interest. For this reason, nucleases must be engineerable – their DNA-binding interfaces need to be amenable to redesign in order to obtain novel target specificities.

Specificity

Perhaps the most important attribute required of a therapeutic-grade nuclease is its genomic specificity. The creation of off-target double-strand breaks may lead to inappropriate gene targeting and knockout, as well as a number of different chromosomal rearrangements.[27–32] The severity of such unintended alterations is highly dependent on the affected locus, however they have the potential to lead to both cytotoxicity and tumorigenesis.[33,34] Achieving single gene cutting in the human genome, approximately $3x10^9$bp in length, would require a nuclease with a specificity equivalent to the precise readout of a single NDA sequence spanning at least sixteen consecutive base-pairs.

Delivery

The problem of how to introduce transgenes into cells safely and effectively has been at the core of gene therapy since its inception. However, whereas supplementing patients with a corrected copy of a gene requires long-term gene expression, nuclease delivery may be transient, as the resultant alteration will be retained at the endogenous genomic locus. This enables nuclease delivery to be performed using non-integrating delivery vectors, and thus circumvents the negative impacts of random gene integration.

There are three common types of vectors used for cellular transgene delivery: plasmid DNA (pDNA), messenger RNA (mRNA) and viral vectors. While pDNA may be the simplest reagent to generate, getting these molecules into the nucleus with high efficiency currently requires toxic chemical methods not suitable for therapeutic use. On the other hand, transfection of mRNA and transduction of non-integrating viral vectors have been used to achieve high levels of transgene expression with much lower toxicity.[35,36]

Several factors affect the deliverability of a transgene, which, in turn, has bearing on the utility of particular nucleases for gene therapy applications. For one, viral vectors are limited in

their DNA packaging capacity, with some commonly used vectors having an upper limit of 4-5kb and shorter transgenes typically yielding better expression.[16,37,38] Another important consideration for proper viral vectorization is the repetitiveness of the transgene, as recombination of repetitive DNA tracts can lead to large deletions in certain types of vectors.[39–41] Lastly, for gene therapy applications, reagents for transgene delivery must be manufactured using costly Good Manufacturing Practices (GMP).[42] For this reason, minimizing the number of delivery molecules, by using monomeric nucleases that would require only a single mRNA species or shorter nucleases that can be combined in parallel on a single viral vector, can be extremely beneficial.

Multiplexing

To date, gene therapy has focused on the delivery or targeting of a single therapeutic gene, due to the difficulties associated with multiplex genome engineering.[43] However, polygenic diseases far outnumber monogenic ones, providing a clear motivation for developing safer multiplex gene therapy methods.[44] Furthermore, multiplex gene knockout could provide a more effective solution for such applications as antiviral treatment, just as drug cocktails do.[25] The advent of nuclease-based genome engineering may provide the specificity needed for safe multiplex gene therapy treatment. For such purposes, the nuclease chosen should be monomeric, to avoid inappropriate dimerization between nuclease halves, and encoded by a short open reading frame (ORF), to enable effective delivery of more than one enzyme simultaneously.

1.4 Designer, rare-cleaving nuclease platforms

Four distinct nuclease platforms have emerged at the forefront of targeted genome engineering; zinc finger nucleases (ZFNs), meganucleases (mn), transcription activator-like

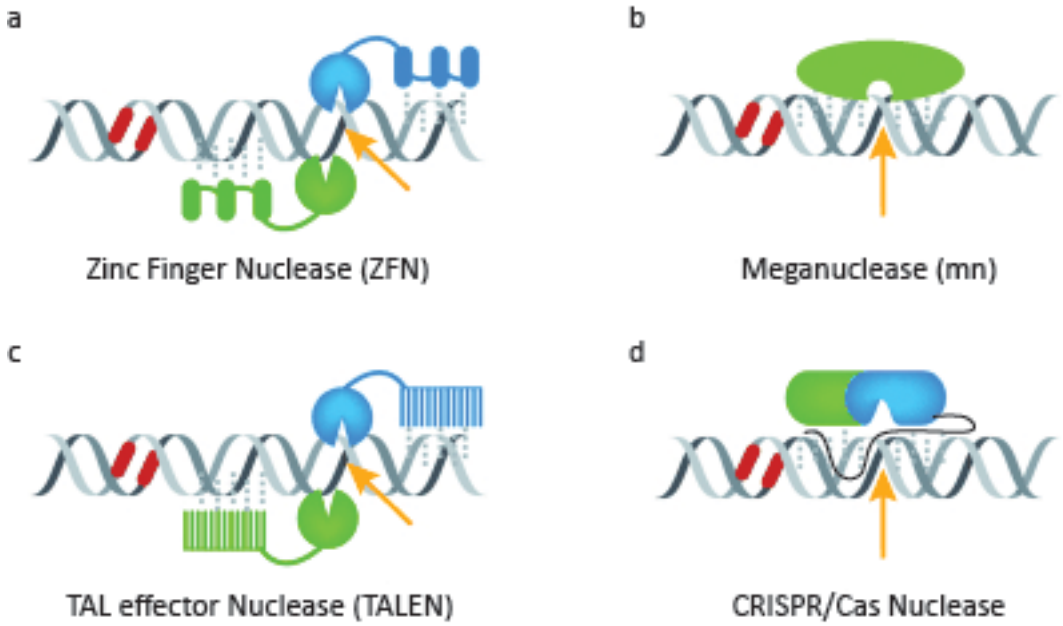

Figure 1: The four commonly used designer, rare-cleaving nuclease platforms – Schematic representation of (a) zinc finger nucleases, (b) meganucleases, (c) TAL effector nucleases and (d) CRISPR/Cas nucleases.

effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR-associated (Cas) nucleases (Figure 1). Each platform provides distinct advantages and disadvantages for gene therapy applications that will be outlined below.

Zinc-finger nucleases

Zinc finger nucleases (Figure 1a) were the first designer nuclease to be developed and tested, originating from the Chandrasegaran laboratory at John Hopkins University.[45] These enzymes are made by fusing together individual zinc finger proteins, which are naturally employed as transcription factors, with the cleavage domain of the Fok I restriction enzyme. Each zinc finger module recognizes a triplet base-pair sequence and modules may be combined to increase the target site recognition length. Because the Fok I cleavage domain must dimerize to cleave its target, two separate ZFN halves, each consisting of zinc fingers fused to Fok I, are required to target a single site.[45,46]

While it was originally believed that assembling zinc finger arrays for targeting extended DNA sequences would be a straightforward task, subsequent examination of the protein-DNA binding interface uncovered synergistic effects between individual fingers that complicated ZFN design.[47] Since the first ZFN was constructed, many advances in both design and selection have been made in an attempt to automate the process of zinc finger assembly.[48–50] However, these attempts have found only moderate success and design of zinc finger arrays towards a given DNA sequence remains challenging.[51,52]

One concern regarding the use of zinc finger nucleases as therapeutic agents is the level of DNA cleavage specificity they can afford. Each nuclease half is typically built with three or four tandem fingers to target 9-12bp sequences on either side of the cleavage site. Despite the length of their full DNA recognition sequence being sufficiently long to achieve single gene targeting, comprehensive studies performed to analyze the specificity of ZFNs, including one currently in use in clinical trials to target the CCR5 gene, have revealed a number of unexpected genomic off-target sites.[28,29,53] Due to the broad range of DNA specificities obtained from selected zinc finger nucleases, future use of these enzymes as gene therapy reagents should require careful examination of all possible genomic targets prior to use in human therapeutics.[33]

Delivery of zinc finger nucleases appears to be relatively straightforward. Each zinc finger half is approximately 1.1kb in length (for a 12bp target), making the total size of genetic material for delivery reasonable. One disadvantage, however, is that because each ZFN is comprised of two halves, some delivery methods will require two separate molecules to be codelivered, with the potential of reducing the nuclease expression efficiency and doubling the cost of generating these reagents.

Lastly, zinc finger nucleases are not ideal reagents for multiplex genome engineering and gene therapy, due in large part to their requirement for dimerization. While the wild-type Fok I cleavage head can indiscriminately dimerize to form active homodimeric nucleases, structure-based redesign has been successfully employed to create obligate heterodimer pairs that exhibit reduced cytotoxicity and off-target cleavage.[54–56] However, coexpression of even two obligate heterodimer pairs leads to cross-reactivity to form inappropriate heterodimeric ZFNs exhibiting cleavage at unintended targets and increased toxicity.[57] Delivery of two separate halves per target becomes increasingly challenging as the number of desired genome modifications increases.

Meganucleases

Meganucleases (Figure 1b) are naturally occurring enzymes found in archaeal, prokaryotic and eukaryotic microorganisms, as well as viruses. They act solely as selfish genetic elements, catalyzing the duplication of their own genetic code at copies of their resident gene lacking the mn ORF. This is achieved by a process called homing, during which the nuclease creates a double-strand break that can be repaired by homologous recombination, with the meganuclease encoding allele serving as a template. Five different families of meganucleases have been identified, exhibiting different proteins structures, mechanisms for catalysis and levels of DNA sequence specificity. The largest and best-characterized of these is the LAGLIDADG meganuclease family, so called for their conserved catalytic motif. LAGLIDADG meganucleases recognize DNA targets approximately 20bp long and are the most specific of the meganucleases, making them the scaffold of choice for targeted genome engineering.[58–60]

The major constraint of using meganucleases as a gene targeting reagent is the engineerability of these enzymes; LAGLIDADG mns are compact proteins that have coupled cleavage and binding activity, making redesign of their native specificity towards a target of

interest a laborious process. In order to more feasibly design active variants, several approaches for improving selection methods by testing extensive nuclease libraries have been developed.[61–64] In addition, efforts have focused on identifying novel meganuclease starting scaffolds that can facilitate redesign by reducing the distance between the starting and desired target specificities.[65–67] Nevertheless, meganuclease engineering still requires significant effort as existing selection methods often yield low affinity variants that necessitate further optimization.

Despite the hurdle of redesigning meganuclease specificity, these enzymes offer significant advantages over other nucleases platforms due to their high sequence specificity. For one, the monomeric, single domain architecture of meganucleases ensures that cleavage occurs in a sequence-specific manner and that target specificity is encoded by a single protein, avoiding issues with inappropriate nuclease dimerization. Furthermore, meganucleases typically exhibit significantly lower toxicity than other nuclease platforms, serving as evidence of reduced off-target cleavage.[68–70] Still, despite their long DNA sequence recognition, meganucleases are not absolutely specific and are capable of genomic off-target cleavage.[27,65,66,71]

The meganuclease architecture also makes it ideal for therapeutic gene delivery, as the monomeric, approximately 1kb ORF can be easily packaged into a single viral vector with adequate space for other gene editing tools. Furthermore, meganuclease can be used for multiplex genome engineering as their short transgene allows for simultaneous delivery of several enzymes and eliminates the unwanted possibility of hybrid dimerization.

Transcription activator-like effector nucleases

Transcription activator-like effector nucleases (Figure 1c) are synthetic proteins that have been developed and described only in the past few years. These nucleases have a similar architecture to zinc finger nucleases, in that they are composed of a DNA-binding domain, a

transcription activator-like (TAL) effector, fused to the cleavage domain of the Fok I restriction enzyme.[72,73] Isolated from plant pathogenic Xanthomonas bacteria, these proteins normally act as virulence factors and contain a DNA-binding domain consisting of repetitive elements.[74–76] These repeats were found to span typically 34 amino acids, with only positions 12 and 13 of these, termed the repeat-variable diresidues (RVDs) showing hypervariability. It was soon discovered that each of these repeat elements conferred binding specificity for a single base on the TAL effector DNA target and that the identity of the RVDs corresponds to the partnering nucleotide.[77,78] Subsequent TAL effector structure determination showed that only residue 13 on each repeat forms direct contacts with the DNA base, while residue 12 acts to stabilize the RVD loop structure.[79,80]

The major significance of these findings is that TAL effectors exhibit a one-to-one relationship between individual repeat elements and target nucleotides that can easily be exploited to obtain novel DNA target specificities. This has enabled rapid design and selection of TALENs targeting a wide range of sequences with an exceptionally high success rate.[81,82] In addition to their high level of engineerability, TALENs have exhibited levels of cleavage activity surpassing that of ZFNs, making this platform very attractive for genome engineering applications.[83] Furthermore, while a detailed analysis of the off-target activity of TALENs remains to be seen, their lowered toxicity compared to rival ZFNs points towards higher cleavage specificity.[70,84,85]

Although TALENs have proven themselves as extremely useful reagents for targeted genome editing, this platform may not lend itself quite as powerfully towards gene therapy applications, because of challenges regarding their cellular delivery. Just like zinc finger nucleases, TALENs consist of two separate halves for a single target locus. Because TAL effectors can be built to

bind any length of DNA, TALENs have been engineered towards targets of variable length; however, most studies employ TALEN halves targeted to sequences 13bp or longer, yielding a DNA ORF close to 3kb in length. For this reason, codelivery of both TALEN halves, as well as other useful factors for increasing genome engineering, becomes problematic due to viral packaging size limitations. In addition, the highly repetitive nature of the TAL effector domain can cause recombination of the TALEN ORF during viral delivery, leading to expression of unwanted protein products that can cause off-target cleavage and increased toxicity.[86] These delivery issues, as well as the capacity of Fok I to form inappropriate dimerization partners, as discussed earlier for ZFNs, also make TALENs a poor reagent choice for multiplex genome engineering.

CRISPR-Cas nucleases

The CRISPR-Cas system (Figure 1d) is the most recent designer nuclease platform to be developed for genome engineering applications. This platform takes advantage of the bacterial adaptive immune system evolved to recognize and cleave invading genetic sequences using a Cas nuclease under the guidance of RNA serving as a DNA-targeting mechanism.[87–89] This system quickly became a favorite in the field of genome engineering, due to the fact that DNA targeting does not require protein engineering, but rather only necessitates the design of an RNA molecule complementary to the sequence of interest. The only design requirement is the need for a protospacer-adjacent motif (PAM, 5'-NGG-3') downstream of the 20bp RNA protospacer target. Many groups have since been able to move this platform into various cell types, including human cells, and successfully modify endogenous targets within the host genome.[90–93] Moreover, multiplex genome engineering can be achieved using CRISPR-Cas nucleases by simply introducing multiple guide RNAs (gRNAs) simultaneously.[93–95]

Unfortunately, much like TALENs, the advantages that CRISPR-Cas nucleases offer towards genome engineering applications do not suffice for therapeutic use. For one, simultaneous delivery of the 4kb Cas9 nucleases along with one or more RNA molecules will prove difficult in a clinical setting, due to both the large nucleases transgene size and the need to get different types of molecules with non-overlapping effective delivery strategies into the desired cell type. Perhaps more importantly, CRISPR-Cas nucleases exhibit an unacceptable level of off-target cleavage, tolerating mutations along the entire length of the protospacer DNA. While the mutation frequency at examined off-target site varied, the rates at some unintended targets rival that of the intended, and indels were observed at targets containing as many as five RNA-DNA mismatches.[94–97]

Other hybrid nucleases

Several other hybrid nucleases have been constructed in an attempt to generate novel reagents with all of the properties desired of therapeutic nucleases. For the most part, these constructions have focused on creating single-chain nucleases with sequence-specific cleavage heads, typically either I-TevI or scPvuII, in order to reduce the possibility of off-target cleavage.[98–103] However, using a active monomeric cleavage domain with only a short DNA recognition site is likely to result in significant off-target cleavage, independent of the activity of the fused DNA binding domain.

1.5 Summary

Targeted genome engineering shows great potential as a therapeutic approach for rectifying genomic abnormalities that lead to disease. However, to minimize potential adverse effects, genetic manipulation must be performed in such a way as to prevent any unintended

modifications. Nucleases can serve as a powerful tool for creating specific genomic changes, due to their capacity to elicit both gene repair and knockout at an endogenous locus in a site-specific manner. Nevertheless, while each of the current designer nuclease platforms displays unique advantages, limitations regarding their engineerability, specificity, deliverability and potential for multiplex use affect their feasible and safe use as agents for gene therapy.

A next generation nuclease platform for gene therapy applications

To address the limitations of available nuclease platform technologies, my thesis project has focused on the development of a novel hybrid nuclease technology (hereafter megaTAL), which combines the advantages of meganuclease specificity with the target-specific affinity and ease of engineering afforded by TAL effector DNA-binding domains. The result, as described throughout this work, is a technology platform that requires reduced engineering efforts, yet allows the generation of single-chain, hyperspecific enzymes compatible with both viral vectorization and multiplexing. The combination of these desirable properties in a single nuclease platform holds significant promise to extend the impact of genome engineering on human therapeutic applications.

Chapter 2

megaTALs: a novel designer nuclease platform

2.1 Introduction

The core hypothesis of this thesis is that by fusing auxiliary DNA-binding domains to meganuclease cleavage domains, we would be able to build novel nuclease reagents that exhibit a high level of sequence-dependent cleavage activity, while lowering the time and resource requirements necessary for engineering of the meganuclease component. The reduced time and resource requirements for meganuclease engineering are predicted to occur due to the partial separation of binding and cleavage functions in the fusion protein; the affinity of the auxiliary DNA binding domain allows the use of a low affinity meganuclease by acting to tether (address) the meganuclease cleavage domain in close proximity to its target. Thus, engineering of the meganuclease protein/DNA interface need only account for efficient cleavage and target specificity, without the need for additionally generating the affinity necessary for sufficient enzyme "dwell" time at the target.

To test this hypothesis, we developed a test set of TAL effector DNA binding domains fused to meganucleases (e.g. "megaTALs") in order to identify an optimal working megaTAL architecture for fusion, and used the protein and DNA properties ascertained from these experiments to explore the relationship between meganuclease affinity and megaTAL cleavage activity.

2.2 Designing and testing a functional megaTAL architecture

The first questions we sought to answer were whether we could identify a megaTAL architecture that does not result in steric hinderance, preventing one or both domains from

functioning properly, and what effect the TAL effector would have on meganuclease cleavage activity. We tested the function of megaTALs using the Traffic Light Reporter (TLR) assay, which provides a readout of both homologous recombination and mutagenic NHEJ. This reporter can be integrated into genomic DNA in order to more realistically assess levels of cellular repair in reporter cell lines.[104] In designing this initial experiment, there were a number of different considerations that needed to be addressed and these are discussed in detail below.

Orientation

MegaTAL nucleases can be designed in two orientations, positioning the TAL effector either N'- or C'-terminally to the meganuclease. We chose to interrogate the former architecture (Figure 2), due to the presence of degenerate repeats at the N-terminus of the TAL effector. These repeats provide non-specific DNA affinity through the interaction of positively-charged protein residues with the negatively-charged DNA backbone and are proposed to seed TAL effector binding as a means to explore DNA for its target site.[105] Thus, fusing the TAL effector C-terminally to a meganuclease would be predicted to have the effect of non-specifically enhancing the meganuclease's affinity for DNA. In contrast, by positioning the TAL effector N-terminally to the meganuclease, the megaTAL would be predicted to function optimally only when the entire TAL effector were efficiently annealed to its cognate target.

TAL effector

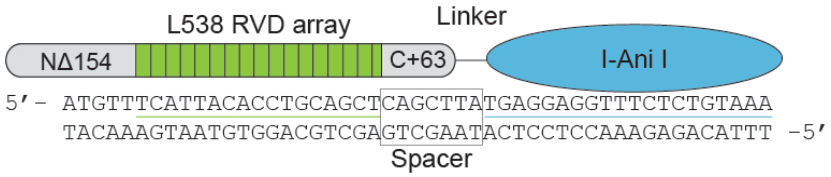

Figure 2: Schematic representation of a megaTAL - The megaTAL architecture involves fusion of a TAL effector with truncated N and C-terminal domains through a short peptide linker to the N-terminus of a meganuclease. Aligned below the megaTAL schematic is the DNA sequence for the L538-I-AniI megaTAL target with a DNA spacer length of seven, the L538 DNA target underlined in green, the I-AniI DNA target underlined in blue and the DNA spacer separating the two outlined with a black box.

We chose a validated TAL effector for testing, to ensure that possible lack of effect could not be caused by use of a non-functional TAL effector. This TAL effector, designated L538, was designed by Sangamo Biosciences to target a 17bp sequence within the CCR5 gene and was shown to be active as part of a TALEN pair.[106]

Meganuclease

To determine whether TAL effectors can be used to rescue activity of low affinity meganucleases, we made megaTAL fusions using variants of the I-AniI meganuclease. These variants all target the same DNA sequence with different affinities and include the wild-type I-AniI mn (K_D = 90 nM, hereafter WT), two experimentally-evolved variants I-AniI F13Y (K_D = 50 nM, hereafter F13Y) and I-AniI Y2 (K_D = 10 nM, hereafter Y2), as well as the catalytically dead variant I-AniI E148D (hereafter E148D).[107]

DNA spacer

In order to identify megaTAL fusions with the highest level of activity, we needed to determine the ideal length of DNA sequence separating the target site of the TAL effector and meganuclease. DNA spacers of 2-16bp were chosen for testing, based on previous data examining spacer length for TALENs, and TLR cell lines were generated harboring target sites consisting of the L538 target separated from the I-AniI target by the variable DNA spacer length. The sequence of the DNA spacer was kept identical to the extent possible across all target sites to minimize any possible effects the DNA spacer sequence itself could have on cleavage activity (Table 1).

Fusion linker

As mentioned briefly above, one potential issue with fusing proteins together is that the individual parts may be linked in such a way as to prevent proper functioning of one or both

Spacer (bp)	Target site
2	TACACATGTACACGTTCATTACACCTGCAGCTTATGAGGAGGTTTCTCTGTAAA
3	TACACATGTACACTTCATTACACCTGCAGCTTTATGAGGAGGTTTCTCTGTAAA
4	TACACATGTACATTCATTACACCTGCAGCTGTTATGAGGAGGTTTCTCTGTAAA
5	TACACATGTACTTCATTACACCTGCAGCTCGTTATGAGGAGGTTTCTCTGTAAA
6	TACACATGTATTCATTACACCTGCAGCTACGTTATGAGGAGGTTTCTCTGTAAA
7	TACACATGTTTCATTACACCTGCAGCTCAGCTTATGAGGAGGTTTCTCTGTAAA
8	TACACATGTTCATTACACCTGCAGCTACAGCTTATGAGGAGGTTTCTCTGTAAA
9	TACACATTTCATTACACCTGCAGCTATCAGCTTATGAGGAGGTTTCTCTGTAAA
10	TACACATTCATTACACCTGCAGCTGATCAGCTTATGAGGAGGTTTCTCTGTAAA
12	TACATTCATTACACCTGCAGCTATGATCAGCTTATGAGGAGGTTTCTCTGTAAA
16	TTCATTACACCTGCAGCTACACATGATCAGCTTATGAGGAGGTTTCTCTGTAAA
scr	ATACACATGTTCATTACACCTGCAGCTACAGCTTAGAGCATGTAATCGTAGTTG

Table 1: TLR Targets with varying DNA spacers (2-16 nucleotides) separating the L538 TAL effector (underlined green) and I-AniI (underlined blue) binding sites tested in the reporter cell lines for cleavage activity with megaTALs and their corresponding standalone meganucleases.

domains due to steric hindrance. To avoid this possibility, we tested four separate protein linkers: the native Fok linker (hereafter Fok) used in many ZFN and TALEN designs, as well as three flexible linkers (hereafter Zn4, Flex1 and Flex2). Furthermore, different protein linkers might have different ideal DNA spacer lengths and could be used interchangeably to achieve the highest level of cleavage for a given locus.

TLR results

Testing of the original megaTAL architecture proved to be very successful and provided a great deal of information regarding important properties of the fused product. MegaTAL fusions were able to increase levels of both homologous recombination and mutNHEJ, when compared to the standalone meganuclease, in an affinity and DNA spacer-dependent manner.

Fusion of the L538 TAL effector to the lowest affinity meganuclease, WT, resulted in up to 18 and 24-fold increases in levels of mutNHEJ and homologous recombination, respectively (Figure 3). A similar effect was observed for megaTALs made with the moderate (F13Y) and high (Y2) affinity mn variants, albeit to a lesser degree (≤13-fold increase for F13Y, ≤4-fold for

Y2) correlating with the intrinsic affinity of the meganuclease variant (Figure 4). Importantly, megaTALs made with all three meganuclease variants of differing affinities were improved to achieve 3-4-fold increases in mutNHEJ and 2-3-fold in homologous recombination over the high affinity standalone Y2 meganuclease. Furthermore, testing of each I-AniI megaTAL against a reporter cell line target consisting of the L538 binding site upstream of a scrambled I-AniI target (Figure 4, labeled scr) showed only background levels of cleavage activity, indicating that docking of the meganuclease in proximity to the TAL effector binding site does not lead to non-specific cleavage. MegaTAL activity was found to be highly spacer-dependent, with the affinity of the starting meganuclease correlating to the range of operational spacer lengths. The WT megaTAL was active at spacers 5-10bp long, while cleavage activity of the Y2 megaTAL was increased across all spacer lengths tested, which we attribute to the N-terminal pseudo repeats of the TAL effector providing sufficient non-specific affinity to affect the on/off rate of the meganuclease.[105] The maximum activity for all three megaTAL variants was achieved with a DNA spacer 7 nucleotides in length (assuming a 20bp meganuclease target).

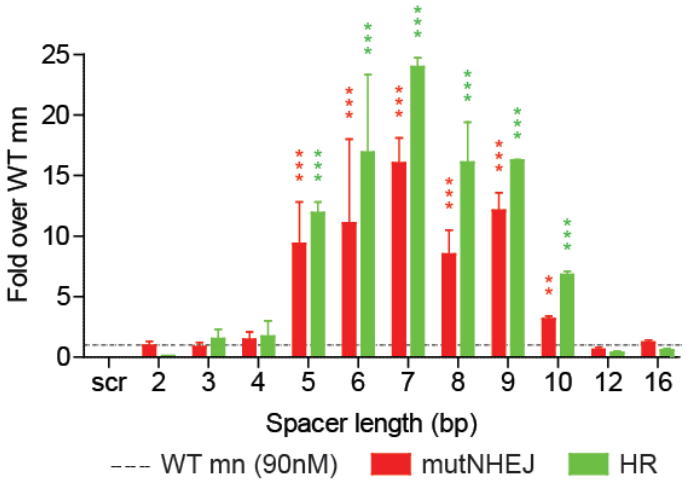

Figure 3: Levels of mutagenic NHEJ (mutNHEJ) and homologous recombination (HR) achieved with the I-AniI megaTAL fusion. Results shown are derived from assays of cleavage activity using TLR reporter 293T cells treated 72 hours with the L538-Zn4-WT megaTAL (mT) or WT I-AniI meganuclease (mn) across targets with different spacer lengths or a 'scrambled' (scr) I-AniI target. Cleavage activity for L538-WT megaTAL for a given spacer length is normalized to the activity of the WT meganuclease level (represented by a dashed line) across all cell lines. For all graphs, error bars indicate s.e.m and P-value comparing megaTALs with their appropriate meganuclease counterpart is indicated by asterix (* signifies P<0.05, ** P<0.005, *** P<.0005).

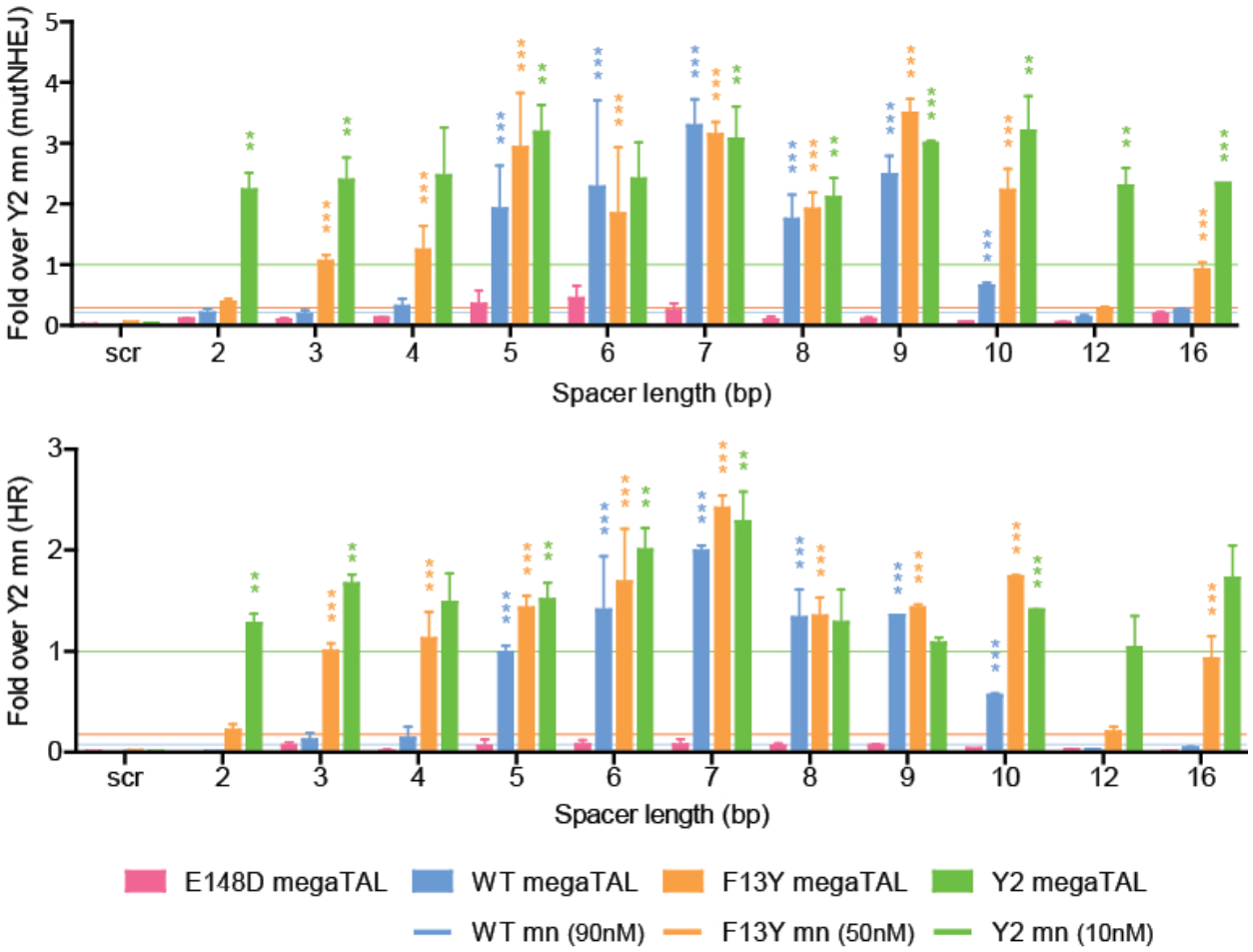

Figure 4: Activity of megaTALs built with I-AniI meganucleases with varying affinities - Levels of mutNHEJ (top) and homologous recombination (bottom) in TLR cell lines containing targets with different DNA spacer lengths. The level of activity for each megaTAL tested was normalized to the activity of Y2 (represented by green line) in the same TLR cell line. The average activity of the WT, F13Y meganucleases across each cell line are represented by blue and orange lines, respectively.

Activity across different DNA spacer lengths tested appears to exhibit periodicity; while the activity of the F13Y megaTAL decreases between 7 and 12bp, it begins to rise back up with a 16bp spacer. Moreover, the minimum activity was observed with spacers 2 and 12bp in length, a distance that corresponds approximately to one turn of the DNA helix. These findings suggest that the effect of DNA spacer length on activity may be due to the relative position of the TAL effector and meganuclease targets along the DNA helix, rather than their distance. Further testing with DNA spacers ranging from approximately 12 to 22bp would provide a clearer view on how spacer length restricts megaTAL activity.

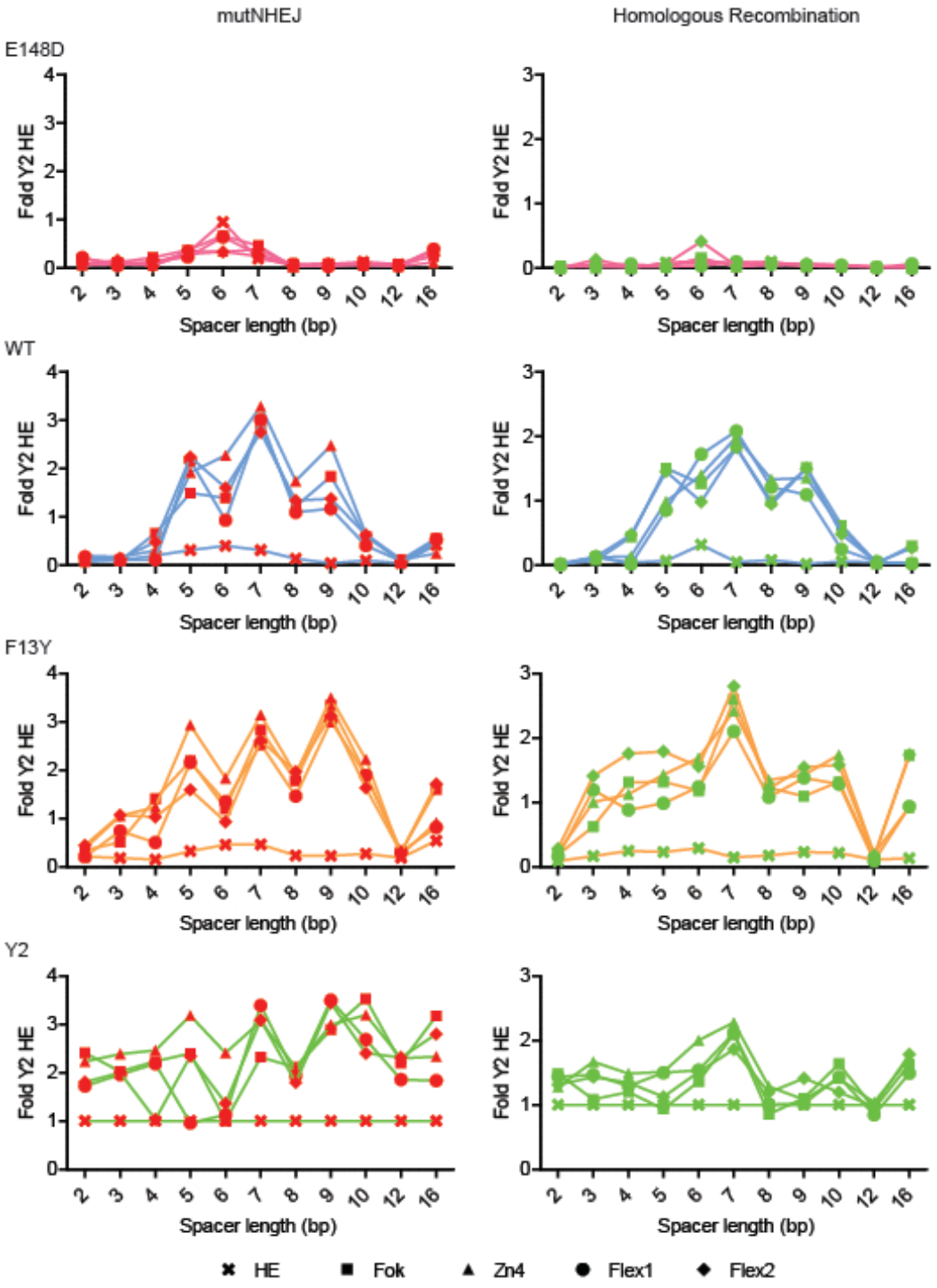

Figure 5: Comparison of cleavage activity of megaTALs made with different protein linkers. Level of mutNHEJ (red symbols, left column) and homologous recombination (green symbols, right column) measured in Traffic Light reporter cell lines with targets from Table 2. Reporter cells were treated with megaTALs made by fusing the L538 TAL effector to either the E148D (pink lines), WT (blue lines), F13Y (orange lines) or Y2 (green lines) variant of the I-AniI meganuclease by four different protein linkers (Fok, Zn4, Flex1 and Flex2). Activity of the standalone meganuclease variant was also tested.

No discernible differences in activity were observed between megaTALs constructed with different protein linkers (Figure 5), which we believe is explained the fact that the C-terminus of the TAL effector is unstructured and thus highly flexible. This flexibility may also account for the apparent wide range of DNA spacer lengths that can be tolerated by megaTAL fusions. We speculate that reconfiguration of the C-terminus of the TAL effector to form a structured linker could provide higher cleavage specificity and would be a valuable approach to designing a 2nd generation megaTAL platform with enhanced performance properties.

2.3 Effect of megaTALs on target site affinity and cleavage activity

Engineering of native meganucleases towards desired sequences frequently results in the isolation of variants that exhibit either a reduced target site affinity or catalytic defects, due to the combinatorial barrier posed by the highly integrated LAGLIDADG meganuclease-DNA binding surface. To save both time and resources when designing megaTALs, it would be advantageous to understand which types of meganuclease variants could be successfully rescued to attain a highly active nuclease. We hypothesized that fusing an auxiliary DNA binding domain to meganucleases would be an effective approach to rescue variants that exhibit low target site affinity, but would be less efficient for rescuing those enzymes with catalytic defects.

In order to discern between these possible effects, we tested the Y2 meganuclease and L538-Y2 megaTAL

Target site	wt base	k_{cat}* (min^{-1})	K_M* (nM)
wild-type	wt	9.4	10
-9A	G	10.5	117
-8G	A	14.3	129
-6T	G	11.3	72
-1G	T	0.2	4.8
+2A	C	0.38	10
+3C	T	0.14	10.4

Data obtained from Thyme et al. (2009).

Table 2: Affinity and catalytic activity of the Y2 meganuclease at I-AniI near native targets - Wild-type and singly-substituted I-AniI targets and previously determined k_{cat}* and K_M* values.

against a series of DNA target variants that harbor single basepair substitutions (relative to the native I-AniI target, wt) for which the k_{cat}* and K_M* of the meganuclease have previously been measured (Table 2).[71] Targets were chosen for which Y2 displays either "high K_M*" (Y2 exhibits a similar k_{cat}*, but reduced affinity relative to wt) or "low k_{cat}*" (Y2 exhibits a reduced k_{cat}*, but comparable binding affinity compared to wt). Cleavage activity was measured using TLR reporter cells with targets comprised of the L538 binding site and the individual high K_M*, low k_{cat}* or wt I-AniI targets, separated by an 8bp spacer.

TLR results

The stand-alone Y2 meganuclease showed very low levels of cleavage activity against all six non-native targets, ranging from background rates to less than 0.2% for either pathway (Figure 6). In contrast, cleavage activity was considerably increased (12-38-fold for HR, 8-27-fold for mutNHEJ) against all high K_M* targets treated with the megaTAL, while little or no effect was observed for the low k_{cat}* targets (0-3-fold for HR and mutNHEJ). However, megaTAL activity against the low k_{cat}* target -1G was found to be significantly increased compared to the standalone meganuclease treatment, suggesting that in some cases a strong target site affinity can partially compensate for low catalytic activity. A significantly higher rate of mutNHEJ was observed against the low K_M* target -8G, without a corresponding increase in levels of HR, indicating both that enzymes combining strong cleavage activity and high affinity may yield very active nucleases and that nucleases with different biochemical properties may display diverse repair pathway biases.

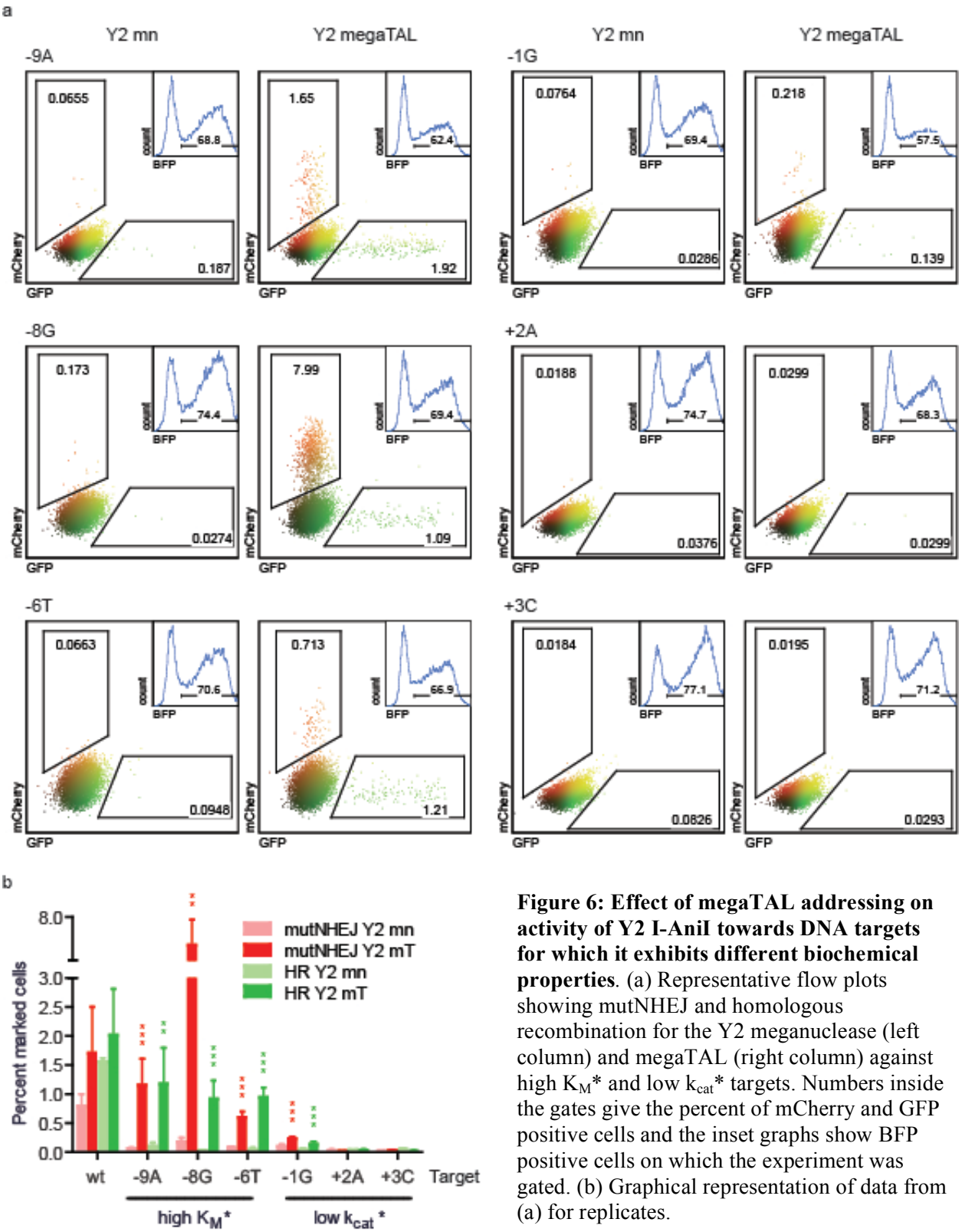

Figure 6: Effect of megaTAL addressing on activity of Y2 I-AniI towards DNA targets for which it exhibits different biochemical properties. (a) Representative flow plots showing mutNHEJ and homologous recombination for the Y2 meganuclease (left column) and megaTAL (right column) against high K_M* and low k_{cat}* targets. Numbers inside the gates give the percent of mCherry and GFP positive cells and the inset graphs show BFP positive cells on which the experiment was gated. (b) Graphical representation of data from (a) for replicates.

2.4 Summary

Fusion of TAL effectors to both high and low affinity meganucleases can yield hybrid nucleases with increased cleavage activity. We found that megaTALs could be made to rescue the activity of both the WT and Y2 I-AniI meganucleases against targets for which no or very low cleavage activity was observed with the standalone meganuclease, due to reduced affinity for the target site. Moreover, megaTALs could further increase levels of both mutagenic NHEJ and homologous recombination for active meganucleases with moderate (F13Y) or high (Y2) affinity for a given target.

MegaTALs made with the three different I-AniI variants yielded equal levels of activity against the native I-AniI target separated from the L538 TAL effector binding site by a 7bp DNA spacer, suggesting that the 16.5 RVD TAL effector was able to compensate for any affinity differences between the variants for targets with an ideal DNA spacer. Disparities in megaTAL activity were observed, however, at targets with non-ideal spacer lengths, which may be due to either a small boost in affinity afforded by non-specific interactions between the N-terminus of the TAL effector and DNA backbone or to the ability of higher affinity meganucleases to function on their own during dissociation of the TAL effector from its target.

We also observed that a megaTAL made with the Y2 meganuclease showed rescued cleavage activity only against singly-substituted targets for which the standalone meganuclease exhibits low affinity and not those for which it exhibits low catalytic activity. Furthermore, levels of mutNHEJ and HR varied among the megaTAL-treated high K_M* targets, suggesting that differences in meganuclease catalytic activity will still affect the outcome of repair.

Taken together, these findings indicate that megaTALs can be consistently constructed to generate highly efficient and specific enzymes using low affinity meganuclease cleavage heads.

These properties suggest that engineering of meganuclease cleavage heads for use in the megaTAL architecture should be significantly less time and resource intensive than engineering of a standalone meganuclease, due to the ease with which TAL effectors can be generated towards almost all DNA sequences and the reduced affinity requirement of the meganuclease domain of the fusion protein.

2.5 Materials & methods

MegaTAL and meganuclease construct generation

MegaTALs were constructed using the Golden Gate assembly strategy previously described by Cermak et al, using an RVD plasmid library and destination vector generously provided by the Voytas lab.[108] The pthX01 destination vector was modified to include a hemagglutinin (HA) tag immediately downstream of the NLS and to yield a NΔ154, C+63 TALEN scaffold. TAL effectors were built using the following RVDs to target each specific nucleotide: A – NI, C – HD, G – NN and T – NG. Following cloning of the TAL effector repeats into the destination vector, the protein linkers (Fok: SQLVKS, Zn4: VGGS, Flex1: GGSG or Flex2: GGSGGGSG) and I-AniI meganuclease variants were cloned in place of the FokI nuclease catalytic domain between the Xba-I and Sal-I restriction sites. All megaTALs, except those used to compare linker activity (Figure 5), were made using the Zn4 protein linker. Control constructs expressing standalone meganuclease variants were made by cloning the nuclease downstream of an NLS and HA tag between the Sbf-I and Sal-I restriction sites. All constructs encode BFP-T2A-nuclease for tracking of nuclease expression during flow cytometry.

Cell line derivation

HEK293T cell lines were generated harboring either the original traffic light reporter (TLR 2.1)[104] or a modified traffic light reporter containing an iRFP gene in place of puromycin (epigenetic TLR, unpublished data), with the appropriate target site embedded within the GFP ORF. Cells were derived as previously described with slight modifications. Briefly, HEK293T cells were transduced with recombinant lentivirus to yield 5-10% transduction, based on either iRFP expression (epigenetic TLR) or cell survival after treatment with puromycin (TLR 2.1). Approximately five days post transduction, cells were sorted for iRFP+/mCherry- populations (epigenetic TLR) or mCherry- populations (TLR 2.1).

Cell sorting and flow cytometry

Cells were analyzed by flow cytometry on the BD LSRII and sorted on the BD FACS ARIAII. Fluorophores were detected using the following lasers and filters: mCherry – excited 561nm, acquired 610/20, mTagBFP – excited 405nm, acquired 450/50, eGFP – excited 488nm, acquired 525/50, iRFP – excited 640nm, acquired 730/45. Data were analyzed using FlowJo software.

Traffic Light Reporter Assay

The traffic light reporter assay was performed as previously described with slight modifications. Cells harboring the TLR 2.1 or epigenetic TLR were plated at $2.0x10^5$ cells/well in a 24-well dish 24h prior to transfection. XtremeGene9 (Roche) was used at 2µl/µg DNA to transfect cells with 0.5µg of both nuclease and GFP donor constructs. Cells were harvested 72h post transfection and read on the flow cytometer. Data were obtained from BFP+ (TLR 2.1) or iRFP+/BFP+ populations (epigenetic TLR).

Statistical analysis

Error bars on graphs represent s.e.m. P-values were calculated using Student's one-tailed unpaired t-test to compare activity of megaTALs with their specific meganuclease counterpart ($P<0.05$ shown as *, $P<0.005$ **, $P<0.0005$ ***).

Chapter 3

megaTALs: a hyperspecific nuclease platform

3.1 Introduction

For therapeutic applications, the importance that reagents exhibit sufficiently high cleavage specificity to ensure minimal off-target effects outweighs considerations of engineerability, as the resources to engineer a nuclease are minimal in comparison to total therapeutic development costs. Currently, none of the available nuclease platforms exhibit complete single target specificity and both off-target cleavage and cytotoxicity have been observed with ZFNs, meganucleases, TALENs and CRISPR/Cas nucleases.[27,29,70] We hypothesized that the hybrid megaTAL architecture would support extremely high cleavage specificity due to the elongated target site provided by fusion of the TAL effector binding domain and meganuclease catalytic domain, coupled with the inherent high cleavage specificity of the meganuclease domain.

3.2 Designing megaTALs to address genomic loci

To evaluate the specificity of the megaTAL architecture, we tested whether a megaTAL nuclease could distinguish between intended and unintended genomic target sites, all cleaved by a common meganuclease cleavage domain. We reasoned that binding of the TAL effector to a target upstream of the desired meganuclease cleavage site could "address" the meganuclease towards that locus, enhancing activity specifically at the addressed locus versus unintended targets lacking the TAL effector docking site.

Identifying candidate genomic targets

Near-native targets of the I-AniI meganuclease (DNA sequences containing 3 or fewer basepair differences from the wild-type I-AniI recognition sequence) were previously identified

in the human genome by Dr. Michael Certo, using the UCSC BLAT tool.[109] Dr. Certo subsequently assessed the level of cleavage activity at these near cognates with the Y2 meganuclease using both episomal and endogenous targets (data not shown). The five sites that displayed the highest levels of cleavage by the endonuclease in those assays (designated +9T, +5A+8T, +2T+9G, +1C+5A and -2A+1G) were chosen for further examination (Table 3).

Building site-specific megaTALs

We chose the +9T and +5A+8T genomic loci for megaTAL targeting, based on the fact that they showed the highest level of DNA cleavage by the Y2 meganuclease as assessed by sequencing at the endogenous loci in Y2-treated cells. Several considerations were taken into account when choosing a TAL effector DNA binding site upstream of the near-native I-AniI target at both loci:

- DNA spacer length: A +9T megaTAL target with a spacer 8 nucleotides in length was chosen based on our original findings that this length spacer gave maximal megaTAL activity; however, it was later discovered that these results were obtained against suboptimal targets, as we neglected to include a thymine base at the 0^{th} position of the TLR targets (data not shown). The +5A+8T megaTAL target was chosen later, once reevaluation of the DNA spacer indicated that a 7bp spacer gives maximal activity, to reflect this new information.
- Length of the TAL effector DNA binding sequence: Since previous experiments had all been performed with the L538 TAL effector that targets a DNA sequence 17bp long, we wanted to choose genomic targets for the +9T and +5A+8T loci of approximately the same length to ensure that the synthetic megaTALs would function optimally. However, as the 16.5 RVD L538 TAL effector was found to maximize affinity, as discussed in the

previous chapter, we felt that TAL effectors with fewer repeats could function equally well.

- Requirement for a thymine at the 0th position of the TAL effector target: Examination of the native TAL effector DNA targets, experiments in which the DNA base just 5' to the TAL effector target (designated the 0th position of the sequence) is varied as well as structural analyses of the TAL effector have all shown that these proteins have a preference for a thymine at 0th position (T_0).[79,110,111] As we observed this same base preference when testing our initial megaTAL architecture, we chose +9T and +5A+8T TAL effector targets such that they would have a 5' T_0.

Name	Genomic sequence	Location
+9T	TGTCATGTAAATAACTGTACTGTATGTTGAGGAGGTTTCTCTGTATA	Ch9: (-)63127729-63127775
+5A+8T	TCTTCTGCTGTTTCATATACCTCTATGTGAGGAGGTTTCTCAGTTAG	Ch15:(-)69553515-69553561
+2T+9G	CATAGGTGACTTTTACTCTCATGATAGTGAGGAGGTTTTTCTGTAGA	Ch12:(+)84384994-84385040
+1C+5A	GTCTTTGAATAAGTCACAATGCTTCTTTGAGGAGGTTCCTCAGTAAA	Ch7: (+)28151155-28151201
-2A+1G	ATGATGTACACTGAGCTCTGAAATGTTGAGGAGGATGCTCTGTAAA	Ch1: (+)223534899-223534945

Table 3: I-AniI human genomic near-native sequences and their location in the genome. Synthetic TAL effectors were built against the +9T and +5A+8T sequences underlined in green and the I-AniI near-natives sequences are underlined in blue.

Given the above restrictions, we chose TAL effector binding targets 14 and 18bp long for the +5A+8T and +9T TAL effectors, respectively, as shown in Table 3. Both synthetic TAL effectors were built and fused to the WT, F13Y and Y2 I-AniI variants for experimental testing.

3.3 Assessing cleavage specificity at Traffic Light Reporter loci

To evaluate megaTAL target specificity, we first generated TLR reporter cells harboring each of the five I-AniI near-native targets (shown in Table 3). These reporter cells were treated with each of the +5A+8T and +9T megaTALs, as well as the standalone I-AniI meganucleases, and assessed for levels of homologous recombination (mutNHEJ was not measured due to the

presence of a stop codon in some TLR lines that prevented readout). The results from these experiments show that the megaTALs built towards a specific near-native site showed significant increases in cleavage activity towards the "addressed" target, compared to the standalone meganuclease, while levels of HR were unaffected at the "unaddressed" targets (Figure 7a). All three megaTAL variants targeted towards the +5A+8T site showed similar levels of homologous recombination, regardless of the affinity of the meganuclease used. In contrast, the +9T megaTAL constructed with the WT I-AniI meganuclease exhibited lower levels of cleavage than

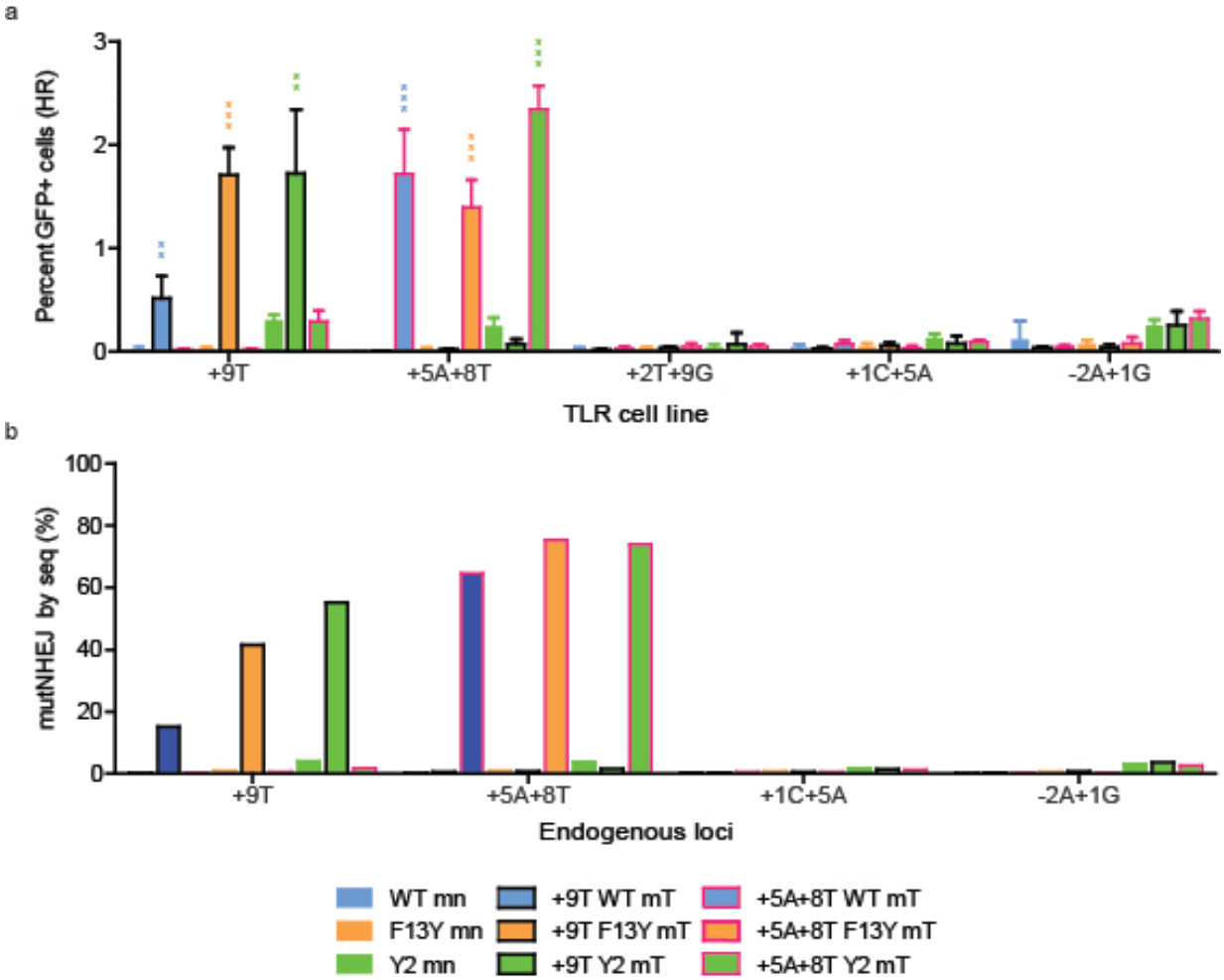

Figure 7: Mutation rates using "addressed" I-AniI megaTALs and "unadressed" meganucleases. The level of mutNHEJ at (a) reporter and (b) endogenous I-AniI near-native targets was determined in 293T cells after 72 of treatment with meganuclease and megaTAL variants based on reporter output (top) and MiSeq sequencing results (bottom).

did the F13Y and Y2 megaTALs, which we attribute to the fact that a non-ideal DNA spacer length was chosen for targeting at that locus.

These results also showed that while the TAL effector supports substantially increased on-target cleavage by localizing the meganuclease to its target, it has little or no effect on cleavage at off-target sites. Therefore, each of the megaTAL variants tested showed comparable levels of homologous recombination to the standalone meganuclease variant at targets unaddressed by the TAL effector. Since the Y2 meganuclease was able to cleave all of the sites with some efficiency, the +5A+8T and +9T megaTALs showed the same level of cleavage activity at the unaddressed sites. However, because the WT and F13Y meganucleases exhibited very low cleavage activity, so did the megaTALs made with these variants at the sites that were not addressed by their TAL effector domain.

From these results, we can thus draw two important conclusions: 1) fusing a TAL effector to a meganuclease increases on-target cleavage dramatically in a site-specific manner and 2) by constructing megaTALs with low affinity meganucleases that show no or little activity on their own, hyperspecific reagents can be generated to yield high on-target cleavage activity with extremely low off-target cleavage.

3.4 Assessing cleavage specificity at endogenous loci

Because therapeutic applications require manipulation of endogenous DNA sequences, we wanted to verify that gene editing results at the endogenous I-AniI near-native genomic targets would parallel what was observed using integrated reporter DNA sequences. To this end, we treated 293T cells with each of the I-AniI meganuclease and +5A+8T or +9T megaTAL variants and performed high-throughput sequencing at each of the five near-native sites tested above to

		mutNHEJ (%) by genomic locus			
Nuclease	mn variant	+9T	+5A+8T	+1C+5A	-2A+1G
-	-	0.14	0.22	0.25	0.16
mn	WT	0.24	0.20	0.29	0.12
	F13Y	0.85	0.95	0.75	0.59
	Y2	3.95	3.85	1.71	3.11
+9T megaTAL	WT	15.24	0.64	0.48	0.42
	F13Y	41.42	0.71	0.63	0.83
	Y2	55.27	1.63	1.39	3.6
+5A+8T megaTAL	WT	0.43	64.6	0.58	0.4
	F13Y	0.66	75.3	0.54	0.49
	Y2	1.76	74.1	1.30	2.63

Table 4: High-throughput sequencing results at endogenous human I-AniI near-native targets in 293T cells treated with meganucleases and megaTALs.

determine the proportion of alleles harboring indels. Sequencing results revealed a single-nucleotide polymorphism at the "+2T+9G" locus of the treated cells that included a third mutation from the native I-AniI target (+2T+7C+9G). As no cleavage activity was observed at this site, data obtained from this locus will not be further discussed.

The results from targeting the endogenous near-native loci mirrored those obtained against the TLR target, with one major difference: on-target activity observed at addressed megaTAL targets was greater than expected based on reporter readout (Figure 7b and Table 4). One potential explanation for this observation may be that the traffic light reporter cells were used to measure a different DNA repair pathway than the high-throughput sequencing analysis (HR versus mutNHEJ), however we are more inclined to speculate that the TLR provides only a comparative and not quantitative measurement of cleavage events due to under reporting of silenced alleles and biases in frame of repair. The percent of modified alleles obtained at the megaTAL addressed +5A+8T and +9T loci reached 75% and 55%, respectively. These sequencing results demonstrate that rates of mutNHEJ obtained using megaTALs rival or surpass those of any other available nuclease platform.

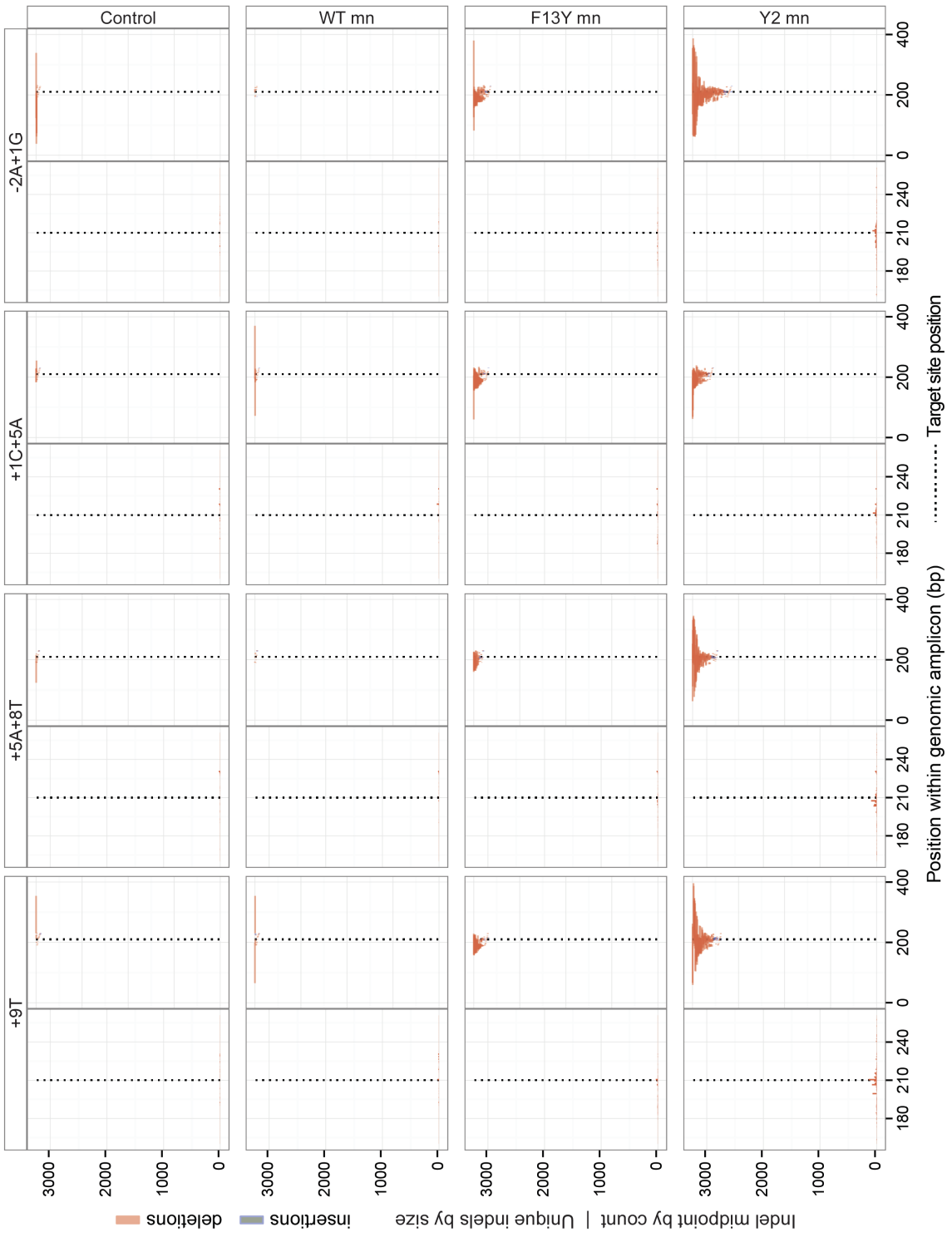

Figure 8: Plots showing the position and size of indels found by high-throughput sequencing of megaTAL and meganuclease treated 293T cells at "addressed" and "unaddressed" loci. For each loci, the panel on the left shows a histogram of indels by their midpoint position and the panel on the right shows a tornado plot of all unique indels that were identified, sorted by insertion or deletion size.

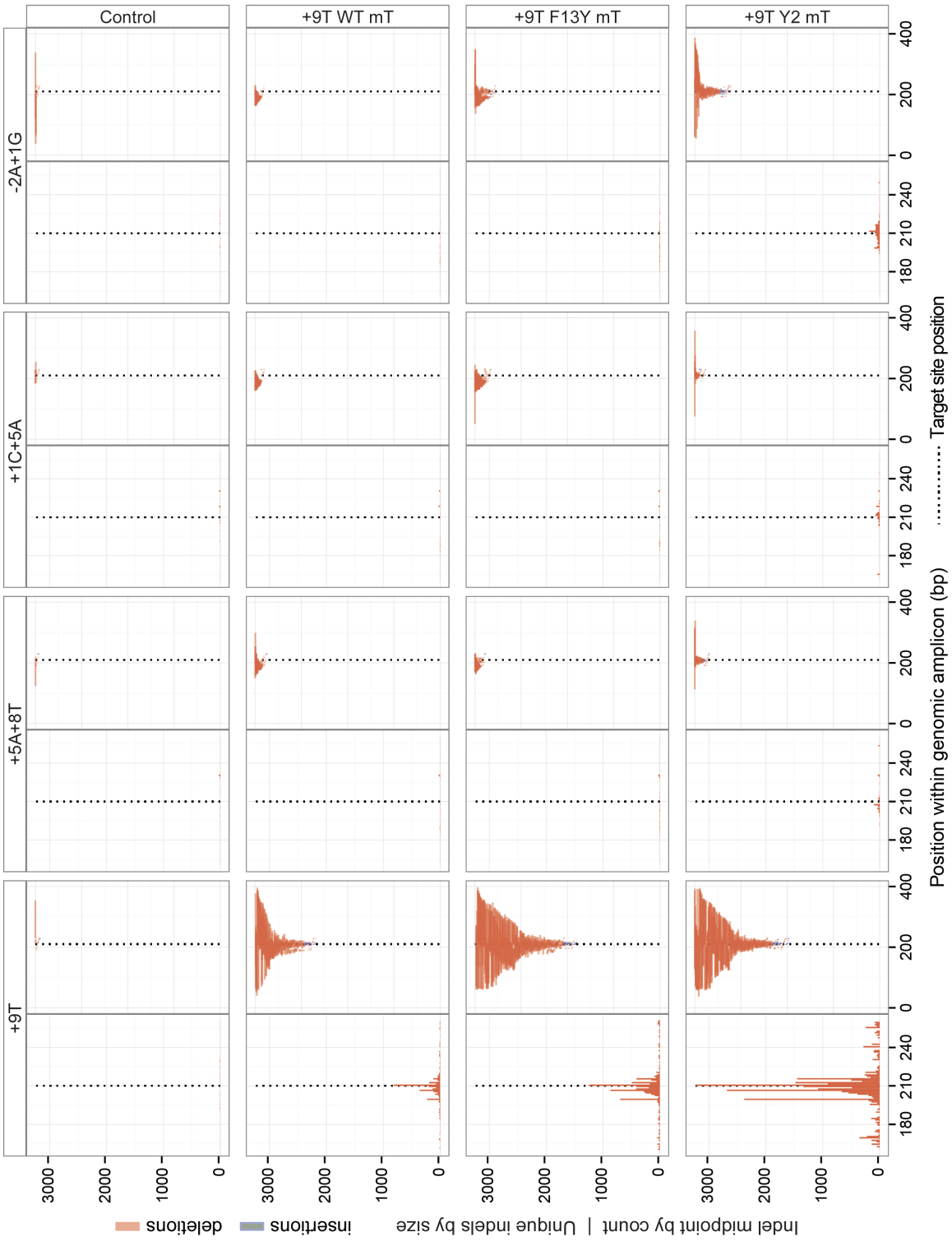

Figure 8, cont.

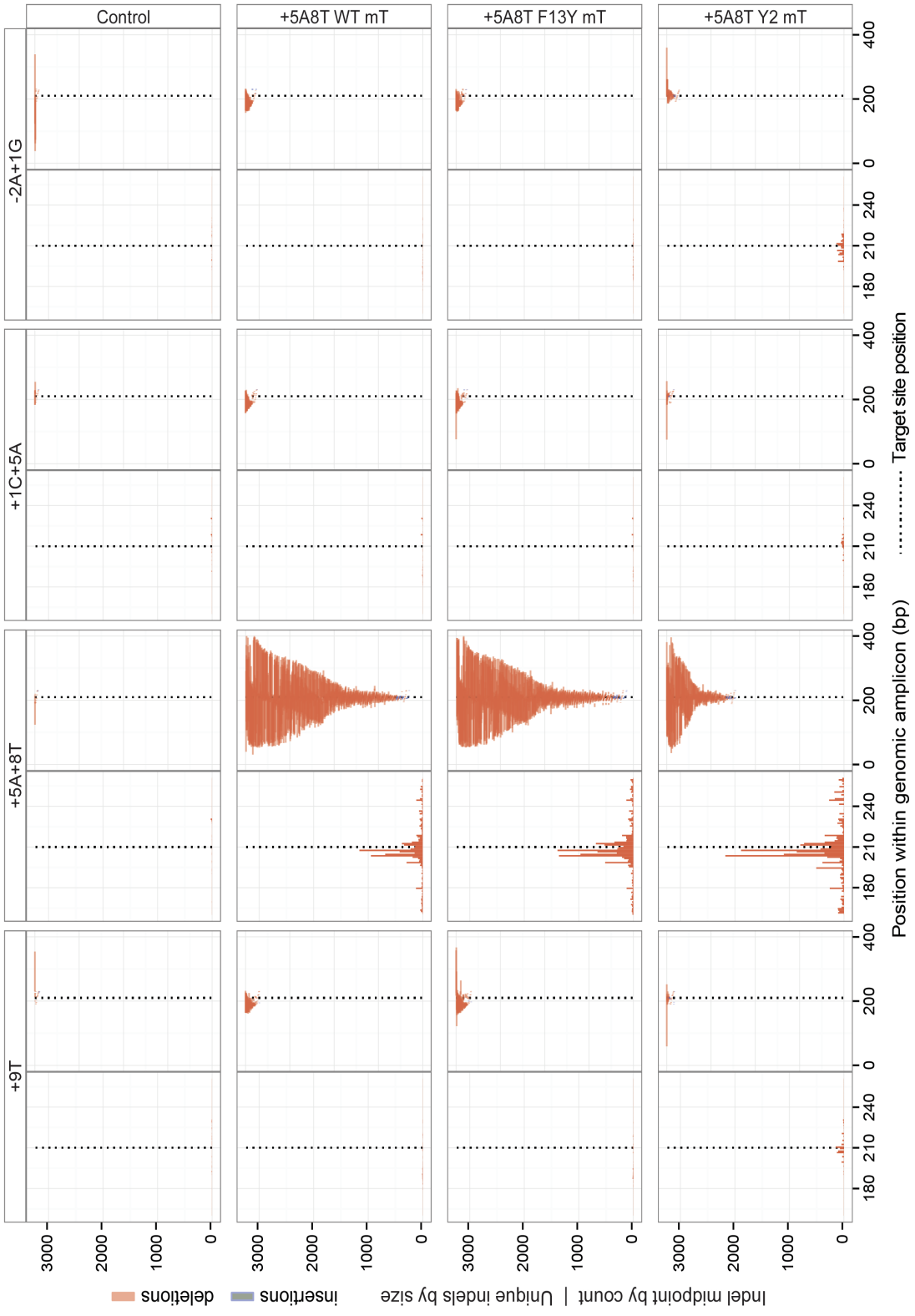

Figure 8, cont.

Additionally, the high throughput data confirmed the observation from reporter cells that megaTALs do not increase cleavage activity at off-target sites, as these megaTAL nucleases showed comparable levels of indel formation to their meganuclease counterparts at unaddressed targets. As suggested from the TLR results, both the Y2 meganuclease and megaTALs resulted in measurable levels of off-target cleavage at each unaddressed site (1-3%), while each of the WT and F13Y nucleases formed indels in less than 1% of sequenced alleles at these loci. Due to the error rates of both the PCR amplification and sequencing steps of the high-throughput sequencing protocol, background rates of 0.1-0.3% indel formation were obtained from untreated cells. Given the measured level of background error, and that sequencing was performed in singlicate, we were unable to accurately determine the level of normal deviation within the experiment. Thus, the absolute specificity of the tested megaTAL variants could not be assessed, however it is clear that modification rates at unaddressed sites after treatment with the WT meganuclease or megaTALs are comparable. Specifically, the +5A+8T megaTAL formed with the WT I-AniI variant achieved 65% on-target modification with <0.4% indel rates above background. Hence,

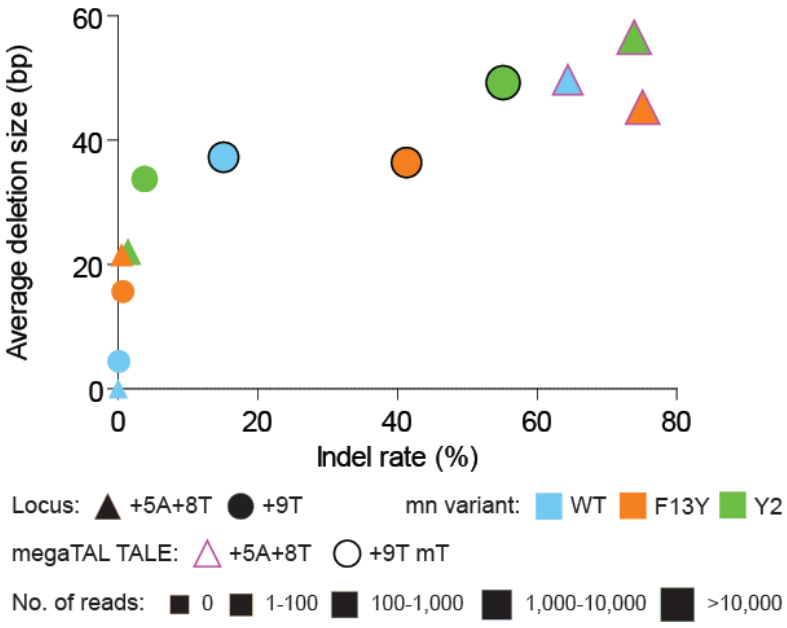

Figure 9: Plot comparing the indel rate and average deletion size at the endogenous +9T and +5A+8T genomic loci after treatment with different nucleases. The addressed megaTALs seem to result in on average larger deletions, however it is unclear whether this may be due to differences in the nuclease architecture, affinity, or number of sequence reads obtained for each.

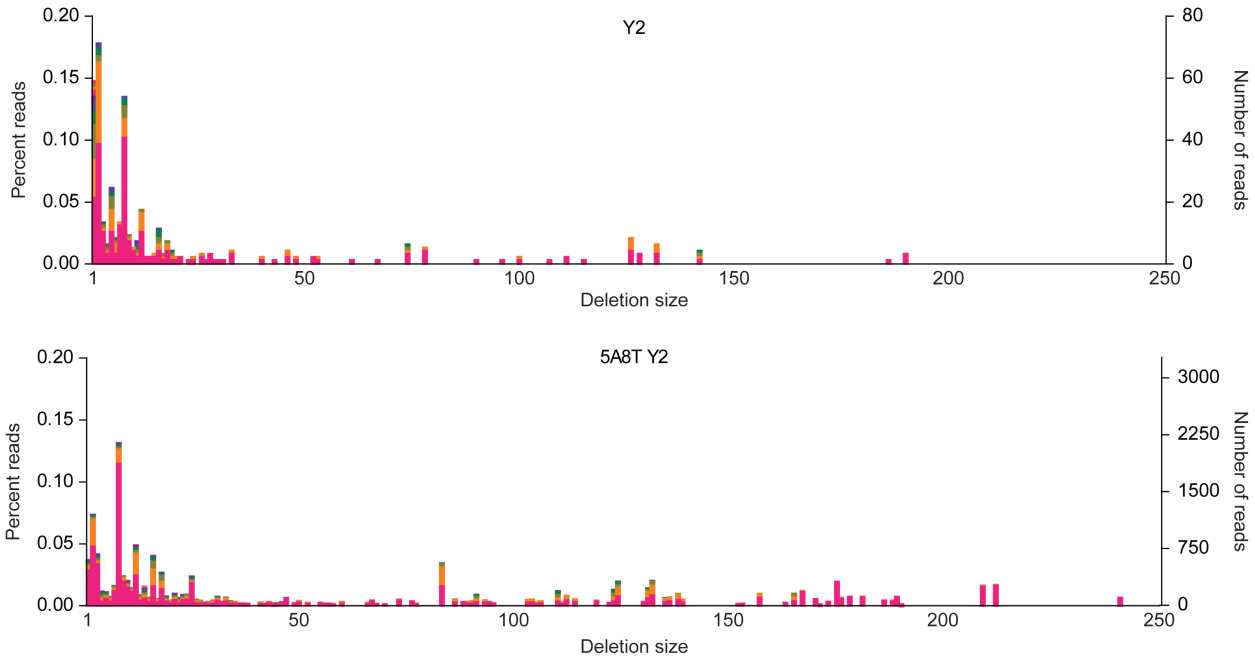

Figure 10: Genomic deletions at the +5A+8T locus in 293T cells after treatment with the addressed megaTAL or unaddressed meganuclease. Analysis of high-throughput sequencing results from Y2 and +5A+8T Y2 treated cells revealed a skew towards larger deletions in the latter. Bars represent the number of reads obtained for unique deletions, where each color represents a unique deletion of the given size.

the high-throughput sequencing data confirm that megaTALs formed with low affinity meganucleases can be used to achieve efficient and hyperspecific genome engineering.

An unexpected observation gathered from the high throughput sequencing data is that the deletions found in megaTAL-treated cells appear to be, on average, larger than those from meganuclease-treated cells (Figure 8, right panels). Further analysis revealed that the average deletion size seems to increase with nuclease activity (Figure 9), although the large discrepancy in the number of deletion events obtained between megaTAL and meganuclease-treated samples prevented any meaningful statistical comparison on deletion size from being performed. A more detailed comparison of the deletions obtained at the +5A+8T locus from cells treated with either the Y2 meganuclease or +5A+8T Y2 megaTAL shows how this skew is resultant from both a proportional loss of very small deletions (1-4bp) and a gain of large deletions (on the order of 100-200bp) (Figure 10). Moreover, the +5A+8T Y2 megaTAL treated cells seem to show a

greater bias towards unique deletions (fewer distinct, recurring deletions), perhaps suggesting that this skew may be due to a bias towards microhomology-mediated end joining. Further study of the deletion distribution using a larger set of megaTALs would help to further clarify whether deletion size may be affected by the architecture of the megaTAL or may simply be related to the overall nuclease activity.

3.5 Summary

High genomic specificity is an extremely desirable property for a therapeutic reagent, with an ideal goal of achieving single gene targeting. Experiments performed to assess the level of sequence-specificity afforded by megaTALs demonstrated that these nucleases do, in fact, exhibit improved specificity with the help of their auxiliary TAL effector DNA binding domain. This added DNA recognition heightens the specificity of active meganucleases by increasing the level of cleavage activity at the addressed site only, resulting in a proportionally greater number of cleavage events at the addressed site only for high on-target cleavage. This level of specificity would be desirable for both therapeutic and research applications in which pools of nuclease-treated cells may be thoroughly examined and only those cells with the desired gene modification(s) can be expanded for use, as with *ex vivo* gene therapy. However, one should be particularly cautious when using a megaTAL made with a meganuclease that exhibits substantial cleavage activity as a standalone nuclease, particularly if there are significant numbers of near-cognate sites, as the added affinity provided by the non-specific DNA binding activity of the TAL effector may increase the number of cleavable near-native sequences, requiring comprehensive analyses of potential genomic targets prior to its use.

For applications that require higher stringency sequence-specific cleavage, megaTALs can be built using meganucleases exhibit no or low *in vivo* activity as standalone enzymes. These meganucleases should be chosen such that they exhibit DNA sequence-specificity and catalytic activity when assessed under conditions in which affinity is not required for target cleavage, such as a tethered *in vitro* assay,[63] but sufficiently low affinity that their capacity for *in vivo* DNA cleavage as a standalone reagent is not detectable. Once fused to a TAL effector that can then rescue the meganuclease affinity deficiency towards the "adressed" target, the resulting megaTAL would be predicted to exhibit a high degree of in vivo activity with essentially no off target activity. The +5A+8T and +9T megaTALs are proof of concept for this approach to creating therapeutic gene editing reagents, as the respective TAL effector fusions to the WT I-AniI mn were found to have extremely high activity at their intended targets and low off-target. Assessing the target specificity of megaTALs made from a panel of meganucleases with different biochemical properties may help identify additional traits that can be used to develop megaTALs with even more consistent and predictable stringent single gene specificity.

3.6 Material & methods

MegaTAL and meganuclease construct generation

MegaTALs were constructed using the Golden Gate assembly strategy previously described by Cermak et al, using an RVD plasmid library and destination vector generously provided by the Voytas lab.[108] The pthX01 destination vector was modified to include a hemagglutinin (HA) tag immediately downstream of the NLS and to yield a NΔ154, C+63 TALEN scaffold. TAL effectors were built using the following RVDs to target each specific nucleotide: A – NI, C – HD, G – NN and T – NG. Following cloning of the TAL effector repeats into the destination

vector, the Zn4 (VGGS) protein linker and I-AniI meganuclease variants were cloned in place of the FokI nuclease catalytic domain between the Xba-I and Sal-I restriction sites. Control constructs expressing standalone meganuclease variants were made by cloning the nuclease downstream of an NLS and HA tag between the Sbf-I and Sal-I restriction sites. All constructs encode BFP-T2A-nuclease for tracking of nuclease expression during flow cytometry.

Cell line derivation

HEK293T cell lines were generated harboring the original traffic light reporter (TLR 2.1)[104], with the appropriate target site embedded within the GFP ORF. Cells were derived as previously described with slight modifications. Briefly, HEK293T cells were transduced with recombinant lentivirus to yield 5-10% transduction, based on cell survival after treatment with puromycin. Approximately five days post transduction, cells were sorted for mCherry- populations.

Cell sorting and flow cytometry

Cells were analyzed by flow cytometry on the BD LSRII and sorted on the BD FACS ARIAII. Fluorophores were detected using the following lasers and filters: mCherry – excited 561nm, acquired 610/20, mTagBFP – excited 405nm, acquired 450/50, eGFP – excited 488nm, acquired 525/50. Data were analyzed using FlowJo software.

Traffic Light Reporter Assay

The traffic light reporter assay was performed as previously described with slight modifications. Cells harboring the TLR 2.1 were plated at $2.0x10^5$ cells/well in a 24-well dish 24h prior to transfection. XtremeGene9 (Roche) was used at 2µl/µg DNA to transfect cells with 0.5µg of both nuclease and GFP donor constructs. Cells were harvested 72h post transfection and read on the flow cytometer. Data were obtained from BFP+ cell populations.

High-throughput sequencing and analysis

HEK 293T cells were plated at $2.0x10^5$ cells/well in a 24-well dish 24h prior to transfection. XtremeGene9 (Roche) was used at 2µl/µg DNA to transfect cells with 0.5µg each nuclease construct. Cells were harvested 72h post transfection and BFP expressing cells were sorted on the flow cytometer. Genomic DNA was extracted from these cells, as well as untreated 293T cells, using the Qiagen Blood and Tissue Kit. PCR of genomic DNA was performed using primers designed to target ~150-200bp on each side of the putative megaTAL cut site and append Illumina sequencing primer adaptors. Amplicons for all samples, including negative controls, were then subjected to a second round of amplification to append library barcodes as well as outer-most flowcell sequences. This setup allowed for library indexing at both the barcode and amplicon level. Final libraries were sequenced on 1.25 runs of a MiSeq (v2) using paired end 250 bp read chemistry with a 9 bp index read. Reads for each de-multiplexed library were aligned to their respective amplicon references using the Phaster read aligner (Phil Green, personal communication) due to its superior ability to accurately align reads with large indel events. Each aligned read pair was then individually genotyped for the presence of indels. Miseq run and data analysis was kindly performed by Andrew Adey in Dr. Jay Shendure's lab at the University of Washington.

Statistical analysis

Error bars on graphs represent s.e.m. P-values were calculated using Student's one-tailed unpaired t-test to compare activity of megaTALs with their specific meganuclease counterpart ($P<0.05$ shown as *, $P<0.005$ **, $P<0.0005$ ***).

Chapter 4

megaTALs: a nuclease platform compatible with both viral and non-viral delivery methods

4.1 Introduction

Therapeutic genome engineering requires a reagent that is deliverable to different cell types with reasonable efficiency and therefore must be compatible with different forms of transgene delivery. The three properties that affect the deliverability of a nuclease platform most are: 1) the number of separate proteins chains required to form a nuclease, 2) the size of the nuclease transgene and 3) the repetitiveness of the transgene sequence. Since megaTALs were developed as single chain proteins, we focused on reducing the size and repetitiveness of their DNA ORF in order to improve the efficiency of viral packing and delivery.

4.2 Reducing the size of the megaTAL transgene

Since our previous experiments had all employed megaTALs built with TAL effectors consisting of 14 or more repeat units, we wanted to determine the minimal number of repeats required for high megaTAL activity. We hypothesized that fewer repeat units would be required to form an active and specific megaTAL than would be similarly required for a TALEN half, due to the fact that the meganuclease cleavage domain can provide DNA target site affinity, while the FokI cleavage domain provides effectively none. By building megaTALs with fewer TAL effector repeats, the size of the megaTAL transgene could potentially be reduced significantly for the vast majority of engineered megaTALs (by approximately 100bp per repeat unit).

We used reporter cell lines previously generated to test the effect of DNA spacer length, as these already contained target inserts with closely matching DNA sequences for which TAL

effectors of increasing RVD array length could be built (Table 5, targets underlined in green). Because TAL effectors have been shown to exhibit polarity, with the N-terminal RVD repeats contributing more to the DNA target affinity than the C-terminal repeats, we ensured that the TAL effector targets used for comparison were matched towards their 5' ends to minimize possible sequence-specific variations in affinity that could contribute to differences in cleavage activity.[112] TAL effectors varying in array size from 2.5 to 10.5 repeat units were built to target corresponding DNA sequences 7bp upstream of the I-AniI target, as shown below. These synthetic TAL effectors were fused to the WT, F13Y and Y2 meganuclease variants and tested alongside the 16.5 RVD L538 megaTALs (discussed in Chapter 2) and the standalone meganucleases, in their appropriate reporter cell line.

No. of RVDs	Target site
2.5	ACATTCATTACACCTGCAGCTGTTATGAGGAGGTTTCTCTGTAAA
3.5	ACTTCATTACACCTGCAGCTCGTTATGAGGAGGTTTCTCTGTAAA
4.5	ATTCATTACACCTGCAGCTACGTTATGAGGAGGTTTCTCTGTAAA
5.5	TTCATTACACCTGCAGCTCAGCTTATGAGGAGGTTTCTCTGTAAA
6.5	TCATTACACCTGCAGCTACAGCTTATGAGGAGGTTTCTCTGTAAA
7.5	CATTACACCTGCAGCTATCAGCTTATGAGGAGGTTTCTCTGTAAA
8.5	ATTACACCTGCAGCTGATCAGCTTATGAGGAGGTTTCTCTGTAAA
10.5	TACACCTGCAGCTATGATCAGCTTATGAGGAGGTTTCTCTGTAAA
16.5	TTCATTACACCTGCAGCTCAGCTTATGAGGAGGTTTCTCTGTAAA

Table 5: Traffic Light Reporter targets used to test the effect of number of TAL effector array units on megaTAL activity, with the TAL effector and I-AniI binding sites underlined in green and blue, respectively.

MegaTALs constructed with the Y2 meganuclease all exhibited similar levels of homologous recombination and mutNHEJ, regardless of the number of RVD repeats in their TAL effector domain (Figure 11). We posit that this result is due to the N-terminal pseudo repeats of the TAL effector that increase the overall DNA affinity of the megaTAL sufficiently to achieve maximal cleavage activity with a meganuclease that exhibits high affinity on its own. MegaTALs constructed with the WT and F13Y meganuclease variants resulted in increased levels of

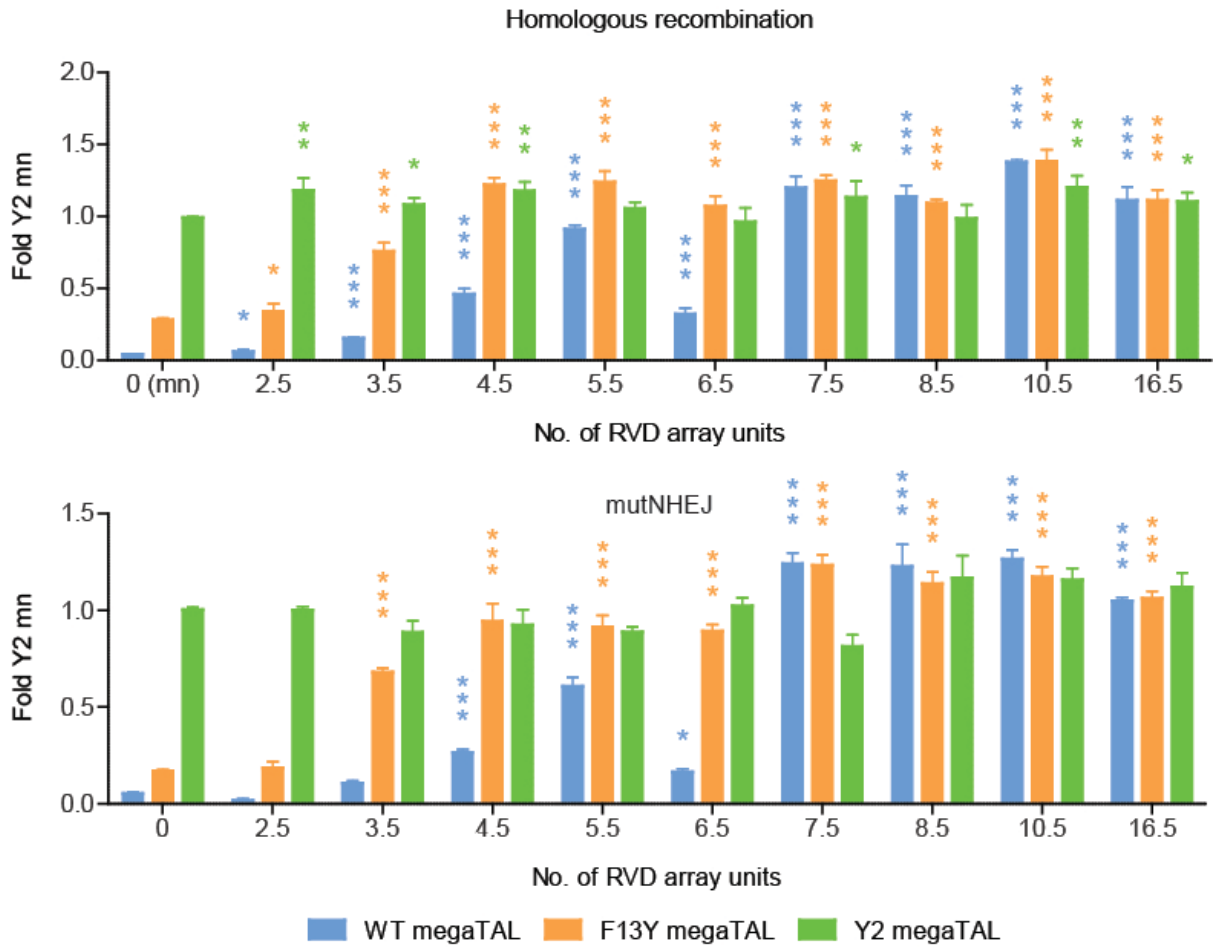

Figure 11: Effect of TAL effector RVD array number on nuclease activity in reporter cells. Level of mutNHEJ (top) and homolgous recombination (bottom) were measured after treatment with WT, F13Y and Y2 I-AniI megaTALs with varying number of TAL effector repeat units (2.5-16.5) or the standalone meganuclease (0).

cleavage activity over the standalone meganucleases with only a 3.5 RVD TAL effector array and cleavage levels rose in correlation with TALE repeat number until the activity reached its maximum with TAL effectors 7.5 and 4.5 RVDs long, respectively.

The one outlier point from this data set was the WT megaTAL made with a 6.5 RVD TAL effector, for which we observed an unexpected decrease in cleavage activity in relation to the other megaTALs tested. We attributed this result to either the presence of a second putative binding target for the 6.5 repeat TAL effector (ACAGCTT) within the TAL effector-spacer region of the intended TAL effector target or to the use of two RVDs at the C-terminus of the

TAL effector that provide low DNA affinity (NG for T and NI for A).[113] We reasoned that the presence of a second cryptic target could result in lowered cleavage activity due to binding of the TAL effector at this site blocking the meganuclease from accessing its DNA target, while the weak RVD-base interactions at the C-terminus of the TAL effector may not provide sufficient affinity for the TAL effector to bind the 3' end of its intended target site and effectively place the meganuclease in the required proximity to its cleavage site.

To investigate these possible causes, we built reporter cells harboring targets similar to the 6.5 RVD site tested (designated A), except either varying in the last nucleotide of the TAL effector binding site (designated C, G and T) or with the cryptic TAL effector binding site ablated (designated A*)(Table 6). Novel 6.5 RVD megaTALs were built with appropriate the RVD within the last repeat of the TAL effector to target each substituted base and were constructed with both the WT and F13Y meganucleases. Cleavage activity at the reporter target in which the cryptic TAL effector binding site was ablated (A*) was not increased compared to the original target (Figure 12). However, substitution of the last base of the TAL effector target resulted in significant differences in levels of homologous recombination and mutNHEJ, with cleavage activity correlating to the reported strength of RVD-base interactions (C > G > A > T).[113] These results suggest that the unexpected activity drop observed with the original 6.5 RVD TAL effector resulted from low affinity of the TAL effector for the 3' end of its target DNA.

Name	Last RVD	Target site
A	NI	TCATTACACCTGCAGCTACAGCTTATGAGGAGGTTTCTCTGTAAA
C	HG	TCATTACACCTGCAGCTCCAGCTTATGAGGAGGTTTCTCTGTAAA
G	NN	TCATTACACCTGCAGCTGCAGCTTATGAGGAGGTTTCTCTGTAAA
T	NG	TCATTACACCTGCAGCTTCAGCTTATGAGGAGGTTTCTCTGTAAA
A*	NI	TCATTACACCTGCAGCTAGCGCTTATGAGGAGGTTTCTCTGTAAA

Table 6: Traffic light reporter target sites for 6.5 RVD array megaTALs tested against targets in which the last base of the TAL effector target is varied (A, C, G and T) or the second cryptic TALE binding site ablated (A*).

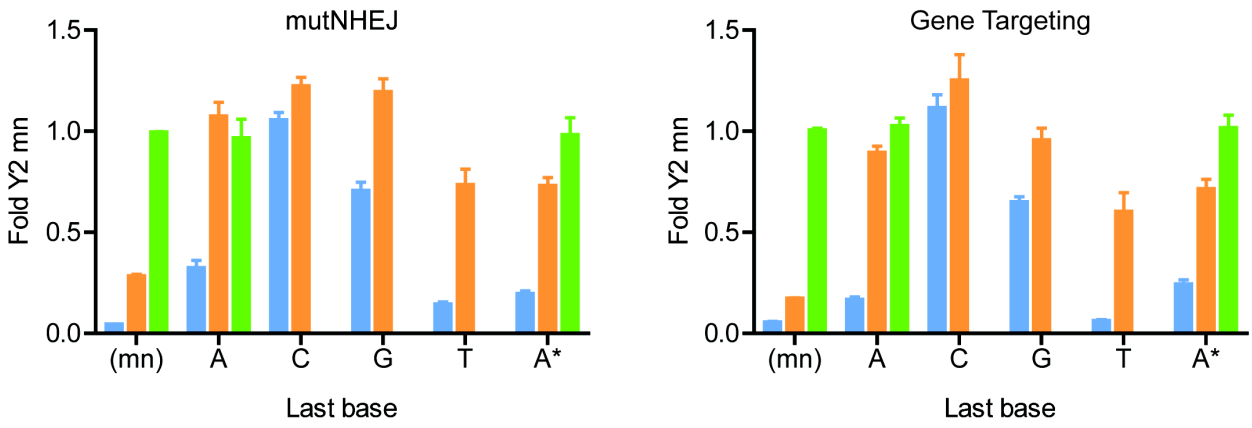

Figure 12: Cleavage activity measured using 6.5 RVD array megaTALS. Levels of mutNHEJ (left) and gene targeting (right) in Traffic Light Reporter cells using WT, F13Y and Y2 I-AniI megaTALs with 6.5 repeat TAL effectors varying in their last RVD or the standalone meganuclease (mn).

However, these findings demonstrate the complexity of TAL effector binding, as one would expect the 7.5 RVD megaTAL constructed with the WT meganuclease to show a similarly low level of cleavage activity, given that its TAL effector target differs from that of the 6.5RVD TALE only by an additional thymine at its 3' end.

Despite our inability to completely define the relationship between RVD-base interactions and cleavage activity, these results have shown that highly active megaTALs can be generated using significantly fewer TAL effector repeat units than previously tested. For even the low affinity WT meganuclease, a 6.5 RVD array provided sufficient supplemental affinity to reach maximal reporter activity. With this in mind, the size of megaTAL transgene can be reduced to approximately 2kb long, compared to 3kb for the original 16.5 RVD L538 megaTAL fusion constructed.

4.3 Precise megaTAL viral packaging and expression using codon diverged TAL effectors

Lentiviral vectors are commonly used for transgene delivery due to their ability to transduce a broad range of both dividing and non-dividing cells, high packaging capacity and low

```
6.5 RVDs GCAGCTA
      L  T  P  D  Q  V  V  A  I  A  S  X  X  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G
NN  CTCACACCTGACCAAGTTGTGGCAATTGCAAGCAATAACGGCGGCAAACAGGCACTCGAAACTGTGCAGCGGCTGCTCCCAGTGCTCTGCCAGGATCACGGG
HD  CTCACTCCCGATCAGGTCGTCGCTATTGCCTCTCATGACGGGGGGAAACAAGCTCTTGAAACAGTCCAGAGGCTTCTCCCTGTTCTGTGCCAAGATCATGGC
NI  TTGACCCCCGACCAGGTTGTTGCCATAGCCTCCAATATCGGAGGAAAACAGGCTCTGGAAACCGTCCAAAGACTCCTTCCCGTCTTGTGCCAGGACCATGGA
NN  CTTACCCCTGATCAAGTGGTGGCCATCGCTAGCAACAATGGGGGGCAAGCAGGCCCTGGAGACCGTGCAACGGCTCCTCCCCGTTCTCTGTCAAGACCATGGG
HD  CTGACACCAGACCAAGTCGTTGCTATCGCCAGCCACGATGGAGGTAAGCAAGCACTTGAGACAGTTCAGAGACTGTTGCCTGTGCTGTGTCAGGACCACGGC
NG  CTGACTCCAGACCAGGTGGTCGCAATCGCATCCAATGGCGGAGGGAAGCAGGCTTTGGAAACGGTTCAAAGGCTGCTGCCAGTCCTTTGCCAAGACCACGGA
HD  TTGACTCCTGATCAGGTAGTTGCAATAGCTTCTCACGACGGCGGGAAACAAGCACTCGAG
```

Figure 13: Alignment of RVD arrays used to generate a 6.5 RVD codon diverged megaTAL. The sequence of each array was diverged so as to reduce the extent of identical DNA stretches across repeats. Alternating codons are shown in red and black with the RVD codons in bold.

immunogenicity.[114] One of the major disadvantages of the TALEN platform is that these nucleases cannot be delivered to cells using lentivirus, due to the rearrangement of the repetitive TAL effector sequence during viral packaging.[86] Delivery of such recombined TALEN ORFs would not only lead to lowered expression of the full-length protein product, but also the expression of recombined nuclease halves with unanticipated sequence specificity that could cause significant off-target cleavage.

Given our finding that a 6.5 RVD TAL effector may be sufficient to transform low affinity meganucleases into highly active enzymes, we hypothesized that megaTALs could be properly packaged and expressed using lentiviral delivery by diverging the DNA sequence of each TAL effector repeat unit to prevent recombination. To test this, we built a codon diverged version of the 6.5(C)-WT megaTAL, discussed in the previous section, by avoiding the use of two identical successive codons across repeats wherever possible (Figure 13). Both integrating and non-integrating lentivirus were produced for the diverged (designated CD) and non-diverged (designated GG for the Golden gate cloning method used to build it) megaTALs, as well as the standalone WT meganuclease. We compared the activity of each nuclease delivered by lentiviral transduction and pDNA transfection using the appropriate reporter cell line. The results confirmed that both the CD and GG megaTALs were able to rescue the activity of the WT meganuclease, with no significant differences in levels of mutNHEJ observed between these nuclease after delivery by plasmid transfection (Figure 14). However, delivery of megaTALs by

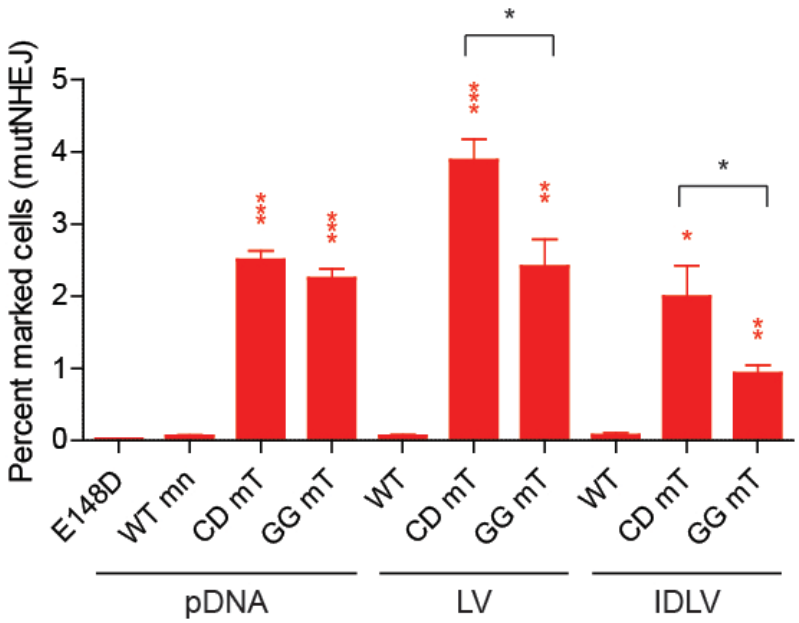

Figure 2: Comparing the effect of delivering diverged or non-diverged megaTAL gene ORFs. The WT meganuclease, diverged and non-diverged 6.5RVD-WT megaTAL were delivered into TLR 293T cells using pDNA, integrating or non-integrating lentivirus and levels of mutNHEJ were assessed 72 hours post delivery. T-test were performed either between the mn and mT (red *) or CD and GG mT (black *).

both integrating and non-integrating lentivirus resulted in approximately 1.5 to 2-fold less activity, respectively, in cells treated with the non-diverged megaTAL compared to the diverged. While we attributed this drop in activity to recombination of the non-diverged megaTAL ORF during viral packaging, the increased levels of mutNHEJ observed in cells treated with the GG megaTAL compared to the standalone WT meganuclease suggest that a portion of the GG megaTAL transgene delivered to cells expressed the full length protein product capable of cleaving the reporter target.

To visualize the level of the full length versus recombined megaTAL nuclease products being expressed under each condition, we collected whole cell lysates from the nuclease-treated reporter cells and ran a western blot to detect expression of nucleases by blotting for the HA tag. Bands corresponding to the expected size of the full-length nuclease protein were observed from each of the samples treated with either the meganuclease or CD megaTAL (Figure 15, top), regardless of the delivery method used. In contrast, blotting of protein collected from GG megaTAL lentiviral transduced cells revealed bands migrating at both the expected protein size and smaller. Protein expressed in the GG megaTAL pDNA transfected cells yielded the correct

sized band, indicating that the smaller protein products observed in LV-treated cells is directly resultant from the delivery vehicle and not intrinsic to the starting sequence of the non-diverged megaTAL. Expression of the GG megaTAL in IDLV treated cells was low, preventing visual assessment of the size of the expressing nuclease products. Taken together, these findings substantiate the assertion that lentiviral delivery of TAL effectors results in recombination of the transgene ORF, yet indicates that divergence of the TAL effector repeat sequences can effectively prevent such recombination.

We next wanted to determine the proportion of lentiviral integrations in which the megaTAL transgene remained intact (no recombination) in cells treated with either the codon diverged or non-diverged megaTALs. To do so, we performed a clonal analysis by amplifying the TAL effector RVD array off of genomic DNA prepped from nuclease-treated reporter cells, using

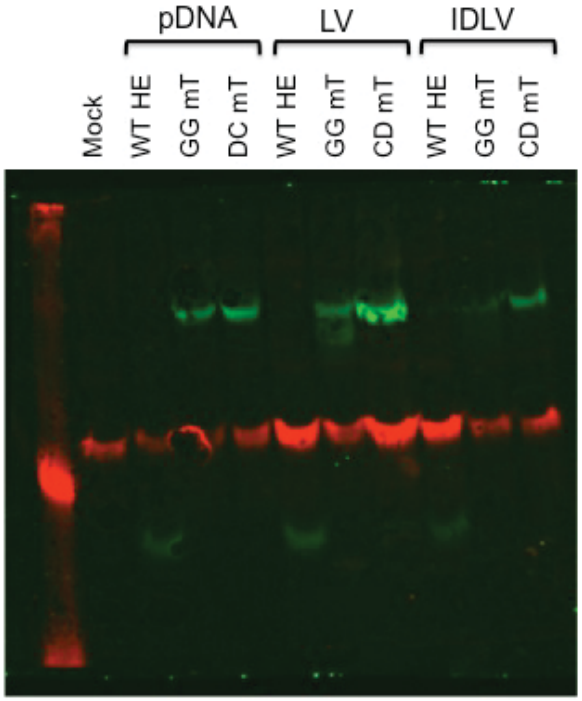

Figure 15: Analysis of the fidelity of transgene packaging after lentiviral delivery of diverged and non-diverged nucleases sequences. (Left) Western blot comparing protein expression (by HA tag labeling, green) in cells transfected with pDNA or transduced with LV or IDLV encoding the WT meganuclease, codon diverged and non-diverged megaTAL. This blot shows streaking corresponding the non-intact delivery of the non-diverged megaTAL by lentiviral delivery. Actin was blotted as a loading control (red). (Bottom) Clonal assessment of TAL effector transgene insertion size revealed that the diverged sequence results in near perfect delivery of nuclease, while the non-divereged version exhibits high levels of erroneous transgene integration.

Treatment	Correct size (%)	Incorrect size (%)	Total count
CD megaTAL	96.6%	3.4%	29
GG megaTAL	21.1%	78.9%	19

primers flanking the TAL effector repeat array. The amplified products were subsequently ligated into the pJET (Clonejet) vector and transformed into bacterial cells so that individual amplicons could be isolated for quantitative characterization of the precision of megaTAL packaging. Colony PCR was performed on individual clones obtained from each ligation reaction to reamplify the TAL effector repeat array and these PCR products were run on a gel to assess amplicon size. Results from this analysis indicate that only 21% of the non-diverged megaTAL delivered by lentiviral transduction encoded the full protein, while nearly all (96.6%) integrations obtained from LV transduction of the codon-diverged nuclease contain the intact ORF (Figure 15, bottom).

4.4 Summary

Assessment of the number of TAL effector RVD repeats needed for reaching maximal cleavage activity in reporter 293T cells revealed that megaTALs exhibit different requirements based on the affinity of the standalone meganuclease. A 6.5 repeat array was sufficient to achieve maximal activity with even the lowest affinity meganuclease tested, the WT variant, allowing for the construction of a highly active genome engineering reagent with only a 2kb transgene. Given that megaTALs made with TAL effectors binding shorter DNA sequences may result in greater off-target activity by increasing the likelihood of tethering a meganuclease in close proximity to a potential cleavage site, we believe that a 6.5RVD megaTAL architecture provides the appropriate balance between decreased transgene size and a reduced potential for off-target cleavage desired for therapeutic use. The reduced size requirement of the megaTAL positively impacts the ability to deliver these nucleases to the necessary cell type by ensuring sufficient levels of expression as well as their compatibility with multiple delivery vectors.

In addition, by reducing the number of TAL effector repeat units, we were able to package megaTALs made with a codon diverged 6.5 RVD array TAL effector into a lentiviral vector with near perfect fidelity. Sequencing of clones obtained from amplification of the TAL array off of codon diverged megaTALs should provide insight into the particular microhomologies that enable recombination and can inform the construction of a next generation codon diverged megaTAL in order to test whether the risk of transgene recombination can be completely ablated. The safe lentiviral delivery of megaTALs would provide an efficient method for transgene delivery that could also be used for applications in which long-term expression of nucleases is desired.

4.5 Materials & methods

MegaTAL and meganuclease construct generation

MegaTALs were constructed using the Golden Gate assembly strategy previously described by Cermak et al, using an RVD plasmid library and destination vector generously provided by the Voytas lab.[108] The pthX01 destination vector was modified to include a hemagglutinin (HA) tag immediately downstream of the NLS and to yield a NΔ154, C+63 TALEN scaffold. TAL effectors were built using the following RVDs to target each specific nucleotide: A – NI, C – HD, G – NN and T – NG. Following cloning of the TAL effector repeats into the destination vector, the Zn4 (VGGS) protein linker and I-AniI meganuclease variants were cloned in place of the FokI nuclease catalytic domain between the Xba-I and Sal-I restriction sites. Control constructs expressing standalone meganuclease variants were made by cloning the nuclease downstream of an NLS and HA tag between the Sbf-I and Sal-I restriction sites. The codon diverged 6.5C-WT megaTAL was constructed by gene synthesizing a TAL effector repeat array

(Invitrogen) that was then cloned between Age-I and Xho-I sites in the Golden Gate destination vector. All constructs encode BFP-T2A-nuclease for tracking of nuclease expression during flow cytometry.

Lentiviral production

Lentivirus (LV) and integration-deficient LV (IDLV) were generated as previously described.[115] Briefly, $9x10^6$cells were transfected in 10cm dishes with 6μg of the nuclease construct, 3μg of psPAX2 viral packaging plasmid and 1.5μg of either pMD2G (for LV) or integration-deficient D64V pMD2G (for IDLV) viral envelope plasmid using PEI (Polysciences). Two days post transfection viral supernatants were collected and concentrated by centrifugation. Viral titers were determined from the population of BFP expressing cells 72hrs after transducing 293T cells with varying volumes of concentrated virus using polybrene.

Cell line derivation

HEK293T cell lines were generated harboring a modified traffic light reporter containing an iRFP gene in place of puromycin (epigenetic TLR, unpublished data), with the appropriate target site embedded within the GFP ORF. Cells were derived as previously described with slight modifications. Briefly, HEK293T cells were transduced with recombinant lentivirus to yield 5-10% transduction, based on iRFP expression. Approximately five days post transduction, cells were sorted for iRFP+/mCherry- populations.

Cell sorting and flow cytometry

Cells were analyzed by flow cytometry on the BD LSRII and sorted on the BD FACS ARIAII. Fluorophores were detected using the following lasers and filters: mCherry – excited 561nm, acquired 610/20, mTagBFP – excited 405nm, acquired 450/50, eGFP – excited 488nm,

acquired 525/50, iRFP – excited 640nm, acquired 730/45. Data were analyzed using FlowJo software.

Traffic Light Reporter Assay

The traffic light reporter assay was performed as previously described with slight modifications. Cells harboring the epigenetic TLR were plated at $2.0x10^5$ cells/well in a 24-well dish 24h prior to transfection/transduction. For pDNA transfection, XtremeGene9 (Roche) was used at 2µl/µg DNA to transfect cells with 0.5µg of both nuclease and GFP donor constructs. For LV and IDLV transduction, cells were transfected at an MOI of 10 with 4µg/ml polybrene. Cells were harvested 72h post transfection and read on the flow cytometer. Data were obtained from iRFP+/BFP+ cell populations.

Western blot analysis

Reporter 293T cells were collected 72hr post treatment and lysed using RIPA lysis buffer with protease inhibitors, then lysates were run on a QiaShredder column (Qiagen). Protein concentrations were determined using the BioRad protein assay reagent and 5-10µg of protein was run on a 10% acrylamide gel alongside Kaleidoscope protein ladder (BioRad) then transferred onto an Immobilo-FL membrane (Millipore) using semi-dry transfer. The membrane was blocked using Licor blocking buffer prior to staining with primary rabbit anti-HA (Cell Signaling Technology) and mouse anti-β-actin monoclonal antibodies and secondary PE-conjugated anti-mouse (Medial & Biological Laboratories) and GFP-conjugated anti-rabbit (Invitrogen) monoclonal antibodies. Membrane fluorescence was then detected using the Licor Odyssey.

PCR analysis of TAL effector size

Reporter 293T cells were collected 72hr after lentiviral transduction and genomic DNA was extracted from these cells using the Qiagen Blood and Tissue Kit. Nested PCR was performed using two sets of primers to amplify the TAL effector repeats from the integrated megaTAL ORFs as well as the original megaTAL constructs for comparison. The PCR products from the LV-treated samples were cloned into the pJET vector (Clonejet) and transformed into DH5α cells. For each sample, 47 colonies were picked and subjected to colony PCR to amplify the pJET insert. PCR products were run on a 1.2% agarose gel for size determination.

Statistical analysis

Error bars on graphs represent s.e.m. P-values were calculated using Student's one-tailed unpaired t-test to compare activity of megaTALs with their specific meganuclease counterpart ($P<0.05$ shown as *, $P<0.005$ **, $P<0.0005$ ***).

Chapter 5

megaTALs: a therapeutic grade nuclease

5.1 Introduction

While each of the experiments described above was carried out using close variants of the native I-AniI meganuclease, genome engineering applications will require that non-native, engineered meganucleases be transformed into functional megaTALs. For this reason, we wanted to test a megaTAL directed towards a translational target of interest to verify that therapeutic reagents could be constructed that exhibit the same properties observed in the proof-of-principle experiments described in previous chapters. Moreover, we wanted to determine whether the benefits of the megaTAL platform could be captured in a primary cell line with relevance to therapeutic applications. To this end, we chose to focus our studies on a variant of the I-OnuI meganuclease designed to knockout the TCRα (T-cell receptor alpha chain) gene.

5.2 Building a TCRα megaTAL

TCRα as a therapeutic target

Knockout of the TCRα gene is highly desired in order to create an anti-cancer reagent that can be administered to a wide number of recipients. T-cells expressing synthetic chimeric antigens receptors (CARs) that recognize tumor associated antigens can be used to target and kill cancerous cells. However, because transplantation of allogeneic donor cells can lead to graft-versus-host disease, knockout of the T-cell receptor (TCR) is first required in order to generate a universal reagent that can be given to multiple recipients.[116] Disruption of the TCRα subunit of the T-cell receptor has been shown to yield this desired TCR knockdown.

I-OnuI	TTTCCACTTATTCAACCTTTTA
TCRα	TgTCtgCcTATTCAcCgaTTTt

Figure 16: Alignment of the DNA specificity of the starting I-OnuI scaffold and evolved TCRα variant. Substitutions from the native target for which the specificity of the enzyme had to be reengineered are shown as lower case letters.

A TCRα meganuclease

A TCRα meganuclease was engineered from the native I-OnuI meganuclease by Dr. Jordan Jarjour of Pregenen, Inc, using a method by which the native enzyme is split into 'modules' by redesigning its specificity towards substitutions within short pockets of the DNA target and the resulting module mutations are combined to target the full desired sequence.[117] An alignment of the native I-OnuI and reengineered TCRα target sites are shown below (Figure 16). Analysis of the meganuclease binding activity by flow cytometry revealed that the enzyme exhibits a high affinity, with a K_D only slightly lower than the original I-OnuI starting scaffold (50 vs 80 pM, Figure 17).

A TCRα megaTAL

Given that the structure of the I-AniI and I-OnuI meganucleases are highly identical, we

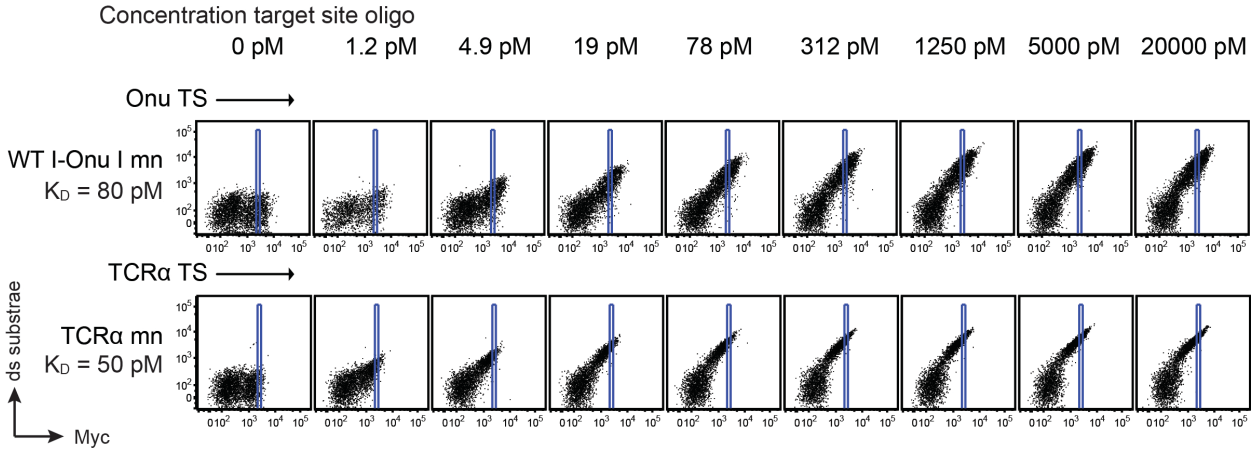

Figure 17: Affinity of the I-OnuI and TCRaα meganucleases for their targets. Increasing amounts (0-2000pM final) of DNA substrates containing the I-Onu I meganuclease and TCRα meganuclease targets (directly labeled with Alexa-647 to indicate binding) was incubated with yeast displaying the appropriate nuclease (labeled with anti-Myc FITC to indicate expression) indicating that the target site affinity of the engineered TCRα meganuclease is similar to the of the wild-type I-Onu I meganuclease.

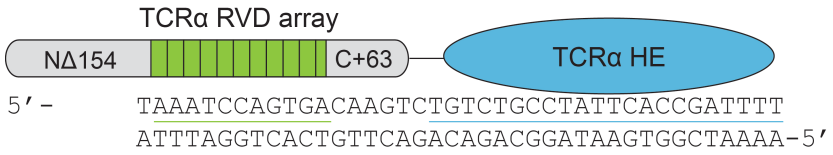

Figure 18: Schematic of the TCRα megaTAL formed by fusing a synthetic TAL effector built towards the TCRα sequence underlined in green to a designed TCRα meganuclease (target underlined in blue).

chose to build a TCRα megaTAL with a 7bp DNA spacer (6bp, assuming a 22bp mn target, as shown in Figure 18). Due to the preference for a thymine at the 0th position of the DNA, there were two appropriate sequences towards which we could build the TAL effector: sequences 7 or 11bp upstream of the DNA spacer. Despite the high affinity of the meganuclease, suggesting that the shorter TAL effector would function as effectively as the longer, we opted to construct a megaTAL with a 10.5 RVD array in order to ensure maximal cleavage activity could be achieved.

5.3 Assessing TCRα knockdown in primary T-cells

T-cell receptor knockout

Primary T-cells were transfected with mRNA encoding either the TCRα megaTAL or the stand-alone meganuclease. In order to increase to efficiency of gene knockout, samples were included wherein either the meganuclease or the megaTAL were co-transfected with mRNA encoding Trex2, a 3'->5' exonuclease that has been shown to significantly increase rates of mutNHEJ at meganuclease cleavage sites.[118,119] Disruption of the TCRα gene was measured by staining for CD3, a multi-protein complex that is associated with the α/β T-cell receptor. Strikingly, while gene disruption occurred at minimal rates in cells treated with the stand-alone meganuclease (1.6%), the TCRα megaTAL exhibited approximately a 20-fold increase in activity at the addressed target site in the TCRα gene (Figure 19). The low cleavage activity of

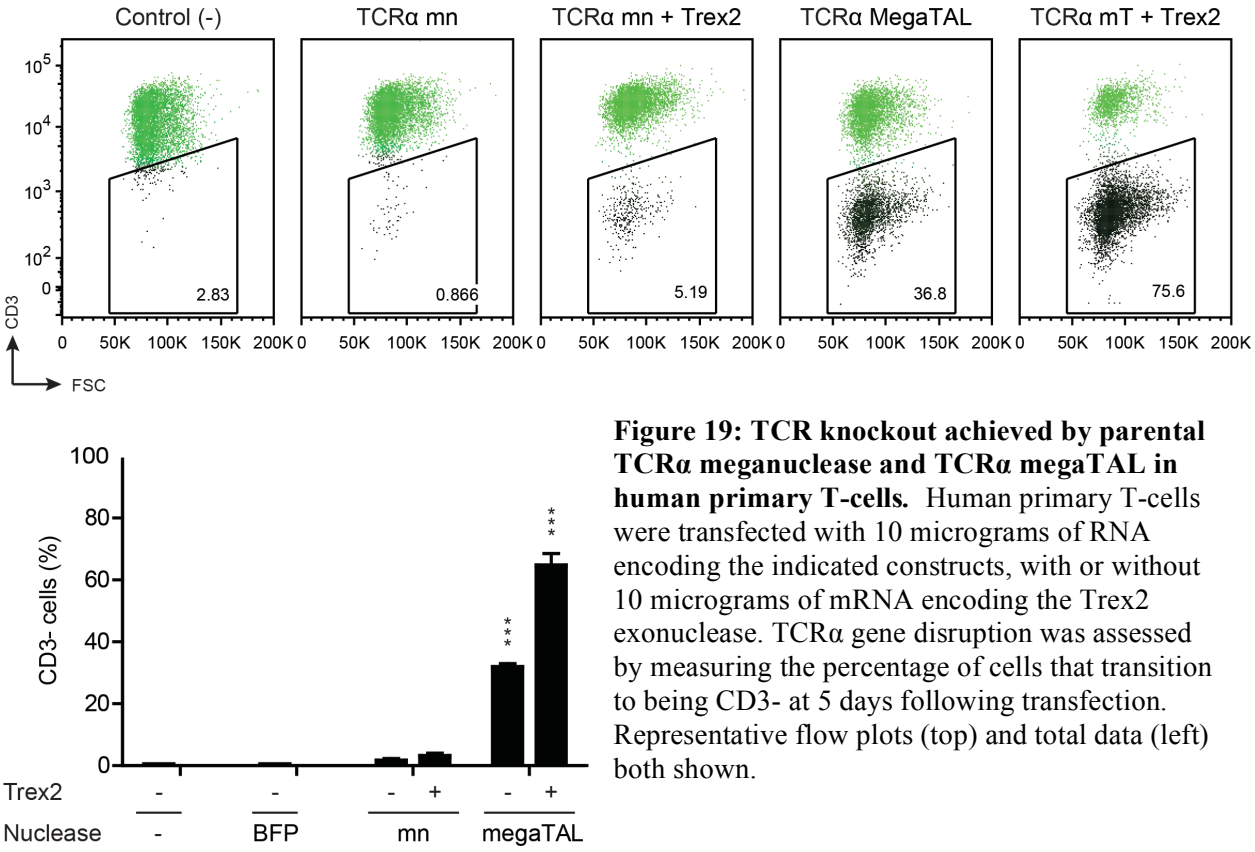

Figure 19: TCR knockout achieved by parental TCRα meganuclease and TCRα megaTAL in human primary T-cells. Human primary T-cells were transfected with 10 micrograms of RNA encoding the indicated constructs, with or without 10 micrograms of mRNA encoding the Trex2 exonuclease. TCRα gene disruption was assessed by measuring the percentage of cells that transition to being CD3- at 5 days following transfection. Representative flow plots (top) and total data (left) both shown.

the meganuclease was somewhat unexpected, given the high starting affinity of the standalone meganuclease and levels of cleavage obtained in 293T cells with lower affinity meganucleases, and underline the importance of testing nuclease reagents in the relevant primary cell type. This reduced activity could be due to lower DNA accessibility, which could be overcome by the added megaTAL affinity, or more robust DSB repair in T-cells, for which scaffolding of the meganuclease by the TAL effector could increase the recurrence of DSB formation and thus likelihood of indel incorporation. Co-transfection of each nuclease with Trex2 further increased the rate of disruption, yielding rates of TCRα gene disruption consistently exceeding 70%. Given that no marker was used to approximate transgene delivery, these values may provide an underestimation of the rates of knockout in nuclease-expressing cells.

Mutation rates at the TCRα locus

Nuclease activity was further examined by high-throughput sequencing of the TCRα locus in FACS sorted CD3- populations of nuclease/Trex2 treated cells and indel rates obtained from these samples were compared to those from untreated T-cells (Table 7). The rate of indel formation at the TCRα locus in cells treated with only the meganuclease was found to be below background (<0.1% vs. 0.2% for control cells). This result was not surprising given that the sample used for sequencing also showed lower levels of CD3 knockout by staining (Figure 19, top), suggesting that the sorted CD3- population consisted of predominantly unstained cells. Other replicates in which the mn produced above background CD3 knockout would have likely have yielded higher rates of indel formation by sequencing. Cells treated with both meganuclease and Trex2 yielded indels at the TCRα locus in 39% of sequenced copies, corresponding to the expected rate of monoallelic disruption given some background, unstained cells contaminating the sorted population. As expected, cells treated with the megaTAL nuclease, both with and without Trex2, contained indels rates greater than 50%, substantiating that observed CD3 knockdown was due to mutations at the expressing, open TCRα allele. Surprisingly, indel rates in megaTAL+/Trex2- treated cells exceeded those in which Trex2 was coexpressed (71% vs 52%, respectively), indicating higher rates of biallelic disruption in the former set. These results may simply be due to sampling error or increased nuclease expression in the singly-transfected cells, however, they may also suggest a more complex mechanism by which megaTALs are able to more readily access the closed TCRα allele in the absence of Trex2.

% mutNHEJ *in vivo* at TCRα and putative off-target loci (CD3- sorted populations)

Locus	Control	TCRα mn	TCRα mn + Trex2	TCRα mT	TCRα mT + Trex2
TCRα	0.225	0.062	38.609	70.599	52.219
OT1	0.352	0.259	0.343	0.378	0.349
OT2	0.659	NA	NA	0.558	0.572
OT3	0.605	0.055	0.578	0.575	0.309
OT4	0.194	0.157	0.177	0.143	0.382
OT5	0.16	0.157	4.052	0.162	0.198
OT6	0.459	0.692	0.407	0.517	0.537
OT7	0.046	NA	0.003	0.032	0.042
OT8	0.113	0.129	0.082	0.159	0.099
OT9	0.121	0.267	0.338	0.237	0.276
OT10	0.225	0.19	0.018	0.361	0.797
OT11	0.137	0.369	0.989	0.162	NA
OT12	NA	NA	NA	NA	NA
OT13	0.17	0.098	0.122	0.12	0.162
OT14	0.779	0.711	0.888	0.749	0.681
OT15	0.896	1.136	0.816	0.934	0.884
OT16	0.214	0.232	0.318	0.233	0.279
OT17	0.55	0.584	0.763	0.629	0.646
OT18	0.387	0.433	0.629	0.347	0.382
OT19	NA	NA	NA	NA	NA
OT20	0.202	0.182	0.295	0.223	0.196
OT21	0.199	0.145	0.144	0.195	0.181
OT22	0.252	0	0.213	0.257	0.232
OT23	0.352	0.383	0.302	0.464	2.078
OT24	0.132	0.121	0.154	0.135	0.148

Table 7: High-throughput sequencing results and analysis of on-target and putative off-target cleavage in T-cells. Summary of high-throughput sequencing results at TCRα and putative off-target loci from T-cells treated with the TCRα meganuclease or megaTAL +/- Trex2.

5.4 Assessing off-target cleavage of therapeutic nucleases

Identifying putative human genomic off-target sites

Using information gathered about the specificity profile of the TCRα meganuclease (data not shown), our collaborators developed a position-specific scoring matrix (PSSM) to identify putative off-target sites within the human genome. The activity of the TCRα meganuclease was tested *in vitro* against a subset of 192 genomic targets identified from the PSSM, using a flow-cytometry based cleavage assay. From these, twenty-four targets for which the meganuclease displayed above background cleavage activity were designated putative off-targets sites and chosen for further examination (Figure 20, Table 8).

TCRα nuclease off-target analysis

High-throughput sequencing of the putative off-targets sites was performed from CD3-nuclease/Trex2 treated samples, as discussed in the previous section. Rates of indel formation were obtained for 22 of these targets which each of the different treatments, as well as control untreated T-cells (Table 7). Results from the OT12 and OT19 loci were thrown out due to high background rates of mutations in all samples, including untreated.

For the majority of the putative off-target sites scrutinized, treatment with either the meganuclease of megaTAL +/- Trex2 did not result in significantly elevated levels of indel formation compared to background levels in untreated cells (with rates from 0.6% below to 0.9% above background, excluding the two samples discussed in the next paragraph). Given the variable background levels obtained across these sequenced loci and that sequencing was performed in singlicate, it was not possible to determine statistically whether nuclease-treated samples with slightly increased indel rates were due to normal variance or to off-target cleavage. Importantly, cells treated with only the TCRα megaTAL exhibited very low levels of indel

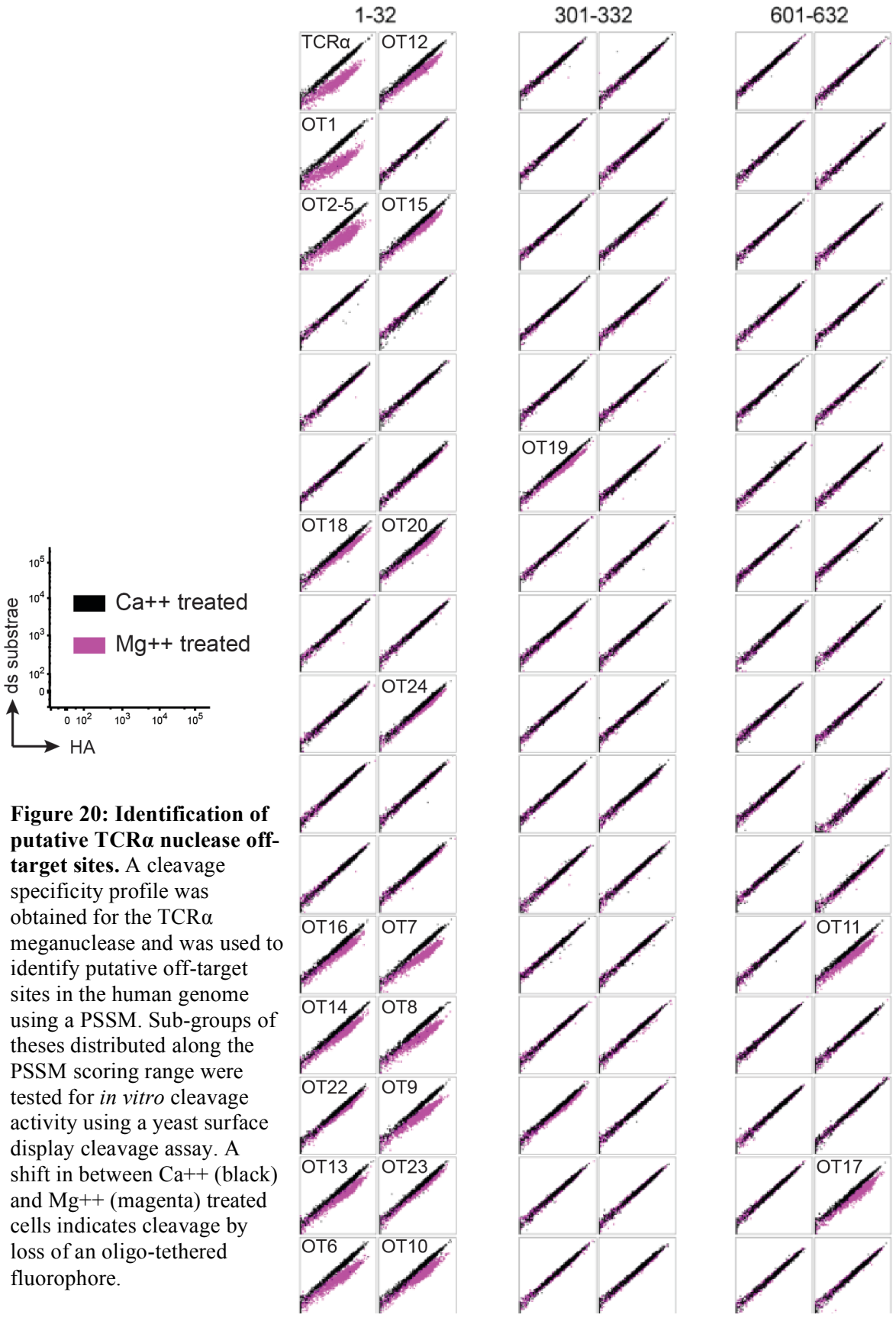

Figure 20: Identification of putative TCRα nuclease off-target sites. A cleavage specificity profile was obtained for the TCRα meganuclease and was used to identify putative off-target sites in the human genome using a PSSM. Sub-groups of theses distributed along the PSSM scoring range were tested for *in vitro* cleavage activity using a yeast surface display cleavage assay. A shift in between Ca++ (black) and Mg++ (magenta) treated cells indicates cleavage by loss of an oligo-tethered fluorophore.

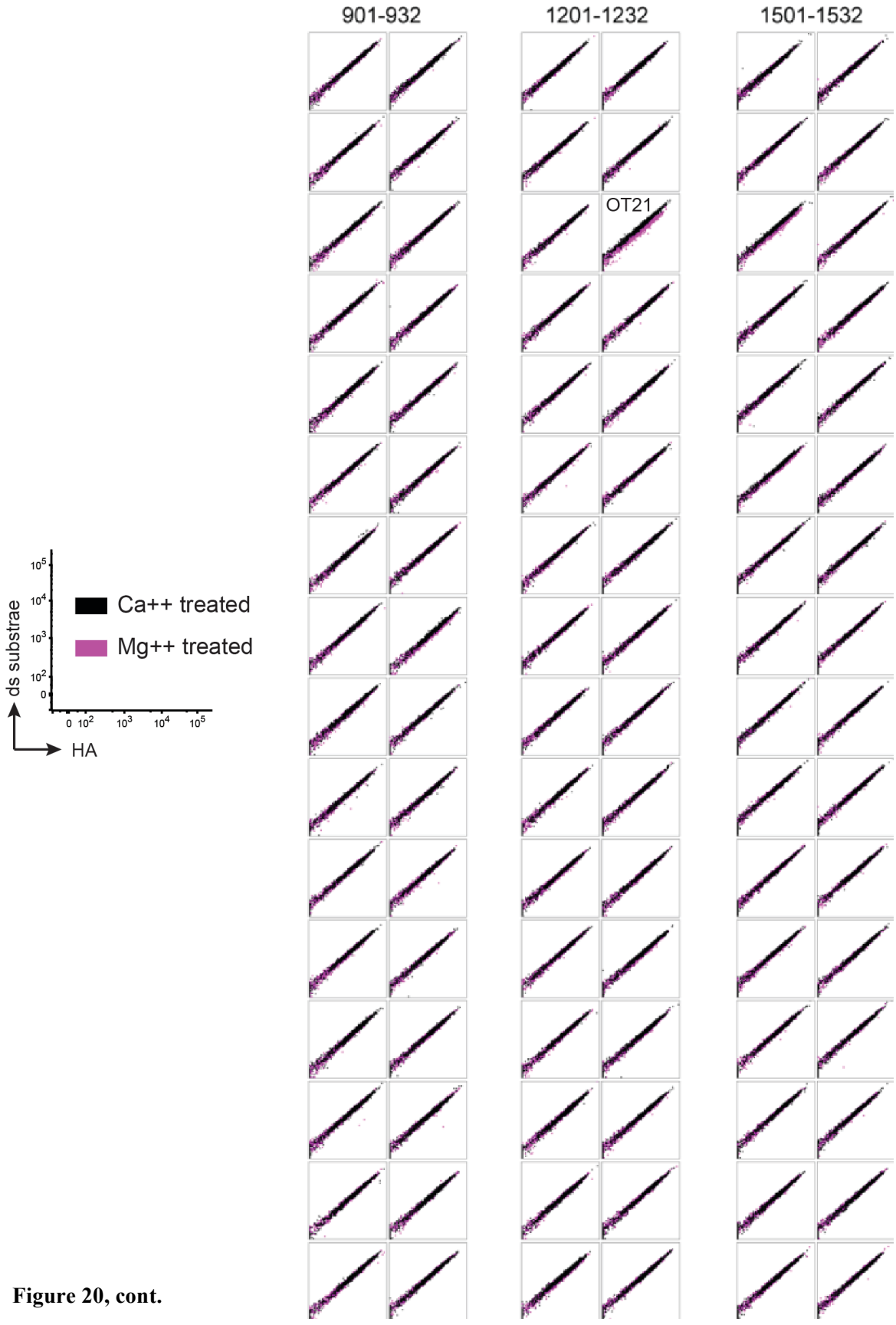

Figure 20, cont.

Putative TCRα nuclease genomic off-targets

Name	Target sequence	Genomic Locus	Ca^{++}/Mg^{++}
TCRα	TGTCTGCCTATTCACCGATTTT	Ch14: (+)23016508-23016529	4.27
OT1	TGTaTGCCTtTTCACCGATTaa	Ch9: (-)21736012-2176033	5.72
OT2	TGTaTGCCTtTTCACCtATTaa	ChX: (+)112764854-112764875	3.74
OT3	TGTaTGCCTtTTCACCtATTaa	Ch20: (+)53071086-53071107	3.74
OT4	TGTaTGCCTtTTCACCtATTaa	Ch9: (+)34907420-34907441	3.74
OT5	TGTaTGCCTtTTCACCtATTaa	Ch9: (-)21119840-21119861	3.74
OT6	TGcCgGCCTtTTCACCcATcTg	Ch7: (-)112542535-112542556	2.92
OT7	GGaaTGCtTATTCACCGtTTTT	Ch14: (+)29466641-29466662	2.73
OT8	TGTgTGCtTATTCcCCtATTTa	Ch4: (-)34480172-34480193	2.68
OT9	TGgCTGCtgATTCcCCtATTTT	Ch2: (-)157897003-157897024	2.48
OT10	aGTgTGCCTATTCACCaATcTc	Ch4: (-)109209885-109209906	2.25
OT11	cacCaGCCTtTTCACCGtTcac	Ch15: (-)33161499-33161520	2.10
OT12	TGTaTGCCTtTTCcCCtATTaa	Ch6: (-)90026596-90026617	2.01
OT13	TGTaTGCCTtTTCcCCtATTaa	Ch10: (+)54726301-54726322	2.01
OT14	TGTaTGCCTtTTCcCCtATTaa	Ch20: (+)53147268-53147289	2.01
OT15	TGTaTGCCTtTTCcCCtATTaa	Ch2: (+)152020105-152020126	2.01
OT16	aGgtTGCCTtTTCACCGtTcTT	Ch1: (+)240704451-240704472	1.94
OT17	aaaaTGCCTtTTCACCtATcac	Ch13: (+)95104630-95104651	1.87
OT18	TtTaTGCCTtTTCACCGtTTTT	Ch4: (-)115725301-115725322	1.74
OT19	gtTtgGCtTATTCACCGATTTc	Ch2: (+)87036974-87036995	1.64
OT20	cGTaTGCCTtTTCcCCtATcTa	Ch5: (+)23787010-23787031	1.60
OT21	atgagGCtTATTCACCccTcaa	Ch2: (-)211575763-211575784	1.53
OT22	TGaaTGCCTtTTCACCtATTaa	Ch20: (-)15561343-15561364	1.50
OT23	gGctTGCCTtTTCACCaATTTT	ChX: (+)92468114-92468135	1.49
OT24	gaTCTGCCTtTTCACCaATTTT	Ch10: (-)109004204-109004225	1.47

Table 8: Table of putative genomic off-target sites that were further analyzed for *in vivo* cleavage with the TCRα meganuclease or megaTAL with the identifying name, target site sequence (lower case letters indicated base substitutions from the TCRα target), genomic position and Ca^{++}/ Mg^{++} ratio (indicating level of cleavage by yeast display).

formation at the putative off-target loci, with rates ranging from 0.1% below to 0.1% above background.

There were only two sites at which greater than 1% above background indel rates were obtained by sequencing - OT5 (3.9% with the meganuclease + Trex2) and OT23 (1.7% with the megaTAL + Trex2). It is not entirely clear, however, that these increased rates are resultant from off-target nuclease cleavage rather than sampling error, as discussed in the two following points. For one, we would expect to observe a similar increase in cleavage activity at OT5 in cells treated with both the TCRα megaTAL and Trex2, given that we have seen no evidence of the megaTAL TAL effector being able to reduce off-target cleavage. Second, we would expect increased levels of cleavage activity at OT23 in megaTAL/Trex2 compared to meganuclease/Trex2 treated cells only if the locus contains a DNA target at which the megaTAL TAL effector could bind to "address" the meganuclease, however, an in-depth search allowing for high RVD-base degeneracy did not identify any such possible binding site. These discrepancies may again simply reflect differences in the accumulation of errors between samples during PCR amplification, or may be resultant from lower expression of the larger megaTAL protein relative to the standalone meganuclease.

5.5 Summary

Fusion of a locus-specific TAL effector to the engineered TCRα meganuclease resulted a hyperspecific reagent capable of capable of extremely high levels of cleavage activity. Despite its high affinity, the standalone TCRα meganuclease yielded low rates of *in vivo* cleavage and gene knockout in T-cells, whereas fusion of a 10.5RVD TAL effector to this endonuclease increased gene disruption to 70% when paired with the Trex2 exonuclease. Furthermore, of the

22 genomic targets found to be cleaved *in vitro* by the meganuclease, only one of these loci showed significant levels of mutation after treatment with both the megaTAL and Trex2. These results support our earlier findings that low activity meganucleases can be transformed into hyperspecific megaTAL reagents for achieving safe and effective therapeutic genome engineering.

5.6 Materials & methods

MegaTAL and meganuclease construct generation

The TCRα meganuclease was generously provided by Pregenen, Inc. The TCRα megaTAL was constructed using the Golden Gate assembly strategy previously described by Cermak et al, using an RVD plasmid library and destination vector generously provided by the Voytas lab.[108] The pthX01 destination vector was modified to include a hemagglutinin (HA) tag immediately downstream of the NLS and to yield a NΔ154, C+63 TALEN scaffold. TAL effectors were built using the following RVDs to target each specific nucleotide: A – NI, C – HD, G – NN and T – NG. Following cloning of the TAL effector repeats into the destination vector, the Zn4 (VGGS) protein linker and TCRα meganuclease were cloned in place of the FokI nuclease catalytic domain between the XbaI and SalI restriction sites and the T7 promoter was cloned in upstream of the transcription start site between the AscI and SbfI restriction sites. A control construct expressing the standalone meganuclease was made by cloning the nuclease downstream of an NLS and HA tag between the SbfI and SalI restriction sites.

In vitro production of synthetic mRNA

We used the T7-Scribe kit (CellScript) for production of synthetic mRNA. Two different templates were utilized for *in vitro* transcription: a linearized plasmid (megaTAL and HE) or a

PCR product containing a T7-promoter (Trex2), generated with the following primers: Trex2 Forward,

GGATCCTAATACGACTCACTATAGGGGCCGCCACCATGTCTGAGCCACCTCGGGCTGAG;

Trex2 reverse,

TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAGGCTTCGAGGCTTGGAC, with T7 promoter underlined. The resultant mRNA was purified with the RNeasy miniprep columns (Qiagen) and a 5' cap was enzymatically added using the vaccinia capping system combined with a Cap 2'-O-Methyltransferase enzyme (both from NEB). A poly(A) tailing kit (Life Technologies) was used to add a polyadenylation tail to 5'-capped mRNA and mRNA was purified using the RNeasy kit. The capped/polyadenylated mRNA was eluted in sterile water and stored at -20°.

Cell culture

We used the EasySEP Human T-cell enrichment kit (Stemcell technologies) to purify T cells from human peripheral blood mononuclear cells (PBMCs) obtained from anonymous donors at the Puget Sound Blood Center. Isolated T-cells were stimulated with CD3/CD28 activation beads (Life Technologies) based on manufacturer's instructions. Activated T-cells were cultured in RPMI supplemented with 10% FCS, 1x GlutaMAX, 55mM b-mercaptoethanol, 100 U/ml penicillin, 100mg/mlstreptomycin, 10mM HEPES and 10ng/ml IL-2 (Biolegend). The beads were magnetically removed after 36hr of incubation and the T-cell resuspended in fresh culture media for 6-12 hours prior to electroporation (48-56hrs total culture time before electroporation).

Human T-cell transfection

The Neon transfection system (Life Technologies) was used for T-cell electroporation, with the 10ml electroporation kit used for all transfections. T-cells were washed with PBS and

resuspended in Buffer T at $2x10^7$ cells/ml, with $2x10^5$ cells used per transfection. Approximately 1-1.5mg of each individual mRNA was added to each sample prior to electroporation. A maximum of 2.5mg of total mRNA was used for samples that had more than mRNA species (ie megaTAL with Trex2). Cells were electroporated with the following conditions: 3 pulses, 1400V/pulse, with 10ms pulse width. After electroporation, cells were immediately dispersed into pre-warmed, antibotic-free T-cell media with IL2 and cultured for 4-6 days, with regular media changes, prior to analysis. Analysis of the T-cell receptor expression was performed using the anti-CD3-AlexaFluor 488 antibody (Biolegend, clone SK7).

Cell sorting and flow cytometry

Cells were analyzed by flow cytometry on the BD LSRII and sorted on the BD FACS ARIAII. Fluorophores were detected using the following lasers and filters: eGFP – excited 488nm, acquired 525/50. Data were analyzed using FlowJo software.

High-throughput sequencing and analysis

Following mRNA transfection, CD3- T-cells were sorted on the flow cytometer. Genomic DNA was extracted from these, as well as control T-cells, using the Qiagen Blood and Tissue Kit. PCR of genomic DNA was performed using primers designed to target ~150-200bp on each side of the putative megaTAL cut site and append Illumina sequencing primer adaptors. Amplicons for all samples, including negative controls, were then subjected to a second round of amplification to append library barcodes as well as outer-most flowcell sequences. This setup allowed for library indexing at both the barcode and amplicon level. Final libraries were sequenced on 1.25 runs of a MiSeq (v2) using paired end 250 bp read chemistry with a 9 bp index read. Reads for each de-multiplexed library were aligned to their respective amplicon references using the Phaster read aligner (Phil Green, personal communication) due to its

superior ability to accurately align reads with large indel events. Each aligned read pair was then individually genotyped for the presence of indels. Miseq run and data analysis was kindly performed by Andrew Adey in Dr. Jay Shendure's lab at the University of Washington.

Statistical analysis

Error bars on graphs represent s.e.m. P-values were calculated using Student's one-tailed unpaired t-test to compare activity of megaTALs with their specific meganuclease counterpart ($P<0.05$ shown as *, $P<0.005$ **, $P<0.0005$ ***).

Chapter 6

Overview of the megaTAL platform

In summary, the megaTAL nuclease architecture described here addresses the key limitations of existing nuclease platforms for therapeutic genome editing. For one, these reagents can be readily generated due to the ease with which TAL effectors can be built and the reduced affinity requirement of the meganuclease domain (Chapter 2). More importantly for therapeutic applications, megaTALs may achieve an extraordinary level of target specificity due to their extended DNA recognition size and linked binding and cleavage. We have shown that by "addressing" megaTALs towards a TAL effector binding site, we can drastically increase the level of on-target cleavage and that megaTALs made with low affinity meganucleases can achieve extremely low rates of off-target cleavage (Chapter 3). Furthermore, we found that because the meganuclease cleavage domain of the megaTAL exhibits sequence-specificity, the number of TAL effector repeats units required to achieve maximal activity could be significantly reduced and allowed for divergence of the repetitive elements to prevent recombination of the nuclease ORF during lentiviral packaging (Chapter 4). Lastly, using the information that was gathered on the ideal properties of a megaTAL fusion protein, we generated a therapeutic megaTAL with the goal of knocking out the TCRα gene in T-cells. As with our initial test set of megaTALs, this reagent rescued activity of an engineered meganuclease to exhibit extremely high on-target cleavage with very low off-target activity (Chapter 5). Based on these results, we conclude that the enhanced properties available with the megaTAL architecture thus set a new standard for efficiency and safety of nuclease-based genome engineering for human therapeutic applications.

References

1. Herring, M. Y. *Genetic Engineering*. (Greenwood Press, 2006).
2. Avery, O. T., MacLeod, C. M. & McCarty, M. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III. *The Journal of Experimental Medicine* **79**, 137–158 (1944).
3. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
4. Weiss, B. & Richardson, C. C. Enzymatic breakage and joining of deoxyribonucleic acid, I. Repair of single-strand breaks in DNA by an enzyme system from Escherichia coli infected with T4 bacteriophage. *Proc. Natl. Acad. Sci. U.S.A.* **57**, 1021–1028 (1967).
5. Lederberg, J. Cell genetics and hereditary symbiosis. *Physiol. Rev.* **32**, 403–430 (1952).
6. Danna, K. & Nathans, D. Specific cleavage of simian virus 40 DNA by restriction endonuclease of Hemophilus influenzae. *PNAS* **68**, 2913–2917 (1971).
7. Jackson, D. A., Symons, R. H. & Berg, P. Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* **69**, 2904–2909 (1972).
8. Cohen, S. N. & Chang, A. C. Recircularization and autonomous replication of a sheared R-factor DNA segment in Escherichia coli transformants. *Proc. Natl. Acad. Sci. U.S.A.* **70**, 1293–1297 (1973).
9. Friedmann, T. & Roblin, R. Gene Therapy for Human Genetic Disease? *Science* **175**, 949–955 (1972).
10. Blaese, R. M. *et al.* T Lymphocyte-Directed Gene Therapy for ADA− SCID: Initial Trial Results After 4 Years. *Science* **270**, 475–480 (1995).
11. Muul, L. M. *et al.* Persistence and expression of the adenosine deaminase gene for 12 years and immune reaction to gene transfer components: long-term results of the first clinical gene therapy trial. *Blood* **101**, 2563–2569 (2003).
12. Cavazzana-Calvo, M. *et al.* Gene Therapy of Human Severe Combined Immunodeficiency (SCID)-X1 Disease. *Science* **288**, 669–672 (2000).
13. Hacein-Bey-Abina, S. *et al.* Sustained Correction of X-Linked Severe Combined Immunodeficiency by ex Vivo Gene Therapy. *New England Journal of Medicine* **346**, 1185–1193 (2002).
14. Gaspar, H. B. *et al.* Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. *The Lancet* **364**, 2181–2187 (18).
15. Hacein-Bey-Abina, S. *et al.* Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J. Clin. Invest.* **118**, 3132–3142 (2008).
16. Sheridan, C. Gene therapy finds its niche. *Nature Biotechnology* **29**, 121–128 (2011).
17. Wu, X., Li, Y., Crise, B. & Burgess, S. M. Transcription start regions in the human genome are favored targets for MLV integration. *Science* **300**, 1749–1751 (2003).
18. Bushman, F. *et al.* Genome-wide analysis of retroviral DNA integration. *Nat. Rev. Microbiol.* **3**, 848–858 (2005).
19. Baum, C., Kustikova, O., Modlich, U., Li, Z. & Fehse, B. Mutagenesis and oncogenesis by chromosomal insertion of gene transfer vectors. *Hum. Gene Ther.* **17**, 253–263 (2006).

20. Nowrouzi, A., Glimm, H., Von Kalle, C. & Schmidt, M. Retroviral vectors: post entry events and genomic alterations. *Viruses* **3**, 429–455 (2011).
21. DeKelver, R. C. *et al.* Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. *Genome Res.* **20**, 1133–1142 (2010).
22. Sadelain, M., Papapetrou, E. P. & Bushman, F. D. Safe harbours for the integration of new DNA in the human genome. *Nat. Rev. Cancer* **12**, 51–58 (2012).
23. Sorrell, D. A. & Kolb, A. F. Targeted modification of mammalian genomes. *Biotechnol. Adv.* **23**, 431–469 (2005).
24. Silva, G. *et al.* Meganucleases and Other Tools for Targeted Genome Engineering: Perspectives and Challenges for Gene Therapy. *Curr Gene Ther* **11**, 11–27 (2011).
25. Schiffer, J. T. *et al.* Targeted DNA Mutagenesis for the Cure of Chronic Viral Infections. *J. Virol.* **86**, 8920–8936 (2012).
26. Rouet, P., Smih, F. & Jasin, M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6064–6068 (1994).
27. Petek, L. M., Russell, D. W. & Miller, D. G. Frequent endonuclease cleavage at off-target locations in vivo. *Mol. Ther.* **18**, 983–986 (2010).
28. Pattanayak, V., Ramirez, C. L., Joung, J. K. & Liu, D. R. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat. Methods* **8**, 765–770 (2011).
29. Gabriel, R. *et al.* An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nature Biotechnology* **29**, 816–823 (2011).
30. Lee, H. J., Kweon, J., Kim, E., Kim, S. & Kim, J.-S. Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. *Genome Res.* **22**, 539–548 (2012).
31. Lee, H. J., Kim, E. & Kim, J.-S. Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res.* **20**, 81–89 (2010).
32. Brunet, E. *et al.* Chromosomal translocations induced at specified loci in human stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 10620–10625 (2009).
33. Cornu, T. I. *et al.* DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Mol. Ther.* **16**, 352–358 (2008).
34. Pierce, A. J. *et al.* Double-strand breaks and tumorigenesis. *Trends in Cell Biology* **11**, S52–S59 (2001).
35. Cevher, E., Demir, A. & Sefik, E.in *Recent Advances in Novel Drug Carrier Systems* (Sezer, A. D.) (InTech, 2012).at <http://www.intechopen.com/books/recent-advances-in-novel-drug-carrier-systems/gene-delivery-systems-recent-progress-in-viral-and-non-viral-therapy>
36. Tavernier, G. *et al.* mRNA as gene therapeutic: How to control protein expression. *Journal of Controlled Release* **150**, 238–247 (2011).
37. O'Connor, T. P. & Crystal, R. G. Genetic medicines: treatment strategies for hereditary disorders. *Nat Rev Genet* **7**, 261–276 (2006).
38. Wu, Z., Yang, H. & Colosi, P. Effect of Genome Size on AAV Vector Packaging. *Mol Ther* **18**, 80–86 (2009).
39. Rhode, B. W., Emerman, M. & Temin, H. M. Instability of large direct repeats in retrovirus vectors. *J. Virol.* **61**, 925–927 (1987).

40. Pathak, V. K. & Temin, H. M. Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: deletions and deletions with insertions. *PNAS* **87**, 6024–6028 (1990).
41. Steinwaerder, D. S., Carlson, C. A. & Lieber, A. Generation of Adenovirus Vectors Devoid of All Viral Genes by Recombination between Inverted Repeats. *J. Virol.* **73**, 9303–9313 (1999).
42. Boeke, A. *et al.* Vector Production in an Academic Environment: A Tool to Assess Production Costs. *Human Gene Therapy Methods* **24**, 49–57 (2013).
43. Meyer, F. & Finer, M. Gene therapy: progress and challenges. *Cell. Mol. Biol. (Noisy-le-grand)* **47**, 1277–1294 (2001).
44. Lvovs, D., Favorova, O. O. & Favorov, A. V. A Polygenic Approach to the Study ?of Polygenic Diseases. *Acta Naturae* **4**, 59–71 (2012).
45. Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *PNAS* **93**, 1156–1160 (1996).
46. Klug, A. The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation. *Annual Review of Biochemistry* **79**, 213–231 (2010).
47. Isalan, M., Choo, Y. & Klug, A. Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *PNAS* **94**, 5617–5621 (1997).
48. Moore, M., Klug, A. & Choo, Y. Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1437–1441 (2001).
49. Maeder, M. L. *et al.* Rapid 'Open-Source' Engineering of Customized Zinc-Finger Nucleases for Highly Efficient Gene Modification. *Molecular Cell* **31**, 294–301 (2008).
50. Sander, J. D. *et al.* Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nature Methods* **8**, 67–69 (2011).
51. Ramirez, C. L. *et al.* Unexpected failure rates for modular assembly of engineered zinc fingers. *Nature Methods* **5**, 374–375 (2008).
52. Isalan, M. Zinc-finger nucleases: how to play two good hands. *Nature Methods* **9**, 32–34 (2012).
53. Perez, E. E. *et al.* Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat. Biotechnol.* **26**, 808–816 (2008).
54. Radecke, S., Radecke, F., Cathomen, T. & Schwarz, K. Zinc-finger Nuclease-induced Gene Repair With Oligodeoxynucleotides: Wanted and Unwanted Target Locus Modifications. *Mol Ther* **18**, 743–753 (2010).
55. Miller, J. C. *et al.* An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.* **25**, 778–785 (2007).
56. Szczepek, M. *et al.* Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat. Biotechnol.* **25**, 786–793 (2007).
57. Söllü, C. *et al.* Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. *Nucleic Acids Res.* **38**, 8269–8276 (2010).
58. Stoddard, B. L. Homing endonuclease structure and function. *Q. Rev. Biophys* **38**, 49–95 (2005).
59. Paques, F. & Duchateau, P. Meganucleases and DNA Double-Strand Break-Induced Recombination: Perspectives for Gene Therapy. *Current Gene Therapy* **7**, 49–66 (2007).
60. Stoddard, B. L. Homing Endonucleases: From Microbial Genetic Invaders to Reagents for Targeted DNA Modification. *Structure* **19**, 7–15 (2011).

61. Doyon, J. B., Pattanayak, V., Meyer, C. B. & Liu, D. R. Directed Evolution and Substrate Specificity Profile of Homing Endonuclease I-SceI. *J. Am. Chem. Soc.* **128**, 2477–2484 (2006).
62. Ashworth, J. *et al.* Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature* **441**, 656–9 (2006).
63. Jarjour, J. *et al.* High-resolution profiling of homing endonuclease binding and catalytic specificity using yeast surface display. *Nucl. Acids Res.* **37**, 6871–6880 (2009).
64. Ashworth, J. *et al.* Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. *Nucl. Acids Res.* **38**, 5601–5608 (2010).
65. Takeuchi, R. *et al.* Tapping natural reservoirs of homing endonucleases for targeted gene modification. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 13077–13082 (2011).
66. Jacoby, K. *et al.* Expanding LAGLIDADG endonuclease scaffold diversity by rapidly surveying evolutionary sequence space. *Nucleic Acids Res.* **40**, 4954–4964 (2012).
67. Szeto, M. D., Boissel, S. J. S., Baker, D. & Thyme, S. B. Mining Endonuclease Cleavage Determinants in Genomic Sequence Data. *J. Biol. Chem.* **286**, 32617–32627 (2011).
68. Pruett-Miller, S. M., Connelly, J. P., Maeder, M. L., Joung, J. K. & Porteus, M. H. Comparison of Zinc Finger Nucleases for Use in Gene Targeting in Mammalian Cells. *Mol Ther* **16**, 707–717 (2008).
69. Pruett-Miller, S. M., Reading, D. W., Porter, S. N. & Porteus, M. H. Attenuation of Zinc Finger Nuclease Toxicity by Small-Molecule Regulation of Protein Levels. *PLoS Genet* **5**, e1000376 (2009).
70. Mussolino, C. *et al.* A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res* **39**, 9283–9293 (2011).
71. Thyme, S. B. *et al.* Exploitation of binding energy for catalysis and design. *Nature* **461**, 1300–1304 (2009).
72. Li, T. *et al.* TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucl. Acids Res.* **39**, 359–372 (2011).
73. Christian, M. *et al.* Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 757–761 (2010).
74. Cook, A. A. & Guevara, Y. G. Hypersensitivity in Capsicum chacoense to Race 1 of the Bacterial Spot Pathogen of Pepper. *Plant Disease* **68**, 329 (1984).
75. Kim, B. S. & Hartman, R. R. Inheritance of a Gene (Bs3) Conferring Hypersensitive Resistance to Xanthomonas campestris pv. vesicatoria in Pepper (Capsicum annuum). *Plant Disease* **68**, 233 (1984).
76. Herbers, K., Conrads-Strauch, J. & Bonas, U. Race-specificity of plant resistance to bacterial spot disease determined by repetitive motifs in a bacterial avirulence protein. *Nature* **356**, 172–174 (1992).
77. Moscou, M. J. & Bogdanove, A. J. A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science* **326**, 1501–1501 (2009).
78. Boch, J. *et al.* Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* **326**, 1509–1512 (2009).
79. Mak, A. N.-S., Bradley, P., Cernadas, R. A., Bogdanove, A. J. & Stoddard, B. L. The Crystal Structure of TAL Effector PthXo1 Bound to Its DNA Target. *Science* **335**, 716–719 (2012).
80. Deng, D. *et al.* Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* **335**, 720–723 (2012).

81. Reyon, D. *et al.* FLASH assembly of TALENs for high-throughput genome editing. *Nature Biotechnology* **30**, 460–465 (2012).
82. Zhang, F. *et al.* Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nature Biotechnology* **29**, 149–153 (2011).
83. Chen, S. *et al.* A large-scale in vivo analysis reveals that TALENs are significantly more mutagenic than ZFNs generated using context-dependent assembly. *Nucl. Acids Res.* (2013).doi:10.1093/nar/gks1356
84. Move over ZFNs. *Nat. Biotechnol.* **29**, 681–684 (2011).
85. Pennisi, E. The Tale of the TALEs. *Science* **338**, 1408–1411 (2012).
86. Holkers, M. *et al.* Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucl. Acids Res.* (2012).doi:10.1093/nar/gks1446
87. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167–170 (2010).
88. Deveau, H., Garneau, J. E. & Moineau, S. CRISPR/Cas system and its role in phage-bacteria interactions. *Annu. Rev. Microbiol.* **64**, 475–493 (2010).
89. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).
90. Hwang, W. Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology* (2013).doi:10.1038/nbt.2501
91. Cho, S. W., Kim, S., Kim, J. M. & Kim, J.-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature Biotechnology* (2013).doi:10.1038/nbt.2507
92. Ding, Q. *et al.* Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell Stem Cell* **12**, 393–394 (2013).
93. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* (2013).doi:10.1126/science.1232033
94. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* (2013).doi:10.1126/science.1231143
95. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).
96. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* (2013).doi:10.1038/nbt.2623
97. Cradick, T. J., Fine, E. J., Antico, C. J. & Bao, G. CRISPR/Cas9 systems targeting β-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res.* (2013).doi:10.1093/nar/gkt714
98. Eisenschmidt, K. *et al.* Developing a programmed restriction endonuclease for highly specific DNA cleavage. *Nucleic Acids Res* **33**, 7039–47 (2005).
99. Fonfara, I., Curth, U., Pingoud, A. & Wende, W. Creating highly specific nucleases by fusion of active restriction endonucleases and catalytically inactive homing endonucleases. *Nucleic Acids Res.* **40**, 847–860 (2012).
100. Schierling, B. *et al.* A novel zinc-finger nuclease platform with a sequence-specific cleavage module. *Nucleic Acids Res.* **40**, 2623–2638 (2012).
101. Kleinstiver, B. P. *et al.* Monomeric site-specific nucleases for genome editing. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 8061–8066 (2012).

102. Kleinstiver, B. P., Wolfs, J. M. & Edgell, D. R. The monomeric GIY-YIG homing endonuclease I-BmoI uses a molecular anchor and a flexible tether to sequentially nick DNA. *Nucleic Acids Res.* **41**, 5413–5427 (2013).
103. Beurdeley, M. *et al.* Compact designer TALENs for efficient genome engineering. *Nat Commun* **4**, 1762 (2013).
104. Certo, M. T. *et al.* Tracking genome engineering outcome at individual DNA breakpoints. *Nature Methods* **8**, 671–676 (2011).
105. Gao, H., Wu, X., Chai, J. & Han, Z. Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell Research* **22**, 1716–1720 (2012).
106. Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nature Biotechnology* **29**, 143–148 (2011).
107. Takeuchi, R., Certo, M., Caprara, M. G., Scharenberg, A. M. & Stoddard, B. L. Optimization of in vivo activity of a bifunctional homing endonuclease and maturase reverses evolutionary degradation. *Nucl. Acids Res.* **37**, 877–890 (2009).
108. Cermak, T. *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucl. Acids Res.* **39**, e82–e82 (2011).
109. Kent, W. J. BLAT—The BLAST-Like Alignment Tool. *Genome Res.* **12**, 656–664 (2002).
110. Römer, P. *et al.* Promoter elements of rice susceptibility genes are bound and activated by specific TAL effectors from the bacterial blight pathogen, Xanthomonas oryzae pv . oryzae. *New Phytologist* **187**, 1048–1057 (2010).
111. Scholze, H. & Boch, J. TAL effector-DNA specificity. *Virulence* **1**, 428–432 (2010).
112. Meckler, J. F. *et al.* Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Res.* **41**, 4118–4128 (2013).
113. Streubel, J., Blücher, C., Landgraf, A. & Boch, J. TAL effector RVD specificities and efficiencies. *Nat. Biotechnol.* **30**, 593–595 (2012).
114. Lundstrom, K. Latest development in viral vectors for gene therapy. *Trends in Biotechnology* **21**, 117–122 (2003).
115. Sather, B. D. *et al.* Development of B-lineage Predominant Lentiviral Vectors for Use in Genetic Therapies for B Cell Disorders. *Molecular Therapy* **19**, 515–525 (2011).
116. Torikai, H. *et al.* A foundation for universal T-cell based immunotherapy: T cells engineered to express a CD19-specific chimeric-antigen-receptor and eliminate expression of endogenous TCR. *Blood* **119**, 5697–5705 (2012).
117. Sethuraman, J., Majer, A., Friedrich, N. C., Edgell, D. R. & Hausner, G. Genes within genes: multiple LAGLIDADG homing endonucleases target the ribosomal protein S3 gene encoded within an rnl group I intron of Ophiostoma and related taxa. *Mol. Biol. Evol.* **26**, 2299–2315 (2009).
118. Certo, M. T. *et al.* Coupling endonucleases with DNA end-processing enzymes to drive gene disruption. *Nature Methods* **9**, 973–975 (2012).
119. Delacôte, F. *et al.* High Frequency Targeted Mutagenesis Using Engineered Endonucleases and DNA-End Processing Enzymes. *PLoS ONE* **8**, e53217 (2013).