Strategies for human genome modification using engineered nucleases and transcription factors

Thesis submitted for the degree of Doctor of Philosophy

By

Elaine Margaret Gourlay

Institute of Cancer Sciences

College of Medical, Veterinary and Life Sciences

University of Glasgow

December 2014

Abstract

VEZF1 is a highly conserved vertebrate transcription factor that is essential for mammalian development. The gene regulatory functions of VEZF1 are largely undetermined. The generation of human cells depleted or absent of VEZF1 would greatly assist the study of VEZF1 functions and mechanism of action. This study makes use of synthetic biology technologies to either repress or knock out *VEZF1* gene transcription to enable further studies of VEZF1 function.

This study explores various strategies to use engineered DNA-binding proteins to direct the repression or mutation of a gene of interest. Zinc Finger (ZF) and Transcription Activator Like Effector (TALE) proteins that specifically recognise DNA sequences at the *VEZF1* gene promoter were constructed using modular or Golden Gate assembly methods. The ability of TALE fusion proteins to function in human cells was studied. An expression vector system was created to assemble TALE Repressor (TALER) fusion proteins. The use of TALERs allowed for the rapid assessment of TALE protein binding at their chromosomal targets in human cells. Transient expression of most of the assembled TALE repressor proteins resulted in reduced VEZF1 transcription. A subset resulted in very substantial VEZF1 repression, making them useful tools for the study of VEZF1 function.

Functional TALE domains were assembled into TALE nuclease (TALEN) fusion proteins. TALEN expression vectors were developed to assemble TALEN proteins with optimised expression, cleavage activity and target specificity. Transient expression of TALEN proteins in human cells was used to direct the cleavage and error-prone DNA repair of the *VEZF1* promoter. Following development of the assays used to detect TALEN-directed mutations, several functional TALEN pairs were identified. Some TALENs resulted in over 65% mutation rates, with some mutations removing the *VEZF1* promoter. These TALENs will be useful for the development of *VEZF1* knock out cell lines. Interestingly, our study reveals a correlation between TALE length and the activity of TALERs and TALENs that should be considered in the future application of TALE proteins.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to thank my Supervisors, Drs Adam & Katherine West, for their advice, support and guidance throughout my PhD. I am grateful to them both for their input, particularly with the design of strategies and their eagle-eyes when it comes to data analyses. I would also like to express my gratitude towards my Advisors, Dr Andrew Hamilton and Prof Marshall Stark for their insights into my project and their advice and encouragement. I thank the BBSRC who funded my studentship.

A word of thanks also goes to Julie Galbraith and Polyomics for analyses using the Agilent DNA Bioanalyser and to Drs Sarah Meek and Susan Jamieson for all their help and support, particularly in the initial stages of my PhD.

The entire West Lab Group, past and present, deserves huge thanks for all of the support they offer to their lab members – I couldn't have done this without the support of my colleagues, you've seen me through the highs and the lows. I have been so lucky to have had the opportunity to work with such wonderful people throughout my PhD – Dr Sabari Chilaka, Dr Narendra Kumar, Aliya Al-Hosni, Ileana Guerrini, Alejandro Rivera, Ohoud Rehibini and Abdul Sindi. Particular thanks go out to Dr Grainne Barkess, Bhoomi Gor, Ronan Mellin and Ileana Guerrini for their contributions to the project. An extra special mention has most definitely been deserved by Dr Grainne Barkess, who has helped me in more ways than she can ever know.

I am privileged to have found some of the best friends anyone could ever ask for – Susan Ridha, Shivam Mishra, Koorosh Korfi, Gokula Mohan, Grainne Barkess, Bhoomi Gor and Carolyn Low. Otto nights, karaoke parties, themed nights and of course Firbush – nothing could have forged a stronger bond between us! A special mention must go to The Bonnie Lassies, Bhoomi and Carolyn, who were always there for a coffee, a long lunch, or a family naan in Akbars!

My deepest thanks must of course go to Lee Brady. I can never express my appreciation for all of your advice, love and support over the years. You always know how to bring a smile to my face. You are my rock, my foundation. I could not have done this without you.

My final thoughts turn to my family. Natalie, you offered a wonderful element of escapism – weekends in Edinburgh with you were a saving grace. Gemma, your hairstyles kept me smiling, and certainly amused the lab, to say the least!

Saving the most important for last, my parents, Shirley and Colin, who have supported me in every aspect and sacrificed so much to give me what I needed. You both have shaped the person I am today. You will always be an inspiration; I dedicate this thesis to you.

“Yatta!”

Author's Declaration

The research reported within this thesis is my own work, except where otherwise stated, and has not been submitted for any other degree.

Elaine Gourlay

Chapter 1

Introduction

1.1 Genome modification

Reverse genetics is the analysis of a gene's function through the modification of its phenotype. With the advent of whole genome sequencing, many genes have been identified through sequence characteristics. While information about the size, sequence, location and the RNA transcript of these genes is known, it is impossible to predict the role of the encoded protein from this data. Phenotypic manipulations can be performed at defined genomic loci using a variety of procedures such as insertional mutagenesis, RNA interference (RNAi), targeted gene repression, mutation or gene knock-in/knockout (Hardy et al., 2010). Introduction of a targeted double stranded break (DSB) at a defined chromosomal location can stimulate repair mechanisms in the cell up to 50,000 fold (Jasin, 1996). This feature can be exploited to perform targeted gene modification by stimulating the cell's natural repair mechanisms at a user-defined genetic locus, resulting in a modification intended to disrupt the gene's capacity to produce its phenotype.

1.2 DNA Repair Processes

Nuclease-mediated genome editing involves inducing targeted double stranded breaks (DSBs) at predefined genomic locations. DNA DSBs are then repaired by either non-homologous end joining (NHEJ) or homology directed repair (HDR) (Figure 1.1). NHEJ is an error prone DNA repair process, which introduces small insertions or deletions in repaired DNA strands (Figure 1.1, left) (Lieber, 2010). HDR-mediated repair involves the use of donor DNA which has sequence homology to the region surrounding the induced DSB. This donor DNA template is then used to faithfully repair DNA DSBs, and can be utilised to promote the insertion of targeted gene mutations (Figure 1.1, middle and right) (Moehle et al., 2007). This thesis aims to promote gene targeting by inducing DSBs at defined locations.

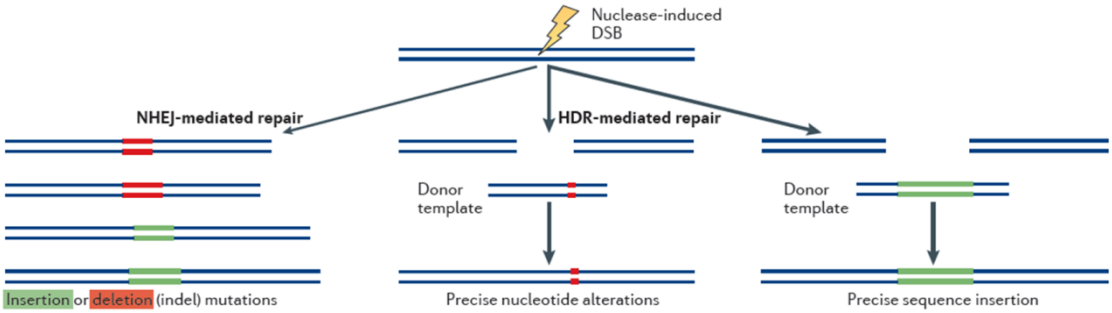

Figure 1.1 Double-strand break repair
DNA double-strand breaks can be repaired via two mechanisms: non-homologous end joining (NHEJ) or homology directed repair (HDR). NHEJ repair is erroneous and introduces mutations in the form of insertions or deletions (indels). HDR is a faithful method of DNA repair, which can be employed to introduce targeted DNA mutations by means of insertion of a donor DNA template (Joung and Sander, 2013).

1.3 Zinc Finger Proteins

Zinc finger proteins can be fused to nucleases, creating zinc finger nucleases (ZFNs). The DNA-binding properties of zinc fingers can be utilised to direct nucleases to induce targeted DSBs at defined locations. The zinc finger was first identified as being involved in protein-DNA interactions by Miller *et al.* (1985) in a study of TFIIIA, a 40 kDa transcription factor involved in the regulation of the 5S RNA genes of *Xenopus laevis*. TFIIIA was found to be composed of a repeating structure containing zinc ions (Miller et al., 1985). This region was responsible for DNA-binding, mediated through globular looped Cys_2-His_2 domains. These domains were described as grasping or gripping the DNA – a description leading to the name zinc "fingers". Since their identification, zinc fingers have been found to be among the most common DNA-binding proteins found in eukaryotes (Emerson and Thomas, 2009, Jacobs and Michaels, 1990, Bateman et al., 2002).

1.3.1 Structure

Each zinc finger is composed of ~30 amino acids coordinated around a zinc ion, forming an independent structural domain (Miller et al., 1985). Each zinc finger module is composed of an α-helix and two anti-parallel β-sheets (Parraga et al., 1988, Lee et al., 1989). In every module, a zinc ion is co-ordinated by two cysteines in the β-sheet and two histidines in the α-helix (Figure 1.2). The zinc finger provides a structurally stable architecture with the ability to harbour various peptide sequences that can interact with DNA (Wolfe et al., 2000). Each zinc finger module binds to its DNA recognition sequence through the α-helix, which fits neatly into the major groove of the DNA. Amino acid residues at positions -1, 3 & 6 (relative to the start of the α-helix) contact 3 bp of DNA

(Figure 1.3), and were initially identified as those responsible for DNA-binding specificity (Pavletich and Pabo, 1991). This was later expanded upon, as residue 2 was also shown to be involved in interactions with a DNA base on the other DNA strand (Fairall et al., 1993, Isalan et al., 1997). Having identified the particular residues involved in DNA-binding, and the overall modular structure of zinc fingers, the idea of generating designer DNA-binding proteins became an attractive and attainable prospect.

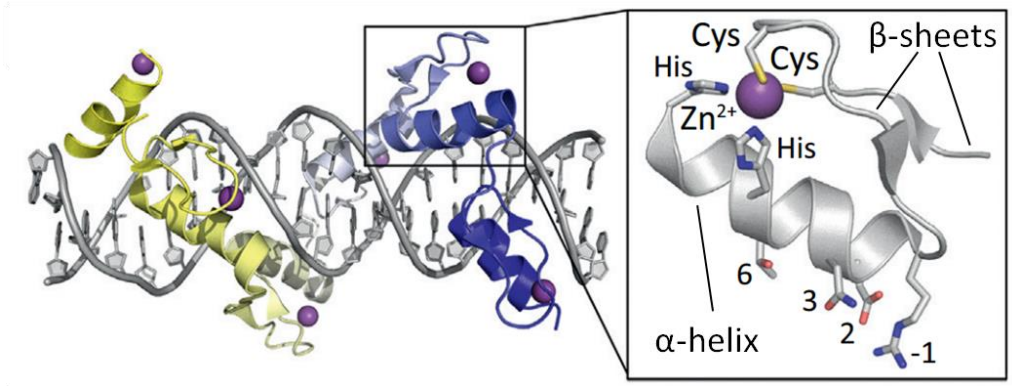

Figure 1.2 Zinc finger structure and DNA binding mechanism
Crystal structure of zinc finger proteins bound to DNA. Each zinc finger is composed of an α-helix and two antiparallel β-sheets and contains a zinc ion between two cysteine and two histidine residues. Each zinc finger contacts 4 bp DNA via residues -1, 2, 3 and 6 of the α-helix (Gaj et al., 2013).

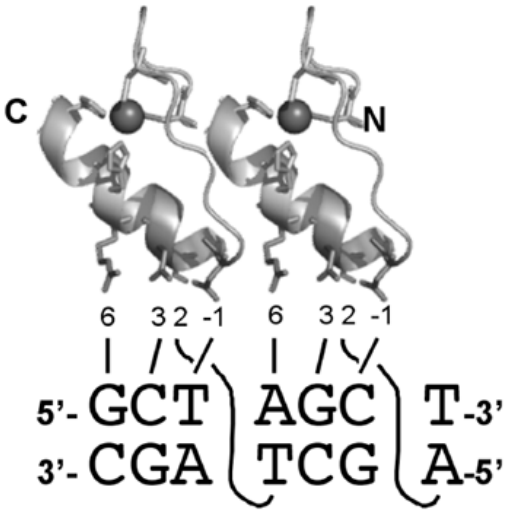

Figure 1.3 Cross-strand DNA contacts involved in zinc finger DNA-binding
Residues -1, 3 and 6 in the zinc finger α-helix recognise 3 bp DNA on one DNA strand. Residue 2 binds to a fourth base on the opposite DNA strand. The overall binding site of a zinc finger protein spans 4 bp and overlaps with the binding site of the preceding finger (Dawson, 2007).

1.3.2 Devising a zinc finger DNA recognition code

Desjarlais and Berg (1993) demonstrated that it was possible to alter the DNA binding characteristics of a zinc finger by changing the DNA-binding residues, and that zinc finger proteins preferentially bind to their designed target sequence, suggesting that it was

possible to assemble custom zinc fingers which would bind to desired DNA sequences. Since then, countless studies have attempted to identify a systematic code of amino acid-DNA binding within a zinc finger protein (reviewed in Papworth et al., 2006). It became apparent, however, that the intricacies involved in the DNA-binding capabilities of a multi-fingered protein were more complex than originally envisaged. A simple recognition code of zinc-finger monomer to DNA triplet, while applicable on its own, appeared to be subject to combinatorial issues when linked together side-by-side, demonstrated in Figure 1.3. The amino acid residue in position 2 makes cross-strand contact with the DNA base immediately adjacent to the three base pair recognition site originally identified (Isalan et al., 1997, Isalan et al., 1998, Cornu et al., 2008, Ramirez et al., 2008).

Several groups began to construct libraries of zinc finger modules which would function together when used as part of a zinc finger protein. The initial studies validated modules which would bind to GNN triplets; libraries of ANN, CNN and selected TNN triplets were developed later (Dreier et al., 2000, Dreier et al., 2001, Dreier et al., 2005, Segal et al., 1999). Zinc finger modules to target all of the 16 possible GNN subsites have been constructed from a variety of different natural sources and artificial development methods, and they have been well characterised in a number of contexts (discussed below).

1.3.3 Assembly of zinc finger proteins

Since the identification of a zinc finger-DNA recognition code and the creation of libraries of zinc finger modules to target predefined DNA triplets in a given position within a zinc finger protein, the development of a DIY assembly kit for the generation of designed zinc finger proteins was in the grasp of non-commercial research labs. A few basic assembly methods were available, but the first comprehensive kit was developed by the Zinc Finger Consortium (www.zincfingers.org).

1.3.3.1 Zinc finger modules

The zinc finger encoding “modules” were characterised by three different groups: the Barbas research group and the companies Sangamo and ToolGen, as detailed below.

1.3.3.1.1 Barbas Modules

The DNA-binding domain of the murine transcription factor Zif268, also known as Egr1, was used as a template to create novel zinc fingers of desired sequence specificity. Zif268 contains an array of three Cys_2-His_2 zinc fingers. The known DNA recognition residues in the α-helix of finger 2 of Zif268 were randomised to generate phage display libraries of zinc finger proteins of varied specificities (Segal et al., 1999, Dreier et al., 2000, Dreier et al., 2001, Dreier et al., 2005). This approach ensured that any novel fingers are capable of working within a zinc finger array. Phage display was used to identify novel zinc fingers that recognise all GNN, most ANN and CNN and some TNN triplet sequences. These zinc finger proteins were then characterised to determine their binding specificities in an ELISA assay. Where any protein demonstrated a high degree of cross reactivity with another DNA binding sequence, it was modified by site-directed mutagenesis in an attempt to generate zinc finger proteins with the highest on-target binding specificity.

The greatest improvements to DNA binding specificity in proteins modified by site-directed mutagenesis were observed when residues 1 and 5, which are not directly involved in DNA recognition, were exchanged (Segal et al., 1999). The combination of phage-display and mutagenesis offers an opportunity to overcome the complexities of DNA-sequence recognition that would not be highlighted when zinc finger modules are constructed using DNA recognition codes, which only consider those residues directly responsible for DNA binding.

The zinc finger modules identified by Barbas and co-workers represent a toolbox capable of constructing many new zinc finger proteins with designed DNA-binding specificities. However, the identification of such context-dependent effects demonstrates the difficulty of treating the DNA-binding specificities of zinc finger proteins in a modular sense. These Barbas modules offer a partially optimised library for use in the construction of designed zinc finger proteins.

1.3.3.1.2 Sangamo Modules

Sangamo Biosciences (Richmond, California) developed zinc finger modules to target most GNN targets using an in-house developed recognition code and systematic testing of the DNA binding abilities of zinc finger modules in electrophoretic mobility shift assays

(EMSAs) (Liu et al., 2002). Every ZF module's ability to bind its predicted DNA target was assessed at each position (F1, F2 or F3) in the zinc finger protein. This method of testing ZF-DNA binding resulted in the first evidence of position effects within a zinc finger array. A particular DNA target would be efficiently bound by a zinc finger module in one position of a ZFP but not in other positions. For example, a zinc finger with the amino acid sequence QSSNLAR effectively bound the triplet GAT when in the F1 position of a ZFP but recognised GAA when in positions F2 or F3. Instead the module TSGNLVR was used when binding to the triplet GAT in positions F2 or F3 (Liu et al., 2002). Sangamo modules offer a well characterised library for use in the construction of designed zinc finger proteins.

1.3.3.1.3 ToolGen Modules

ToolGen, a bioscience company based in Seoul, Korea, took the approach to characterise the properties of naturally occurring human zinc finger domains (Bae et al., 2003). Zinc finger 3 of Zif268 was exchanged with isolated human zinc fingers and the resulting arrays were tested for binding to all 64 possible triplets *in vitro*. The performance and specificity of these novel finger combinations was confirmed using a gene reporter assay in transfected cells. The results indicated that a zinc finger proteins composed of naturally occurring fingers were more likely to be functional than their synthetic counterparts. Proteins composed of both naturally occurring and engineered zinc fingers performed moderately, but not as well as proteins composed exclusively of human zinc fingers (Bae et al., 2003). ToolGen modules offer a well characterised library for use in the construction of designed zinc finger proteins, especially when the modules are placed at the end of an array.

1.3.3.1.4 Recommended Module Sets

Despite the efforts of several groups to identify and characterise zinc fingers that recognise known DNA triplets, the performance of zinc fingers in the context of zinc finger nucleases can be unpredictable. One large study of 315 ZFN pairs created by modular assembly found that 44% demonstrated cleavage activity *in vitro* but only 7.3% were functional nucleases at their intended targets in cultured HEK293 cells (Kim et al., 2009). The study further reinforced the concept that modular assembly could produce highly active zinc finger nucleases, provided that the module sets were chosen from validated sets. This study was used to identify a reliable subset of 37 zinc fingers, comprised mostly

of ToolGen modules, (Table 1.1) that could be used in different arrays without risk of incompatibility. These recommended modules should be preferred when designing zinc finger nucleases.

ZF module number	Source	DNA recognition residues	DNA recognition sequence
ZF107	ToolGen	CSNR	GAC
N/A	ToolGen	DSAR	GTC
ZF108	**ToolGen**	**DSCR**	**GCC**
N/A	ToolGen	DSNR	GAC
ZF112	ToolGen	ISNR	GAT
ZF113	ToolGen	KSNR	GAG
N/A	ToolGen	QNTQ	ATA
ZF117	ToolGen	QSHR	GGA
ZF120	ToolGen	QSHV	YGA
ZF121	ToolGen	QSNI	AAA
ZF123	ToolGen	QSNR	GAA
ZF125	ToolGen	QSNV	CAA
ZF126	ToolGen	QSSR	GYA
ZF128	ToolGen	QTHQ	AGA
ZF130	ToolGen	RDER1	GYG
N/A	ToolGen	RDER2	GYG
ZF131	ToolGen	RDHR	GGG
ZF132	**ToolGen**	**RDHT**	**YGG**
ZF133	ToolGen	RDKR	AGG
ZF134	ToolGen	RSHR	GGG
N/A	ToolGen	VDYK	TAT
ZF137	ToolGen	VSNV	AAT
ZF139	**ToolGen**	**VSTR**	**GCT**
ZF140	ToolGen	WSNR	GGT
ZF52	Sangamo	RDNQ	AAG
ZF61	Barbas	DGHR	GGC
ZF75	Barbas	DGNV	AAC
ZF78	**Barbas**	**SADR**	**ACA**
ZF81	Barbas	TLDR	ACT
ZF90	Barbas	SKAE	CAC
ZF91	**Barbas**	**RDNE**	**CAG**
ZF93	Barbas	THSE	CCA
ZF95	Barbas	RDTE	CCG
ZF96	Barbas	TNSE	CCT
ZF98	**Barbas**	**HGHE**	**CGC**
ZF100	Barbas	SRTA	CGT
ZF104	Barbas	RDNT	TAG

Table 1.1 Recommended zinc finger modules (Kim et al., 2009)
Zinc finger module numbers refer to numbering scheme of Zinc Finger Consortium Modular Assembly Kit. Modules indicated in bold were chosen to be used in this study (Chapter 3).

1.3.3.2 Modular assembly

The Zinc Finger Consortium developed a standardised kit and protocol for the straightforward assembly of zinc finger arrays (Wright et al., 2006). The pre-characterised zinc fingers, from the Sangamo, ToolGen and Barbas module sets, were distributed as a set of 141 plasmids which each encoded a single zinc finger module. The zinc finger modules were amenable to being cloned together sequentially via standard restriction enzyme based cloning techniques to create a multi-fingered array of any desired length (Figure 1.4).

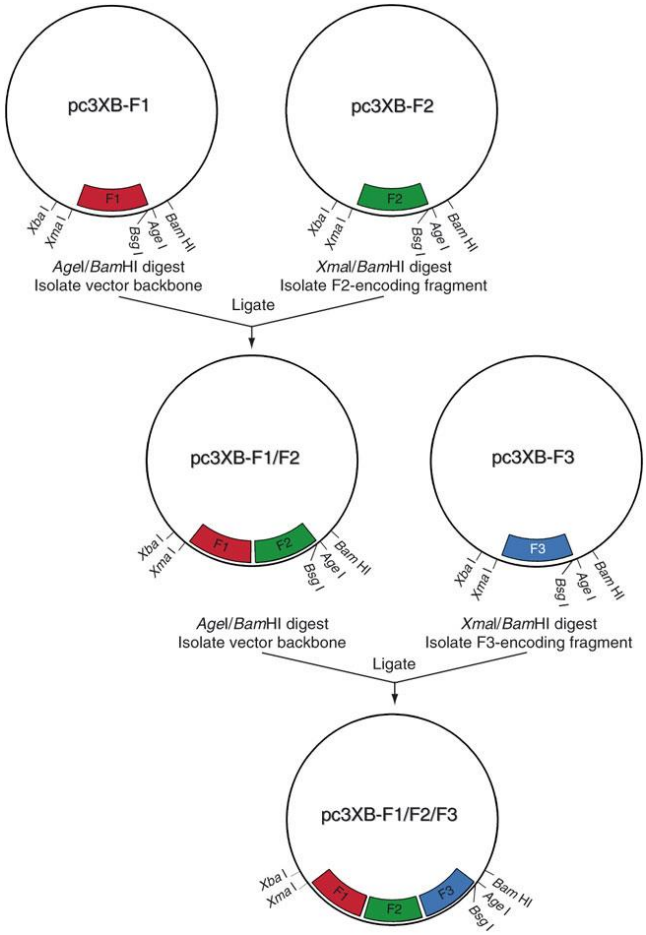

Figure 1.4 Overview of Modular Assembly Cloning Procedure
Plasmids encoding individual zinc finger expression sequences are digested and ligated together to create the desired expression sequence for a zinc finger array (Wright et al., 2006).

Modular assembly is a straightforward and simple method to assemble custom ZFN expression vectors. However, success rates of modularly assembled ZFNs are reportedly low. ZF proteins shown to bind the intended DNA target in a bacterial two-hybrid assay have been reported to have up to a 76% failure rate when used *in vivo* (Ramirez et al., 2008, Kim et al., 2009). This high failure rate is likely to be due to the lack of consideration for context dependent effects faced when linking multiple zinc finger modules side by side in an array (Elrod-Erickson et al., 1998, Wolfe et al., 1999, Cornu et al., 2008).

Selecting ZF modules from a recommended module set (Kim et al., 2009) can more reliably generate functional ZFNs. Context dependent behaviour of zinc finger has been taken into account in further assembly methods, OPEN and CoDA (described below) and these methods offer a more stringent selection and testing processes when compared to modular assembly.

1.3.3.3 OPEN

Oligomerised pool engineering (OPEN) was developed by the Zinc Finger Consortium (Maeder et al., 2008, Maeder et al., 2009) following the reported low efficacy of ZFNs made by modular assembly (Cornu et al., 2008, Pruett-Miller et al., 2008, Ramirez et al., 2008). OPEN involves greater stringency in the selection of zinc finger proteins, resulting in the production of zinc fingers with greater activity and lower toxicity than those assembled via modular assembly methods. OPEN employs a series of pools, each containing zinc finger modules that target a 3 bp subsite in a given position of a zinc finger protein (F1, F2 or 3). Zinc finger expression sequences are PCR amplified from a pool of fingers which bind the desired 3 bp target and PCR products are stitched together randomly, creating zinc finger arrays to target the desired DNA target sequence (Figure 1.5 A). The arrays are then subjected to two rounds of selection in bacterial reporter assays, permitting selection of active arrays and quantifying their DNA binding affinities, governing the identification of highly active zinc finger proteins (Figure 1.5 B).

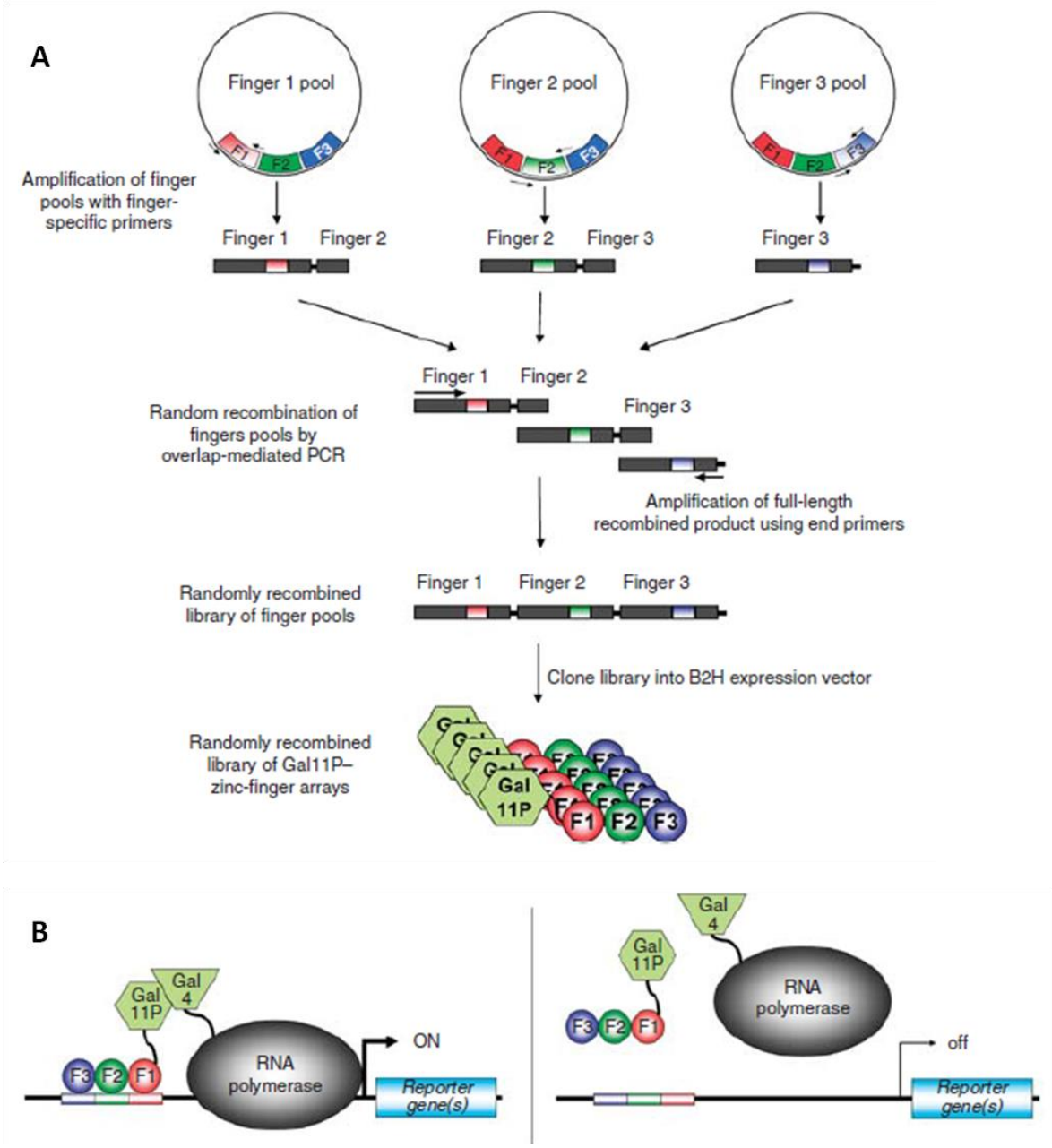

Figure 1.5 Overview of OPEN method for construction of zinc finger proteins
(A) Schematic representation of assembly of randomised zinc finger pools by overlapping PCR. Pools of zinc fingers to target a given 3 bp subsite in a predetermined position of a zinc finger array are amplified via PCR and each PCR product is "stitched" together to create an array. The arrays are expressed as a fusion to Gal11P to be tested in a bacterial two-hybrid system. (B) Bacterial two-hybrid selection of assembled zinc finger arrays. Zinc finger binding activates a reporter gene which can be used to select and quantify the efficacy of zinc finger arrays which bind to their target site (Maeder et al., 2009).

ZFNs assembled by OPEN have been reported to have equal or greater *in vitro* efficacy than ZFNs constructed by modular assembly (Townsend et al., 2009, Lee et al., 2010). However, the overall likelihood of constructing ZFNs that will actually function at the intended target site *in vivo* is reportedly ~67%, which is much greater than would be achieved with modular assembly (Maeder et al., 2008). OPEN, while providing a reliable and effective method for assembling ZFNs, is highly labour intensive, is not amenable to

the production of multiple ZFNs in parallel, and interpretation of selection results can require considerable expertise (Kim et al., 2010, Sander et al., 2011b). I attempted to create bacterial two-hybrid selection vectors to test the binding affinities of ZF arrays assembled in this study. However, the cloning procedure was found to be highly inefficient, further adding to the labour intensive, low throughput nature of this procedure. This, and the subsequent development of a higher-throughput assembly method, has restricted the broad application of OPEN.

1.3.3.4 CoDA

Expanding on the current zinc finger assembly methods available, the Zinc Finger Consortium developed CoDA (Context Dependent Assembly). CoDA addressed combinatorial issues associated with combining zinc finger modules into an array, rather than treating zinc fingers as individual modules (Sander et al., 2011b). CoDA makes use of libraries of three fingered arrays in which the N- and C-terminal modules have been previously characterised in an array with a common F2 module (Figure 1.6).

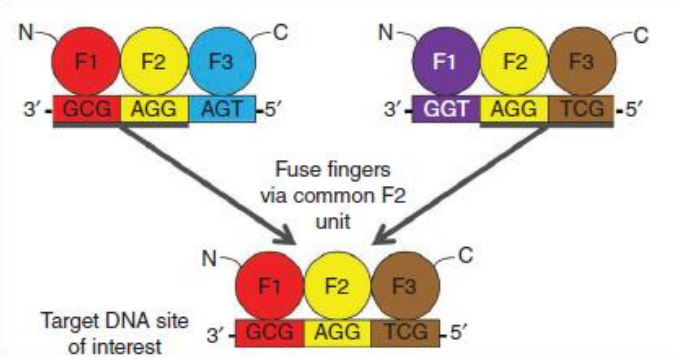

Figure 1.6 Principle of CoDA
Two separate zinc finger arrays, each binding different 9 bp target sites are shown, which each share a common F2 module. A new zinc finger array to bind a different DNA sequence can be assembled by joining together the F1 module from one array and the F3 module from the other, with the common F2 module (Sander et al., 2011b).

Half of ZFNs assembled by CoDA have been demonstrated to be functional at their intended target (Sander et al., 2011b). CoDA has provided a rapid and non-labour intensive procedure to assemble ZFNs whilst considering context dependent issues.

1.3.4 Zinc finger nucleases

Designing and constructing zinc finger arrays which bind to predefined sequences is with the ultimate aim of directing an effector domain, such as a nuclease, to the target sequence. The induction of targeted DSBs via custom nucleases can be utilised to create

mutations in a gene of interest. Kim and Chandrasegaran (1994) demonstrated the first fusion of a DNA-binding protein to the non-specific cleavage domain of *Fok*I. The drosophila Ubx homeodomain, which recognises a 9 bp DNA sequence through a helix-turn-helix motif, was fused to *Fok*I to create a chimeric restriction endonuclease with designed sequence specificity. This fuelled the construction of zinc finger nucleases (Kim et al., 1996). Expression sequences of previously characterised zinc finger proteins were cloned into plasmid vectors immediately upstream of the cleavage domain of *Fok*I nuclease, separated by a glycine linker. These zinc finger nucleases were demonstrated to cleave λ-DNA with sequence specificity (Kim et al., 1996). ZFNs were shown initially to promote homologous recombination of extrachromosomal DNA in *Xenopus* oocyte nuclei. Pairs of ZFN recognition sites, located on opposite DNA strands, and separated by a 6-8 bp spacer (Figure 1.7) were found to promote optimal cleavage of DNA substrate (Bibikova et al., 2001). This study was extended to demonstrate ZFN-mediated mutations in chromosomal DNA of *Drosophila* and offered the first evidence of permanent genetic alterations induced by ZFNs (Bibikova et al., 2002). Numerous studies then commenced to develop and assess mutations induced by DNA cleavage with customisable zinc finger nucleases (ZFNs).

ZFNs were first demonstrated to induce gene targeting in human cells via NHEJ-mediated repair of a stably integrated GFP transgene in HEK293 cells (Porteus and Baltimore, 2003). ZFNs have now been utilised to target a range of genes in various cell types such as mammalian cells, including ES cells, and in many organisms for example *Drosophila*, sea urchins, zebrafish, mice and pigs (reviewed in Hauschild-Quintern et al., 2013). Targeting efficiencies with ZFNs are variable due to differences in ZFN functionality and target site accessibility, however ZFNs generally stimulate gene targeting at a rate where modifications occur in around 10 % of transfected cells (Carroll, 2011).

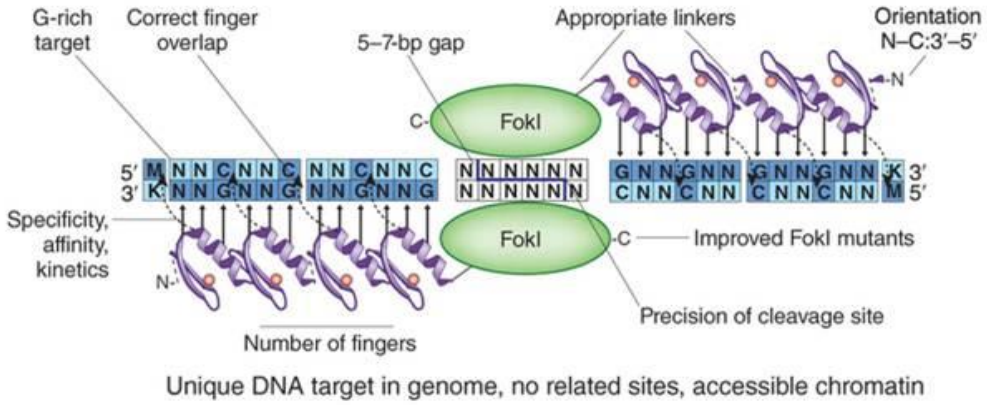

Figure 1.7 ZFNs DNA binding mechanism
Schematic of a pair of four-fingered ZFNs binding to DNA. Each zinc finger module contacts 4 bp DNA and module target sites overlap by 1 bp. Zinc finger arrays are fused to the non-specific DNA cleavage domain of *Fok*I nuclease. ZFNs in a pair are separated by a spacer of 5 – 7 bp, where *Fok*I dimerisation and DNA cleavage occurs, resulting in a 4 bp 5'overhang (Isalan, 2012).

1.3.4.1 Optimisations to ZFN architecture

1.3.4.1.1 Linker choice and spacing between ZFN pairs

Various studies have investigated the optimal conditions for ZFN activity. In agreement with studies of the natural *Fok*I enzyme, it was revealed that *Fok*I domains in ZFNs required dimerisation in order to perform DNA cleavage (Bitinaite et al., 1998, Smith et al., 2000). Cleavage was observed where the binding sites of the ZFNs were positioned in an inverted orientation from each other, as shown in Figure 1.7 (Smith et al., 2000, Bibikova et al., 2001). The optimal length of the DNA spacer between a zinc finger pair has also been subject to much debate. The current consensus is a 6 bp spacer using the common architectures employed by the Zinc Finger Consortium (Wright et al., 2006, Maeder et al., 2009, Sander et al., 2011b). However, this seems to be subject to the length and composition of the amino acid linker between ZF DNA binding domain and *Fok*I nuclease. Studies suggest that the commonly used four amino acid linker promotes cleavage at both 5- and 6 bp spacers, although spacers of 6 bp are favoured (Handel et al., 2009, Wilson et al., 2013). Another study suggests that a six amino acid linker is sufficient to restrict DNA cleavage to sites separated by a 6 bp spacer (Shimizu et al., 2009).

1.3.4.1.2 *Fok*I variants

Improvements to promote specific ZFN cleavage at the desired target site have focussed around modification of the *Fok*I domain and its dimerisation interface. The development

of obligate heterodimer mutants of *Fok*I has improved the activity and faithfulness of ZFNs (Figure 1.8). Dimerisation of the two domains in WT *Fok*I is not selective for heterologous zinc finger DNA binding domains. The result is that dimerisation can occur between two "left" or two "right" ZFNs at off-target sites containing relevant binding sites, or perhaps by weak interactions between one DNA-bound and one unbound ZFN (Smith et al., 2000, Miller et al., 2007). Introduction of obligate heterodimer mutations in the *Fok*I domains mean that a "left" ZFN is only able to interact with the "right" ZFN when the relevant cognate mutations are found in the "right" domain, and that it is not possible for two "left" or two "right" domains to facilitate DNA cleavage (Figure 1.8). Obligate heterodimer ZFNs were demonstrated to produce lower levels of cellular cytotoxicity, as measured by γH2AX staining (Doyon et al., 2011). ZFNs with the obligate heterodimer architecture did not promote cleavage at off target locations which had been shown to be cleaved by homodimer versions of the same ZFNs (Doyon et al., 2011). A summary of the most commonly adopted *Fok*I obligate heterodimer mutants developed to date is given in Table 1.2.

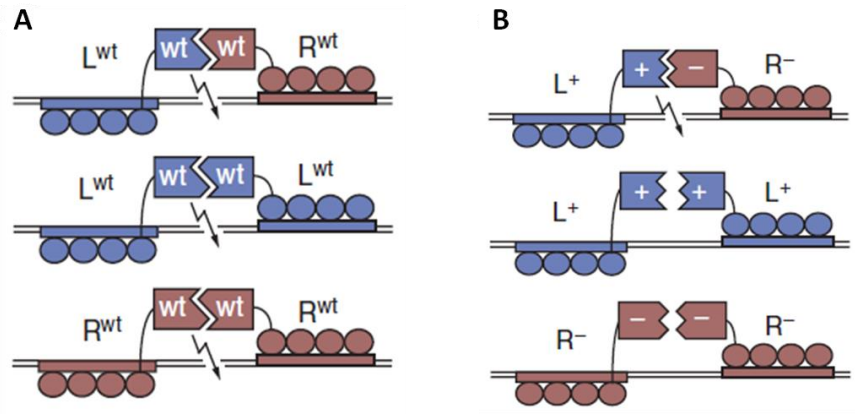

Figure 1.8 *Fok*I obligate heterodimers offer improved specificity
(A)Top: Intended on-target activity of a ZFN pair involves interaction of a left (L) and a right (R) ZFN across a spacer where *Fok*I nuclease dimerises and cleaves DNA. Middle and bottom: off-target DNA cleavage mediated by the dimerisation of either two left (L) or two right (R) ZFNs where homodimer WT *Fok*I domains are used. (B) Obligate heterodimer FokI variants promote the dimerisation of heterologous ZFN pairs (top) and exclude the dimerisation of homodimer species (middle and bottom)(Miller et al., 2007).

Variant	Mutations (Left/Right)	References
EL/KK	Q486E, I499L / E490K, I538K	(Miller et al., 2007)
D/R	R478D / D483R	(Szczepek et al., 2007)
ELD/KKR	Q486E, I499L, N496D / E490K, I538K, H537R	(Doyon et al., 2011)
DD/RR	R478D, N496D / D483R, H537R	(Doyon et al., 2011)

Table 1.2 Obligate heterodimer *Fok*I mutants

Despite the obvious advantage of obligate heterodimer *Fok*I to reduce off-target cleavage events and the associated cytotoxicity, this came at the cost of a threefold reduction in cleavage activity of obligate heterodimers when compared to WT *Fok*I (Miller et al., 2007). This loss of activity observed was restored by the development of a variant with enhanced cleavage activity, which showed no increase in cytotoxicity (Guo et al., 2010). This variant, dubbed "Sharkey", contained the mutations S418P and K441E on both sides of the *Fok*I dimer and has been successfully employed in combination with obligate heterodimer variants (Doyon et al., 2011, Perez-Pinera et al., 2012).

1.3.4.1.3 Number of zinc fingers

Adding additional zinc fingers to a ZFN increases the length of the DNA binding domain. Designing ZFNs that target longer DNA sequences was postulated to increase on-target specificity by reducing the likelihood that ZFN binding sites would occur elsewhere in the genome by chance. A study tested the DNA cleavage abilities of ZFNs composed of variable numbers of zinc fingers ranging from 1 zinc finger module to 6 zinc finger modules per array (Shimizu et al., 2011). This study systematically tested ZFNs targeting the same genomic locus but with additional ZF modules per array and utilised the same *Fok*I variant in all ZFNs tested. The results demonstrated that the use of ZFNs with 5 or 6 modules per array was in fact detrimental to ZFN activity. ZFNs containing one or two modules were not active but ZFNs with 3 or 4 modules per array demonstrated the greatest cleavage activity. The reduced cleavage with longer arrays may be due to the use of an inappropriate linker to promote optimal ZF-DNA contact or perhaps due to confounding combinatorial effects of zinc finger modules. Many studies have demonstrated the effective use of 3- and 4-fingered ZFNs, and reports on the activity levels of 3- versus 4-fingered ZFNs have been inconclusive (Urnov et al., 2005, Durai et al., 2005, Pruett-Miller et al., 2008, Kim et al., 2009). This may be due to differences in target site accessibility, DNA-binding affinities, or combinatorial effects of zinc fingers in an array, as very few direct and systematic comparisons have been made. In general, four fingered ZFNs have been demonstrated to promote an equal, if not greater, level of specific DNA cleavage than three fingered ZFNs (Durai et al., 2005, Kim et al., 2009, Shimizu et al., 2011, Pattanayak et al., 2011).

1.3.5 Applications

The prospect of inducing targeted DSBs in cells presents a paradigm shift for the potential of customised gene therapy. ZFNs have the potential to offer a precise and permanent solution to gene correction in disease studies. To date, notable applications of ZFN-mediated gene therapy include corrections of mutations associated with X-SCID (Urnov et al., 2005), Parkinson's disease (Soldner et al., 2011), haemophilia B (Li et al., 2011a) and sickle cell disease (Sebastiano et al., 2011, Zou et al., 2011) via HDR gene targeting methods in cultured cells and mouse models. NHEJ-mediated gene editing using ZFNs has been employed to confer resistance to HIV-1 via disruption of the *CCR5* gene (Perez et al., 2008, Holt et al., 2010). This is part of an ongoing clinical trial (NCT00842634) and initial results suggest that viral load has dropped to below a detectable limit in 3 of 7 subjects, all of whom entered the trial with a heterozygous deletion of *CCR5*. One patient has remained aviremic and continues to possess an undetectable HIV viral load despite the discontinuation of HIV therapy (Sangamo Biosciences, 2013).

1.3.6 Limitations

Despite the obvious potential of ZFN technology, some pitfalls remain in the design and development. A substantial proportion of assembled ZFN pairs fail (Ramirez et al., 2008) and, for this reason, it is advisable that multiple ZFN pairs are assembled and tested for any intended target. While advanced construction methods such as OPEN and CoDA and the use of recommended module sets have improved success rates of assembled ZFNs, failure of ZFN arrays is most likely due to a combination of context dependent effects within a zinc finger protein, DNA-binding affinity and target site accessibility due to chromatin environment (Cornu et al., 2008, van Rensburg et al., 2013).

Obligate heterodimer *Fok*I variants have reduced levels of off target cleavage observed (Doyon et al., 2011) but the level of cytotoxicity observed can obviously be variable between different zinc finger arrays.

ZFNs offer a robust and widely attainable platform for customised gene editing, which have been demonstrated to be highly efficient in a variety of contexts. However, the lack of a definitive DNA-recognition code, leading to combinatorial issues, high failure rate of

assembled nucleases and cytotoxicity arising from off target cleavage, have resulted in concerns about the ease of production of reliable zinc finger nucleases.

1.4 TALE Proteins

Another way to target a nuclease to a particular DNA sequence is using transcription activator like effectors (TALEs). TALEs are virulence factors from the plant pathogen *Xanthomonas spp*. They promote virulence in host cells by mimicking eukaryotic transcription factors via DNA binding through a repetitive amino acid domain (Kay et al., 2007). Some TALE proteins are virulent and others are avirulent, they can modify gene expression to either promote or evade the plant immune response (Kay et al., 2005). Each TALE is composed of a repeat domain, a nuclear localisation signal (NLS) and a transcriptional activation domain (AD) (Figure 1.9 A). The repeat domain contains multiple repeats of a 34 amino acid DNA binding domain, each of which recognises one DNA base. The DNA binding domain is composed of two closely positioned α-helices in a helix-loop-helix configuration, with two variable amino acids – the repeat variable diresidue (RVD) - that recognise a single base pair of DNA (Van den Ackerveken et al., 1996, Zhu et al., 1998, Schornack et al., 2006). It is the second of the two residues that makes direct contact with the DNA (Boch et al., 2009, Moscou and Bogdanove, 2009). Each repeat is named according to the two amino acids in the RVD. For example, repeat NN contains two asparagines in the RVD. Each repeat domain is tightly packed in close proximity to the adjacent repeat in order to bind DNA in a modular fashion (Figure 1.9 B).

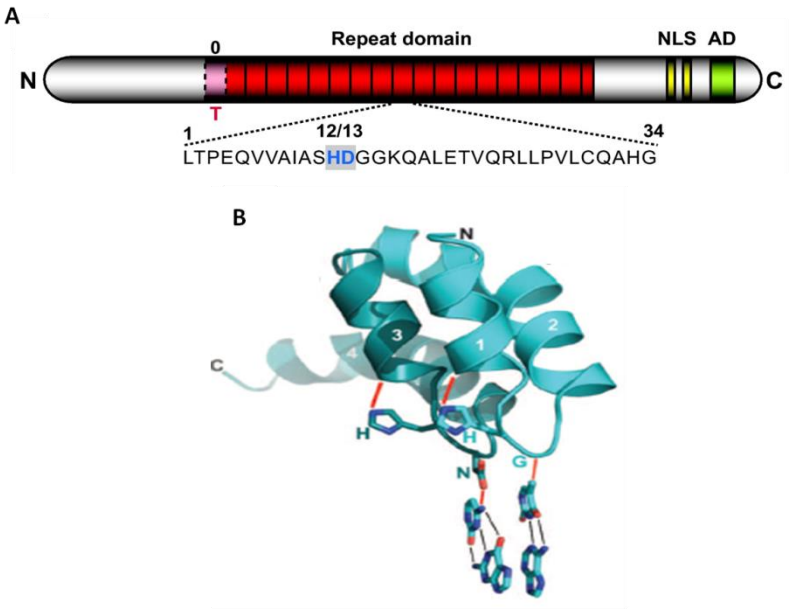

Figure 1.9 TALE Structure
(A) TALEs are composed of a repeating amino acid domain, nuclear localisation signal (NLS) and transcriptional activation domain (AD). Each repeat is composed of 34 amino acids and residues at positions 12/13 comprise the repeat variable diresidue (RVD, highlighted), responsible for DNA binding specificity (Boch et al., 2009). (B) Crystal structure of two TALE domains interacting with DNA. Each RVD binds 1 bp DNA (Mak et al., 2012).

1.4.1 DNA recognition code

Observations of TALE-DNA interactions had demonstrated that TALE proteins which contained similar RVD codes had the capability to activate each other's target genes (Schornack et al., 2008). This led to the hypothesis that modularity existed between TALE RVDs and DNA bases. A DNA-RVD code was established by Boch, Muscou and Bogdanove in two separate studies (Table 1.3) (Boch et al., 2009, Moscou and Bogdanove, 2009). While some flexibility exists between RVDs and bp (for example, NN recognises G/A), TALEs generally appear to be highly specific for their intended target. Studies have demonstrated that RVD NK can bind to G (Morbitzer et al., 2010, Miller et al., 2011) but TALE nucleases (TALENs) employing this RVD are less active than ones which use NN to target G (Huang et al., 2011). NH has subsequently been identified as binding to G and has been demonstrated to target G with higher efficiency than NK (Cong et al., 2012, Streubel et al., 2012). In addition to the recognition code, TALE binding sites are normally preceded by a thymine base, which can interact with the protein region preceding the RVD repeats (Bogdanove et al., 2010).

RVD	NI	HD	NG	IG	NN	NK	NH
DNA specificity	A	C	T	T	G or A	G	G

Table 1.3 DNA binding specificity of TALE RVDs

1.4.2 TALE nucleases (TALENs)

Following the use of ZFNs as tools to direct targeted genetic modifications, the desire to construct custom TALE nucleases (TALENs) was a natural progression. Both technologies rely on the same fundamental principle – a DNA-binding domain with bespoke target specificity, fused to the non-specific nuclease domain *Fok*I. However, the simple and modular TALE-DNA recognition code (Boch et al., 2009, Moscou and Bogdanove, 2009) means that TALENs provide the potential to create custom nucleases without the complexities faced in the design of custom ZFNs.

Naturally occurring TALE proteins were demonstrated to be able to induce DSBs when fused to *Fok*I (Christian et al., 2010, Li et al., 2011b). TALENs constructed using the RVD recognition code could initiate targeted DSBs in an *in vivo* yeast assay (Christian et al., 2010). The design or identification of TALEN target sites was approached in a similar fashion to ZFN targets. TALE binding sites are positioned in an inverted orientation to each other, centred around a spacer, where DNA cleavage by *Fok*I occurs (Figure 1.10). Typically, each TALE is composed of 15 – 20 RVDs targeting 15 – 20 bp (so the total recognition sequence for a TALEN is 30 – 40 bp long). The TALEN pair is separated by a spacer of 15 – 30 bp, which is much longer than the 6 bp spacer typically used by ZFNs. This is due to the length of the amino acid linker between the TALE and *Fok*I domains (Christian et al., 2010, Miller et al., 2011, Cermak et al., 2011).

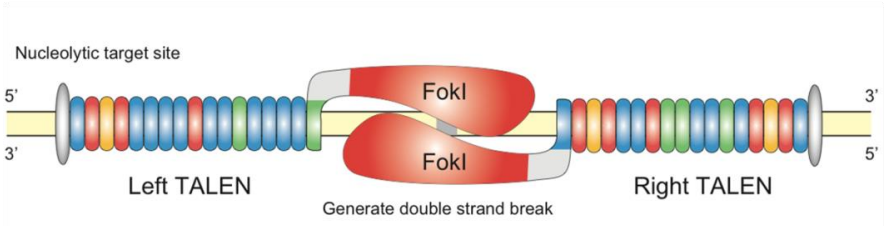

Figure 1.10 DNA binding and cleavage by TALENs
Schematic of a TALEN pair binding to DNA. Each TALE RVD module contacts 1 bp DNA. TALE DNA-binding domains are fused to the non-specific DNA cleavage domain of *Fok*I nuclease. A TALEN pair is separated by a spacer of 15 – 30 bp, where *Fok*I dimerisation and DNA cleavage occurs (Sanjana et al., 2012).

1.4.3 Optimisations of TALE architectures

1.4.3.1 *Fok*I nuclease domain

TALENs have employed the same *Fok*I domain modifications as were developed for ZFNs. The use of an obligate heterodimer *Fok*I domain in TALENs is now becoming widely adopted (Huang et al., 2011, Hockemeyer et al., 2011, Cade et al., 2012) and is considered the most efficient adaptation to reduce off target cleavage resulting from the dimerisation of two “left” or two “right” DNA binding monomers (Figure 1.8).

1.4.3.2 RVD domain choice

DNA contacts from TALE RVDs are made via the second residue in the RVD. The first residue in the RVD stabilises the conformation of the RVD containing loop. DNA binding affinities of some RVDs are stronger than others. RVD NI makes van der Waals contacts with DNA base adenine, as do RVDs NG and HG with thymine. Hydrogen bonds, which are much stronger than van der Walls, are formed between HD and C bases and also between NN and G or A bases. It has been suggested that TALENs composed of a greater number of stronger RVDs are more efficient and have greater target specificity than those comprising weaker RVDs (Streubel et al., 2012, Sun and Zhao, 2013).

The lack of a definitive RVD choice for G may promote off-target specificity. The RVD NK was identified as a substitute RVD to target G bases with greater specificity than NN RVDs, showing a much stronger preference for guanine than adenine (Morbitzer et al., 2010, Miller et al., 2011). However, employment of the RVD NK in TALEN pairs was demonstrated to lower TALEN activities in plants, mammalian cells and zebrafish embryos (Cong et al., 2012, Streubel et al., 2012, Huang et al., 2011). The RVD NH was demonstrated to be specific for guanine bases, without sacrificing TALEN activity (Cong et al., 2012, Streubel et al., 2012). This RVD is now available as part of the Golden Gate TALEN assembly kit.

1.4.3.3 Target sequence requirements

The majority of natural TALE binding sites are preceded by a 5’ T. Custom TALEs that bind to a target site preceded by a T have been demonstrated to be much more highly active than those that do not (Lamb et al., 2013). The N-terminal region immediately before the

repeat RVDs has been demonstrated to facilitate some interactions with the DNA sequence preceding the TALE binding site (usually a 5' T) (Mak et al., 2012). TALE design tools now incorporate this parameter into target site identification (Doyle et al., 2012). Recent studies have designed alternative TALE architectures to avoid this constraint (Lamb et al., 2013, Tsuji et al., 2013, Sun et al., 2012). Further recommendations are that RVDs in positions one and two should not be those binding T and A, respectively, and that recognition sites should not end with a G (Boch et al., 2009, Moscou and Bogdanove, 2009, Cermak et al., 2011).

1.4.3.4 N- and C-terminal truncations

The first 152 residues of the TALE N-terminus have been demonstrated to be responsible for transport of a TALE into a plant cell but dispensable for mediating TALE binding (Szurek et al., 2002). A truncation of these residues, named Δ152, was shown to be sufficient for the assembly of functional TALE transcriptional activators (Miller et al., 2011). It was reasoned that shortening of the C-terminus would provide an optimal framework for expressing TALEs fused to *Fok*I. Truncation of the C-terminus to +63 residues was also demonstrated to produce highly functional TALEN pairs (Miller et al., 2011). This "Miller" architecture comprising truncated N and C termini has been widely adopted in TALE assembly kits (Sanjana et al., 2012).

1.4.4 Applications

To date, TALENs have been employed to facilitate gene editing in human somatic cells, iPSCs, *Drosophila*, zebrafish, rats and livestock (Christian et al., 2010, Hockemeyer et al., 2011, Liu et al., 2012, Sander et al., 2011a, Tesson et al., 2011, Carlson et al., 2012). TALENs have been utilised in conjunction with HDR to repair a mutation in the human β-globin gene responsible for sickle cell disease (Sun et al., 2012, Voit et al., 2014). TALENs and a homology-containing donor vector were delivered to human cells, where it was demonstrated that TALEN-mediated DSBs were repaired via HDR insertion of the donor, correcting the mutation responsible for sickle cell disease. TALENs have also been employed to create a large animal model for atherosclerosis utilising NHEJ-mediated mutations in the low density lipoprotein detector (LDLR) gene (Carlson et al., 2012).

These studies paved the way for development and assessment of custom TALEs in both somatic cells and induced pluripotent stem cells (IPSCs) (Miller et al., 2011, Hockemeyer et al., 2011) and in model organisms such as *Drosophila*, zebrafish and rats (Liu et al., 2012, Huang et al., 2011, Sander et al., 2011a, Tesson et al., 2011). TALENs have been demonstrated to induce mutations at a frequency of around 20% on average (Chen et al., 2013).

1.4.5 TALE Transcriptional Modifiers

TALEs function as transcriptional modulators in nature, altering the gene expression of host genes in infected plants (Szurek et al., 2002, Büttner and Bonas, 2002). Following the identification of the TALE-DNA binding recognition code, engineered TALEs were demonstrated to activate plant genes via a designed DNA binding domain (Morbitzer et al., 2010). Subsequently, designed TALEs were demonstrated to activate mammalian target genes via fusion of a TALE-DNA binding domain with a transcriptional regulator (Miller et al., 2011, Geissler et al., 2011). TALEs can be used to direct a transcriptional effector domain (ED) to alter expression of a gene of interest by designing a TALE-DNA binding domain that binds to the promoter of the target gene (Figure 1.11). Transcriptional activation mediated by custom TALEs has been performed via fusion to activation domains such as VP64 and NF-KB p65 (Zhang et al., 2011, Maeder et al., 2013b). Repressor domains, SID and KRAB, have been employed to facilitate targeted gene repression with custom TALEs (Cong et al., 2012, Garg et al., 2012).

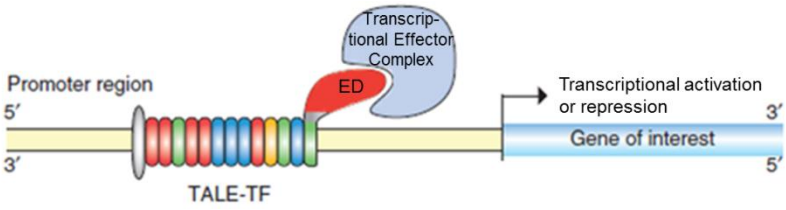

Figure 1.11 TALE Transcriptional Effector activity
Schematic of a TALE effector binding to DNA. A TALE DNA-binding domain is fused to a transcriptional effector domain (ED). Binding of the effector domain to a target sequence in a gene promoter recruits a transcriptional effector complex, which is composed of factors that facilitate transcriptional activation or repression of the gene of interest, immediately downstream of the TALE target site(adapted from Sanjana et al., 2012).

1.4.5.1 VP64-mediated activation

The VP64 activation domain is tetrameric version of the VP16 transcription factor from Herpes Simplex Virus (HSV) Type 1 (Beerli et al., 1998). VP16 was found to be one of the

most potent transcriptional activation domains and was widely used until the adaptation of the tertameric version VP64, which was demonstrated to promote higher levels of gene activation (Beerli et al., 1998). The transcriptional activation domain can be directed to specific gene promotes via a DNA binding domain, in this case a TALE domain (Figure 1.11) (Zhang et al., 2011, Miller et al., 2011). The VP64 activation domain mediates gene transcription through interactions with many other transcription factors, for example TFIIA, B, D, F and H (Hall and Struhl, 2002, Xiao et al., 1994, Kobayashi et al., 1995), recruitment of histone acetyltransferases such as pCAF and p300 (Tumbar et al., 1999) and recruitment of SAGA and SWI/SNF complexes (Hall and Struhl, 2002, Hirai et al., 2010). The interaction and recruitment of these partners promotes the acetylation of histones and widespread chromatin decondensation, leading to active gene transcription (Utley et al., 1998, Carpenter et al., 2005, Tumbar et al., 1999).

1.4.5.2 SID-mediated repression

The SIN Interaction Domain (SID) from the transcriptional repressor Mad1 mediates gene repression by the recruitment of the SIN3A co-repressor complex (Ayer et al., 1996). SID can be directed to DNA via fusion to a heterologous DNA-binding domain, such as a TALE (Figure 1.11). Components of the SIN3A repressor complex are recruited in the presence of SID. Gene repression is mediated via a histone deacetylase complex composed of histone deacetylases HDAC1 and HDAC2, sin associated proteins SAP18 and SAP30 and retinol binding protein, and RBBP4 and RBBP7 (Kadamb et al., 2013). Factors such as KDM5A, SUV39H1 and ESET are recruited by the SIN3A complex, resulting in demethylation of lysine 4 of histone H3 (KDM5A) and methylation of lysine 9 of H3 (SUV39H1, ESET) which promote the formation of an inactive chromatin state (van Oevelen et al., 2008, Yang et al., 2003).

1.4.5.3 KRAB-mediated repression

The Krüppel Associated Box (KRAB) repressor has also been utilised to mediate gene repression as a fusion to TALE DNA-binding domains (Figure 1.11). The KRAB repression domain originates from KRAB-containing zinc finger transcription factors and mediates transcriptional repression in a similar manner to SID, as described previously (Margolin et al., 1994). KRAB recruits the nucleosome remodelling deacetylase (NuRD) complex, which contains factors such as TIF1β (also called KAP1), CHD3 and histone methyltransferase

SET/DB1. KRAB-mediated recruitment of chromatin modifying factors promotes heterochromatin spreading (Groner et al., 2010). Transcriptional gene silencing mediated by KRAB results in the deposition of repressive histone marks, such as methylation of lysine 9 on histone H3 (H3K9me) and deacetlyation of lysines 9 and 14 of histone H3 (H3K9 and H3K14, respectively) (Ivanov et al., 2007).

1.4.5.4 Transcriptional Modification by TALE Transcription Factors

TALE transcription factors have been employed in a variety of cell types from plants to human cell lines and IPSCs (Morbitzer et al., 2010, Zhang et al., 2011, Zhang et al., 2013) and have also been demonstrated to modify transcription in *Drosophila* (Crocker and Stern, 2013). TALEs have been used to direct transcription factors to a variety of genes, including *Oct4*, *Sox2*, *Nanog*, and *VEGFA* (Bultmann et al., 2012, Zhang et al., 2011, Maeder et al., 2013b) and also non-coding regions such as miR-302/367 (Maeder et al., 2013b, Zhang et al., 2013). In addition to the traditional approach of targeting gene promoters, studies have also shown that efficient gene knockdown can be achieved via targeting TALE Repressor proteins to specific gene enhancers (Gao et al., 2014, Crocker and Stern, 2013). Interestingly, it has been found that the use of multiple TALE transcriptional activators targeting the same regions simultaneously has a synergistic effect and this can be utilised to achieve greater levels of target gene expression (Maeder et al., 2013b, Perez-Pinera, 2013).

While studies described above have achieved high levels of transcriptional modification at target genes, it has also been demonstrated that TALE transcriptional effectors can fail to significantly alter gene expression in some cases, summarised by Maeder *et al.* (2013b). The factors governing the optimal conditions for TALE mediated transcriptional regulation are still somewhat unclear. Many of the issues encountered with TALENs, such as binding site location, target site competition and methylation of DNA target sequence are also known to impair the functionality of TALE transcription factors (Bultmann et al., 2012, Werner and Gossen, 2014, Maeder et al., 2013b). Future studies will no doubt focus on outlining a set of recommended guidelines, similar to that developed for the design of TALENs and ZFNs.

1.4.6 Assembly of custom TALEs

Following directly on from the construction of ZFNs, TALEs were first tested in conjunction with *Fok*I nuclease to determine if the DNA binding domain of a TALE could be synthetically constructed and fused with another effector module to direct the production of targeted double strand breaks (DSBs). In a preliminary study, five custom TALEs to target DNA sequences from *Arabidopsis* and zebrafish were constructed and expressed as fusions with *Fok*I (Christian et al., 2010). The propensity of custom TALENs to induce DSBs was tested in an *in vivo* yeast assay. Two TALENs demonstrated robust nuclease activity and one showed modest levels of cleavage. This study demonstrated that custom TALEs could be designed to target novel DNA sequences and could be used in conjunction with nucleases to mediate gene targeting in a similar fashion to ZFNs. This fuelled the development of comprehensive assembly kits, similar to those developed for ZFN construction. Unlike ZFN assembly methods, there have been no context dependent issues described with TALEN assembly. Highly functional TALEs have been produced using each of the assembly methods described below. All of the kits described are available from www.addgene.org.

1.4.6.1 Golden Gate assembly (Voytas Kit)

The Golden Gate (GG) TALEN assembly kit comprises 86 plasmids; each RVD is provided in multiple plasmids with different overhang sequences, each one corresponding to a different position of that RVD within a 10-repeat array (Cermak et al., 2011). GG utilises type IIS restriction enzymes, which cleave a number of bases away from their recognition sequence. This creates DNA sticky ends that are unique to each position within the array, and thus are only compatible with ligation of the RVDs in the desired arrangement (Figure 1.12). This is a rapid assembly platform that allows digestion and ligation to occur in one reaction and facilitates assembly of multiple TALENs simultaneously. TALENs constructed via this method were demonstrated to be active in a yeast DNA cleavage assay and were effective at inducing targeted mutations in human cells and *Arabidopsis* protoplasts.

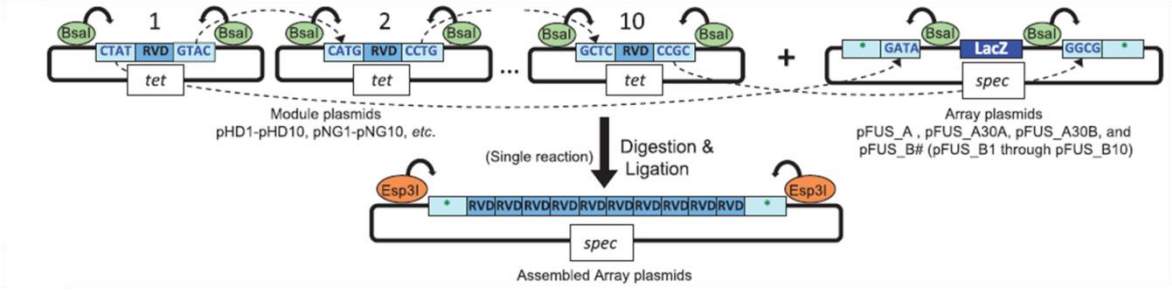

Figure 1.12 Principle of golden gate TALEN assembly
Plasmids encoding TALE RVD modules are digested using type IIS restriction enzymes creating compatible overhangs for ligation in the desired assembly. Digestion and ligation takes place in a single reaction as the plasmid overhangs are designed such that the reaction can only progress in one direction (Cermak et al., 2011).

1.4.6.2 REAL Assembly

Restriction Enzyme And Ligation (REAL) assembly employs standard cloning techniques to assemble TALEs from a kit of 28 plasmids each of which encode a single TALE RVD expression sequence (Figure 1.13 A) (Sander et al., 2011a). This method has been recently updated, called REAL-FAST, which incorporates plasmids containing up to four pre-assembled TALE RVDs, enabling higher throughput with fewer cloning steps (Figure 1.13 B) (Reyon et al., 2012a). Both of these methods rely on standard cloning techniques and TALE assembly requires the use of highly repetitive sequences, which can be prone to recombination.

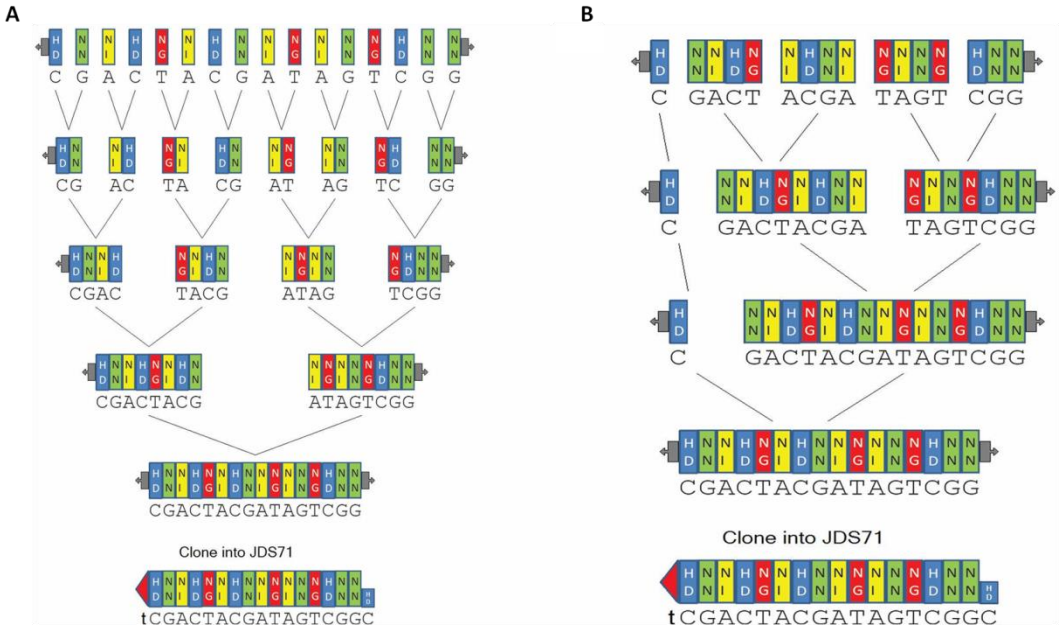

Figure 1.13 Overview of REAL and REAL-FAST TALE assemblies
Schematic showing the stages involved in (A) REAL and (B) REAL-FAST assembly methods. (A) Individual TALE RVD modules are digested and ligated together in stages to create a final TALE array. (B) Pre-assembled TALEs of up to 4 RVD modules are digested and ligated together to create a final TALE array (Reyon et al., 2012a).

TALENs assembled via REAL assembly have been demonstrated to induce high rates of NHEJ-mediated mutations in somatic zebrafish cells (Sander et al., 2011a). However, the TALENs assembled in this study were designed to target DNA sequences that have been previously targeted using ZFNs, so it is not an unbiased study.

1.4.6.3 PCR-based assembly (Zhang kit)

The TAL Effector Toolbox kit is a PCR-based assembly procedure that involves addition of position-specific Type IIS recognition sequences to each RVD using PCR, followed by restriction digestion-based cloning to assemble the RVDs in the desired order (Figure 1.14) (Zhang et al., 2011). To reduce issues faced with cloning of highly repetitive sequences, the RVD expression sequences have been codon optimised. However, PCR amplification of RVD expression sequences must be performed with high fidelity. This method provides TALE RVD sequences that are human codon optimised, and it is amenable to the assembly of multiple TALEs at a time. The Toolbox kit has now been extended to include expression vectors for assembly of TALE nucleases and TALE transcription activators for mammalian expression (Sanjana et al., 2012).

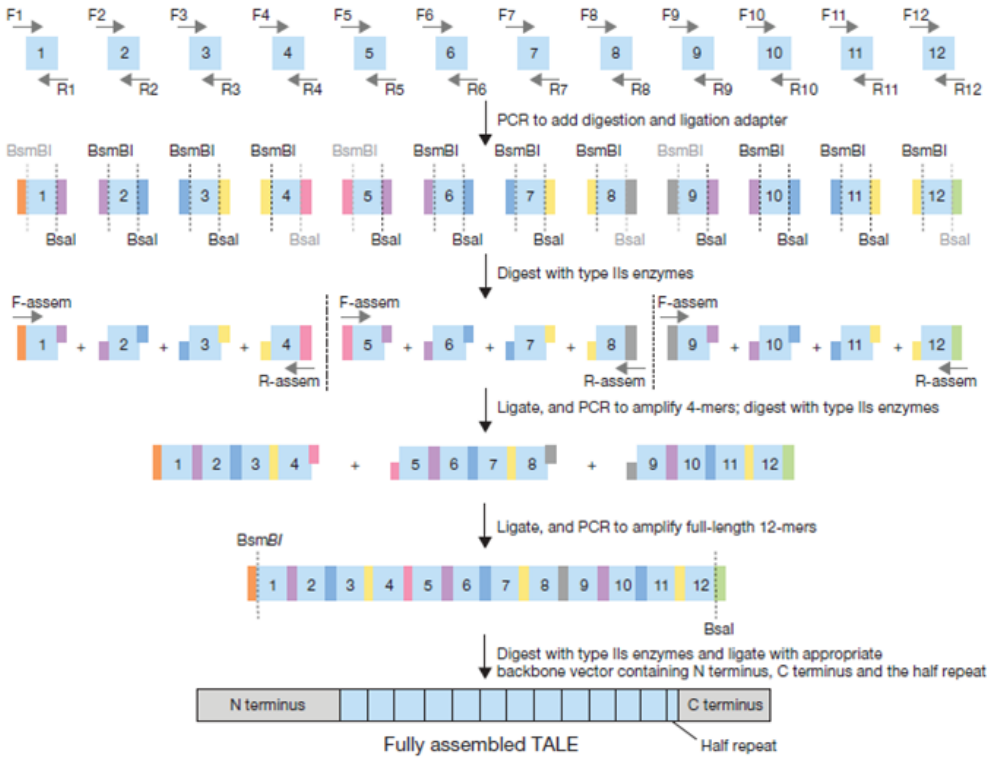

Figure 1.14 Schematic of TAL Effector Toolbox Kit assembly method
Overview of PCR-based TALE assembly. Each TALE RVD is PCR amplified from a plasmid provided in the kit in order to add adaptors for subsequent cloning using Type IIS restriction enzymes. As described in GG assembly, the RVDs can only be assembled in one predefined order. RVDs are ligated to create 4mers, which are PCR amplified to add adapters for ligation, as before. Final TALE arrays are assembled by ligation of digested 4mers (Zhang et al., 2011).

TALEs assembled in the study which first describes this construction method were fused to a transcriptional activator and assessed for their ability to activate an exogenous fluorescence reporter gene (Zhang et al., 2011). All TALEs were demonstrated to activate reporter gene expression by binding to their intended target sequence. However, the levels of gene activation varied largely between different TALEs. This may have been due to differences in TALE-DNA interactions, as the TALEs were designed to target a range of binding sites, varying largely in sequence composition and GC content (Zhang et al., 2011).

1.4.6.4 FLASH assembly

While many manual and modular TALE assembly procedures had been developed (described above), there was a lack of any high-throughput publically available systems. Fast Ligation-based Automatable Solid-phase High-throughput (FLASH) assembly is an automatable system that can enable the construction of up to 96 TALE arrays in one day (Reyon et al., 2012b). TALE RVD modules are provided as a set of 376 plasmids, which contain individual and pre-assembled RVDs (Figure 1.15 A). The first TALE RVD is PCR amplified using a biotinylated forward primer, is attached to streptavidin coated magnetic beads, and then subsequent RVDs are ligated in sequence (Figure 1.15 B). FLASH utilises solid-phase synthesis by means of magnetic beads, which avoids any requirement for analysis of intermediate constructs. However, the initial preparation of RVD modules still relies upon PCR, restriction digest and column purification of RVD expression sequences, and for this reason the system is not fully automatable.

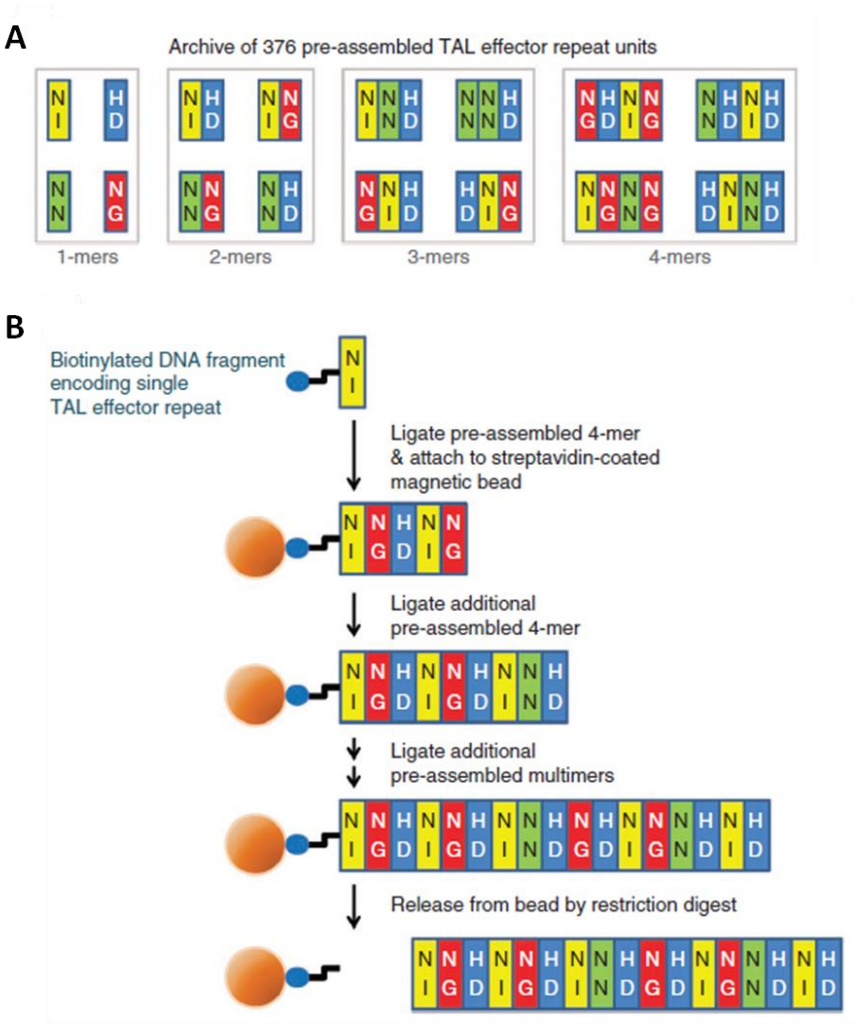

Figure 1.15 FLASH assembly methods to create custom TALEs
Overview of FLASH assembly kit and method. (A) The kit contains plasmids containing TALE RVDs either as individual RVDs (1-mers), 2-mers, 3-mers or 4-mers to be utilised in TALE assembly. (B) FLASH assembly procedure. The first TALE RVD is labelled on the 5' end with biotin (blue circle). A pre-assembled 4-mer is then ligated on to the first RVD and then attached to a streptavidin bead (orange circle). Subsequent RVDs are ligated until the desired array is assembled. The full length TALE array is then cleaved from the bead by restriction digest (Reyon et al., 2012b).

FLASH has been reported to have an ~88% success rate in producing functional TALEs (Reyon et al., 2012b). However, FLASH has been demonstrated to produce TALENs which were not active. Inactivity may have been due to target site accessibility due to localised chromatin structure or DNA modifications. It is possible that inactivity of these TALENs was as a result of low expression or inefficient folding of TALE proteins. The outcome showing production of inactive TALENs by this method is perhaps due to the high-throughput nature of FLASH assembly, permitting the generation of better quality statistics demonstrating the success rate of TALE assembly.

1.4.7 Comparison to ZFN technology

While some ZFNs have been demonstrated to induce mutations at a similar rate to TALENs (Beumer et al., 2008, Beumer et al., 2013), TALENs are generally considered to be more mutagenic than ZFNs (Chen et al., 2013, Beumer et al., 2013, Voit et al., 2014). TALENs have been demonstrated to promote fewer off-target events than ZFNs and are considered to promote the formation of larger deletions, which may be due to the larger spacer when compared to ZFNs (Beumer et al., 2013). There are still unanswered questions when it comes to TALEN design and success rate as it is unclear why certain TALEN pairs are non-functional, particularly those which target a site previously targeted by a ZFN. Certain target sites may only be accessible to smaller proteins, making ZFNs more favourable in these instances (Beumer et al., 2013). TALE activity has been demonstrated to be reduced at methylated targets, while there have been no comprehensive studies to support or refute this effect in ZFN targeting (Bultmann et al., 2012). Due to the greater mutagenicity and lower off-target activity, coupled with the ease of TALEN construction and the greater likelihood of assembling functional nucleases, TALENs are now considered the custom nuclease of choice, having now out-performed ZFNs. However, ZFNs still hold a major advantage over TALENs as they have been demonstrated in human trials (see section 1.3.5) and for this reason, ZFNs still remain a major player for use in human gene editing.

1.5 VEZF1

In this study, I intend to use custom nucleases as an alternative to RNAi to create targeted modifications at the *VEZF1* gene promoter. VEZF1 (Vascular Endothelial Zinc Finger 1) is a highly conserved vertebrate transcription factor that is essential for the development of the vascular system (Zou et al., 2010). VEZF1 is expressed broadly across all somatic cell types and seems to have its major role in the vascular system, where it is involved in regulating the development of endothelial cells (Xiong et al., 1999, Aitsebaomo et al., 2001, Zou et al., 2010) and also in the regulation of the angiogenic and lymphangiogenic phases in wound healing in adults (Gerald et al., 2013). Knockout of VEZF1 results in embryonic lethality in mice, due to disrupted organisation of the extracellular matrix and defective adhesion between vascular endothelial cells which leads to extensive haemorrhaging. The action of VEZF1 in regulation of vascular development has been

demonstrated to be in a dose dependent manner, with heterozygous embryos displaying haploinsufficienciency (Kuhnert et al., 2005).

VEZF1 has been demonstrated to have binding sites in the promoters of active genes, enhancer elements and barrier elements (Koyano-Nakagawa et al., 1994, Miyashita et al., 2004, Aitsebaomo et al., 2001, Zou et al., 2010). VEZF1 contains six C_2H_2 zinc finger (ZF) domains that are likely to be responsible for its DNA binding activity. VEZF1 has been found to bind to sequences containing runs of G found at promoter and insulator elements (Koyano-Nakagawa et al., 1994, Aitsebaomo et al., 2001, Dickson et al., 2010).

VEZF1 has been shown to have transcription activity in promoter assays *in vitro*, binding at the promoters of genes such as *interleukin-3 (IL-3), endothelin-1 (ET-1)* and *stathmin/OP18* (Koyano-Nakagawa et al., 1994, Miyashita et al., 2004, Aitsebaomo et al., 2001, Zou et al., 2010). Insulator elements are defined by two characteristics – the ability to act as an enhancer-blocker when positioned between an enhancer and a promoter and preventing the spreading of chromosomal silencing by acting as a barrier (Yusufzai and Felsenfeld, 2004). The HS4 insulator element is a DNAse I hypersensitive region of the chicken *β-globin* locus (Chung et al., 1993). The enhancer-blocking activity of HS4 is mediated by the protein CTCF, while its barrier activity is associated with VEZF1 (Bell et al., 1999). This role of VEZF1 in barrier activity has been shown to be associated with the protection of DNA from *de novo* methylation. It has also been shown that VEZF1 has a role in protecting CpG island promoters from *de novo* DNA methylation and therefore preventing gene silencing (Dickson et al., 2010). Counter intuitively, VEZF1 also has a role in promoting DNA methylation in mouse ES cells as it is required for the expression of the *de novo* DNA methyltransferase *Dnmt3b* gene (Gowher et al., 2008).

In a recent study, VEZF1 was shown to form a complex with RhoB and play a role in the regulation of angiogenesis and lymphangiogenesis (Gerald et al., 2013). The VEZF1-RhoB transcription factor complex participates in ensuring the delay in lymphangiogenesis and promotes successful wound healing. In a mouse model of oxygen induced retinopathy, DNA-binding by VEZF1 was prevented using a small molecule inhibitor and the emergence of blood vessels with a normal morphology was increased in mice which had been treated with the inhibitor molecule. This suggests that the VEZF1-RhoB complex may serve as a potential therapeutic target for development of treatment against numerous diseases

which contain a vascular component. Targeting the VEZF1-DNA binding interface could promote the formation of normal vasculature following ischaemia.

Due to the conflicting and varied roles suggested for VEZF1, a robust knockdown or knockout is required. RNAi-mediated knockdown of *VEZF1* by up to 70% of WT expression levels has resulted in depletion of VEZF1 at around half of its putative binding sites (Strogantsev, 2009). *VEZF1* knockdown to greater levels was not attainable by RNAi prior to the inception of this project. Therefore, I intend to perform targeted modifications of the *VEZF1* gene promoter using custom designed ZFNs, TALE Repressors and TALENs.

1.6 Aims and objectives of this thesis

This project was established with the view of designing a robust and targeted gene knockout system to investigate *VEZF1* gene function. It is proposed to target the *VEZF1* gene promoter using custom nucleases and design a strategy for future generation of VEZF1 knockout cell lines via insertion of a transgene system. The general aims of this work were:

1. Assemble custom ZFNs and TALENs to create targeted DSBs at the *VEZF1* gene promoter.
2. Create a modified expression vector to express TALE Repressors and assess the level of *VEZF1* gene repression directed by custom TALE proteins.
3. Design and assemble a targeting vector system and perform targeted transgene insertion at the site of nuclease-mediated DSBs in the *VEZF1* gene promoter.
4. Characterise the rate at which custom nucleases create targeted mutations at the *VEZF1* promoter, and analyse the types of mutations induced.
5. Assay the phenotype of cells which have been targeted using custom nucleases and repressors.

Chapter 2

Materials & Methods

2.1 Cell Lines

Cell Line	Reference	Source
HEK293 T17		ATCC® CRL-11268™
K562	(Lozzio and Lozzio, 1975)	Prof Tony Green, School of Clinical Medicine, University of Cambridge, via Dr David Vetrie, Institute of Cancer Sciences Epigenetics Unit, University of Glasgow

2.2 Reagents

2.2.1 Cell culture reagents

Product	Manufacturer	Product Code
DMEM Medium (1X) + GlutaMAX™ - I	Life Technologies	10564029
DMSO	Riedel-de Haen	60153
Foetal Bovine Serum	Sigma	F9665
Lipofectamine® 2000 Transfection Reagent	Life Technologies	11668019
Optimem® Reduced Serum Medium - I	Life Technologies	31985088
Penicillin/Streptomycin	Sigma	P0781
RPMI Medium 1640 (1X) + GlutaMAX™ - I	Life Technologies	72400146
Trypsin	Life Technologies	12604013

2.2.2 Antibodies

Target Protein	Manufacturer	Product Number
Anti-rabbit IgG-HRP	Pierce	1858415
VEZF1	In-house raised	(Gowher et al., 2008)

2.2.3 Enzymes

Enzyme	Manufacturer	Product Number
*Afl*II	NEB	R0520
*Age*I	NEB	R0552
*Asc*I	NEB	R0558
*Bam*HI	NEB	R3136
*Bgl*II	NEB	R0144
*Bsa*I	NEB	R0535
*Bsr*BI	NEB	R0102
*Bsr*GI	NEB	R0575
*Bst*EII-HF	NEB	R3162
CIP	NEB	M0290
*Eco*RI HF	NEB	R3101
*Esp*3I (*Bsm*BI)	NEB	R0580
*Hind*III HF	NEB	R3104
*Kpn*I HF	NEB	R3142
*Nco*I	NEB	R0193
*Nhe*I HF	NEB	R3131
*Not*I HF	NEB	R3189
Maximase Polymerase	Transgenomic	703205
*Msp*I	NEB	R0106
OneTaq Polymerase	NEB	M0480
Pfu Polymerase	Promega	M7741
Plasmid Safe™ DNase	Epicentre	E3101K
Platinum Taq Polymerase	Life Technologies	109668018
*Pml*I	NEB	R0532
*Rsa*I	NEB	R0167
RNase H	NEB	M0297
*Sac*I HF	NEB	R3156

*Stu*I	NEB	R0187
T4 DNA Ligase	Invitrogen	15224041
T7 Endonuclease I	NEB	M0302
Taq DNA Polymerase	NEB	M0267
*Xba*I	NEB	R0145
*Xho*I	NEB	R0146
*Xma*I	NEB	R0180

2.2.4 Chemicals & reagents

Reagent	Manufacturer	Product Number
1 kb DNA Ladder	NEB	N32315
100 bp DNA Ladder	Life Technologies	10787-018
100 mM dNTP set	Life Technologies	10297018
AEBSF	Sigma	A8456
Agar	Formedium	AGA02
Agarose	Roche	11388991001
Ampicillin	Sigma	A0166
BCP	Sigma	B9673
BSA 10 mg/ml	NEB	B9001S
DH5α Subcloning Efficiency Competent Cells	Life Technologies	18265017
DTT	Melford Laboratories Ltd	MB1015
Ethidium Bromide	Sigma	E1510
FastStart Universal SYBR Green Master (Rox)	Roche	04913914001
IPTG	Sigma	I5502
LB Broth	Sigma	L1900
Leupeptin	Sigma	L2884
NuPAGE® Antioxidant	Life Technologies	NP0005
NuPAGE® LDS sample Buffer (4X)	Life Technologies	NP0007

NuPAGE® MOPS SDS Running Buffer (20X)	Life Technologies	NP0001
NuPAGE® Sample Reducing Agent (10X)	Life Technologies	NP0004
NuPAGE® Transfer Buffer (20X)	Life Technologies	NP0006
Pepstatin	Sigma	P5318
Phosphate buffered saline (PBS)	Sigma	D8662
Ponceau S	Sigma	P7170
Puromycin	Sigma	P8833
Safeview	ABM	G108
Sodium dodecyl sulphate	Sigma	L3771
Spectinomycin	Sigma	S4014
SuperSignal West Dura Extended Duration Substrate	Thermo	34075
Trizol	Life Technologies	15596026
Tween® 20	Sigma	P1379
X-gal	Fisher Scientific	7240-90-6

2.2.5 Reagent Kits

Reagent	Manufacturer	Product Number
DNeasy Blood & Tissue Kit	Qiagen	69581
Gateway® LR Clonase® II Enzyme Mix	Life Technologies	11791020
MinElute PCR Purification Kit	Qiagen	28004
NanoOrange® Protein Quantification Kit	Life Technologies	N-6666
pGEM®-T Easy Vector System I	Promega	A1360
QiaPrep Miniprep Spin Kit	Qiagen	27104
QiaQuick Gel Extraction Kit	Qiagen	28704
Superscript III Kit	Life Technologies	12574026
Surveyor® Mutation Detection Kit	Transgenomic	706020

2.2.6 Oligonucleotide primers

Custom oligonucleotide primers were purchased from Biomers or MWG Eurofins. Sequences are listed in appendix.

2.2.7 Plasmids

Zinc finger plasmids were obtained from The Zinc Finger Consortium (www.zincfingers.org) Modular Assembly Kit (Wright et al., 2006), purchased from www.addgene.org.

Golden Gate TALEN Assembly kit was developed by The Voytas Lab (Cermak et al., 2011) and purchased from www.addgene.org.

Plasmid	Source	References
pFOKI-EL (pMLM292)	www.addgene.org	Addgene plasmid 21873
pFOKI-KK (pMLM290)	www.addgene.org	Addgene plasmid 21872
pBlueScript_Sharkey	GeneCust (synthesis)	(Guo et al., 2010)
pCAG-T7_TALEN	www.addgene.org	Addgene plasmid 37184
pTALETF_v2	www.addgene.org	Addgene plasmid 32186
pUC57_p2A_SID	GeneCust (synthesis)	(Cong et al., 2012)
pUC57_TALE NT	ShineGene (synthesis)	(Miller et al., 2011)
pUC57_TALE CT WT	ShineGene (synthesis)	(Miller et al., 2011)
pUC57_TALE CT ELD-S	ShineGene (synthesis)	(Miller et al., 2011)
pUC57_TALE CT KKR-S	ShineGene (synthesis)	(Miller et al., 2011)
pSHTLR3	GeneCust (synthesis)	Ileana Guerrini
pSHTLR5	GeneCust (synthesis)	Ileana Guerrini
pSHTLRmid	In-house assembly	Ileana Guerrini
pDESTR4-R3	In-house assembly	Ileana Guerrini
pTRIPZ	Thermo	RHS4740
pEGFP-N1	CloneTech	6085-1

2.3 Buffers

2.3.1 Buffers for Gel Electrophoresis

5X TBE Buffer

1.1 M Tris
900 mM Boric Acid
25 mM EDTA

6X Gel Loading Buffer

30 % Glycerol
0.02 % Bromophenol Blue
0.02 % Xylene Cyanol

2.3.2 Buffers for Western Blotting

Cell Lysis Buffer (50 ml)

4.5 ml 0.5 M Tris-HCl pH 8.5
100 µl 0.5 M EDTA
5 ml 10% SDS
5 ml Glyerol

Immediately prior to use, 10 ml buffer was pipetted into a new tube and supplemented with the following:

25 mg/ml AEBSF
0.5 mg/ml Pepstatin
0.7 mg/ml Leupeptin

1X Transfer Buffer (1L)

50 ml 20X NuPAGE® Transfer Buffer
1ml NuPAGE® Antioxidant
100 ml Methanol

10X TBST (500ml)

250mM Tris pH8
1.5M NaCl
27mM KCl
2% Tween

Blocking Buffer (100 ml)

5% Non-fat Dried Milk Powder (Marvel)
To 100ml with 1x TBST

2.4 Modular Assembly of ZFNs

Modular assembly was carried out using the Zinc Finger Consortium Modular Assembly Kit (www.zincfinges.org, www.addgene.org) and construction was carried out according to the protocol by Wright *et al.* (2006).

2.4.1 ZFN designs using Zinc Finger Targeter (ZiFiT) software

400 bp of the *VEZF1* core promoter, 200 bp either side of the TSS, was searched using ZiFiT (Sander et al., 2007). Suggested ZFN design outputs were chosen based on the number of zinc fingers in the array (3 or 4), the zinc finger module source and genomic location with respect to VEZF1 TSS and nucleosome positioning.

2.4.2 Restriction digest of zinc finger plasmids for modular assembly

1 µg vector backbone plasmid DNA was digested as follows:

10 x NEB Buffer 1	4 µl
*Age*I (50 U/µl)	1 µl
*Bam*HI HF (20 U/µl)	1 µl
BSA (10mg/ml)	1 µl
DNA	1 µg
H_2O	to 40 µl

Reactions were incubated in a 37°C hotblock for 3 hours. 1µl CIP was added to each reaction and incubated at 37°C for a further 10 minutes. Digest products were resolved using a 0.8% TBE-agarose gel stained post run with 1 µg/ml ethidium bromide. The desired band was visualised using a UV luminometer and was excised using a scalpel. The digested plasmid DNA was purified using a QiaQuick Gel Extraction Kit.

2µg vector insert plasmid DNA was digested as follows:

10 x NEB Buffer 4	4 µl
*Bam*HI HF (20 U/µl)	1 µl
*Xma*I (10 U/µl)	1 µl
BSA (10 mg/ml)	1 µl
DNA	2 µg
H_2O	to 40 µl

Reactions were incubated in a 37°C hotblock for 3 hours. Digest products were resolved using a 2% TBE-agarose gel stained post run with 1 µg/ml ethidium bromide. The desired band was visualised using a UV luminometer and was excised using a scalpel. The digested plasmid DNA was purified using a QiaQuick Gel Extraction Kit. Digested DNA fragments were ligated as outlined in 2.4.8.

2.4.3 Analytical Restriction Digest to Screen ZF Assemblies

2µg plasmid DNA was digested as follows:

10 x NEB Buffer 4	4 µl
*Bam*HI HF (20 U/µl)	1 µl
*Xba*I (20 U/µl)	1 µl
BSA (1 mg/ml)	1 µl
DNA	2 µg
H_2O	to 40 µl

Reactions were incubated at 37°C for 3 hours. The digest products were resolved on a 2% TBE-agarose gel to confirm if cloning had been successful. The gel was stained prior to running with 2µl Safeview per 100 ml TBE buffer. Gels were imaged using a Fujifilm FLA-5000 scanner.

2.4.4 Restriction Digest of Zinc Finger Plasmids and *Fok*I Expression Plasmids for Cloning ZFNs

1µg vector backbone or 2µg vector insert plasmid DNA was digested as follows:

10 x NEB Buffer 4	4 µl
*Bam*HI HF (20 U/µl)	1 µl
*Xba*I (20 U/µl)	1 µl
BSA (1 mg/ml)	1 µl
DNA	1 or 2 µg
H_2O	to 40 µl

Reactions were incubated in a 37°C hotblock for 3 hours. 1µl CIP was added to each digest of vector backbone and incubated at 37°C for a further 10 minutes. Digest products were resolved using a 0.8% TBE-agarose gel (backbones) or 2% TBE-agarose gel (inserts). Gels were stained post run with 1 µg/ml ethidium bromide. The desired bands were visualised using a UV luminometer and was excised using a scalpel. The digested plasmid DNA was purified using a QiaQuick Gel Extraction Kit and fragments were ligated together following the procedure described in section 2.4.8.

2.4.5 Screening ZFs Ligated to *Fok*I

2 µg plasmid DNA was digested as follows:

10 x NEB Buffer 4	4 µl
*Eco*RI HF (20 U/µl)	1 µl
*Nhe*I HF (20 U/µl)	1 µl
BSA (1 mg/ml)	1 µl
DNA	2 µg
H_2O	to 40 µl

Reactions were incubated at 37°C for 3 hours. The digest products were resolved on a 2% TBE-agarose gel, stained with Safeview, to confirm if cloning had been successful. Gels were imaged using a Fujifilm FLA-5000 scanner.

2.4.6 Sharkey *Fok*I variant assembly in EL and KK *Fok*I

1 µg vector backbone (pFOKI-EL or pFOKI-KK) or 2 µg vector insert (pBluescript_Sharkey) plasmid DNA were digested as follows:

10 x NEB Buffer 4	4 µl
*H*indIII HF (20 U/µl)	1 µl
BSA (1 mg/ml)	1 µl
DNA	1 or 2 µg
H_2O	to 40 µl

Reactions were incubated in a 37°C hotblock for 3 hours. 1µl CIP was added to each digest of vector backbone and incubated at 37°C for a further 10 minutes. Digest products were resolved using a 0.8% TBE-agarose gel (backbones) or 2% TBE-agarose gel (inserts). Gels were stained with Safeview. The desired bands were visualised using a UV luminometer and excised using a scalpel. The digested plasmid DNA was purified using a QiaQuick Gel Extraction Kit. Sharkey was ligated into FokI vectors as described in section 2.4.8.

2.4.7 Analytical Restriction Digests – Confirmation of Sharkey Assembly

2 µg plasmid DNA was digested as follows:

10 x NEB Buffer 1	4 µl
*Age*I (50 U/µl)	1 µl
*Bam*HI HF (20 U/µl)	1 µl
BSA (1 mg/ml)	1 µl
DNA	2 µg
H_2O	to 40 µl

Reactions were incubated in a 37°C hotblock for 3 hours. The digest products were resolved on a 1% TBE-agarose gel to confirm if cloning had been successful. Gels were stained with prior to running with 2 µl Safeview per 100 ml TBE buffer. Gels were imaged using a Fujifilm FLA-5000 scanner.

2.4.8 Ligation of Restriction Fragments

Ligation reactions were set up using 50 ng of plasmid vector backbone and the corresponding nanogram amount of vector insert to gain the desired ratio of backbone:insert. Typical ratios used were 3:1, 1:1 1:3 & 1:10. Calculations were based on the following equation:

$$\text{ng insert} = \left\{ \frac{\text{ng backbone x size insert (kilobase pairs)}}{\text{size backbone (kilobase pairs)}} \right\} \text{x ratio}$$

5 x Reaction Buffer	4 µl
Vector Backbone	50 ng
Vector Insert	as above
T4 DNA Ligase (5 U/µl)	1 µl
H_2O	to 20 µl

Additional controls were set up which included a backbone only ligation, without addition of insert, to test the degree of self-ligation occurring within the reaction. Reactions were incubated at 15°C overnight and transformed into DH5α competent cells the following day.

2.4.9 Transformation of Competent Cells

DH5α competent cells were thawed on ice and 50 µl per transformation was aliquoted into pre-chilled eppendorfs. To this, 2.5 µl DNA was added and the reaction was incubated on ice for 30 minutes. The cells were heat shocked at 42°C for 30 seconds in a water bath and returned to ice for 2 minutes. 950 µl SOC was added to each reaction under sterile conditions and the reactions were incubated for 1.5-2 hours at 37°C with shaking at 225 rpm. 50 µl and 100 µl of each reaction was spread onto pre-warmed LB agar plates containing 100 µg/ml ampicillin. The plates were incubated at 37°C overnight.

2.4.10 Plasmid Isolation

Several colonies were isolated from a transformation reaction using a sterile pipette tip. This was then used to inoculate a 5ml LB broth containing 100 µg/ml ampicillin in a 50 ml falcon tube. The culture was grown at 37°C with shaking at 225 rpm overnight. 1ml of the culture was transferred to a 1.5 ml eppendorf and the bacterial cells recovered by centrifugation at 13000 rpm for 5 minutes. The supernatant was discarded and plasmid DNA was isolated from the bacterial cell pellet using a QiaPrep Miniprep Spin Kit, following the standard protocol.

2.4.11 Preparation of Glycerol Stocks

0.5 ml of an overnight bacterial culture was added to 0.5 ml 50% glycerol and stored at -80°C.

2.5 Golden Gate TALE Assembly

Golden Gate assembly was carried out using the Voytas Lab Assembly Kit (www.addgene.org) and construction was carried out according to protocol by Cermak *et al.*(2011).

2.5.1 TALE Nucleotide Targeter (TALE-NT)

400 bp of the *VEZF1* core promoter, 200 bp either side of the TSS, was searched using TALE-NT (Doyle et al., 2012). Suggested TALE design outputs were chosen based on the number of RVDs in the array, the spacer between TALE pairs (15 or 30 bp) and genomic location with respect to *VEZF1* TSS and nucleosome positioning.

2.5.2 Golden Gate Assembly of pTAL_A, B & C vectors

TALE RVDs are assembled in pTAL vectors up to 10 RVDs at a time. The first 10 RVDs are assembled into pTAL_A, the next 10 into pTAL_B and the final 10 into pTAL_C. Plasmids for all four RVDs in any position of a TALE array are selected from the kit and assembled into the necessary pTAL vectors. The Golden Gate reactions were set up as follows:

10 x DNA Ligase Buffer	2 µl
BSA (2 mg/ml)	1 µl
pTAL vector	150 ng
RVD DNA	150 ng each RVD plasmid, up to 10 plasmids
*Bsa*I (10 U/µl)	1 µl
T4 DNA Ligase (5 U/µl)	2 µl
H_2O	to 20µl

Samples were subjected to a simultaneous digestion and ligation reaction using a thermocycler programmed to perform the following temperature cycles:

Temperature	Time	Number of Cycles
37°C	5 minutes	10
16°C	10 minutes	
50°C	5 minutes	1
80°C	5 minutes	1

Reactions were then treated with Plasmid Safe Nuclease treatment to remove unligated linear dsDNA. 1 µl 10mM ATP and 1 µl Plasmid Safe Nuclease (10 U/µl) were added to reactions and incubated at 37°C for 1 hour.

Golden Gate reactions were then transformed into DH5α competent cells as detailed in section 2.4.9. Transformation reactions were plated on LB Agar plates containing 50 µg/ml spectinomycin, 20 mg/ml X-gal and 0.1M IPTG to perform blue/white screening of

clones containing inserted RVDs. The plates were incubated at 37°C overnight and were then incubated at 4°C overnight to allow blue colour to fully develop.

2.5.3 Colony PCR to confirm assembly of RVDs in pTAL vectors

Three white colonies from each assembly were selected and subjected to screening by colony PCR. Isolated colonies were inoculated into 50 µl H_2O and a 5 µl aliquot was taken into the following PCR reaction:

10 x ThermoPol Buffer	5 µl
10 mM dNTPs	1 µl
10 µM F Primer pCR8_F1	1 µl
10 µM R Primer pCR8_R1	1 µl
Colony suspension	5 µl
NEB Taq Polymerase (5 U/µl)	0.25 µl
H_2O	to 50 µl

PCR was performed using a thermocycler programmed to perform the following temperature cycles:

Temperature	Time	Number of Cycles
95°C	30 seconds	1
95°C	30 seconds	33
55°C	30 seconds	
68°C	1.75 minutes	
68°C	5 minutes	1

PCR products were resolved using a 1% TBE-agarose gel stained with Safeview. Gels were imaged using a Fujifilm FLA-5000 scanner. Colony suspensions from correct clones identified were cultured overnight by inoculating the remaining suspension into 2 ml LB containing 50 µg/ml spectinomycin and shaking at 37°C for 16 hours. Plasmid DNA was isolated following the procedure outlined in section 2.4.10.

2.5.4 Restriction digest to confirm pTAL assemblies

Miniprepped DNA was analysed by restriction digestion, set up as follows:

10 x NEB Buffer 4	2 µl
BSA (1 mg/ml)	1 µl
DNA	1 µg
*Afl*II (20 U/µl)	0.5 µl
*Xba*I (20 U/µl)	0.5 µl
H_2O	to 20 µl

Reactions were incubated in a 37°C hotblock for 3 hours. Digest products were visualised using a 1% TBE-agarose gel stained with Safeview. Gels were imaged using a Fujifilm FLA-5000 scanner. Correct clones were confirmed by sequencing using primers pCR_8F1 and pCR_8R1.

2.6 Construction of SID-Repressor Vector

2.6.1 Creation of pRFPturbo

RFPturbo was PCR amplified from pTRIPZ using primers which facilitated the addition of restriction enzyme recognition sites. PCR reaction was set up as follows:

10 x Buffer	5 µl
10 mM dNTPs	4 µl
10 µM F Primer RFP5_Nhe_Age	1 µl
10 µM R Primer RFP3_Bam_Not	1 µl
pTRIPZ	50 ng
Pfu Polymerase (2.5 U/µl)	0.5 µl
H_2O	to 50 µl

PCR was performed using a thermocycler programmed to perform the following temperature cycles:

Temperature	Time	Number of Cycles
95°C	2 minutes	1
95°C	30 seconds	33
55°C	30 seconds	
72°C	2.5 minutes	
72°C	5 minutes	1

PCR products were resolved using a 1% TBE-agarose gel stained with Safeview. The desired band was visualised using a UV luminometer and was excised using a scalpel. The PCR product was purified using a QiaQuick Gel Extraction Kit.

1 µg purified PCR product (insert) and 1 µg of pEGFP-N1 (backbone) were digested as follows:

10 x NEB Buffer 4	4 µl
*Nhe*I HF (20 U/µl)	1 µl
*Not*I HF (20 U/µl)	1 µl
BSA (1 mg/ml)	1 µl
DNA	1 µg
H_2O	to 40 µl

Digest reactions were incubated at 37°C for 3 hours in a thermocycler. 1µl CIP was added to digestion of pEGFP-N1 vectors and incubated at 37°C for a further 10 minutes. pEGFP-N1 digest products were resolved using a 0.8% TBE-agarose gel stained pre-run with Safeview. The desired band was visualised using a UV luminometer and was excised using a scalpel. The digested plasmid DNA was purified using a QiaQuick Gel Extraction Kit. Digested RFPturbo PCR product was purified using a QiaQuick PCR Purification Kit.

Fragments were ligated together following the procedure described in section 2.4.8.

2.6.2 Confirmation of pRFPturbo assembly

1µg plasmid DNA was digested as follows:

10 x NEB Buffer 4	4 µl
*Nhe*I HF (20 U/µl)	1 µl
*Not*I HF (20 U/µl)	1 µl
BSA (1 mg/ml)	1 µl
DNA	1 µg
H_2O	to 40 µl

Reactions were incubated at 37°C for 3 hours. The digest products were resolved on a 1% TBE-agarose gel to confirm if cloning had been successful. The gel was stained prior to running with 2µl Safeview per 100 ml TBE buffer. Gels were imaged using a Fujifilm FLA-5000 scanner.

2.6.3 Construction of pRFP_SID

The sequence of pSID4X (Cong et al., 2012) was utilised as a template to synthesise p2A-SID. p2A-SID sequence was flanked by a 5' *Bam*HI and a 3' *Not*I restriction site recognition sequences. p2A-SID synthesised sequence (in pUC57_p2A_SID) was digested and cloned into pRFP_TURBO to assemble pRFP_SID. Digest reactions of 2 µg pRFP_TURBO or 1 µg synthesised cassette were set us as follows:

10 x NEB Buffer 4	2 µl
BSA (1 mg/ml)	1 µl
DNA	1 or 2 µg
*Bam*HI HF (20 U/µl)	0.5 µl
*Not*I HF (20 U/µl)	0.5 µl
H_2O	to 20 µl

Digest reactions were incubated in a 37°C hotblock for 3 hours. 1 µl CIP was added to each digest of vector backbone and incubated at 37°C for a further 10 minutes. Digest products were resolved using a 0.8% TBE-agarose gel (backbones) or 2% TBE-agarose gel (inserts). Gels were stained with Safeview. The desired bands were visualised using a UV luminometer and excised using a scalpel. The digested plasmid DNA was purified using a QiaQuick Gel Extraction Kit. Extracted DNA fragments were ligated following the procedure in section 2.4.8.

2.6.4 Confirmation of pRFP_SID Assembly

2 µg plasmid DNA was digested as follows:

10 x NEB Buffer 4	2 µl
BSA (1 mg/ml)	1 µl
DNA	2 µg
*Nhe*I HF (20 U/µl)	0.5 µl
*Not*I HF (20 U/µl)	0.5 µl
H_2O	to 20 µl

Reactions were incubated in a 37°C hotblock for 3 hours and digest products were visualised using 1% TBE-agarose gel stained with Safeview. Gels were imaged using a Fujifilm FLA-5000 scanner.

TALE N- and C- termini from the vector pCAG_T7_TALEN were cloned into pRFP_SID to create pRFP-TALE-SID, which is capable of participating in Golden Gate Assembly.

2.6.5 Assembly of pRFP-TALE-SID

A digest reaction of 1 µg pRFP_TALE-SID was set up as follows:

10 x NEB Buffer 3	2 µl
pRFP_TALE-SID	2 µg
*Bgl*II (10 U/µl)	0.5 µl
H_2O	to 20 µl

A digest reaction of 2 µg pCAG_T7 was set up as follows:

10 x NEB Buffer 3	2 µl
BSA (1 mg/ml)	1 µl
pCAG_T7	2 µg
*Bam*HI (20 U/µl)	0.5µl
*Bgl*II (10 U/µl)	0.5 µl
H_2O	to 20 µl

Digest reactions were incubated in a 37°C hotblock for 3 hours. 1 µl CIP was added to digest of vector backbone and incubated at 37°C for a further 10 minutes. Digest products were resolved using 0.8% TBE-agarose gel (backbone) or 1% TBE-agarose gel (insert). Gels were stained with Safeview prior to running. The desired bands were visualised using a UV luminometer and excised using a scalpel. The digested plasmid DNA was purified using

a QiaQuick Gel Extraction Kit. Extracted DNA fragments were ligated following the procedure in Section 2.4.8.

2.6.6 Confirmation of pRFP-TALE-SID Construction

1 µg plasmid DNA was digested as follows:

10 x NEB Buffer 4	2 µl
BSA (1 mg/ml)	1 µl
DNA	2 µg
*Nhe*I HF (20 U/µl)	0.5 µl
*Not*I HF (20 U/µl)	0.5 µl
H_2O	to 20 µl

Digest reactions were incubated in a 37°C hotblock for 3 hours. Digest products were resolved using a 1% TBE-agarose gel stained with Safeview. The gel was imaged using a Fujifilm FLA-5000 scanner to confirm if cloning had been successful.

2.7 Golden Gate Assembly of TALE RVDs into SID-Repressor Vectors

2.7.1 Golden Gate TALE Repressor Assembly

TALE RVD pTAL_A, B & C vectors (section 2.5.2) and pRFP-TALE-SID were combined with pLR (last RVD repeat vector) in a Golden Gate reaction to assemble all TALE RVDs together in SID repressor vector (section 2.6). Reactions were set up as follows:

10 x DNA Ligase Buffer	2 µl
pTAL_A, B & C	150 ng each
pLR	150 ng
pRFP-TALE-SID	75 ng
*Esp*3I (10 U/µl)	1 µl
T4 DNA Ligase (5 U/µl)	2 µl
H_2O	to 20µl

Samples were subjected to a simultaneous digestion and ligation reaction using a thermocycler programmed to perform the following temperature cycles:

Temperature	Time	Number of Cycles
37°C	5 minutes	5
16°C	10 minutes	
37°C	15 minutes	1
80°C	5 minutes	1

Golden Gate reactions were then transformed into DH5α competent cells as detailed in section 2.4.9. Transformation reactions were plated on LB Agar plates containing 50 µg/ml ampicillin. Blue/white screening was performed by addition of 100 µl 20 mg/ml X-gal and 100 µl 10mM IPTG to plates and allowing them to dry before plating bacteria. The plates were incubated at 37°C overnight and were then incubated at 4°C overnight to allow blue colour to fully develop.

2.7.2 Confirmation of TALE Assembly in Repressor Vector

Plasmid DNA isolated from white colonies (see section 2.4.10) was screened to confirm insertion of TALE RVDs into pRFP-TALE-SID vector. Restriction digest reactions were carried out as follows:

10 x NEB Buffer 4	4 µl
*Bam*HI HF (20 U/µl)	1 µl
*Stu*I (10 U/µl)	1 µl
DNA	2 µg
H_2O	to 40 µl

Reactions were incubated in a 37°C hotblock for 3 hours. The digest products were resolved on a 1% TBE-agarose gel, stained with Safeview, to confirm if cloning had been successful. Gels were imaged using a Fujifilm FLA-5000 scanner.

2.8 Construction of Advanced Nuclease Expression Vectors for TALENs

2.8.1 Modification of pTALETF_v2 to form pTALETFΔhygro

A portion of pTALETF_v2 had which contained an *Esp*3I site had to be deleted. Restriction digest to remove this vector sequence was carried out as follows:

10 x NEB Buffer 1	2 µl
BSA (1 mg/ml)	1 µl
*Pml*I (20 U/µl)	0.5 µl
DNA	2 µg
H_2O	to 20 µl

Reactions were incubated in a 37°C hotblock for 3 hours and digest products were visualised using 0.8% TBE-agarose gel stained with Safeview. The desired band was visualised using a UV luminometer and excised using a scalpel. The digested plasmid DNA was purified using a QiaQuick Gel Extraction Kit. Extracted DNA fragment was recircularised using blunt ended ligation.

2.8.2 Blunt Ended Ligation

pTALETFΔhygro was recircularised by blunt ended ligation using the following procedure:

10 x T4 DNA Ligase Buffer	2 µl
pTALETFΔhygro	100 ng
T4 DNA Ligase (5 U/µl)	1 µl
H_2O	to 20 µl

The reaction was incubated at 23°C in a thermocycler for 2 hours.

2.8.3 Construction of designed TALEN N-terminus (pTALv2_NT)

TALEN N-terminus sequence was constructed using DNA synthesis and was flanked by *Sac*I and *Not*I restriction sites (pUC57_TALE NT). Restriction digest of 1 µg pTALETF_v2 (backbone) and 2 µg pUC57_TALE NT (insert) were digested using the following set up:

10 x NEB Buffer 4	2 µl
BSA (1 mg/ml)	1 µl
*Not*I HF (20 U/µl)	0.5 µl
*Sac*I HF (20 U/µl)	0.5 µl
DNA	1 or 2 µg
H_2O	to 20 µl

Reactions were incubated in a 37°C hotblock for 3 hours. Digested DNA was run on 0.8% (backbone) or 2% (insert) TBE-agarose gels stained with Safeview and visualised using a UV luminometer. Desired bands were excised using a scalpel and purified by QiaQuick Gel Extraction. Fragments were ligated together following the protocol outlined in section 2.4.8.

2.8.4 Confirmation of designed TALEN N-terminus assembly

Insertion of designed TALE N-terminus was confirmed by performing two restriction enzyme digestions, as follows:

10 x NEB Buffer 4	2 µl
BSA (1 mg/ml)	1 µl
*Not*I HF (20 U/µl)	0.5 µl
*Sac*I HF(20 U/µl)	0.5 µl
DNA	1 µg
H_2O	to 20 µl

10 x NEB Buffer 3	2 µl
BSA (1 mg/ml)	1 µl
*Nco*I (10 U/µl)	0.5 µl
DNA	1 µg
H_2O	to 20 µl

Reactions were incubated at 37°C in a hotblock for 3 hours and digest products were analysed using 1% TBE-agarose gel electrophoresis. Gels were stained with Safeview and imaged using a Fujifilm FLA-5000 scanner.

2.8.5 Construction of designed TALEN C-terminus

Sequences encoding the TALEN C-terminus and wt or mutant codon optimised *Fok*I, flanked by two *Bsr*GI restriction sites, were constructed using DNA synthesis (pUC57_TALE CT WT, pUC57_TALE CT ELD-S and pUC57_TALE CT KKR-S). Restriction digest of 1 µg pTALv2_NT (backbone) and 2 µg C-terminus-FokI (insert) were digested using the following set up:

10 x NEB Buffer 2	2 µl
BSA (1 mg/ml)	1 µl
*Bsr*GI (10 U/µl)	0.5 µl
DNA	1 or 2 µg
H_2O	to 20 µl

Reactions were incubated in a 37°C hotblock for 3 hours. Digested DNA was run on 0.8% (backbone) or 1% (insert) TBE-agarose gels, stained with Safeview, and visualised using a UV luminometer. Desired bands were excised using a scalpel and purified by QiaQuick Gel Extraction. Fragments were ligated together following the protocol outlined in section 2.4.8.

2.8.6 Confirmation of designed TALEN C-terminus assembly

Insertion of designed TALE C-terminus was confirmed by performing two restriction enzyme digestions, as follows:

10 x NEB Buffer 3	2 µl
BSA (1 mg/ml)	1 µl
*Nco*I (10 U/µl)	0.5 µl
DNA	1 µg
H_2O	to 20 µl

10 x NEB Buffer 4	2 µl
*Stu*I (10 U/µl)	0.5 µl
DNA	1 µg
H_2O	to 20 µl

Reactions were incubated at 37°C in a hotblock for 3 hours and digest products were analysed using 1% TBE-agarose gel electrophoresis. Gels were stained with Safeview and imaged using a Fujifilm FLA-5000 scanner.

2.9 Assembly of TALE RVDs into Nuclease Expression Vectors

2.9.1 Golden Gate TALEN Assembly (pTALEN)

TALE RVD pTAL_A, B & C vectors (section 2.5.2) and pTALv1-FokIWT/pTALv2-FokIWT/pTALv3-FokI ELDS or KKRS were combined with pLR (last RVD repeat vector) in a Golden Gate reaction to assemble all TALE RVDs together in a TALEN expression vector (section 2.8). Reactions were set up as follows:

10 x DNA Ligase Buffer	2 µl
pTAL_A, B & C	150 ng each
pLR	150 ng
pTALv1/v2/v3	75 ng
*Esp*3I (10 U/µl)	1 µl
T4 DNA Ligase (5 U/µl)	2 µl
H_2O	to 20µl

Samples were subjected to a simultaneous digestion and ligation reaction using a thermocycler programmed to perform the following temperature cycles:

Temperature	Time	Number of Cycles
37°C	5 minutes	5
16°C	10 minutes	
37°C	15 minutes	1
80°C	5 minutes	1

Golden Gate reactions were then transformed into DH5α competent cells as detailed in section 2.4.9. Transformation reactions were plated on LB Agar plates containing 50 µg/ml ampicillin, 20 mg/ml X-gal and 0.1M IPTG to perform blue/white screening of clones containing inserted RVDs. The plates were incubated at 37°C overnight and were then incubated at 4°C overnight to allow blue colour to fully develop.

2.9.2 Confirmation of TALE Assembly in Nuclease Vectors

Plasmid DNA isolated from white colonies (see section 2.4.10) was screened to confirm insertion of TALE RVDs into pTALv1/v2/v3 vectors. Restriction digest reactions were carried out as follows:

10 x NEB Buffer 4	4 µl
BSA (1 mg/ml)	1 µl
*Sac*I HF (20 U/µl)	1 µl
*Xho*I (20 U/µl)	1 µl
DNA	2 µg
H_2O	to 40 µl

Reactions were incubated in a 37°C hotblock for 3 hours. The digest products were resolved on a 1% TBE-agarose gel, stained with Safeview, to confirm if cloning had been successful. Gels were imaged using a Fujifilm FLA-5000 scanner.

2.10 Construction of pTRGT VEZF1

2.10.1 PCR of hVEZF1 homology arms for pSHTLR cloning

Homology arms were obtained by performing PCR of regions surrounding *VEZF1* gene promoter from genomic DNA isolated from K562 cells. Primers used to amplify homology arms contained att sites for use in Gateway® Gene Cloning (Life Technologies). Homology arms were inserted into pDONR vectors by BP reaction Gateway Cloning (Grainne Barkess). Homology arms were PCR amplified from pDONR vectors.

<u>5' Homology Arm</u>

PCR reaction was set up as follows:

10 x Buffer	5 µl
10 mM dNTPs	4 µl
10 µM F Primer VEZF1_5'arm_H3F	1 µl
10 µM R Primer VEZF1_5'arm_AscR	1 µl
pDONR_VEZF1 5'arm	10 ng
Pfu Polymerase (2.5 U/µl)	0.5 µl
H_2O	to 50 µl

PCR was performed using a thermocycler programmed to perform the following temperature cycles:

Temperature	Time	Number of Cycles
95°C	2 minutes	1
95°C	30 seconds	33
55°C	30 seconds	
72°C	2.5 minutes	
72°C	5 minutes	1

PCR products were resolved using a 1% TBE-agarose gel stained with Safeview. The desired band was visualised using a UV luminometer and was excised using a scalpel. The PCR product was purified using a QiaQuick Gel Extraction Kit.

3' Homology Arm

PCR reaction was set up as follows:

10 x Buffer	5 µl
DMSO	2.5 µl
10 mM dNTPs	1 µl
10 µM F Primer VEZF1_3'arm_H3F	1 µl
10 µM R Primer VEZF1_3'arm_AscR	1 µl
pDONR_VEZF1 5'arm	10 ng
Pfu Polymerase (2.5 U/µl)	0.5 µl
H_2O	to 50 µl

PCR was performed using a thermocycler programmed to perform the following temperature cycles:

Temperature	Time	Number of Cycles
95°C	2 minutes	1
95°C	30 seconds	33
60°C	30 seconds	
72°C	2.5 minutes	
72°C	5 minutes	1

PCR products were resolved using a 1% TBE-agarose gel stained with Safeview. The desired band was visualised using a UV luminometer and was excised using a scalpel. The PCR product was purified using a QiaQuick Gel Extraction Kit.

2.10.2 Restriction digestion of hVEZF1 homology arms for pSHTLR cloning

1 µg purified PCR product (insert) and 1 µg of either pSHTLR3 or pSHTLR5 (backbones) were digested as follows:

10 x NEB Buffer 4	4 µl
*Asc*I (10 U/µl)	1 µl
*Hind*III HF (20 U/µl)	1 µl
DNA	1 µg
H_2O	to 40 µl

Digest reactions were incubated at 37°C for 3 hours in a thermocycler. 1µl CIP was added to digestion of pSHTLR vectors and incubated at 37°C for a further 10 minutes. pSHTLR digest products were resolved using a 0.8% TBE-agarose gel stained pre-run with Safeview. The desired bands were visualised using a UV luminometer and was excised using a scalpel. The digested plasmid DNA was purified using a QiaQuick Gel Extraction Kit. Digested homology arm PCRs were purified using a QiaQuick PCR Purification Kit.

5' VEZF1 homology arm was cloned into pSHTLR3 and 3' VEZF1 homology arm was cloned into pSHTLR5. Fragments were ligated together following the procedure described in section 2.4.8.

2.10.3 Confirmation of pSHTLR-Homology arm cloning

1µg plasmid DNA was digested as follows:

10x NEB Buffer 4	4 µl
*Asc*I (10 U/µl)	1 µl
*Hind*III HF (20 U/µl)	1 µl
DNA	1 µg
H_2O	to 40 µl

Reactions were incubated at 37°C for 3 hours. The digest products were resolved on a 1% TBE-agarose gel to confirm if cloning had been successful. The gel was stained prior to running with 2µl Safeview per 100 ml TBE buffer. Gels were imaged using a Fujifilm FLA-5000 scanner.

2.10.4 Assembly of pTRGT_VEZF1

Gateway Clonase LR reaction to combine pSHTLR5, pSHTLR3 and pSHTLRmid was set up as follows:

pSHTLR_5'homology arm	10 fmol
pSHTLR_3'homology arm	10 fmol
pSHTLRmid	10 fmol
pDESTR4-R3	20 fmol
H_2O	to 8 µl
LR Clonase II	2 µl

Reactions were incubated overnight at 25°C in a thermocycler. The clonase reaction was stopped by addition of 1 µl Proteinase K and incubating at 37°C for 10 minutes.

2.10.5 Confirmation of pTRGT_VEZF1 assembly

1µg plasmid DNA was digested as follows:

10 x NEB Buffer 4	4 µl
*Kpn*I HF (20 U/µl)	1 µl
DNA	1 µg
H_2O	to 40 µl

Reactions were incubated at 37°C for 3 hours. The digest products were resolved on a 1% TBE-agarose gel to confirm if cloning had been successful. The gel was stained prior to running with 2µl Safeview per 100 ml TBE buffer. Gels were imaged using a Fujifilm FLA-5000 scanner.

2.11 Cell Culture

2.11.1 Culturing Cell Lines

Cell lines were cultured with 5% CO_2 at 37°C in vented flasks or dishes.

<u>Human Embryonic Kidney Cell Line – HEK293 T/17 Cells</u>

HEK293 T/17 cells were cultured in 90% DMEM-Glutamax, 10% FBS, 1% Penicillin-Streptomycin solution. Cultures were passaged once cell density reached ~80% at a ratio of 1/4 to 1/5.

Human Erythroleukemia Cell Line – K562 Cells

K562 cells were cultured in 90% RPMI 1640 + GlutaMAX™, 10% FBS, 1% Penicillin-Streptomycin solution. Cultures were passaged three times weekly, once confluency was around 70-80% at a ratio of 1/5.

2.11.2 Passaging Cell Lines

Human Embryonic Kidney Cell Line – HEK293 T/17 Cells

Media was removed from the flask by pipetting and the cells were rinsed with PBS. The cells were dislodged from the flask by addition of trypsin and incubation at 37°C for 3 minutes. Trypsin was deactivated by addition of 10 volumes of fresh media and the cells carried over into a falcon tube by pipetting. Cells were pelleted by centrifugation at 1100 rpm for 5 minutes. The supernatant was removed before the cell pellet was resuspended in fresh media and used to seed a new flask.

Human Erythroleukemia Cell Line – K562 Cells

Cells were removed from the flask by pipetting and transferred to a falcon tube. The cells were collected by centrifugation at 1100 rpm for 5 minutes and the supernatant removed before resuspending the pellet in fresh media and using this to seed a new culture.

2.11.3 Cell Cryopreservation

Confluent cell cultures were collected by centrifugation at 1100 rpm for 5 minutes and cell pellets resuspended in 90% FBS, 10% DMSO in aliquots containing ~$1x10^7$ cells. Cells were gradually frozen in CoolCell® (Biocision) chambers in -80°C freezers overnight before being transferred to liquid nitrogen vapour phase for long term storage. Cells were thawed rapidly at 37°C and added to fresh media where they were pelleted by centrifugation at 1100 rpm for 5 minutes. The pellet was then resuspended in 10 ml fresh media and used to seed a 25 cm^2 flask.

2.11.4 TALE Repressor Transfections

The day before transfection, $7.5x10^6$ HEK293 T/17 cells per well were seeded in a T75 flask with 16ml media which did not contain antibiotics. K562 cells to be transfected were counted and $7.5x10^6$ cells seeded on the day of transfection. Transfections were performed using Lipofectamine 2000 reagent. 15.6 µg plasmid DNA was combined in a

total volume of 2 ml in Optimem. 78 µl Lipofectamine 2000 was combined with 1872 µl Optimem and incubated at room temperature for 5 minutes. The DNA/Optimem and Lipofectamine/Optimem mixtures were then combined and incubated at room temperature for 20 minutes before being added dropwise to cells. The cells were incubated for three days before cell pellets were collected for analysis.

2.11.5 ZFN & TALEN Transfections

The day before transfection, $1x10^6$ HEK293 T/17 cells per well were seeded in a 6 well plate with 2ml media which did not contain antibiotics. K562 cells to be transfected were counted and $1x10^6$ cells seeded on the day of transfection. Transfections were performed using Lipofectamine 2000 reagent. In total, 4 µg DNA was transfected: 2µg each ZFN or TALEN expression vector or, in the case of HDR experiments involving the targeting vector, a targeting vector/nuclease expression vector ratio of 30/1 was used (3874 ng pTRGT_VEZF1 + 63 ng each ZFN or TALEN) (Connelly et al., 2010). Plasmid DNA was combined in a total volume of 250 µl in Optimem. 10 µl Lipofectamine 2000 was combined with 240 µl Optimem and incubated at room temperature for 5 minutes. The DNA/Optimem and Lipofectamine/Optimem mixtures were treated as outlined in section 2.11.4.

2.11.6 FACS Analysis

Approximately $1x10^6$ cells were collected by centrifugation at 1000 rpm for 5 minutes. The cell pellet was resupened in 500µl fresh culture medium and FACS analysis was performed using an Attune Acoustic Focusing Cell Cytometer (Life Technologies). Parameters were optimised for GFP and RFP signals from HEK293 T/17 cells. Fluorochromes were excited by a laser at 488 nm. Voltage parameters were used as follows: RFP detection – FSC 2100 mV, SSC 2650 mV, BL1 1100 mV, BL2 1500 mV, BL3 2550 mV; GFP detection – FSC 2100 mV, SSC 2650 mV, BL1 1100 mV, BL2 2200 mV, BL3 2550 mV. Data from 20,000 cells was gathered to estimate the percentage of GFP+ cells within the population. The data collected was gated for analysis of live cells. Fluorescence intensity of GFP+ or RFP+ cells was compared to that of wild type cells using Attune Cytometric Software (Life Technologies).

2.12 Protein Isolation & Analysis

2.12.1 Preparation of Whole Cell Extracts

Approximately $5x10^7$ cells were collected by centrifugation at 1000 rpm for 5 minutes. The supernatant was removed and the pellet washed with PBS. Cell pellets were resuspended in 350 µl cell lysis buffer and pellets dislodged by vortexing. The samples were freeze-thawed twice and then sonicated 5 x 5 seconds, until they became runny and clear. The samples were centrifuged at 14,000 rpm for 5.5 minutes to separate cellular debris from protein samples. Supernatant was removed and stored at -20°C.

2.12.2 Protein Quantification

Protein samples were quantified using NanoOrange Protein Quantification Kit. A standard curve was prepared using BSA dilutions following the protocol in the kit. Samples were prepared following the standard protocol and compared to the standard curve using a Stratagene Mx3000P thermocycler.

2.12.3 Detection of Target Proteins using Western Blotting

A 4 – 12 % Bis-Tris precast gel cassette (Invitrogen) was placed into a running tank and the tank was filled with MOPS SDS Running Buffer (1x). The protein samples were prepared by addition of 4x LDS sample buffer and 10x sample reducing agent (to final concentration of 1x) and were heated at 70°C for 10 minutes before being loaded into the wells. Proteins were resolved by running the gels at 200 V for 45 minutes.

1x Transfer Buffer was prepared as outlined in section 2.3.2. Blotting pads and filter paper were soaked in 1x Transfer buffer and a PVDF membrane was prepared by soaking in 100% methanol, followed by soaking in distilled H_2O and finally soaked in 1x transfer buffer.

The gel was removed from the cassette after running and the wells and foot were removed. The transfer stack was prepared and secured in the gel tank. The inner chamber of the gel tank (containing the transfer stack) was immersed in 1x transfer buffer, while the outer chamber was filled with distilled H_2O. The transfer was carried out

at 30V for 90 minutes. Verification of transfer was performed by Ponceau staining of the membrane. The membrane was then destained by several washes in distilled H_2O.

The PVDF membrane was blocked by shaking for 2x 30 minutes at room temperature in 1x blocking buffer. The membrane was then incubated overnight in a sealed Kapak bag, rocking gently at 4°C with primary antibody, diluted 1/2000 in blocking buffer. The membrane was washed four times for 15 minutes in 1xTBST whilst rocking. The membrane was then incubated with secondary antibody, diluted 1/2000 in blocking buffer, whilst rocking, for 2 hours at room temperature and was then washed again as before. SuperSignal West Dura Extended Duration Substrate (Pierce) ECL detection reagent was prepared by mixing the luminal enhancer and peroxide solutions in equal proportions. This solution was then applied directly to the PVDF membrane to visualise the proteins. The membrane was incubated for 5 minutes and the resulting chemiluminescent signal was detected using a CCD camera imaging system (Fuji).

2.13 RNA Isolation & Gene Expression Analysis

2.13.1 RNA Extraction

Approximately 7.5 x 10^6 cells were harvested by centrifugation and the supernatant removed from the cell pellet. The pellet was then resuspended in 1 ml Trizol and incubated at room temperature for 5 – 10 minutes. 100 µl BCP (1-bromo-3-chloropropane) was added and the reactions were shaken vigorously and then incubated at room temperature for 15 minutes. The reactions were then centrifuged at 12,000 g for 15 minutes at 4°C to separate RNA, DNA and cellular proteins. After centrifugation, the supernatant was transferred to a fresh tube. 500 µl of ispopropanol was added to each new tube and the reactions were vortexed before incubation at room temperature for 10 minutes and then at -20°C for 20 minutes. RNA was collected by centrifugation at 12,000g for 15 minutes at 4°C and the supernatant was removed. The RNA pellet was washed using 1 ml of 75% ethanol and then pelleted again at 7,500g for 5 minutes at room temperature. The supernatant was again discarded and the wash repeated as before. Following the second wash, the pellet was air dried before being resuspended in 50µl molecular grade H_2O.

2.13.2 Assessment of RNA integrity

RNA quality was determined by running 2 µg of each RNA preparation on 1% TBE-agarose gel. Gels were stained post-run using 1 µg/ml ethidium bromide and destained using TBE buffer. Gels were then imaged using a Fujifilm FLA-5000 scanner.

2.13.3 cDNA Synthesis

cDNA synthesis was carried out by combining 400 ng RNA in the following reaction:

RNA (100 ng/µl)	4 µl
dNTPs (10 mM)	0.5 µl
Random Primers (50 ng/µl)	1 µl
H_2O	1 µl

Reactions were incubated at 65°C for 5 minutes and then placed on ice for 2 minutes before addition of the following reagents:

5X First Strand Synthesis (FS) Buffer	2 µl
RNase OUT	0.5 µl
DTT	0.5 µl
SuperScript III Reverse Transcriptase	0.5 µl

Reactions were subjected to the following temperature conditions in a thermocycer:

Temperature	Time
25°C	10 minutes
50°C	50 minutes
85°C	5 minutes

Template RNA was digested by addition of 1 µl RNase H. Reactions were incubated in a 37°C hotblock for 20 minutes.

2.13.4 qRT-PCR

Gene expression analysis was performed using FastStart Universal SYBR Green. qRT-PCR reactions were performed in triplicate using previously validated primer sets which are widely used in the lab. The reactions were carried out as follows:

2X SYBR Green Master Mix	10 µl
F & R Primer Mix (4 µM each primer)	4 µl
cDNA	4 µl
H_2O	1 µl

Real time PCR reactions were carried out on a Stratagene Mx3000P thermocycler programmed to perform the following temperature cycles:

Temperature	Time	Number of Cycles
95°C	10 minutes	1
95°C	15 seconds	40
60°C	30 seconds	

A dissociation curve was performed upon completion of each real-time PCR in order to measure the dissociation temperature of PCR products between 60 °C – 95 °C, and thus detect primer dimers.

Ct values were generated using the MxPro version 4.0 software.

2.13.5 qRT-PCR Analysis

Average Ct values were calculated for each sample, which was analysed in triplicate. Error values were expressed as a standard deviation from the mean. Expression levels of *VEZF1* were analysed by normalising to the expression of housekeeping gene, β-actin. This generated a ΔCt value:

$$\Delta Ct = Ct \text{ (gene of interest)} - Ct \text{ (housekeeping gene)}$$

Changes in gene expression were detected by calculating a ΔΔCt value using the following equation:

$$\Delta\Delta Ct = \Delta Ct \text{ (treated sample)} - \Delta Ct \text{ (WT sample)}$$

Fold change in gene expression was calculated as follows:

$$\text{Fold change} = N^{(-\Delta\Delta Ct)}$$

Where ‘N’ = PCR amplification efficiency (typically = 2)

2.14 NHEJ Assays

2.14.1 *CelI*/T7 EndonucleaseI Control Assays

2.14.1.1 Preparation of Control Assay Substrate DNA

PCR of Control DNA

PCR was performed from plasmids pFOKI-EL and pFOKI-ELS. Reactions were set up as follows:

10 x PCR Buffer	5 µl
50 mM $MgCl_2$	1.5 µl
10 mM dNTPs	1 µl
10 µM F Primer FokI_CelI_Cont_F	1 µl
10 µM R Primer FokI_CelI_Cont_R	1 µl
DNA	100 ng
Platinum Taq (5 U/µl)	0.2 µl
H_2O	to 50 µl

PCR was performed using a thermocycler programmed to perform the following temperature cycles:

Temperature	Time	Number of Cycles
94°C	1 minute	1
94°C	30 seconds	30
47°C	30 seconds	
72°C	1 minute	
72°C	5 minutes	1

PCR products were resolved using a 1% TBE-agarose gel stained with Safeview. Desired bands were visualised using a UV luminometer and excised using a scalpel. The digested plasmid DNA was purified using a QiaQuick Gel Extraction Kit.

Denature & re-annealing of control DNA

Purified PCR products from pFOKI-EL and pFOKI-ELS were mixed in predefined ratios to create mismatched DNA. This was added to the following reagents:

10 x Maximase Polymerase Buffer	5 µl
25 mM $MgSO_4$	4 µl
10 mM dNTPs	1 µl
DNA	2 µg
H_2O	to 50 µl

Reactions were incubated in a thermocycler programmed to perform the following temperature cycles:

Temperature	Time
90°C	3 minutes
65°C	10 minutes
37°C	10 minutes
22°C	10 minutes

2.14.1.2 Digestion of control DNA with *Cel*I

Re-annealed DNA was digested in the following reaction:

10 x Maximase Polymerase Buffer	2.5 µl
150 mM $MgCl_2$	1.25 µl
Enhancer	1 µl
DNA	up to 400 ng
*Cel*I	1 µl
H_2O	to 25 µl

Digest reactions were incubated in a thermocycler at 42°C and 2.5 µl Stop Solution was added to each sample following incubation. Results of control assay were analysed immediately by agarose gel electrophoresis of the samples using 2% TBE-agarose gels stained with Safeview. Gels were then imaged using a Fujifilm FLA-5000 scanner.

2.14.1.3 Digestion of control DNA with T7 EndonucleaseI

Re-annealed DNA was digested in the following reaction:

10 x NEB Buffer 2	2.5 µl
BSA (1 mg/ml)	1 µl
DNA	up to 400 ng
T7EI (10 U/µl)	1 µl
H_2O	to 25 µl

Digest reactions were incubated in a thermocycler at 37°C and 2.5 µl 250 mM EDTA was added to each sample following incubation. Results of control assay were analysed immediately by agarose gel electrophoresis of the samples using 2% TBE-agarose gels stained with Safeview. Gels were then imaged using a Fujifilm FLA-5000 scanner.

2.14.2 RFLP Assays

Genomic DNA was isolated from HEK 293 and K562 cells which had been transfected with vectors expressing TALENs. DNA extraction was performed using the standard protocol from the DNeasy Blood & Tissue Kit.

A 1310 bp region of the *VEZF1* promoter was PCR amplified using the following set up:

5 x GC Buffer	10 µl
GC Enhancer	5 µl
10 mM dNTPs	1 µl
10 µM F Primer hVEZF1pro-374F	2 µl
10 µM R Primer hVEZF1pro+895R	2 µl
DNA	100 ng
One Taq (20 U/µl)	0.25 µl
H_2O	to 50 µl

PCR reactions were incubated in a thermocycler programmed to perform the following temperature cycles:

Temperature	Time	Number of Cycles
94°C	30 seconds	1
94°C	30 seconds	30
56°C	30 seconds	
68°C	1.5 minutes	
68°C	5 minutes	1

An aliquot of each PCR reaction was visualised using 1% TBE-agarose gel stained with Safeview and imaged using a Fujifilm FLA-5000 scanner. The remainder of the PCR reaction was purified using the standard protocol of the MinElute PCR purification kit.

2.14.2.1 RFLP Assay 1 Digestion

Purified PCR product was subject to restriction digest as follows:

10 x NEB Buffer 4	1 µl
BSA (1 mg/ml)	0.5 µl
DNA	250 ng
*Bsr*BI (10 U/µl)	0.5 µl
*Bst*EII HF (20 U/µl)	0.5 µl
H_2O	to 10 µl

Digest reactions were incubated in a thermocycler at 37°C for 3 hours and the enzymes were heat inactivated by incubation at 80°C for 20 minutes.

2.14.2.2 RFLP Assay 2 Digestion

Purified PCR product was subject to restriction digest as follows:

10 x NEB Buffer 4	1 µl
BSA (1 mg/ml)	0.5 µl
DNA	250 ng
*Bst*EII HF (20 U/µl)	0.5 µl
*Rsa*I (10 U/µl)	0.5 µl
H_2O	to 10 µl

Digest reactions were incubated in a thermocycler at 37°C for 3 hours and the enzymes were heat inactivated by incubation at 80°C for 20 minutes.

2.14.2.3 RFLP Assay 3 Digestion

Purified PCR product was subject to restriction digest as follows:

10 x NEB Buffer 4	1 µl
BSA (1 mg/ml)	0.5 µl
DNA	250 ng
*Msp*I (20 U/µl)	0.5 µl
*Sac*I HF(20 U/µl)	0.5 µl
H_2O	to 10 µl

Digest reactions were incubated in a thermocycler at 37°C for 3 hours and the enzymes were heat inactivated by incubation at 80°C for 20 minutes.

2.14.2.4 DNA Bioanalyser analysis of RFLP fragments

DNA bioanalyser analysis was carried out using a DNA-1000 kit for the Agilent 2100 Bioanalyser by Julie Galbraith, Glasgow Polyomics.

2.14.2.5 Agarose gel electrophoresis analysis of RFLP fragments

Digested DNA samples were analysed by agarose gel electrophoresis. Half of each RFLP digest (5 µl) was run on TBE-agarose gels. RFLP assays 1, 2 and 3 were analysed on 2%, 1.5% and 2.5% agarose gels respectively. Samples were loaded in an equal volume (5 µl) gel loading buffer to prevent sample loss. Gels were stained post-run with 1 µg/ml ethidium bromide and destained using TBE buffer. Gels were then imaged using a Fujifilm FLA-5000 scanner.

2.15 Sequence Analysis of TALEN-treated DNA

2.15.1 PCR Amplification of TALEN-treated DNA

Genomic DNA from both WT HEK 293 cells and HEK 293 treated with TALEN 15/16 v2 was PCR amplified using the following protocol:

5 x GC Buffer	10 µl
GC Enhancer	5 µl
10 mM dNTPs	1 µl
10 µM F Primer hVEZF1pro-374F	2 µl
10 µM R Primer hVEZF1pro+895R	2 µl
DNA	100 ng
One Taq (20 U/µl)	0.25 µl
H_2O	to 50 µl

PCR reactions were incubated in a thermocycler programmed to perform the following temperature cycles:

Temperature	Time	Number of Cycles
94°C	30 seconds	1
94°C	30 seconds	30
56°C	30 seconds	
68°C	1.5 minutes	
68°C	5 minutes	1

An aliquot of each PCR reaction was visualised using 1% TBE-agarose gel stained with Safeview and imaged using a Fujifilm FLA-5000 scanner. The remainder of the PCR reaction was purified using the standard protocol of the MinElute PCR purification kit.

2.15.2 A-tailing of PCR products

A-tailing reactions were set up as detailed below:

10 x ThermoPol Buffer	1 µl
50 mM $MgCl_2$	1 µl
10 mM dATP	1 µl
DNA	218 ng
NEB Taq Polymerase (5 U/µl)	1 µl
H_2O	to 10 µl

Reactions were incubated in a hotblock at 70°C for 30 minutes. Amount of PCR product DNA required for ligation was calculated using the equation in section 2.4.8 and DNA was used directly from this reaction. Ligation reactions were set up as detailed in section 2.4.8.

2.15.3 Confirmation of insertion of TALEN-treated DNA into pGEM-T Easy

Ligations were transformed as detailed in section 2.4.9 and transformation reactions were plated on LB Agar plates containing 50 µg/ml ampicillin. Blue/white screening was performed by addition of 100 µl 20 mg/ml X-gal and 100 µl 10mM IPTG to plates and allowing them to dry before plating bacteria. The plates were incubated at 37°C overnight and were then incubated at 4°C overnight to allow blue colour to fully develop.

Several white colonies were analysed for the presence of inserted TALEN-treated DNA by colony PCR. Individual colonies were responded in 50 µl H_2O and a 5 µl aliquot was taken into the following PCR reaction:

10 x ThermoPol Buffer	2.5 µl
10 mM dNTPs	1 µl
10 µM F Primer M13F	0.5 µl
10 µM R Primer M13R	0.5 µl
Colony suspension	5 µl
NEB Taq Polymerase (5 U/µl)	0.2 µl
H_2O	to 25 µl

PCR was performed using a thermocycler programmed to perform the following temperature cycles:

Temperature	Time	Number of Cycles
94°C	2 minutes	1
94°C	30 seconds	34
55°C	30 seconds	
72°C	1 minute	
72°C	10 minutes	1

PCR products were resolved using a 1% TBE-agarose gel stained with Safeview. Gels were imaged using a Fujifilm FLA-5000 scanner. Amplification of inserted genomic PCR product was difficult due to the GC rich and complex nature of the inserted genomic DNA sequence. Therefore, clones which did not result in production of a PCR amplicon were further screened by restriction enzyme digest of plasmid DNA.

Colony suspensions from potentially correct clones were cultured overnight by inoculating the remaining suspension into 2 ml LB containing 50 µg/ml ampicillin and shaking at 37°C for 16 hours. Plasmid DNA was isolated following the procedure outlined in section 2.4.10.

Plasmid DNA of potentially positive clones was screened by restriction enzyme digest as detailed below:

10 x NEB CutSmart Buffer	2.5 µl
DNA	800 ng
*Not*I HF (20 U/µl)	0.5 µl
*Rsa*I (10 U/µl)	1 µl
H_2O	to 25 µl

Digest reactions were incubated in a hotblock at 37°C for 3 hours. Digest products were analysed using 1% TBE-agarose gel electrophoresis. Gels were stained with Safeview and imaged using a Fujifilm FLA-5000 scanner.

2.15.4 Sequence Analysis of TALEN-treated DNA

Plasmid DNA samples were sequenced by Source Bioscience using M13F and M13R primers. Five clones containing WT HEK 293 DNA were analysed and compared to generate a consensus WT sequence with which to compare TALEN-treated DNA samples. Sequencing of clones containing TALEN-treated DNA was performed in duplicate and each run analysed comparatively to create a consensus sequence for that sample. Sequence analyses from all TALEN-treated clones were compared to WT consensus and mutations identified by alignments using MultAlin online software (http://multalin.toulouse.inra.fr/multalin/).

Chapter 3

Zinc Finger Nucleases to Modify the *VEZF1* Gene Promoter

3.1 Introduction

VEZF1 is a highly conserved vertebrate transcription factor that is essential in mice and is reported to have roles in vascular and haematopoietic development (Kuhnert et al., 2005, Zou et al., 2010). Recent ChIP-seq analyses in human cells have found that VEZF1 is bound at the promoters of most transcriptionally active genes in addition to many cell type-specific enhancer elements (Low, 2013). RNA interference has been used to knockdown VEZF1 expression as an approach to studying VEZF1 function. However, this approach only leads to modest reductions in VEZF1 binding *in vivo* (Strogantsev, 2009). It is therefore desirable to knock out the *VEZF1* gene in human somatic cells to study VEZF1's gene regulatory functions.

Gene targeting in somatic cell types is a very inefficient process. The creation of a double stranded DNA break (DSB) at a target locus promotes gene targeting via the homology directed DSB repair pathway (Brenneman et al., 1996, Choulika et al., 1995, Sargent et al., 1997). Zinc finger nucleases (ZFNs) are engineered chimeric proteins that fuse zinc finger DNA binding domains with an endonuclease domain. The choice of pre-characterised zinc fingers of known DNA sequence specificity allows for the creation of endonucleases that can target a chosen genetic locus. In this chapter, I describe the design and assembly of custom zinc finger nucleases (ZFNs) that target the human *VEZF1* gene promoter region and exon 1. Designed ZFNs were assembled using the Modular Assembly Kit (Wright et al., 2006), which was available at the time of starting this project. Purchasing an assembly kit was considered more cost effective, as it allowed us to construct multiple ZFN pairs and provided the option to create reagents for other genes of interest in our lab. The cost of purchasing a pair of pre-assembled and validated ZFNs is now around £4500 (http://www.sigmaaldrich.com/catalog/product/sigma/cstzfnd?lang=en&region=GB) but this price was much higher when beginning this study. Assembled ZFNs will be subsequently used to promote gene targeting (Chapters 5 & 6).

3.2 Aims of this chapter

- Design custom ZFNs that target the *VEZF1* promoter
- Construct ZFNs using the Zinc Finger Consortium Modular Assembly Kit
- Develop expression vectors that produce obligate heterodimer ZFNs with enhanced cleavage properties

3.3 Design of Custom Zinc Finger Nucleases

3.3.1 Zinc Finger Nuclease Design using Zinc Finger Targeter (ZiFiT)

ZiFiT is a web based tool (www.zifit.partners.org/ZiFiT_v3.3) which enables users to identify characterised zinc finger modules that can be used in ZFN designs for any given target sequence (Sander et al., 2007). Created in collaboration with the Zinc Finger Consortium (www.zincfingers.org), it includes a database of the properties of the modular assembly zinc fingers (Wright et al., 2006). ZiFiT searches can be restricted to suggest zinc finger proteins (ZFPs) from only one or all three module sets.

A 400 bp DNA sequence covering 200 bp either side of the transcription start site of the human *VEZF1* gene was searched using ZiFiT (Sander et al., 2007) to design two pairs of three-fingered ZFNs. I decided to utilise Sangamo modules that recognise GNN triplets, as these were the most well-characterised for use at the ends of, and within, zinc finger arrays (Section 1.3.3.1.2, (Liu et al., 2002)). During the construction of these ZF arrays, a study identified a number of zinc finger modules that consistently performed well in ZFNs (Section 1.3.3.1.4, (Kim et al., 2009)). A pair of four-fingered ZFNs was designed from this recommended module set. ZFN design details are outlined in Table 3.1. The *VEZF1* promoter is largely uncharacterised and promoter regulatory elements have not yet been identified. The three ZFN pairs described above were chosen as they recognise different locations around the transcription start site and 5' untranslated region (UTR) of the *VEZF1* gene. Different target locations were chosen to account for differences in sequence accessibility. The ZFN pairs were chosen as they bind the VEZF1 promoter between two well positioned nucleosomes in a nucleosome depleted region (NDR, Figure 3.1).

Design name	ZF module numbers	DNA recognition residues	DNA recognition sequence	DNA binding configuration
ZFN_35 **ZFN_36**	07-23-42 06-30-50	RDDR-RDDR-RDDR EGTR-DSAR-QSAR	GCG-GCG-GCG GCC-GTC-GTT	3 2 1 CGCCGCCGCCGCTTT**GTTGTCGCC** **GCGGCGGCG**GCGAAACAACAGCGG 1 2 3
ZFN_37 **ZFN_38**	11-26-42 11-23-36	RDHR-DSHR-RDDR RDHR-RDDR-QGNR	GGG-GGC-GCG GGG-GCG-GAA	3 2 1 CCCGCCCGCCCCGGC**GAAGCGGGG** **GGGCGGGCG**GGGCCGCTTCGCCCC 1 2 3
ZFN_43 **ZFN_44**	098-091-132-078 108-139-108-108	HGHE-RDNE-RDHT-SADR DSCR-VSTR-DSCR-DSCR	CGC-CAG-YGG-ACA GCC-GCT-GCC-GCC	4 3 2 1 GCGCTGCCATGTTGAGGA**GCCGCCGCTGCC** **CGCGACGGTACA**ACTCCTCGGCGGCGACGG 1 2 3 4

Table 3.1 Zinc Finger Nuclease Designs
ZFN designs using ZiFiT. The module reference numbers for the modular assembly kit are indicated for each ZF array (Wright et al., 2006). The amino acids at α-helix positions -1, 2, 4 & 6 that recognise DNA bases are shown for each ZF module. The left ZFN recognises the bottom DNA strand and the right binds to the top strand. ZF modules are numbered with the first finger being furthest away from the cleavage domain. The DNA recognition sequence and binding configuration of each finger module to its target site are also shown.

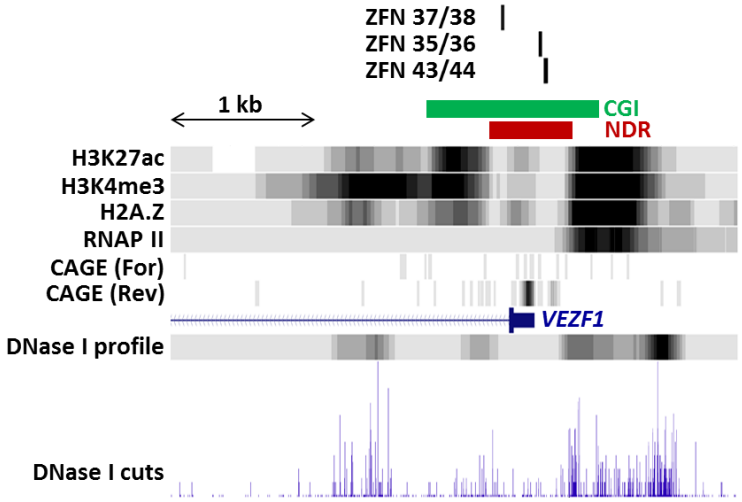

Figure 3.1 ZFN sites locate within a nucleosome depleted region of the *VEZF1* promoter
UCSC Genome browser view of Human hg19 chr17:56063000-56067000. The Refseq annotation of the *VEZF1* gene is shown in blue, with Exon 1 depicted as a wide bar, preceded by a 5'UTR. Intron 1 is depicted by a line with chevrons indicating the direction of the gene. A 1221 bp CpG island is depicted by a green bar. CAGE-seq data from K562 cells (FANTOM5:DRR008730) indicates that *VEZF1* is a unidirectional promoter, with the K562 TSS located within the annotated 5'UTR. ENCODE ChIP-seq analysis of H3K27ac (wgEncodeEH000043), H3K4me3 (wgEncodeEH000048) and H2A.Z (wgEncodeEH001038) indicate that highly positioned nucleosomes enriched for these modifications flank a ~560 bp nucleosome depleted region (NDR, red bar). The cleavage profile and individual cuts of DNAse I (wgEncodeEH000480) are shown beneath, indicating that DNAse I cuts either side of the positioned nucleosomes. The locations of the ZFN recognition sites are shown above (black bars).

3.4 Construction of Designed Zinc Finger Expression Sequences

3.4.1 Modular Assembly of Designed Zinc Fingers

Complete zinc finger expression sequences were constructed via modular assembly (Wright et al., 2006). Each zinc finger module is joined onto an array individually, starting with the addition of ZF2 onto ZF1, followed by the addition of ZF3 and ZF4 (Figure 3.2). Each ZF module plasmid is flanked by a common set of restriction enzyme recognition sequences that permit the sub-cloning of ZF modules in a manner that retains the open reading frame of a ZF array. The receiving plasmid, for ZF1 for example, is opened using *Age*I and *Bam*HI. The donating insert, ZF2 for example, is excised using *Xma*I and *Bam*HI. The compatible CCGG 5' overhangs of *Age*I and *Xma*I permit ligation of ZF2 onto ZF1. The insert fragment carries an *Age*I recognition site, which is used in repeat rounds to add ZF3, ZF4 and so on. The addition of each zinc finger module adds 96 bp to the receiving vector. Analysis of the *Xba*I – *Bam*HI fragment size is used to confirm the inclusion of each ZF module after each assembly step. Plasmids containing 1, 2, 3 or 4 ZF modules yield a 96, 192, 288 or 384 bp *Xba*I – *Bam*HI fragment, respectively (Figure 3.3). The completed zinc finger arrays 35, 36, 37, 38, 43 and 44 (Table 3.1) were checked by Sanger sequencing (Appendix II) prior to sub-cloning into ZFN expression vectors, as described below.

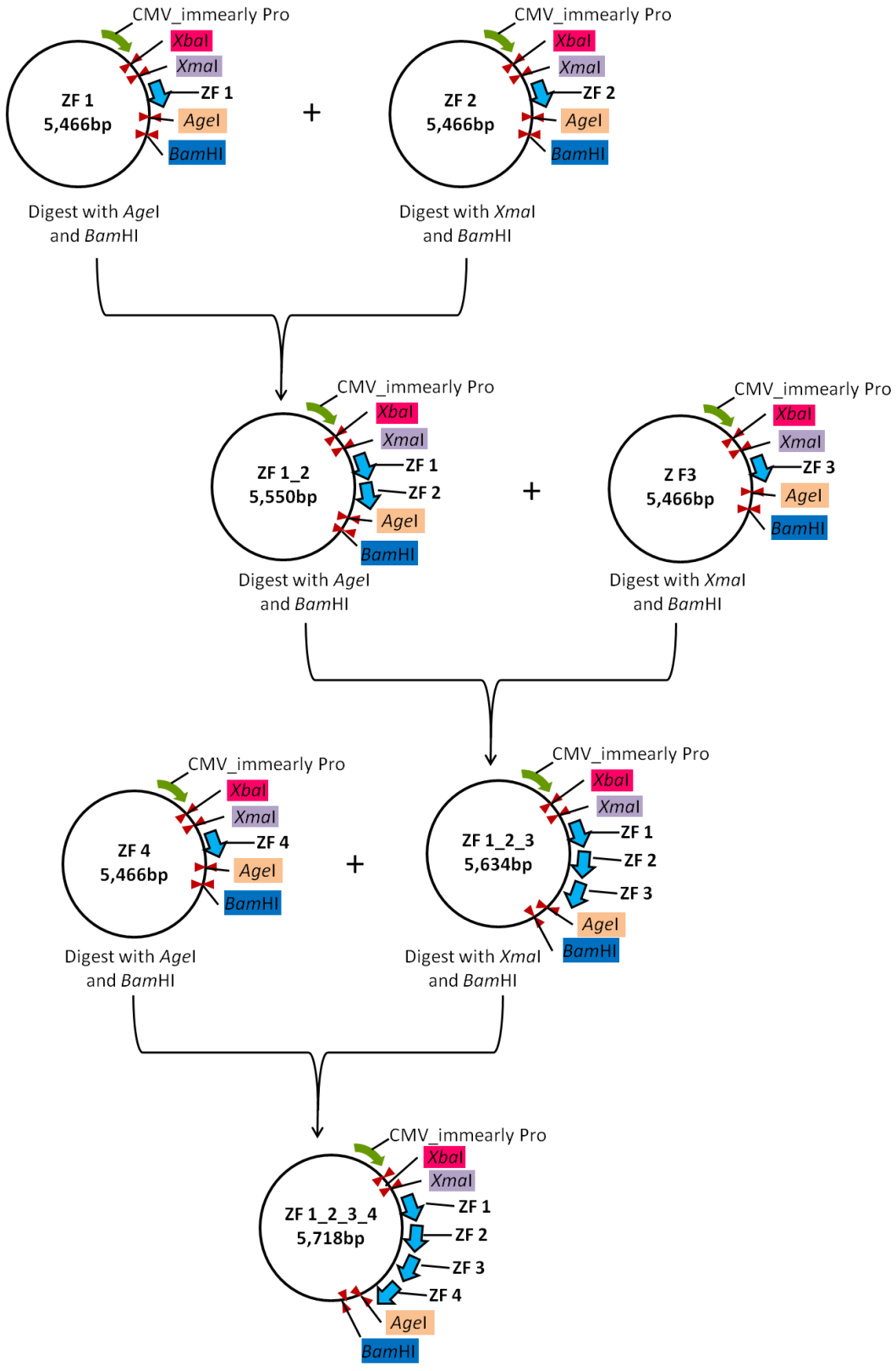

Figure 3.2 Construction of Zinc Finger Arrays by Modular Assembly
Schematic showing the stepwise assembly of four finger array plasmid. Receiving plasmids are opening using *Age*I and *Bam*HI. Donating fragments are excised using *Xma*I and *Bam*HI. Compatible *Age*I and *Xma*I overhangs allow for assembly. The carry over of a new *Age*I recognition site with each insert allows for repeated round of assembly. *Xba*I and *Bam*HI are used to excise the whole ZF array for size analysis by agarose gel electrophoresis.

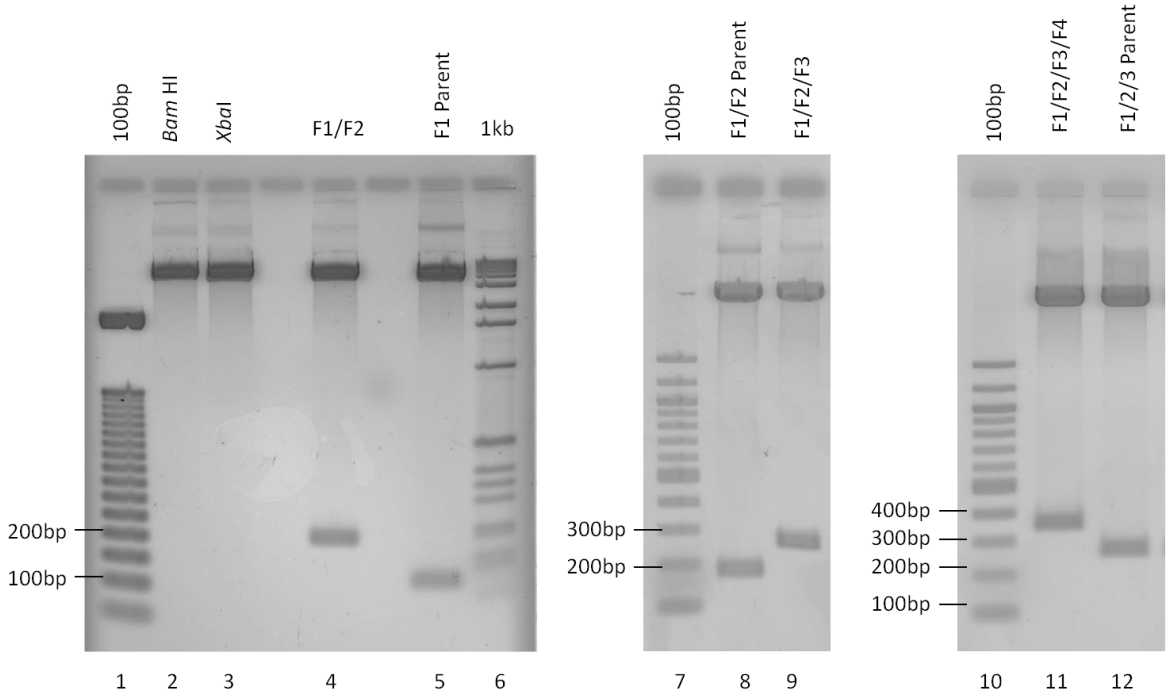

Figure 3.3 Cloning of Zinc Finger Arrays
Agarose gel electrophoresis of *Xba*I – *Bam*HI restriction fragments from ZF array plasmids. Plasmids containing 1 (lane 5), 2 (lane 4), 3 (lane 9) or 4 (lane 11) ZF modules yield a 96, 192, 288 or 384 bp *Xba*I – *Bam*HI fragment, respectively. Single digest controls with *Bam*HI and *Xba*I individually (Lanes 2 & 3) confirm each enzyme's activity by linearising the plasmid.

3.5 Construction of Designed Zinc Finger Nucleases

3.5.1 *Fok*I domain mutagenesis

ZFNs utilise the cleavage domain of the Type II restriction enzyme *Fok*I, which has no DNA sequence specificity of its own (Li et al., 1992). *Fok*I functions as a homodimer when mediating endonuclease activity (Bitinaite et al., 1998). A number of groups have undertaken mutagenesis of *Fok*I such that it only functions as an obligate heterodimer (Section 1.3.4.1.2). Obligate heterodimer ZFNs have been found to have greatly reduced off-target cleavage compared to their wild type *Fok*I counterparts (Miller et al., 2007). *Fok*I expression plasmids pMLM290 (Addgene 21872) and pMLM292 (Addgene 21873) which encode the "KK" (E490K, I538K) and "EL" (Q486E and I499L) heterodimers mutants were acquired. The KK ZFN is employed on the left and the EL ZFN is employed to the right of the target cleavage site. The EL/KK heterodimer mutations have been shown to reduce *Fok*I cleavage activity by three fold (Miller et al., 2007). One study described *Fok*I cleavage domain mutations that increased endonuclease activity without any observed increase in cytotoxicity (Guo et al., 2010). This and subsequent studies have shown that

these “Sharkey” mutations (S418P and K441E) can be used to increase the performance of heterodimer *Fok*I mutant ZFNs (Doyon et al., 2011, Perez-Pinera et al., 2012).

I decided to combine the Sharkey mutations with both the EL and KK *Fok*I mutants to allow the study of whether these mutations can result in better ZFN performance. The inclusion of the Sharkey mutations required two single base pair changes at codon 418, TCC -> CCC, (S418P) and at codon 441, AAA -> GAA (K441E) (Figure 3.4). Codons 418 and 441 lie within a 274 bp *Hind*III restriction fragment, so the Sharkey mutant fragment was therefore generated via DNA synthesis. The Sharkey fragment was sub-cloned into the EL and KK *Fok*I plasmids using the *Hin*dIII restriction sites. The resulting constructions were checked for the correct orientation of the Sharkey fragment using *Bam*HI and *Age*I digestion (Figure 3.5). The KK-S and EL-S encoding plasmids (Table 3.2) were checked by Sanger sequencing (Appendix II) prior to the creation of ZFN expression vectors, as described below.

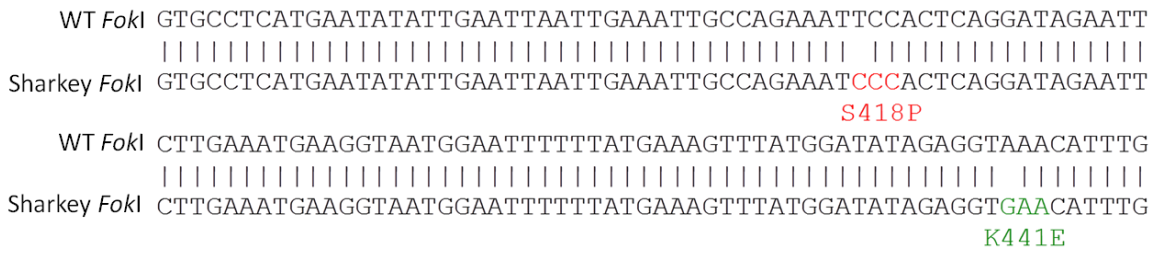

Figure 3.4 Modification of the *Fok*I coding sequences to incorporate Sharkey mutations

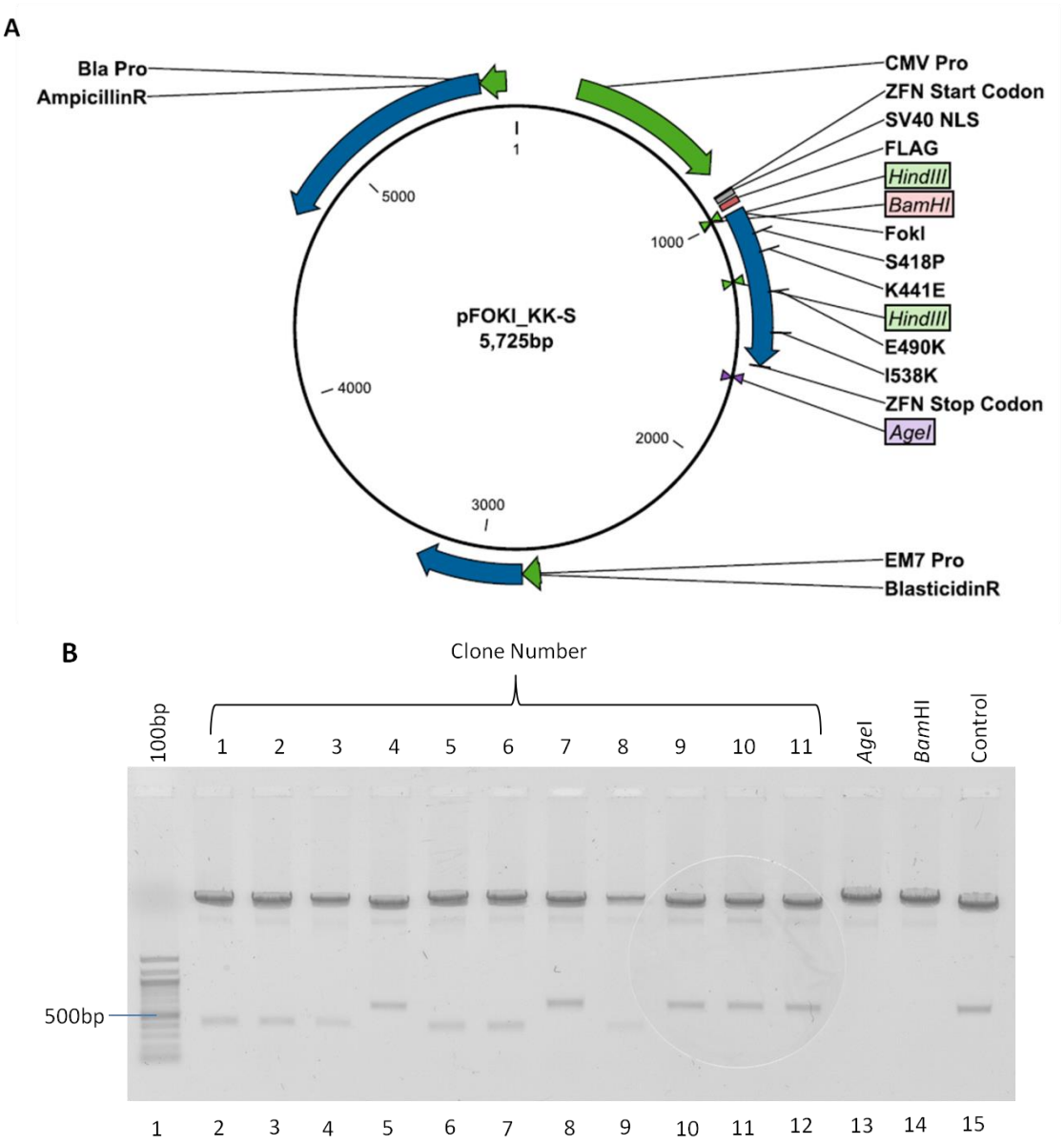

Figure 3.5 Construction of plasmids encoding Sharkey *Fok*I heterodimer mutants
A) Scale plasmid map of pFOKI_KK-S. The KK obligate heterodimer (E490K, I538K) and Sharkey (S418P, K441E) *Fok*I mutations are indicated. B) Agarose gel electrophoresis of pFOKI_KK-S plasmid clones digested with *Age*I and *Bam*HI. Sharkey fragments correctly oriented between the *Hind*III sites result in 654 bp *Age*I - *Bam*HI fragment (Lanes 5, 8, 10 – 12). Single digest controls with *Age*I and *Bam*HI individually (Lanes 13 & 14) confirm each enzyme's activity by linearising the plasmid. The control is the parent plasmid pFOKI_KK.

Plasmid Name	Source	*Fok*I Mutation
pFOKI_KK	pMLM290 (Addgene 21872)	E490K, I538K
pFOKI_EL	pMLM292 (Addgene 21873)	Q486E, I499L
pFOKI_KK-S	Modification of pMLM290	E490K, I538K, S418P, K441E
pFOKI_EL-S	Modification of pMLM292	Q486E, I499L, S418P, K441E

Table 3.2 Plasmids encoding mutant *Fok*I domains

3.5.2 Construction of Zinc Finger Nuclease Expression Vectors using Assembled ZF Domains

The three pairs of ZF arrays were subcloned into the pFOKI vectors via the *Xba*I and *Bam*HI restriction sites. The insertion of the ZF arrays into the resulting pZFN vectors was checked by restriction analysis with *Nhe*I and *Eco*RI (Figure 3.6). The ZFN expression vectors created are summarised in Table 3.3. The pFOKI vectors, which are derived from pMLM290/292, already contain a CMV promoter, SV40 enhancer and polyA elements. The pZFN vectors are therefore ready for the expression of ZFNs in mammalian cells.

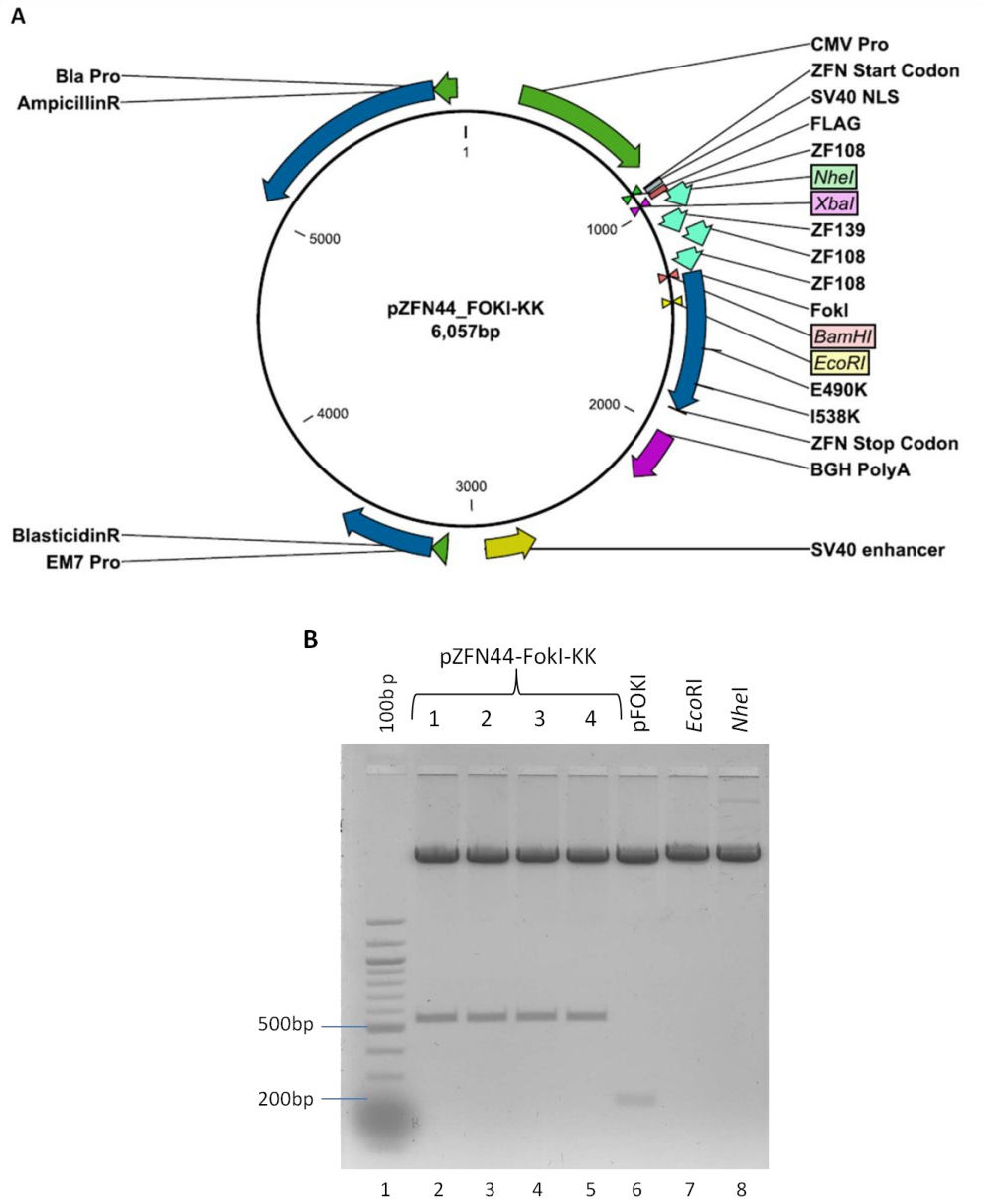

Figure 3.6 Construction of zinc finger nuclease expression vectors
A) Scale map of pZFN44-FokIKK. The location of the sequences encoding the zinc fingers 108, 139, 108 and 108 and the *Fok*I expression cassette. B) Agarose gel electrophoresis of pZFN44-FokIKK plasmid clones digested with *Eco*RI and *Nhe*I. Plasmids containing the four finger ZF44 contain a 590 bp fragment (lanes 2-5). The parental pFOKI-KK plasmid contains a 206 bp fragment (lane 6). Single digest controls with *Eco*RI or *Nhe*I alone (Lanes 7 and 8) confirm each enzyme's activity by linearising the plasmid.

Plasmid Name	*Fok*I Mutation	DNA binding configuration
pZFN35-FokI-KK	E490K, I538K	3 2 1 CGCCGCCGCCGCTTT**GTTGTCGCC** **GCGGCGGCG**GCGAAACAACAGCGG 1 2 3
pZFN36-FokI-EL	Q486E, I499L	
pZFN37-FokI-KK	E490K, I538K	3 2 1 CCCGCCCGCCCCGGC**GAAGCGGGG** **GGGCGGGCG**GGGCCGCTTCGCCCC 1 2 3
pZFN38-FokI-EL	Q486E, I499L	
pZFN43-FokI-KK	E490K, I538K	4 3 2 1 GCGCTGCCATGTTGAGGA**GCCGCCGCTGCC** **CGCGACGGTACA**ACTCCTCGGCGGCGACGG 1 2 3 4
pZFN44-FokI-EL	Q486E, I499L	
pZFN43-FokI-KKS	E490K, I538K, S418P, K441E	
pZFN44-FokI-ELS	Q486E, I499L, S418P, K441E	

Table 3.3 Mammalian expression plasmids encoding ZFNs assembled in this study

3.6 Discussion

This chapter has described the design and construction of mammalian expression plasmids that encode zinc finger nucleases that target the human *VEZF1* promoter. The ZiFiT design tool was straightforward to use and enabled the design of ZFNs that use specific sets of zinc finger modules that are reported to function in the context of ZFNs in mammalian cells. Standardised plasmid libraries and protocols were utilised to assemble zinc finger arrays from individual modules selected in the design process. Custom zinc finger arrays were cloned into obligate heterodimer *Fok*I nuclease expression vectors, and one pair was assembled into an advanced *Fok*I nuclease architecture that should improve cleavage activity. These ZFNs are later used in experiments to modify the VEZF1 gene promoter via HDR and NHEJ (Chapters 5 & 6).

While straightforward, modular assembly is a laborious and time consuming method of constructing zinc finger arrays as each finger must be added in sequence. The repetitive nature of the zinc finger array sequences can be prone to recombination during the cloning stages, especially when the array is longer than three ZF modules. In fact, greater difficulties were faced as lytic phage contamination accompanied our use of the zinc finger consortium plasmid library. The contamination prevented successful plasmid culture through the West laboratory for over four months until new methods for decontamination and waste control, together with physical separation of three areas for media preparation, inoculation and plasmid preparation were established in the

laboratory. Great care must be taken when receiving plasmids and bacterial stocks from outside laboratories.

Since the advent of this project, there have been major advances in assembling zinc finger arrays. Construction methods such as OPEN (Maeder et al., 2009) and CoDA (Sander et al., 2011b) greatly accelerate assembly, regardless of whether low throughput in an academic group or high throughput in a facility are required. Furthermore, the falling cost of DNA synthesis enables individual groups to create ZFN expression constructs more readily.

More importantly, a new technology arose during the course of this work, where DNA-binding domains from the TALE transcription factors have been fused to *Fok*I domain to create TALE nucleases (TALENs). TALENs have been found to be more reliable than ZFNs and the proteins have a simple DNA-recognition code, simplifying the construction process. The development and application of TALE proteins that target the human *VEZF1* promoter is described in the following chapters.

Chapter 4

Development of TALE proteins that target the *VEZF1* gene

4.1 Introduction

Transcription activator like effectors (TALEs) are a class of DNA-binding proteins from *Xanthomonas* spp. which permit fusion with a variety of protein domains to direct site-specific DNA cleavage or gene regulatory functions (Section 1.4). TALE proteins contain tandem arrays of ~34 amino acid repeats, where each repeat interacts with a single DNA base pair. Two amino acids in each repeat, called the repeat variable diresidue (RVD), mediate base interactions. The TALE repeats are modular, such that TALE proteins can be designed to target any sequence of choice. TALE protein fusions with the catalytic domain of the *Fok*I endonuclease, known as TALE nucleases (TALENs), have been shown to be efficient at directing specific targeted genetic mutations in a wide variety of species including mammals (Section 1.4.2). The simplicity of the TALE DNA recognition code, where only 4 repeats are required, coupled with the high reliability of TALE proteins are major advantages compared to zinc fingers. Furthermore, the ability to assemble long TALE proteins with over 20 bp of specificity greatly reduces confounding off-target effects (Miller et al., 2011, Beumer et al., 2013, Mussolino et al., 2011).

A number of molecular cloning strategies have been devised to allow the rapid assembly of repetitive TALE protein expression constructs, whether for low throughput individual groups or for high throughput facilities (Section 1.4.6). In this chapter, the use of one of these strategies to create TALE proteins that target the human *VEZF1* promoter is described. I developed expression vectors that allow the creation of TALE-repressor proteins that are designed to give a readout of TALE binding performance at the *VEZF1* promoter in human cells. Functional TALE-repressors will also be a useful resource for the knockdown of VEZF1 expression. Functional TALE domains could be subsequently transferred into TALE-nuclease expression vectors for use in *VEZF1* gene knockout strategies.

4.2 Aims of this chapter

- Design and construct custom TALE DNA-binding domains that target the human *VEZF1* gene promoter

- Develop expression vectors for the creation of TALE-repressor proteins

- Determine the performance of TALE-repressor proteins in human cells

4.3 Design of TALE DNA-binding domains

Previous studies have indicated that there are some target DNA sequence constraints, particularly at the start of a TALE repeat, which should be considered to increase the chance of obtaining high affinity TALE proteins (Boch et al., 2009, Moscou and Bogdanove, 2009, Cermak et al., 2011). TALE protein recognition sites tend to start with a T then the first two RVD domains should be not T and not A, respectively. Furthermore, the most functional TALEs tend not have recognition sites that end with a G (Cermak et al., 2011). Given that the end goal was to create functional TALENs, the TALE DNA-binding proteins were designed so they could function both individually in TALE repressors, and as in pairs in TALENs. A number of studies published at the time recommended that TALE domains be 15 - 20 RVDs long and suggested various spacer sizes ranging from 15 - 31 bp in size (Cermak et al., 2011, Miller et al., 2011, Mussolino et al., 2011, Li et al., 2011b). The TALE-NT webtool (www.tale-nt.cac.cornell.edu) was used to design TALENs using these guidelines within a 400 bp region of the human *VEZF1* promoter (200 bp either side of the TSS) (Doyle et al., 2012) to generate a design output of all possible TALE domains that could target this region.

I chose to develop TALE proteins of 16 - 29 repeats in length, to investigate whether longer TALEs result in better performance, either by increased affinity or specificity for the target sequence. Furthermore, I selected to test TALEN designs where the DNA spacer between the left and right TALEN targets was either 15 or 30 bp in length. While a 15 bp spacer had been reported by various groups as being the optimal distance for *Fok*I dimerisation between two TALEs (Cermak et al., 2011, Miller et al., 2011), larger spacers of up to 30 bp had also been demonstrated to be advantageous (Li et al., 2011b). Ten TALE protein designs were chosen, which can be combined into six TALEN pairs (Table 4.1). The TALE designs were selected as they bind in the nucleosome depleted region of the *VEZF1* promoter (Figure 4.1). In total, the selected TALENs are design to cleave at three different locations within the *VEZF1* promoter and incorporate different TALE lengths and TALEN spacers (Figure 4.2). TALE pair 15/16 straddles the predicted TSS of *VEZF1* and the remaining five TALE pairs bind at in the 5' untranslated region (UTR). All of these TALENs are compatible with the gene targeting strategy to knockout *VEZF1* described in Chapter 5.

Design	RVD Sequence	DNA Recognition Sequence	Configuration Left-X-Right
TALE_13	NN NG HD NN HD HD NG HD HD NG HD HD NN NN HD NG	TALE 13 |||||||||||||||| GTCGCCTCCTCCGGCT---AGTGGGGAGGAGGGGGGTCGGCCGCCGC CAGCGGAGGAGGCCGA---TCACCCCTCCTCCCCCCAGCCGGCGGCG |||||||||||||||||||||||||||||||| TALE 14	16-30-28
TALE_14	NN HD NN NN HD NN NN HD HD NN NI HD HD HD HD HD HD NG HD HD NG HD HD HD HD NI HD NG		
TALE_15	HD NN HD HD NN HD HD NN HD NN NN NG NN NN NG NN HD NN NG	TALE 15 |||||||||||||||||| CGCCGCCGCGGTGGTGCGT---AAGTGGCCGCGCTCGGCGCTGCC GCGGCGGCGCCACCACGCA---TTCACCGGCGCGAGCCGCGACGG |||||||||||||||||||||||| TALE 16	19-15-23
TALE_16	NN NN HD NI NN HD NN HD HD NN NI NN HD NN HD NN NN HD HD NI HD NG NG		
TALE_17/19	NN HD HD NN HD HD NN HD HD NN HD HD NN HD HD NN HD HD NN HD NG NG NG NN NG NG NN NG	TALE 17/19 |||||||||||||||||||||||||| GCCGCCGCCGCCGCCGCCGCTTTGTTGT---AGGAGGCTCCCCGAGCGGGGGG**AGTGGGG** CGGCGGCGGCGGCGGCGGCGAAACAACA---TCCTCCGAGGGGCTCGCCCCCC**TCACCCC** ||||||||||||||||||||||||||||||||| TALE 18 **TALE 20**	28-15-22
TALE_18	HD HD HD HD HD HD NN HD NG HD NN NN NN NN NI NN HD HD NG HD HD NG		
TALE_20	HD HD HD HD NI HD NG HD HD HD HD HD HD NN HD NG HD NN NN NN NN NI NN HD HD NG HD HD NG		28-15-29
TALE_21/23	NN NG NG NN NG HD NN HD HD NG HD HD NG HD HD NN NN HD NG	TALE 21/23 |||||||||||||||||| GTTGTCGCCTCCTCCGGCT---AGGAGGGGGGTCGGCCGCCGCAGCC**ATGG** CAACAGCGGAGGAGGCCGA---TCCTCCCCCCAGCCGGCGGCGTCGG**TACC** |||||||||||||||||||||||||||||||||| TALE 22 **TALE 24**	19-30-25
TALE_22	NN NN HD NG NN HD NN NN HD NN NN HD HD NN NI HD HD HD HD HD HD NG HD HD NG		
TALE_24	HD HD NI NG NN NN HD NG NN HD NN NN HD NN NN HD HD NN NI HD HD HD HD HD HD NG HD HD NG		19-30-29

Table 4.1 Designs for TALENs that target the human *VEZF1* gene promoter

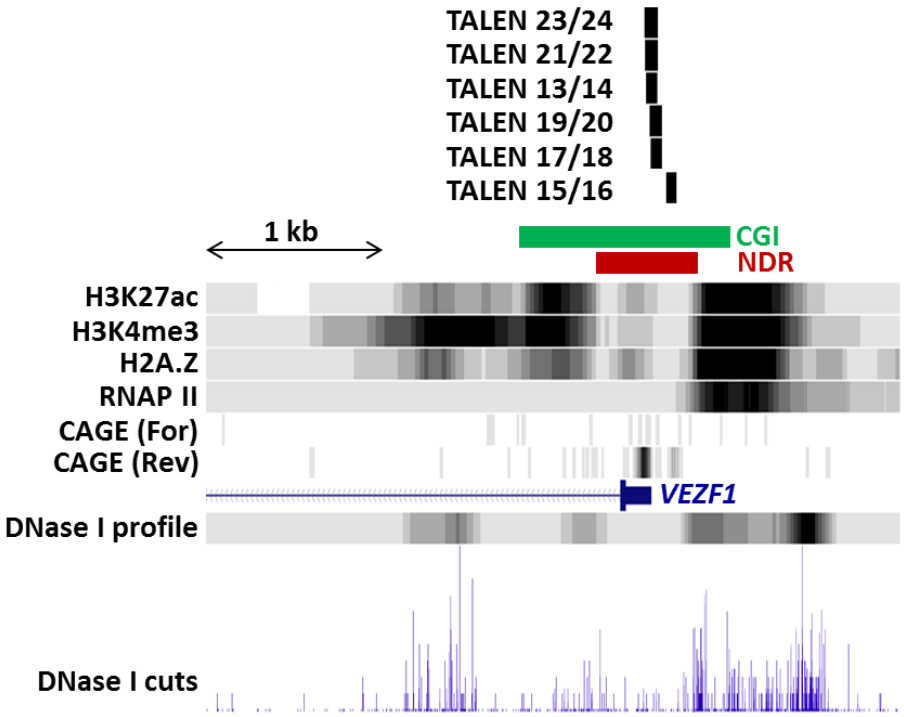

Figure 4.1 TALEN sites locate within a nucleosome depleted region of the *VEZF1* promoter
UCSC Genome browser view of Human hg19 chr17:56063000-56067000. The Refseq annotation of the *VEZF1* gene is shown in blue, with Exon 1 depicted as a wide bar, preceded by a 5'UTR. Intron 1 is depicted by a line with chevrons indicating the direction of the gene. A 1221 bp CpG island is depicted by a green bar. CAGE-seq data from K562 cells (FANTOM5:DRR008730) indicates that *VEZF1* is a unidirectional promoter, with the K562 TSS located within the annotated 5'UTR. ENCODE ChIP-seq analysis of H3K27ac (wgEncodeEH000043), H3K4me3 (wgEncodeEH000048) and H2A.Z (wgEncodeEH001038) indicate that highly positioned nucleosomes enriched for these modifications flank a ~560 bp nucleosome depleted region (NDR, red bar). The cleavage profile and individual cuts of DNAse I (wgEncodeEH000480) are shown beneath, indicating that DNAse I cuts either side of the positioned nucleosomes. The locations of the TALEN recognition sites are shown above (black bars).

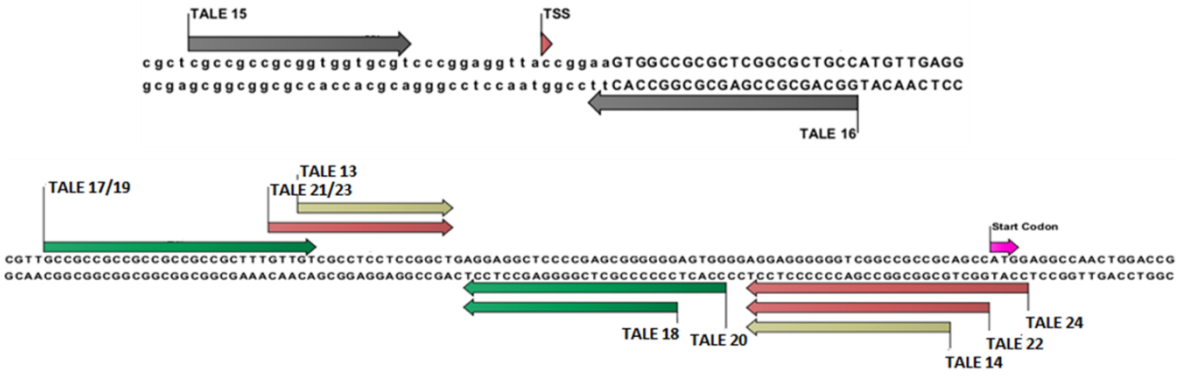

Figure 4.2 TALE target sites at the human *VEZF1* gene promoter
Cartoon demonstrating the location of TALEs within the *VEZF1* gene. Designed TALEs can be used individually as TALE repressors or as pairs (as indicated by colour coding) when co-expressed as TALENs.

4.4 Assembly of Designed TALEs

The Golden Gate (GG) assembly method, developed by the Voytas group, was chosen to enable the straightforward assembly of TALE repeats into the TALE protein PthXoI (Cermak et al., 2011). This system assembles up to ten repeats at a time in a single GG reaction. Up to three sets of arrays are joined with a last repeat and the rest of the TALE encoding sequences in a second GG reaction to create TALE proteins up to 31 repeats in length. This GG assembly system utilises the type IIS restriction enzymes *Bsa*I or *Esp*3I, which cleave a number of bases away from their recognition site. Both enzymes generate 4 base overhangs of any sequence. The overhang sequences can therefore be used to direct the ordered ligation of a mixed reaction (Figure 4.3). Another favourable property of these two enzymes is that they are active in T4 DNA ligase buffer, which allows rounds of sequential digestion and ligation reactions to be performed in the same tube. The assembly vectors are designed in a way that the *Bsa*I or *Esp*3I recognition sequences are not retained upon assembly, so the GG reaction can only proceed in the desired direction.

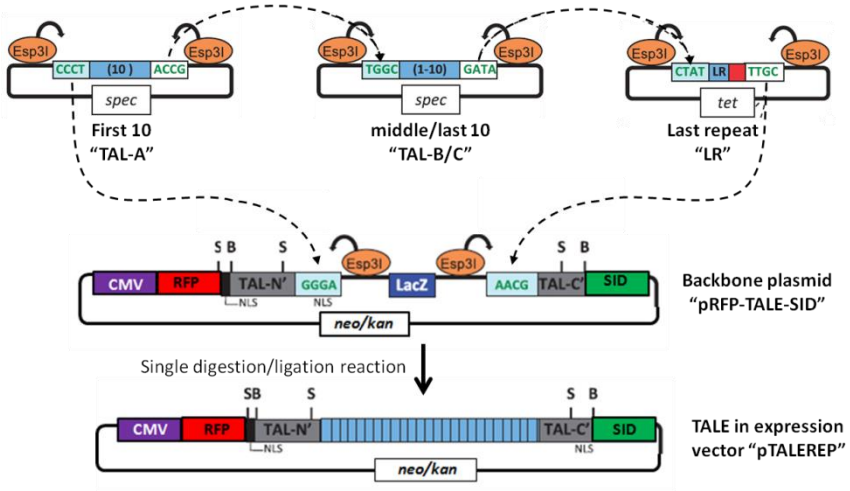

Figure 4.3 Golden Gate Assembly
Golden Gate (GG) assembly employs type IIS restriction enzymes, which cleave outside of their recognition sites. The vectors in the GG kit are designed such that digest with Esp3I creates complementary overhangs which will assemble the TALE repeats in the correct order. The first 10 RVDs are assembled into TAL-A, the next 10 are assembled into TAL-B and any additional RVDs in TALEs more than 20 RVDs long are assembled into TAL-C. TAL-A/B/C vectors are then digested and RVDs are combined with pRFP-TALE-SID to create full length TALEs in pTALEREP expression vector (adapted from Cermak et al., 2011).

The Voytas GG system consists of plasmids encoding the four TALE modules NI, HD, NN and NG (which recognise A, C, G and T, respectively) with flanking overhang sequences that direct assembly into positions 1 to 10 of an array (40 plasmids). There are also array, last repeat and expression vector plasmids (Figure 4.4). The assemblies of the first, second and an optional third set of TALE repeats were named TAL-A, TAL-B and TAL-C. In

order to prevent errors at the bench, the output from the TALE-NT website was used to create a spreadsheet file that summarised the picking of plasmids to be added into the GG reactions (Figure 4.5).

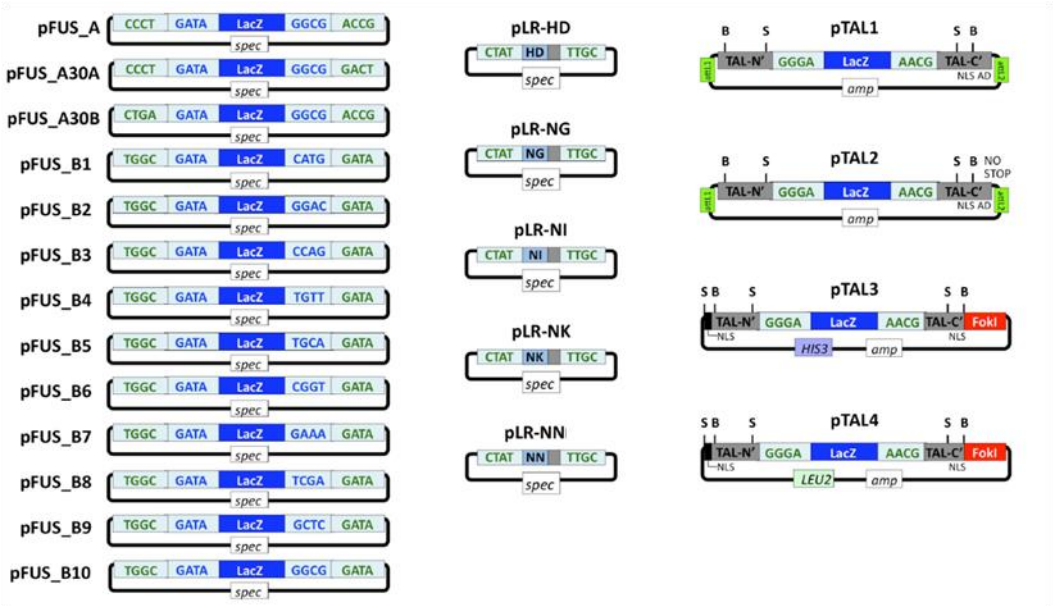

Figure 4.4 Golden Gate array, last repeat and backbone plasmids
Schematic of plasmid vectors in the GG kit for assembly of TALEs. To assemble TALEs of fewer than 20 RVDs, the first 10 RVDs are assembled in pFUS_A and the remainder are assembled in a pFUS_B vector. Assembly of TALEs containing more than 20 RVDs in total is carried out by inserting the first 10 RVDs into pFUS_A30A. RVDs 11-20 of long TALE domains are assembled into pFUS_A30B and the remainder are constructed in a pFUS_B vector. The choice of pFUS_B depends on the number of RVDs in the assembly (e.g. to assemble 5 RVDs use pFUS_B5). pFUS assemblies are carried out by digest with *Bsa*I, sequences of overhangs created are in blue. The pFUS RVD assemblies can be combined, with pLR (last repeat) into pTAL vectors by digest with *Esp*3I, sequences of overhangs created are in green (Cermak et al., 2011).

GG assembly reactions were performed on a thermocycler (10 cycles of 37°C/5minutes, 16°C/10minutes) and then transformed into competent cells and transformants selected via blue/white screening. White colonies were selected and subjected to screening by colony PCR to determine the number of RVD domains contained within the vector (Figure 4.6 A). Colony PCR of TALE repeat assemblies with Taq polymerase results in a laddered banding pattern due to termination of the PCR product as a result of the repetitive nature of the sequence. Nevertheless, the main product represented the expected length for each TALE repeat for the majority of clones. The only exceptions are plasmid clones from failed assemblies, which appear as re-ligated empty pTAL vector (e.g. Figure 4.6 A, lanes 6 and 7). One plasmid clone for each assembly was checked for the correct insert length by restriction digestion (Figure 4.6 B). This confirmed the integrity of the TALE repeats and that aborted elongation was responsible for the laddered colony PCR products. In total, 27 pTAL plasmids containing the TALE repeats were constructed and these were next

combined in the final stage of GG assembly to create full length TALEs in expression vectors. The identity of each repeat in the assembly was confirmed by Sanger sequencing (Figure 4.7).

Golden Gate TALE Assembly 1 of 4

"TAL-A" TALE 1-10 assembly

Assembly name	Target name	Size	1µl array *	Water	BsaI	BSA 2mg/ml	T4 Ligase	10X Ligase buffer	1µl TALE RVDs 1-10									
									1	2	3	4	5	6	7	8	9	10
TAL-A 013	VEZF1_001 L1	16	pFUS-A	3	1	1	2	2	NN	NG	HD	NN	HD	HD	NG	HD	HD	NG

Golden Gate TALE Assembly 2 of 4

"TAL-B" TALE 11-20 assembly

Assembly name	Target name	Size	1µl array *	Water	BsaI	BSA 2mg/ml	T4 Ligase	10X Ligase buffer	1µl TALE RVDs 1-10									
									1	2	3	4	5	6	7	8	9	10
TAL-B 013	VEZF1_001 L1	16	pFUS-B - 5	8	1	1	2	2	HD	HD	NN	NN	HD	0	0	0	0	0

Golden Gate TALE Assembly 4 of 4

TALEREP vector assembly

Assembly name	Target name	Size	1µl first array	1µl second array	1µl third array	Last repeat	Expression plasmid	Water	Esp3I	T4 Ligase	10X Ligase buffer
pTALEREP 013	VEZF1_001 L1	16	TAL-A 013	TAL-B 013		pLR NG	pRFP_TALE_SID	12	1	2	2

Figure 4.5 Plasmid Pick Sheet for Golden Gate Assembly of TALEREP plasmids
Example of a plasmid pick sheet used to set up Golden Gate reactions. Plasmids containing each RVD in the desired position in the final TALE array are combined in a tube with pTAL_A, pTAL_B or final expression plasmid with restriction enzyme *Bsa*I and T4 DNA ligase to undergo a Golden Gate reaction. The pick sheet format above allows for quick and easy setup.

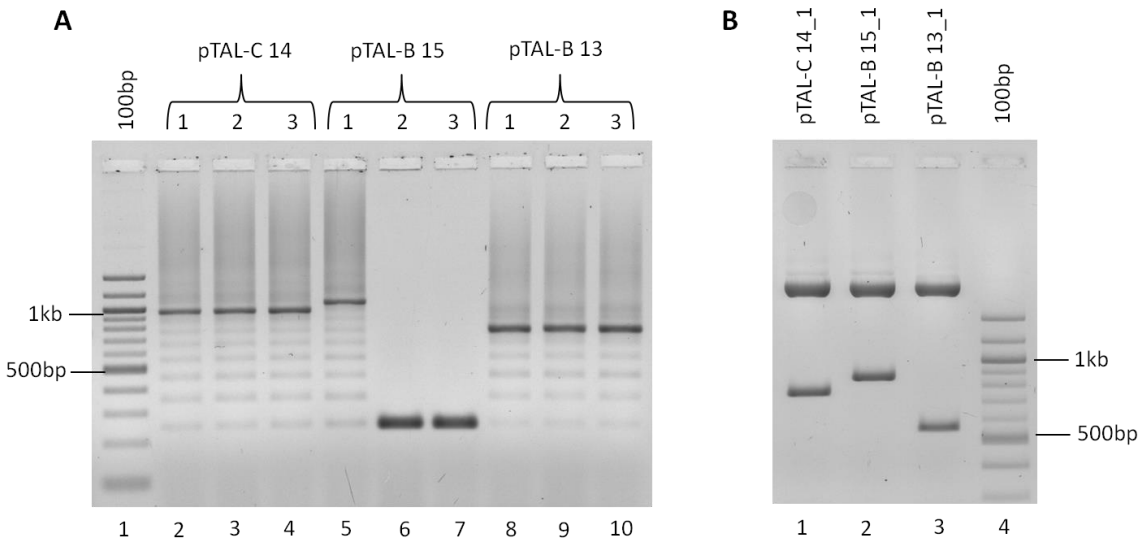

Figure 4.6 Golden Gate assembly of TALE repeats
(A) Agarose gel electrophoresis of TALE repeat length after GG assembly. Colony PCR using the primers pCR8_F1 and pcR8_R1 from three transformants resulting from three pTAL-B GG assembly reactions are shown. Colony PCR from cells containing empty pFUS-B vector yield a 250 bp product, as found in lanes 6 and 7. Successful GG assembly adds 100 bp per TALE repeat. The expected PCR product sizes for 7-mer (lanes 2-4) 8-mer (lane 5) and 5-mer (lanes 8-10) repeats are 950, 1050 and 750 bp, respectively. Aborted PCR products of 100 bp decreasing sizes are likely to be due to the repetitive nature of the TALE repeat template. (B) Agarose gel electrophoresis of pTAL-B/C plasmids prepared from colonies analysed in part A. pTAL-B/C plasmids containing 7 (lane 18 (lane 2) or 5 (lane 3) TALE repeats were digested with *Afl*II and *Xba*I to excise760 860 and 560 bp repeat fragments, respectively.

```
>pTAL-A-13 (10) NN-NG-HD-NN-HD-HD-NG-HD-HD-NG YES
NNNNNNNNNTNNNNNNNNGTTTTNNNNTCACGACGTTGTAAAACGACGGCCAGTCTTAAGCGTCTCCCCCTGAACC
TGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTT
GCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGGCAAGCAA
GCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTA
TCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGG
CCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTG
TTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGC
AAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGC
TATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCAT
GGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGC
TGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAA
GCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTG
GCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACC
ATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCG
GCTGTTGCCGGTGCTGTGCCAGGACCATGGCCGAGACGTCTAGACANCNGGANNNAAATGCNCGACTTCNCTGCTA
CCCNNNNNNNNNGACNNNNNNCGNNNAANGNTNANNNNNANCCNNNNGNNNNNNCNGNTTCGNNGTANCNGNANGC
AGTANCNNATGNNNNNNNNCNCTNNNNNANNNNNNCNNANNNNNNGGNGGNNACGNNCCNNNNNNNNNNNNNNNNN
NNNGNNNNTNNANNNGNTTTTNNNGGNNNNNNN
```

Figure 4.7 Sequence analysis of TALE assemblies
An example Sanger sequencing analysis of a pTAL plasmid resulting from GG assembly of TALE repeats. pTAL-A-13, a plasmid consisting of the first set of 10 repeats for the TALE 13 protein assembled into pFUS_A, was sequenced using the primers pCR8_F1 and pCR8_R1. The expected RVD composition is listed at the top. Bases highlighted in yellow have been assigned following inspection of the chromatogram data.

4.5 Development of a TALE transcriptional repressor expression vector

While the Voytas Golden Gate system allows for the rapid assembly of TALE repeats, it lacks a mammalian expression vector that the repeats can be assembled into to form functional TALE effector proteins. Rather than employ bacterial or yeast one hybrid and related assays to score for the DNA-binding activity of TALE proteins, I preferred to assess the performance of TALE proteins at their intended targets in human cells. I decided to construct a TALE repressor expression vector that is compatible with the Voytas GG assembly system. This ability of TALE-repressor expression to mediate the repression of *VEZF1* would both confirm the ability of TALE proteins to bind at their intended target and offer another tool for *VEZF1* knockdown.

Szurek *et al.* (2002) demonstrated that the TALE protein AvrBs3 with a truncated version of the N-terminus (Δ152) was fully capable of reaching its intended DNA binding site and so the amino acids deleted in this truncation were dispensable for DNA binding activities of TALEs. Miller *et al.* (2011) employed this truncation to develop highly active TALEN nuclease architectures. This study also demonstrated that the truncation of the C-terminus of TALE protein AvrBs3 to 63 amino acid residues in length (+63 truncation) resulted in TALENs with a higher DNA-binding activity than full length TALE proteins (Miller et al., 2011). I sought to incorporate these truncations into our TALE repressor vectors.

Cong *et al.* (2012) recently tested the performance of several transcription repression domains when fused to TALE proteins in HEK293 cells. The SIN Interaction Domain (SID) was reported to mediate the strongest repression of the target gene. The SID domain is a fragment of the Mad1 transcription factor, which mediates the recruitment of the SIN3A/B co-repressor complex (Ayer et al., 1996). The SIN3 proteins act as a scaffold to bring together many co-repressor proteins that mediate histone modification and chromatin remodelling. The repression activities include histone deacetylation (HDAC1/2), histone H3K4 demethylation (KDM5A), H3K9 methylation (SUV39H1, ESET) (van Oevelen et al., 2008, Yang et al., 2003). I therefore decided to develop an expression vector that encodes TALE-SID proteins following the TALE repressor architecture described by Cong *et al.* In order to track TALE-repressor expression by flow cytometry, I

decided to express turboRFP and the TALE-SID proteins from the same transcript, using an intervening 2A self-cleavage peptide sequence (Figure 4.8).

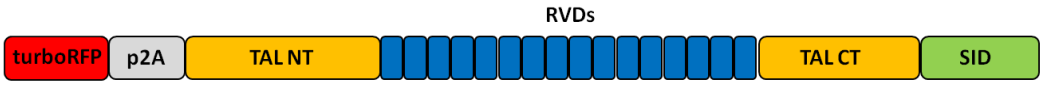

Figure 4.8 TALE repressor expression cassette
Schematic representation of the proteins encoded by pTALEREP vectors. Abbreviations: turbo RFP – red fluorescent protein, p2A – porcine 2A self cleaving peptide, TAL NT – TALE Δ152 N-terminus, RVD – repeat variable diresidue, TAL CT – TALE +63 C-terminus, SID – SIN interaction domain.

Three rounds of DNA sub-cloning were required to produce the pRFP-TALE-SID vector to be used in GG assembly reactions. Firstly, the sequences encoding turboRFP were PCR amplified from the lentiviral vector pTRIPZ (Thermo) using the primers RFP5_Nhe_Age and RFP3_Bam_Not, which were then sub-cloned into pEGFP-N1 using *Nhe*I and *Not*I. This creates pRFPturbo, which offers CMV-driven RFP expression (Figure 4.9 and Figure 4.10). In the second step, a 180 bp fragment encoding a P2A peptide and the SID domain flanked by restriction sites was synthesised (Figure 4.11 A). This fragment was inserted into pRFPturbo using *Bam*HI and *Not*I to create pRFP-SID (Figure 4.11 B and Figure 4.12 A). Lastly, a 1027 bp *Bgl*II-*Bam*HI encoding TALE N- and C-termini with an internal lacZ fragment was sub-cloned from pCAG_T7_TALEN (Addgene 37184) into the single *Bgl*II site of pRFP-SID to form pRFP-TALE-SID (Figure 4.13). Restriction digestion and Sanger sequencing was used to check each stage of sub-cloning (Figure 4.12 B, not shown).

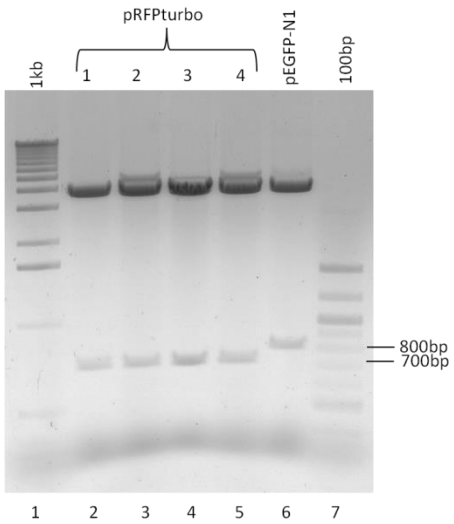

Figure 4.9 Construction of pRFPturbo
Agarose gel electrophoresis of restriction digested plasmid DNA from transformants gained following ligation of RFPturbo into pEGFP-N1 to create pRFPturbo. Digest of pRFPturbo with *Nhe*I & *Not*I results in 722 bp band (lanes 2 – 5). Digest of pEGFP-N1 with *Nhe*I & *Not*I results in 814 bp band (lanes 2 – 5).

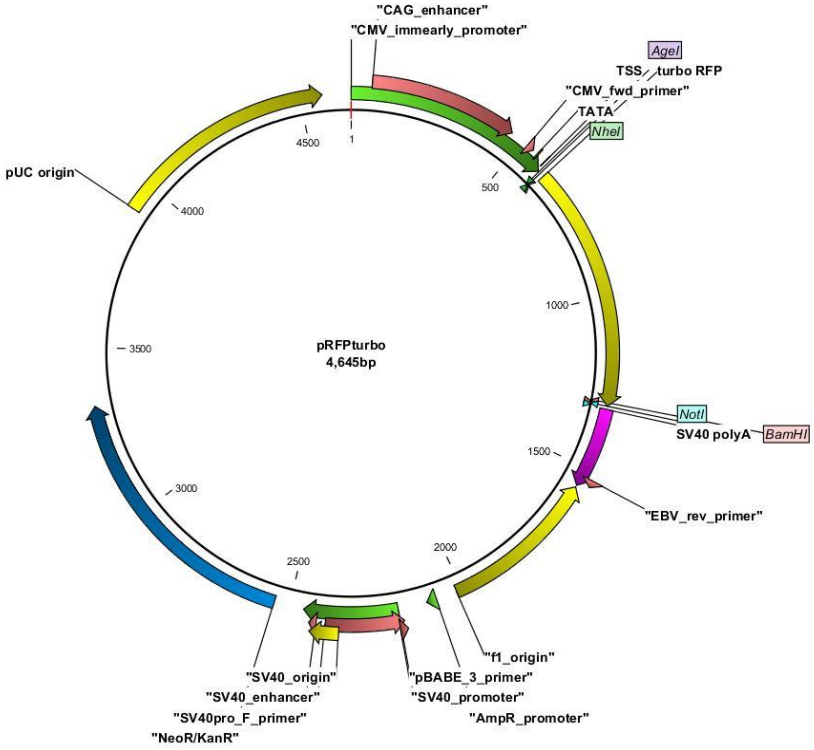

Figure 4.10 Plasmid map of pRFPturbo
Scale map of pRFPturbo. Sequences encoding turboRFP (derived from pTRIPZ, Thermo) are located in between the *Nhe*I and *Not*I sites. TurboRFP is expressed from the CMV promoter (green) using a downstream SV40 enhancer (red).

A

>P2A_SID
GGATCCGGCGCCACCAACTTCAGCCTGCTGAAGCAGGCCGGCGACGTGGAGGAGAACCCCGG
CCCCAGATCTATGAACATCCAGATGCTGCTGGAGGCCGCCGACTACCTGGAGCGCCGCGAGC
GCGAGGCCGAGCACGGCTACGCCAGCATGCTGCCCTGAATCGATGCGGCCGCGATC

B

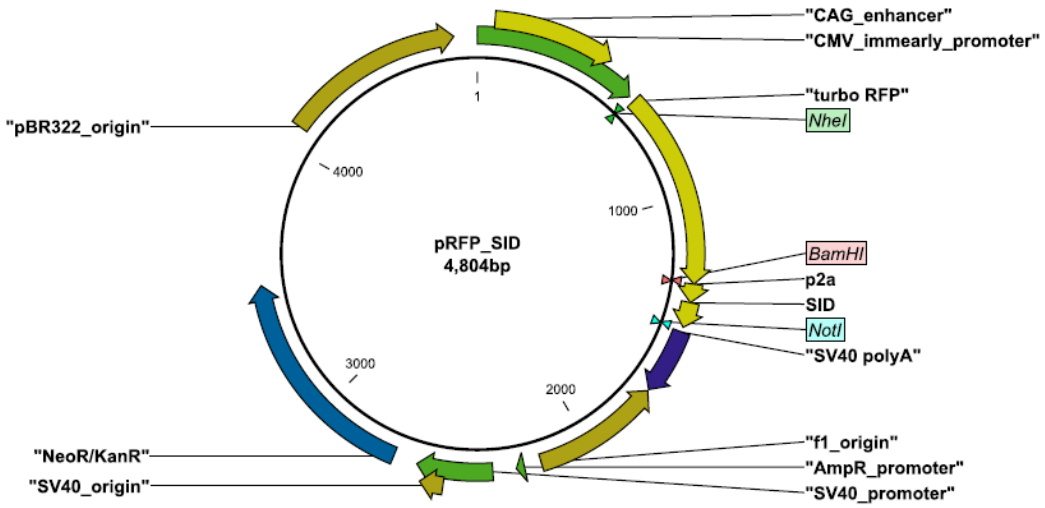

Figure 4.11 Plasmid map of pRFP-SID
A) Synthetic DNA fragment encoding a P2A peptide (brown, derived from (Kim et al., 2011)) and the SID domain (blue, derived from Cong *et al.* (2012)). *Bam*HI (orange), *Bgl*II (green), *Cla*I (red) and *Not*I (purple) restriction sites used for sub-cloning are shown. B) Scale map of pRFP-SID. Sequences encoding the P2A-SID peptide are located in between the *Bam*HI and *Not*I sites.

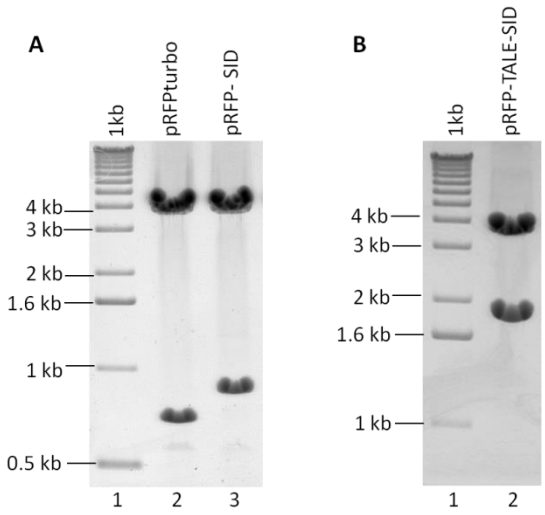

Figure 4.12 Restriction analysis of plasmid constructs
(A) Agarose gel electrophoresis of restriction digested plasmid DNA from transformants gained following ligation of P2A_SID into pRFPturbo to create pRFP-SID. Digest of pRFPturbo with *Nhe*I & *Not*I results in 722 bp band (lane 2). Addition of P2A_SID adds 159 bp, therefore positive clones result in 881bp product (lane 3). (B) Agarose gel electrophoresis of restriction digested plasmid DNA from transformants gained following ligation of TALE N- and C-termini to create pRFP-TALE-SID, which adds 1027 bp. Digest of pRFP-TALE-SID with *Nhe*I & *Not*I results in 1908 bp band (lane 2).

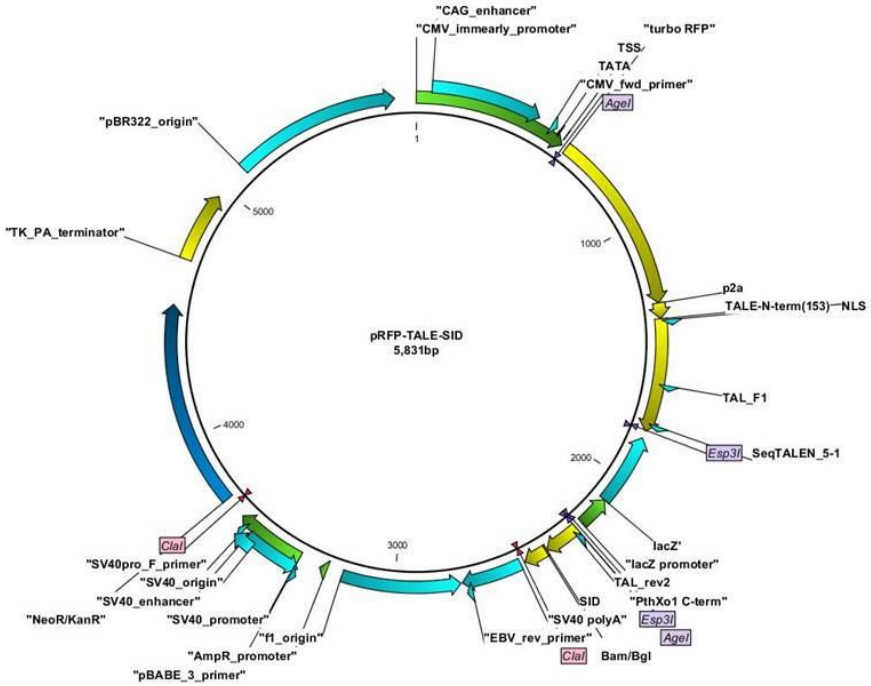

Figure 4.13 Plasmid map of pRFP-TALE-SID
Scale map of pRFP-TALE-SID. Sequences encoding the N- and C-termini of the PthXo1 TALE protein (derived from pCAG_T7_TALEN) are located between the P2A and SID coding sequences. A lacZ gene for blue white selection is located within the TALE domain. The *Esp*3I sites are compatible with the Voytas GG assembly of TALE repeats (Cermak et al., 2011), which would remove the lacZ gene.

4.6 Assembly of TALE repressor expression vectors

The production of TALE effector expression plasmids requires the GG assembly of TALE repeats from pTAL vectors (Section 4.4) into an expression plasmid, in this case pRFP-TALE-SID (Section 4.5). The Voytas GG assembly is arranged such that cleavage of the vectors with the type IIS restriction enzyme *Esp*3I generates cohesive ends that facilitate the ordered unidirectional assembly of TALE domains into the pRFP-TALE-SID vector. A plasmid pick sheet just as was designed for use in construction of the pTAL_A/B/C plasmids (Figure 4.5) was devised to aid in the reaction set up for the final step GG assembly of TALE repressors (Figure 4.14). The resulting pTALEREP vectors encode a fully assembled TALE-SID protein, which is co-expressed with turboRFP from a shared CMV promoter (Figure 4.15). Plasmid DNA was prepared from white colonies resulting from transformation of GG assembly reactions. These pTALEREP plasmids were checked for the correct assembly of TALE repeats by restriction digestion with *Stu*I and *Bam*HI (Figure 4.16).

TALEREP vector assembly													
Assembly name	Target name	Size	1µl first array	1µl second array	1µl third array	Last repeat	Expression plasmid	Water	Esp3 I	T4 Ligase	10 X Ligase buffer
pTALEREP 013	VEZF1_001 L1	16	TAL-A 013	TAL-B 013		pLR NG	pRFP_TALE_SID	12	1	2	2
pTALEREP 014	VEZF1_001 R2	28	TAL-A 014	TAL-B 014	TAL-C 014	pLR NG	pRFP_TALE_SID	12	1	2	2
pTALEREP 015	VEZF1_008 L1	19	TAL-A 015	TAL-B 015		pLR NG	pRFP_TALE_SID	12	1	2	2
pTALEREP 016	VEZF1_008 R2	23	TAL-A 016	TAL-B 016	TAL-C 016	pLR NG	pRFP_TALE_SID	12	1	2	2
pTALEREP 017	VEZF1_052 L1	28	TAL-A 017	TAL-B 017	TAL-C 017	pLR NG	pRFP_TALE_SID	12	1	2	2
pTALEREP 018	VEZF1_052 R2	22	TAL-A 018	TAL-B 018	TAL-C 018	pLR NG	pRFP_TALE_SID	12	1	2	2
pTALEREP 019	VEZF1_053 L1	28	TAL-A 019	TAL-B 019	TAL-C 019	pLR NG	pRFP_TALE_SID	12	1	2	2
pTALEREP 020	VEZF1_053 R2	29	TAL-A 020	TAL-B 020	TAL-C 020	pLR NG	pRFP_TALE_SID	12	1	2	2
pTALEREP 021	VEZF1_062 L1	19	TAL-A 021	TAL-B 021		pLR NG	pRFP_TALE_SID	12	1	2	2
pTALEREP 022	VEZF1_062 R2	25	TAL-A 022	TAL-B 022	TAL-C 022	pLR NG	pRFP_TALE_SID	12	1	2	2
pTALEREP 023	VEZF1_063 L1	19	TAL-A 023	TAL-B 023		pLR NG	pRFP_TALE_SID	12	1	2	2
pTALEREP 024	VEZF1_063 R2	29	TAL-A 024	TAL-B 024	TAL-C 024	pLR NG	pRFP_TALE_SID	12	1	2	2

Figure 4.14 Plasmid pick sheet for the assembly of pTALEREP TALE repressor expression plasmids

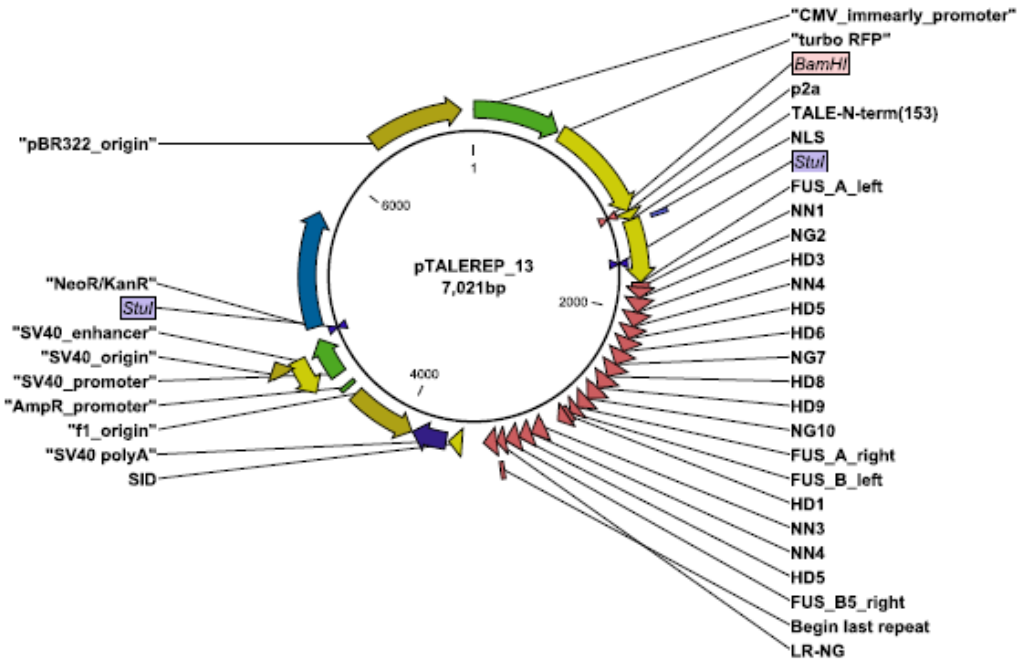

Figure 4.15 Plasmid map of pTALEREP-13
Scale map of pTALEREP-S-13 as an example of pTALEREP vectors. pTALEREP vectors encode a fully assembled TALE-SID protein that is co-expressed with turboRFP from a shared CMV promoter. *Stu*I and *Bam*HI sites used to confirm TALE assembly are shown. TALE 13 contains 16 RVD TALE repeats (red triangles).

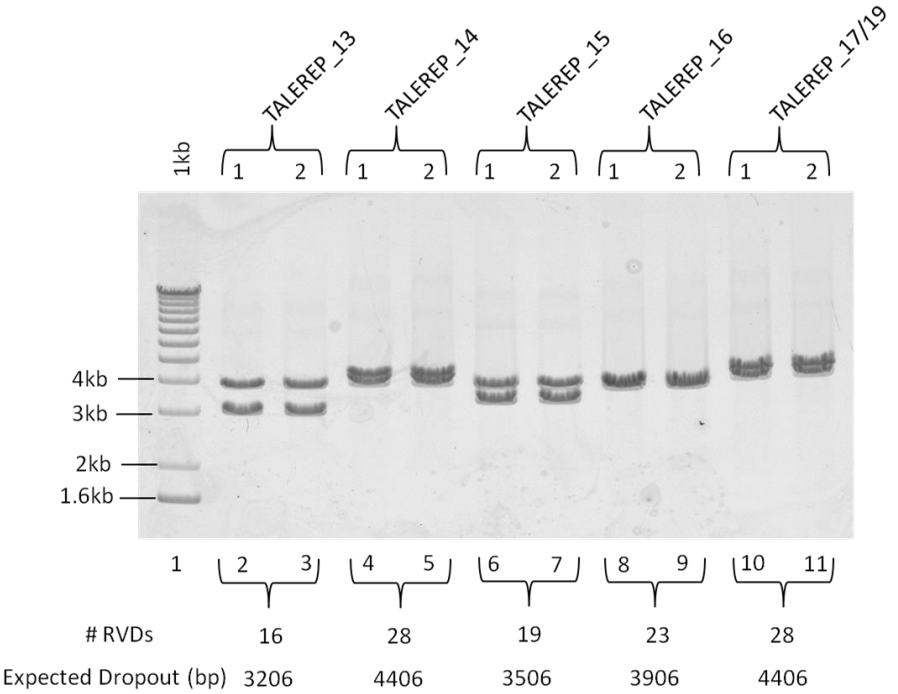

Figure 4.16 Restriction analysis of pTALEREP plasmids
Agarose gel electrophoresis of pTALEREP plasmids digested with *Stu*I and *Bam*HI. Digestion cleaves the plasmid backbone into 3450 bp and 350 bp (not shown) fragments and drops out the full length TALE repressor coding sequences (sizes indicated below lanes). Each of the vectors shown contains the expected size of dropout fragment.

4.7 TALE repressor-mediated silencing of the human *VEZF1* gene

HEK293 cells were transfected with individual pTALEREP plasmids. Two independent bioreplicates of TALE repressor transfection were performed. The level of turbo RFP expression three days post-transfection was monitored by flow cytometry. Approximately 40-50% of cells have high levels of RFP expression at this time (Table 4.2), although a further 30-40% have low levels of RFP expression (Figure 4.17). Cells were harvested and RNA and protein were isolated to investigate the effect of TALE repressor activity on *VEZF1* expression.

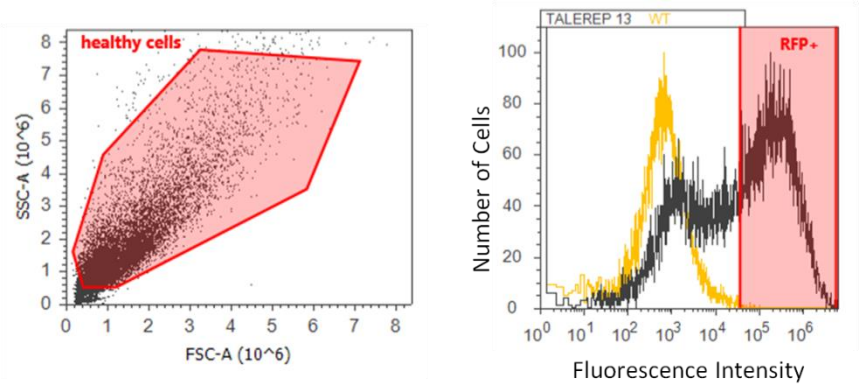

Figure 4.17 Analysis of RFP fluorescence in cells transfected with pTALEREP plasmids
An example flow cytometry analysis of RFP fluorescence in HEK293 cells. Forward scatter (FSC) and side scatter (SSC) plot is shown on the left and fluorescence analysis on the right. Healthy cells are gated in the FSC/SSC plot to exclude debris. Fluorescence plots of HEK293 cells are shown before (yellow) and after transfection with pTALEREP plasmid (black). This example is 3 days after transfection of pTALEREP-13, first replicate. In this example, 69% of events had more fluorescence than non-transfected HEK293 cells.

Sample	% RFP +	
	Run 1	Run 2
WT	<1	<1
TALEREP 13	69	44
TALEREP 14	58	30
TALEREP 15	54	22
TALEREP 16	63	30
TALEREP 17/19	71	23
TALEREP 18	62	50
TALEREP 20	58	59
TALEREP 21/23	66	59
TALEREP 22	69	48
TALEREP 24	59	50

Table 4.2 Quantification of RFP expression in cells transfected with pTALEREP plasmids

Quantitative RT-PCR analyses revealed that the expression of six of the ten TALE repressors resulted in reduced *VEZF1* RNA expression levels (Figure 4.18). The expression of TALEREP 13 or TALEREP 21/23 resulted in substantial *VEZF1* silencing, to averages of 17.5 and 23.5% of wild type levels, respectively. The expression of TALEREP 15 or TALEREP 22 resulted in moderate *VEZF1* repression, to averages of 74.5 and 86.5% of wild type levels, respectively. The expression of TALEREP 16 or TALEREP 18 resulted in silencing of *VEZF1* to as much as 59 and 55% of wild type, but this repression was not consistent between replicates. The expression of TALEREPs 14, 17/19, 20 or 24 had no effect on *VEZF1* mRNA levels.

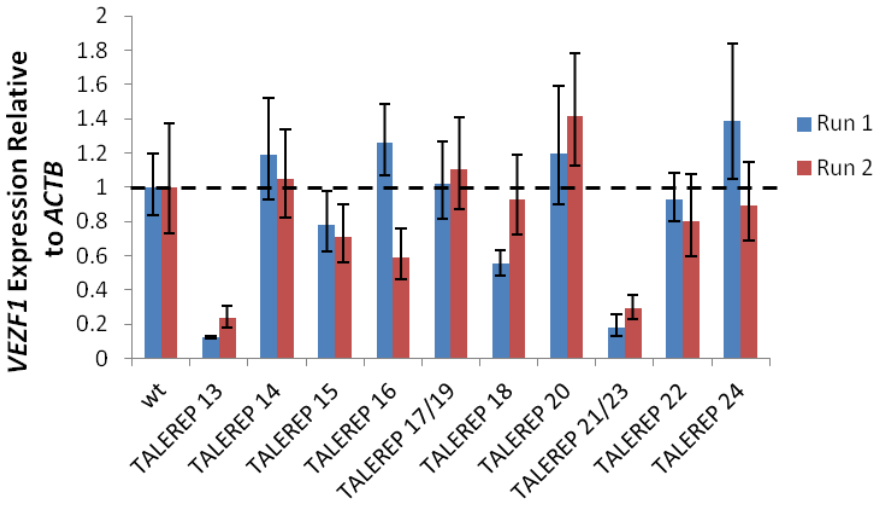

Figure 4.18 Analysis of TALE repressor action on *VEZF1* gene expression
Quantitative RT-PCR analysis of *VEZF1* mRNA expression levels in HEK293 cells. The expression of *VEZF1* in each cell line is normalised against *ACTB* expression and then adjusted to the level of *VEZF1* expression in untransfected HEK293 cells. RNA levels were quantified three days after transfection of individual TALEREP plasmids. The results from two independent bioreplicate transfections are shown (runs 1 and 2). Statistical analyses haven not been carried out on this data as the experiments were performed in duplicate. Dotted line shows level of *VEZF1* expression level of WT cells. Error bars represent the standard deviation between three technical qRT-PCR replicates.

It is possible that while TALE repressors can result in silencing of *VEZF1* transcription, post-transcriptional regulatory mechanisms may maintain VEZF1 protein levels. Western blotting analysis was therefore used to confirm that the expression of TALE repressors does indeed knock down VEZF1 protein levels within three days after transfection. Protein levels from Run 1 were quantified relative to the Ponceau stain (Figure 4.19 A), however, this is not a validated method of protein quantification. Samples from Run 2 (Figure 4.19 B) were not quantified due to the high level of background staining in the Ponceau. The expression of TALEREP 13 and TALEREP 21/23 both resulted in substantial

knock down of VEZF1 protein levels (Figure 4.19), which is consistent with the RNA analyses.

However, the moderate level of *VEZF1* silencing at the RNA level following the expression of TALEREPs 15, 16 or 18 did not significantly impact VEZF1 protein levels (Figure 4.19). The inconsistencies between VEZF1 RNA and protein levels are likely to be due to different modes of action of each TALEREP in some cases and delayed effects of RNA knock down on VEZF1 protein levels in other cases (see Discussion).

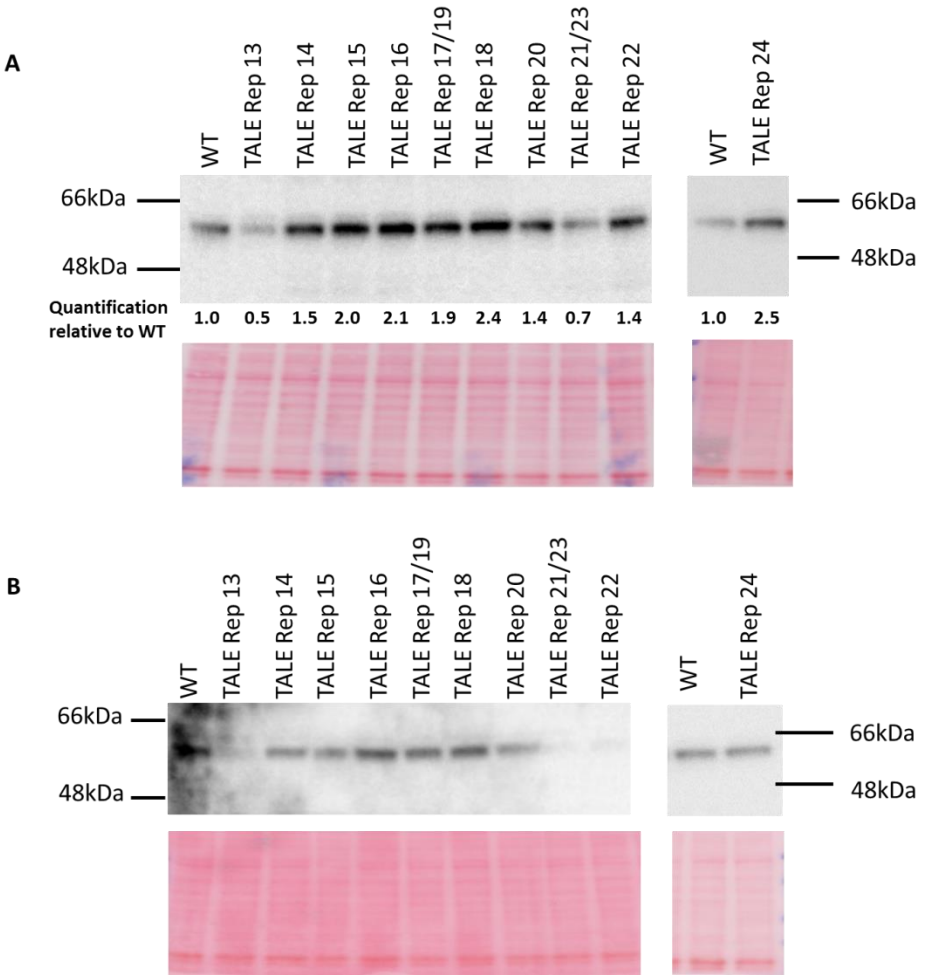

Figure 4.19 TALE repressor expression reduces VEZF1 protein levels
Western blot analysis of HEK293 cell extracts three days after transfection of pTALEREP plasmids. Ponceau staining of the membrane after transfer is a loading control. The analyses of two bioreplicates (A and B) are shown. Sample set A were quantified and levels compared to WT, using the intense band at the bottom of the Ponceau stain to normalise for loading (please note that the use of Ponceau as a loading control is not a validated method of protein quantification).

4.8 Discussion

This chapter has described the design and construction of ten TALE DNA-binding proteins that specifically target *VEZF1* gene promoter sequences of 16 to 29 bp in length. An expression vector was created that allows the production of TALE repressor proteins. The expression of TALE repressor proteins in a human cell line was found to rapidly silence VEZF1 RNA and protein levels. The TALE repressors were found to be a simple approach to determine the performance of TALE domains in a chromatin context. Two of the TALE repressors are powerful tools to knock down VEZF1 expression in future studies.

A Golden Gate assembly system was used to assemble TALE repeats. The system proved to be simple and reliable and relatively time efficient given the number of constructs assembled. Importantly, no evidence of unwanted recombinants of the TALE repeats was observed at any stage. There are more rapid methods of TALE construction available to the academic community, but these systems require liquid handling robots as they rely on larger vector libraries and/or work on solid phase supports (Reyon et al., 2012b) The other GG assembly system of interest to individual research groups is one developed by the Zhang group, which relies on a library of PCR primers rather than plasmids, as components for assembly reactions (Zhang et al., 2011). I opted against this system as sub-optimal PCR could reduce the performance of the system and/or introduce error.

Another key advantage of stepwise GG assembly systems is that pre-assembled TALE DNA-binding domains can be quickly inserted into different expression vectors to produce TALE effector proteins such transcription activators and repressors or site-specific nucleases for example. At the time of release of Voytas GG assembly system there was no mammalian-based vector platform for creating TAL effectors or nucleases. I developed a mammalian expression vector to produce TALE repressor proteins using this system. Six of the ten TALE repressor proteins produced in this study were found to direct repression of *VEZF1.* The TALEREP 13 and TALEREP 21/23 proteins were more potent repressors than the other TALEREPs. Interestingly, these two proteins target the same genomic site at the *VEZF1* promoter, with TALEREP 21/23 recognising an additional 3 bp target (Figure 4.2). The superior performance of these two TALEREPs not only indicates that their target sequence is accessible, but that TALEREP binding may interfere with the binding of a key

transcription factor at the *VEZF1* promoter. Alternatively, this position is optimal for SIN3-mediated repression of the *VEZF1* promoter.

The TALEREPs 15, 16, 18 and 22 resulted in moderate repression of VEZF1. The degree of knockdown of VEZF1 RNA and protein levels varied between the TALEREPs. This is probably due to the fact that RNA and proteins samples were both collected three days after transfection. If a TALEREP has slow kinetics of repression due to low target site accessibility, RNA levels may be just beginning to reduce at day 3. Protein levels may be unaffected at this time. SIN3 acts as a scaffold to recruit co-repressors that both remove histone modifications that are favourable to transcription (H3/H4 acetylation and H3K4 methylation) and add marks that are repressive to transcription (H3K9 methylation and possibly DNA methylation) (van Oevelen et al., 2008, Yang et al., 2003). It is possible that analyses of VEZF1 expression after longer periods of exposure to TALEREPs will reveal greater degrees of repression with more of the TALEREPs. Nevertheless, the aim of this study was to identify TALE domains that can access their sequence readily after transient transfection. This information can be taken into consideration when developing TALE nuclease strategies using these TALE proteins.

Interestingly, there appears to be a correlation between level of *VEZF1* knockdown mediated by a TALE Repressor and TALE domain length. The longest TALEs, found in TALEREPs 14, 17/19, 20 and 24, recognise the longest sites at 28 or 29 bp, but had little effect on *VEZF1* expression. TALEREPs 16 and 18, which recognise 22 and 23 bp sites, provided moderate levels of VEZF1 repression. TALEREP 15, which binds a 19 bp site, performed consistently to mediate a moderate knockdown. The greatest level of *VEZF1* repression was observed using the shortest TALE domains. TALEREPs 13 and 21/23 recognise 16 and 19 bp sites, respectively. The correlation between decreasing TALE length and increasing repression performance may be a coincidence, and the results may differ if the effects of TALEREPs are studied for longer periods, given the inconsistencies observed between biological replicates (e.g. TALEREP 16) and in the disagreeing levels of repression in RNA versus protein expression (e.g. TALEREP 22, run 2). However, it is possible that the reduced, or slower, performance of TALEREPs with increasing binding site length may be due to competition for binding with endogenous transcription factors. Further studies are required to address this possibility. Researchers may need to balance the improved site specificity of longer TALE proteins with reduced DNA accessibility. A

comparison of *in vitro* DNA-binding affinity of each TALE protein with their *in vivo* performance would be helpful to address this possibility.

From this work, it has been demonstrated that at least six of the ten TALEs designed and constructed are functional at their intended chromosomal targets after a short period of expression. These TALE DNA-binding domains should be useful in strategies to create TALE nucleases that can direct genetic alteration of the *VEZF1* gene to create a knockout of VEZF1 expression in somatic human cells, which is described in Chapters 5 and 6.

Chapter 5

Production of TALENs to Modify the *VEZF1* Gene Promoter

5.1 Introduction

Transcription activator like effectors (TALEs) are a class of DNA binding proteins from the plant pathogen *Xanthomonas Spp.* (Boch et al., 2009). Unlike other DNA-binding domains, the RVD domains of TALE proteins recognise a single DNA base, which allows the modular assembly of engineered proteins of designed DNA sequence specificity. I have designed and assembled 10 TALE domains to target DNA sequences of the *VEZF1* gene promoter (Chapter 5). Following the successful validation of *VEZF1* promoter targeting, the next aim is to fuse TALE domains to the catalytic domain of the *Fok*I endonuclease to create TALE nucleases (TALENs) that can be employed to mutate the promoter of *VEZF1*.

The ability to knockout the *VEZF1* gene would greatly assist the study of VEZF1 function. VEZF1 is broadly expressed across many cell types and is known to bind at the promoters of many genes (Section 1.5) VEZF1 has been demonstrated to be essential in embryonic vascular development (Kuhnert et al., 2005, Zou et al., 2010) and also in the regulation of the angiogenic and lymphangiogenic phases in wound healing in adults (Gerald et al., 2013). Its mechanism of action is unknown due to the difficulties faced in generating a sufficient level of *VEZF1* knockdown using RNAi. Previous efforts to generate VEZF1 knockdown lines have resulted in reduction of VEZF1 expression by up to 80% but this resulted in only partial depletion of VEZF1 binding at its target sites *in vivo* (Strogantsev, 2009).

Murine *Vezf1* is essential for cell survival and a recent study by a colleague has revealed that human VEZF1 is bound at the promoters of the majority of transcribing genes (Low, 2013), there is a possibility that human VEZF1 promoter mutation may impact upon normal cell growth. The straightforward delivery of either ZFNs or TALENS into cells would result in a variety of uncontrolled *VEZF1* mutations as a result of double stranded break repair by non-homologous end joining (NHEJ). NHEJ results in mutations that are usually small insertions or deletions (Durai et al., 2005). It may be challenging to isolate *VEZF1* null cells from a population of heterozygous and silent mutations, particularly if *VEZF1* null cells have a reduced growth phenotype.

It is therefore more desirable to undertake a gene targeting approach where homology-directed repair (HDR) of a DSB using a designed targeting vector is used to create desired mutations that can be readily selected for. This strategy will require the assembly of a targeting vector that contains a selectable transgene cassette flanked by regions of sequence homology to sequences surrounding the *VEZF1* promoter. HDR-mediated insertion of this transgene cassette would result in complete loss of the *VEZF1* promoter and first exon on targeted alleles. This approach can be used to create and screen for *VEZF1* null cells and/or create inducible *VEZF1* null cells.

A number of different TALEN architectures have been described, where truncations of the TALE domain or mutations of the *Fok*I nuclease domains have been developed to improve DNA-binding activity or target specificity (Section 1.3.4.1.2). There were no vectors developed for the mammalian expression of TALENs using the Voytas Golden Gate assembly system as the time of starting this study. I therefore needed to develop expression vectors that were compatible with this system and incorporated the latest developments in TALEN architecture.

5.2 Aims of this chapter

- Design and construct mammalian expression vectors that assemble validated TALE domains into TALE Nucleases (TALENs)

- Devise a strategy for performing homology directed repair (HDR) at the *VEZF1* promoter using a targeting vector

- Develop ZFN/TALEN-stimulated HDR based targeting as an approach to generating designed *VEZF1* gene mutations

5.3 Development of TALE Nuclease Expression Vectors

A Golden Gate assembly system was utilised to assemble TALE domains that were then validated to target the VEZF1 promoter in cultured human cells in the context of monomeric TALE repressor proteins (Chapter 4). Destination vectors for the Voytas Golden Gate (GG) Assembly system that enable mammalian expression of TALENs were not available at the time of completing this study (Cermak et al., 2011). I considered that the following requirements should be met in order to maximise the performance of TALENs:

1. A strong mammalian promoter should be used
2. The TALE domain should contain the Miller truncations NΔ152 and C+63 for maximal nuclease activity in mammalian cells (Szurek et al., 2002, Miller et al., 2011)
3. Epitope tags should be added for Western blotting analysis of TALEN expression, if required
4. The cloning vector should be compatible with the Voytas GG TALE assembly system, with a selection marker for correct cloning events
5. Mutations should be incorporated into the *Fok*I nuclease domain to allow for obligate heterodimer function and enhanced cleavage
6. The TALEN coding sequences should be optimised for human codons

In order to meet these requirements, new TALEN expression vectors had to be assembled. It was clear that the coding sequences for the N- and C-termini of the TALE protein needed to be generated by DNA synthesis in order to meet our requirements. I decided that the TALE transcription activator plasmid, pTALETF_v2 would be the best starting point as it contained a strong CMV promoter, SV40 polyA and enhancer for mammalian expression and a central ccdB negative selection marker that can be used to screen for GG insertions (Sanjana et al., 2012). Firstly, I needed to delete a portion of the pTALETF_v2 plasmid which contained an *Esp*3I site, as this is the restriction site that is used in the Voytas GG TALE assembly system (Figure 5.1). A 1,032 bp fragment containing an unnecessary hygromycin resistance gene was excised using *Pml*I and the plasmid was re-circularised by blunt ended ligation (pTALETFΔhygro).

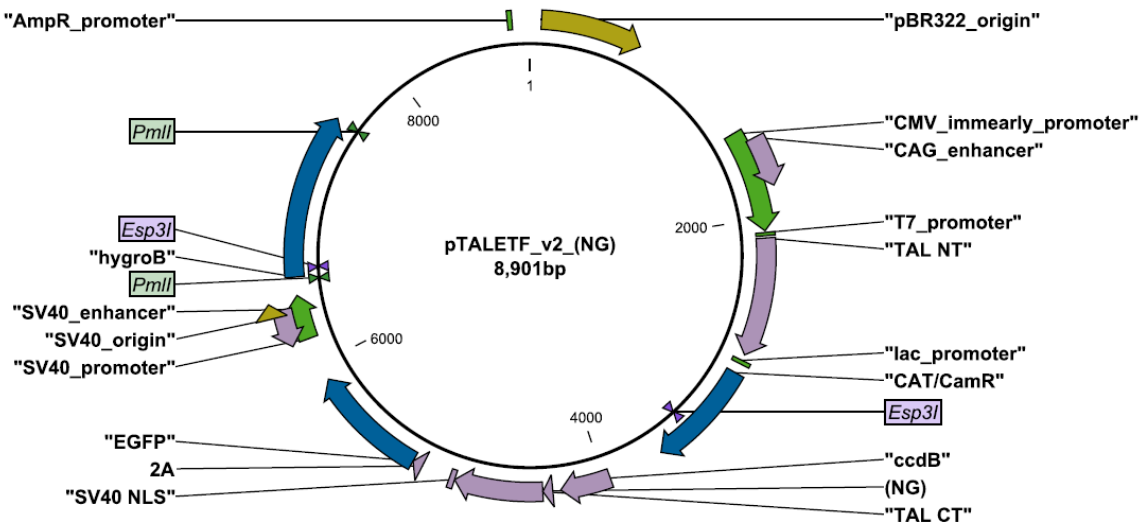

Figure 5.1 Plasmid Map of pTALETF_v2
Scaled map of pTALETF_v2 showing CMV promoter and CAG/T7 enhancer from where transcription of inserted TALEs initiates, TAL N- & C-termini and SV40 NLS. Hygromycin resistance gene (*hygroB*) containing an *Esp*3I site, and *Esp*3I site used in Golden Gate Assembly are also indicated. *Pml*I restriction sites shown were used to remove this region.

5.3.1 Creation of a designed TALE N-terminus

In order to obtain a TALE Δ152 N-terminus coding sequence that i) has an additional 3x FLAG-tag, ii) is human codon optimised and iii) has an *Esp*3I site compatible with the Voytas GG system, I had to replace the TALE N-terminus sequences already present between the *Sac*I and *Not*I restriction sites in pTALETF_v2 with a designed fragment (Figure 5.2) to create the vector pTALv2_NT (Figure 5.3). The insertion of the *Sac*I - *Nco*I TALv2_NT fragment was confirmed by restriction digestion of pTALv2_NT plasmid DNA with *Sac*I & *Not*I (Figure 5.4, Lanes 2 – 7) and by digest with *Nco*I (Figure 5.4, Lanes 8 – 13).

A

B

>TALv2NT SacI-NotI CMVpro T7 Kozak FLAGx3 NLS TALΔ152 **SapI AgeI**
vector **Esp3I-CUT**
GAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGG
GGCCACCATGGACTATAAGGATCATGACGGCGACTACAAGGACCACGACATCGACTACAAGGACGACG
ACGACAAGATGGCCCCCAAGAAGAAGCGCAAGGTGGGCATCCACGGCGTGCCCGCCGCCGTGGACCTG
CGCACCCTGGGCTACAGCCAGCAGCAGCAGGAGAAGATCAAGCCCAAGGTGCGCAGCACTGTCGCCCA
GCACCACGAGGCCCTGGTGGGCCACGGCTTCACCCACGCCCACATCGTGGCCCTGAGCCAGCACCCCG
CTGCACTGGGCACCGTGGCCGTGAAGTACCAGGACATGATCGCC**GCTCTTC**CGGAAGCCACCCACGAG
GCCATCGTGGGCGTGGGCAAGCAGTGGAGCGGCGCCCGCGCCCTGGAGGCCCTGCTGACCGTGGCCGG
CGAGCTGCGCGGCCCCCCCCTGCAGCTGGACACCGGCCAGCTGCTGAAGATCGCCAAGCGCGGCGGCG
TGACCGCCGTGGAGGCCGTGCACGCCTGGCGCAACGCCCTG**ACCGGT**GCCC**CCCT**G**GAGACG**ctgaca
gagacc**GCGGCCGC**

Figure 5.2 Design of the synthesised TALE v2 N-terminus fragment TALv2NT
A) Cartoon of the designed TALE N-terminus. B) Sequence of the synthesised *Sac*I-*Not*I TALv2NT fragment. The humanised AvrBs3 TALE coding sequence matched that of TALE effector vectors developed by the Zhang lab (Sanjana et al., 2012). The 5' end of the TALv2_NT fragment contained the remainder of the CMV promoter (green) after the *Sac*I cloning site (light blue) followed by i) a T7 promoter (brown) for future *in vitro* transcription work, ii) a Kozak translation initiation site (red), iii) a 3x FLAG-tag epitope (mid blue) to monitor TALE protein expression and iv) a nuclear localisation signal (NLS, grey). The 3' end of the TALv2NT fragment ends with an *Esp*3I site that creates the same CCCT overhang utilised in the Voytas GG system (underlined).

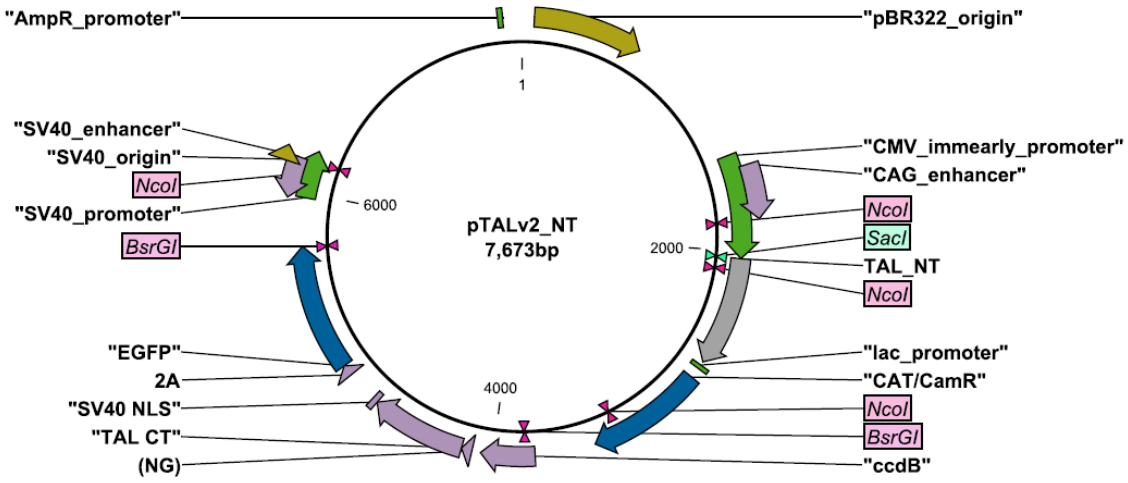

Figure 5.3 Plasmid Map of pTALv2_NT
Scaled map of pTALv2_NT showing the modified TALv2 N-terminus inserted using restriction sites *Sac*I & *Not*I. Insertion of the N-terminus expression cassette was confirmed using *Nco*I, which is also annotated. TAL C terminus and EGFP were subsequently replaced with the modified TALv2 C-terminus using *Bsr*GI (section 5.3.2).

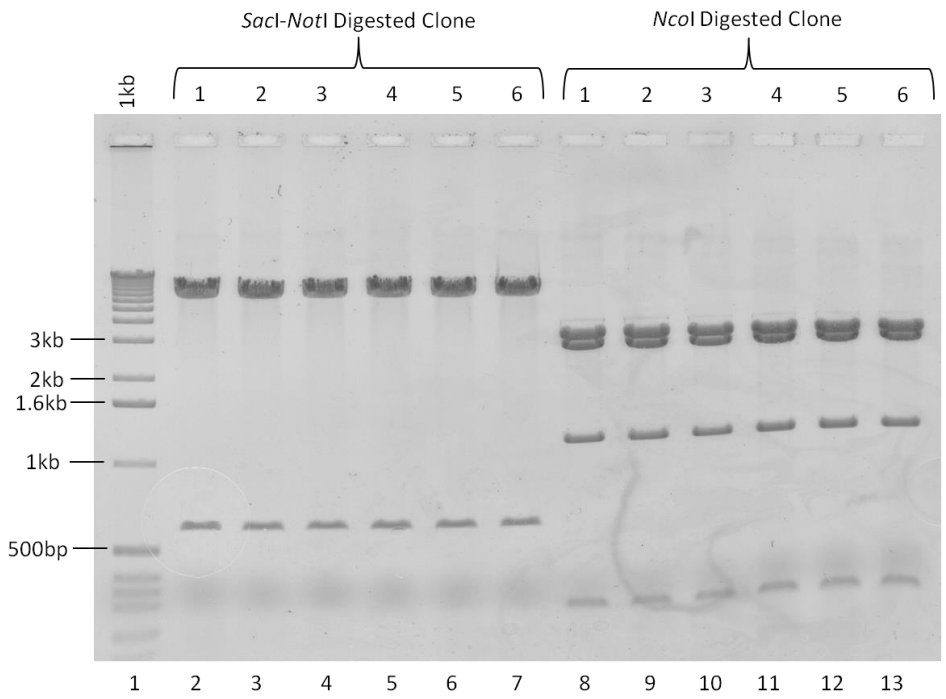

Figure 5.4 Restriction analysis of pTALv2_NT clones
Agarose gel electrophoresis of restriction digested plasmid DNA from transformants gained following ligation of the TALv2_NT into pTALETF_v2. Correct insertion of the TALv2NT fragment results in the excision of a 619 bp fragment with *Sac*I & *Not*I, whereas the original vector would yield an 811 bp fragment. All six clones analysed had the expected insert for TALv2NT (lanes 2-7). Insertion of the TALv2NT fragment introduces an additional *Nco*I site. All six clones resulted in the excision of 1167 & 277 bp fragments expected for pTALv2_NT (lanes 8 – 13).

5.3.2 Creation of designed TALEN C-termini

In order to create a TALEN C-terminus that i) encoded the TALE +63 C-terminus fused to the wild type *Fok*I nuclease domain, ii) is human codon optimised and iii) is has an *Esp*3I site compatible with the Voytas Golden Gate Assembly system, the incompatible TALE C-terminus and EGFP sequences already present between the two *Bsr*GI restriction sites in pTALv2NT (Figure 5.3) had to be replaced with a designed TALv2CT-FokIWT fragment (Figure 5.5) to create the vector pTALv2-FokIWT (Figure 5.6).

A

B

>TALv2CT-FokWT **BsrGI** ccdB **Esp3I** TAL_CT FokIWT **XhoI** **BsaI**

TGTACAGAGTGATATTATTGACACGCCCGGGCGACGGATGGTGATCCCCCTGGCCAGTGCACGTCTGCTG
TCAGATAAAGTCTCCCGTGAACTTTACCCGGTGGTGCATATCGGGGATGAAAGCTGGCGCATGATGACCA
CCGATATGGCCAGTGTGCCGGTCTCCGTTATCGGGGAAGAAGTGGCTGATCTCAGCCACCGCGAAAATGA
CATCAAAAACGCCATTAACCTGATGTTCTGGGGAATATAATGTCAGGCTCCCTTATACACAGCCAGTCT
GCAGGTCGACCGTCTCCAACGACCACCTGGTGGCCCTGGCCTGCCTGGGCGGACGCCCtGCtCTGGACGC
CGTGAAGAAGGGCCTGCCCCACGCCCCCGCCCTGATCAAGCGCACCAACCGCCGCATCCCCGAGCGCACC
AGCCACCGCGTGGCCGGCAGCCAGCTGGTGAAGAGtGAaCTGGAGGAGAAGAAGAGCGAGCTGCGCCACA
AGCTGAAGTACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCCGCAACAGCACCCAGGACCGCAT
CCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACCGCGGCAAGCACCTGGGCGGCTCG
AGAAAGCCCGACGGCGCCATCTACACCGTGGGCAGCCCCATCGACTACGGCGTGATCGTGGACACCAAGG
CCTACAGCGGCGGCTACAACCTGCCCATCGGCCAGGCCGACGAGATGCAGCGCTACGTGGAGGAGAACCA
GACCCGCAACAAGCACATCAACCCCAACGAGTGGTGGAAGGTGTACCCCAGCAGCGTGACCGAGTTCAAG
TTCCTGTTCGTGAGCGGCCACTTCAAGGGCAACTACAAGGCCCAGCTGACCCGCCTGAACCACATCACCA
ACTGCAACGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGCGGCGAGATGATCAAGGCCGGCACCCT
GACCCTGGAGGAGGTGCGCCGCAAGTTCAACAACGGCGAGATCAACTTCTAATGTACA

Figure 5.5 Design of the synthesised TALEN C-terminus fragment TALv2CT-FokIWT
A) Cartoon of the designed TALEN C-terminus. B) Sequence of the synthesised *Bsr*GI TALv2CT-FokIWT fragment. The humanised AvrBs3 TALE coding sequence matched that of TALE effector vectors developed by the Zhang lab (Sanjana et al., 2012). The 5' end of the TALv2CT-FokIWT fragment contained the remainder of the ccdB selection marker (blue) *Bsr*GI cloning site (purple) followed by i) an *Esp*3I site that creates the same AACG overhang utilised in the Voytas Golden Gate system, ii) the humanised AvrBs3 TALE +63 C-terminus coding sequence (orange) as used previously (Sanjana et al., 2012), iii) a humanised coding sequence for the wild type *Fok*I cleavage domain (green). Silent mutations in the *Fok*I domain were included to create a *Xho*I restriction site (dark green) that could be used to exchange *Fok*I domain mutation in future studies.

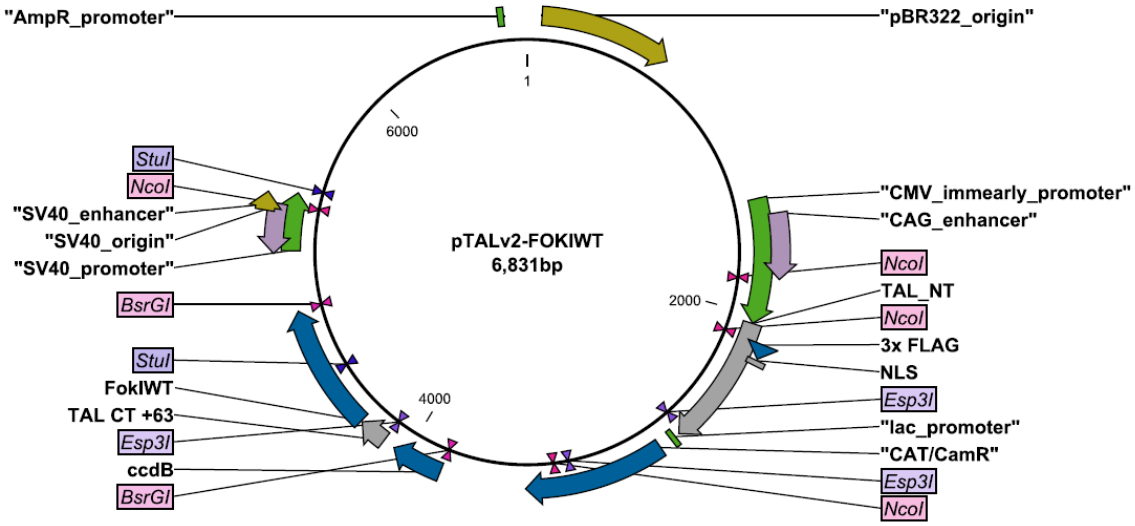

Figure 5.6 Plasmid Map of pTALv2-FokIWT
Scaled map of pTALv2-FokIWT showing the previously modified TALE N-terminus and the present insertion of the TALEN C-terminus at *Bsr*GI. Insertion of the C-terminus was confirmed using *Nco*I & *Stu*I, which are also annotated.

The correct insertion of the TALv2CT-FokIWT fragment into pTALv2-FokIWT was confirmed by restriction digestion of the plasmid with *Nco*I, which excises the inserted TALEN C-terminus sequence from the plasmid backbone (Figure 5.7, Lanes 2 – 9). The correct orientation of the cloned TALEN C-terminus had to be determined as the cloning strategy involved a pair of compatible *Bsr*GI sites. Restriction digestion with *Stu*I, which has a recognition sequence present once in the cloned TALEN C-terminus and once in the plasmid backbone identified that 4 of the 8 clones analysed had the TALEN C-terminus cloned in the desired orientation (Figure 5.7, Lanes 10 – 17). Clones in the correct orientation can be seen in lanes 10, 14, 15 & 17.

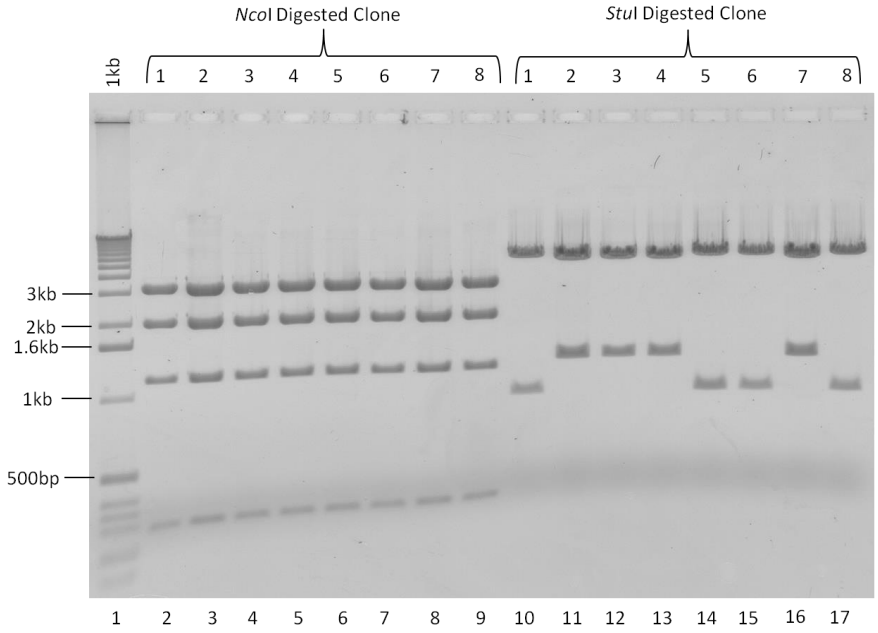

Figure 5.7 Restriction analysis of pTALv2-Fok1WT clones
Agarose gel electrophoresis of restriction digested pTALv2-FokIWT plasmid DNA clones resulting from ligation of the TALv2CT-FokIWT fragment into pTALv2_NT. Digestion with *Nco*I confirmed that all eight clones contained the TALv2CT-FokIWT fragment as each had the expected pattern of 3326, 2001, 1167 and 277 bp digestion products (lanes 2 – 9). Digestion with *Stu*I identified that only four of the clones contained the TALv2CT-FokIWT fragment in the desired orientation, where a 911 bp product is excised (lanes 10, 14, 15 & 17).

Two further versions of the TALEN C-terminus were created that incorporate previously characterised mutations of *Fok*I that i) cause it to function as an obligate heterodimer and ii) increase cleavage activity. This obligate heterodimer mutations are the same as those employed in ZFN construction (Chapter 3), namely Q486E and I499L ("EL") in the right hand monomer and E490K, I538K ("KK") in the left hand monomer. However, I have also included the N496D and H537R mutations to create "ELD" and "KKR". The D/R mutations used in addition to EL/KK mutations have been shown to have greater cleavage activity than either WT or EL/KK versions of *Fok*I (Doyon et al., 2011). In addition, I have added

the so-called Sharkey mutations S418P and K441E to both ELD and KKR as the Sharkey mutations have been reported to result in improved cleavage activity of *Fok*I heterodimer mutants (Guo et al., 2010). The resulting heterodimer TALENs ELD-S and KKR-S are hereafter referred to as version 3 (v3) TALENs.

A

B

>TALv2CT-FokELD-S (Q486E,I499L,N496D - S418P,K441E)
BsrGI ccdB **Esp3I** TAL_CT FokI **XhoI** **BsaI**
TGTACAGAGTGATATTATTGACACGCCCGGGCGACGGATGGTGATCCCCCTGGCCAGTGCACGTCTGC
TGTCAGATAAAGTCTCCCGTGAACTTTACCCGGTGGTGCATATCGGGGATGAAAGCTGGCGCATGATG
ACCACCGATATGGCCAGTGTGCCGGTCTCCGTTATCGGGGAAGAAGTGGCTGATCTCAGCCACCGCGA
AAATGACATCAAAAACGCCATTAACCTGATGTTCTGGGGAATA**TAA**ATGTCAGGCTCCCTTATACACA
GCCAGTCTGCAGGTCGAC**CGTCTC**CAACGACCACCTGGTGGCCCTGGCCTGCCTGGGCGGACGCCCtG
CtCTGGACGCCGTGAAGAAGGGCCTGCCCCACGCCCCCGCCCTGATCAAGCGCACCAACCGCCGCATC
CCCGAGCGCACCAGCCACCGCGTGGCCGGCAGCCAGCTGGTGAAGAGtGAaCTGGAGGAGAAGAAGAG
CGAGCTGCGCCACAAGCTGAAGTACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCCGCAACC
CCACCCAGGACCGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACCGCGGC
GAGCACCTGGGCGG**CTCGAG**AAAGCCCGACGGCGCCATCTACACCGTGGGCAGCCCCATCGACTACGG
CGTGATCGTGGACACCAAGGCCTACAGCGGCGGCTACAACCTGCCCATCGGCCAGGCCGACGAGATGG
AGCGCTACGTGGAGGAGAACCAGACCCGCGACAAGCACCTGAACCCCAACGAGTGGTGGAAGGTGTAC
CCCAGCAGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGCGGCCACTTCAAGGGCAACTACAAGGCCCA
GCTGACCCGCCTGAACCACATCACCAACTGCAACGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCG
GCGGCGAGATGATCAAGGCCGGCACCCTGACCCTGGAGGAGGTGCGCCGCAAGTTCAACAACGGCGAG
ATCAACTTC**TAA**TGTACA

>TALv2CT-FokKKR-S (E490K,I538K,H537R - S418P,K441E)
BsrGI ccdB **Esp3I** TAL_CT FokI **XhoI** **BsaI**
TGTACAGAGTGATATTATTGACACGCCCGGGCGACGGATGGTGATCCCCCTGGCCAGTGCACGTCTGC
TGTCAGATAAAGTCTCCCGTGAACTTTACCCGGTGGTGCATATCGGGGATGAAAGCTGGCGCATGATG
ACCACCGATATGGCCAGTGTGCCGGTCTCCGTTATCGGGGAAGAAGTGGCTGATCTCAGCCACCGCGA
AAATGACATCAAAAACGCCATTAACCTGATGTTCTGGGGAATA**TAA**ATGTCAGGCTCCCTTATACACA
GCCAGTCTGCAGGTCGAC**CGTCTC**CAACGACCACCTGGTGGCCCTGGCCTGCCTGGGCGGACGCCCtG
CtCTGGACGCCGTGAAGAAGGGCCTGCCCCACGCCCCCGCCCTGATCAAGCGCACCAACCGCCGCATC
CCCGAGCGCACCAGCCACCGCGTGGCCGGCAGCCAGCTGGTGAAGAGtGAaCTGGAGGAGAAGAAGAG
CGAGCTGCGCCACAAGCTGAAGTACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCCGCAACC
CCACCCAGGACCGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACCGCGGC
GAGCACCTGGGCGG**CTCGAG**AAAGCCCGACGGCGCCATCTACACCGTGGGCAGCCCCATCGACTACGG
CGTGATCGTGGACACCAAGGCCTACAGCGGCGGCTACAACCTGCCCATCGGCCAGGCCGACGAGATGC
AGCGCTACGTGAAGGAGAACCAGACCCGCAACAAGCACATCAACCCCAACGAGTGGTGGAAGGTGTAC
CCCAGCAGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGCGGCCACTTCAAGGGCAACTACAAGGCCCA
GCTGACCCGCCTGAACCGCAAGACCAACTGCAACGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCG
GCGGCGAGATGATCAAGGCCGGCACCCTGACCCTGGAGGAGGTGCGCCGCAAGTTCAACAACGGCGAG
ATCAACTTC**TAA**TGTACA

Figure 5.8 Design of the synthesised heterodimer TALEN C-terminus fragments
A) Cartoon of the designed TALEN C-termini. B) Sequence of the synthesised *Bsr*GI TALv2CT-FokELD-S and TALv2CT-FokKKR-S fragments. The composition of the fragments is as described for TALv2CT-FokIWT in Figure 5.5, except that the indicated mutations are present in the *Fok*I coding sequence.

The ELD-S and KKR-S forms of *Fok*I were created by DNA synthesis of the TALEN C-terminus fragments TALv2CT_FokELD-S and TALv2CT_FokKKR-S (Figure 5.8). These were cloned into the *Bsr*GI restriction sites of pTALv2NT as described for TALv2CT-FokIWT above (not shown) to create the vectors pTALv3-FokIELD-S and pTALv3-FokIKKR-S. During

the course of this work, an additional destination vector for the Voytas GG system that creates a wild type *Fok*I TALEN for mammalian expression called pCAG-T7-TALEN became available (Addgene.org, plasmid 37184). I acquired this vector as other groups have begun to use it to produce functional TALENs, and refer to pCAG-T7-TALEN as pTALv1-FokIWT (Table 1.1).

Name	FokI	Human codon optimised	References
pTALv1-FokIWT	Wild type	No	Pelczar (Unpublished)
pTALv2-FokIWT	Wild type	Yes	(Miller et al., 2011, Sanjana et al., 2012)
pTALv3-FokI-ELDS	S418P, K441E, Q486E, N496D and I499L	Yes	(Miller et al., 2007, Guo et al., 2010, Doyon et al., 2011)
pTALv3-FokI-KKRS	S418P, K441E, E490K, H537R and I538K	Yes	(Miller et al., 2007, Guo et al., 2010, Doyon et al., 2011)

Table 5.1 Destination vectors for mammalian TALEN expression
All of these vectors create TALEN assemblies with Δ152 TALE N-termini and +63 C-termini.

5.3.3 Assembly of Designed TALE RVD Domains into Nuclease Expression Vectors

All ten of the TALE domains that target the human *VEZF1* gene promoter were assembled into pTALv1, pTALv2 or pTALv3 vectors as described in Table 5.2. A plasmid pick sheet was used to set up Golden Gate assembly reactions (similar to that shown in Figure 4.14) of pTAL A, B and C constructs of TAL RVDs and the desired *Fok*I expression vector to assemble pTALEN vectors. An example TALEN vector is shown in Figure 5.9. The correct assembly of TALE RVDs into pTAL plasmids was confirmed by *Sac*I – *Xho*I digestion of the resulting pTALEN plasmids. An example is shown in Figure 5.10. The pTALEN plasmids can be combined in transfection experiments to produce 18 TALEN pairs. These TALEN pairs are subsequently referred to by their original TALE names followed by the TALEN version, for example TALEN 13/14_v1 or 13/14_v2 etc.

v1 TALEN (standard homodimer)	v2 TALEN (optimised homodimer)	v3 TALEN (optimised heterodimer)	Configuration (L-Space-R)	DNA Recognition Sequence
TALE13-WT-L + TALE14-WT-R	TALE13-WT*-L + TALE14-WT*-R	TALE13-ELDS*-L + TALE14-KKRS*-R	16-30-28	TALE 13 GTCGCCTCCTCCGGCT---AGTGGGGAGGAGGGGGGTCGGCCGCCGC CAGCGGAGGAGGCCGA---TCACCCCTCCTCCCCCCAGCCGGCGGCG TALE 14
TALE15-WT-L + TALE16-WT-R	TALE15-WT*-L + TALE16-WT-R	TALE15-ELDS*-L + TALE16-KKRS*-R	19-15-23	TALE 15 CGCCGCCGCGGTGGTGCGT---AAGTGGCCGCGCTCGGCGCTGCC GCGGCGGCGCCACCACGCA---TTCACCGGCGCGAGCCGCGACGG TALE 16
TALE17/19-WT-L + TALE18-WT-R	TALE17/19-WT*-L + TALE18-WT*-R	TALE17/19-ELDS*-L + TALE18-KKRS*-R	28-15-22	TALE 17/19 GCCGCCGCCGCCGCCGCCGCTTTGTTGT---AGGAGGCTCCCCGAGCGGGGGGAGTGGGG CGGCGGCGGCGGCGGCGGCGAAACAACA---TCCTCCGAGGGGCTCGCCCCCCTCACCCC TALE 18 TALE 20
TALE17/19-WT-L + TALE20-WT-R	TALE17/19-WT*-L + TALE20-WT*-R	TALE17/19-ELDS*-L + TALE20-KKRS*-R	28-15-29	
TALE21/23-WT-L + TALE22-WT-R	TALE21/23-WT*-L + TALE22-WT*-R	TALE21/23-ELDS*-L + TALE22-KKRS*-R	19-30-25	TALE 21/23 GTTGTCGCCTCCTCCGGCT---AGGAGGGGGGTCGGCCGCCGCAGCCATGG CAACAGCGGAGGAGGCCGA---TCCTCCCCCCAGCCGGCGGCGTCGGTACC TALE 22 TALE 24
TALE21/23-WT-L +TALE24-WT-R	TALE21/23-WT*-L +TALE24-WT*-R	TALE21/23-ELDS*-L +TALE24-KKRS*-R	19-30-29	

Table 5.2 TALEN assemblies that target the human *VEZF1* gene promoter

The TALE domains characterised in Chapter 4 (Table 4.1) were assembled into pTALv1-FokIWT (v1), pTALv2-FokIWT (v2) or pTALv3-FokI-ELDS/KKRS (v3) using the Voytas Golden Gate system. The different versions of TALENs are v1 = wild type FokI, non-optimised ("WT"), v2 = wild type FokI, human codon optimised ("WT*") and v3 = heterodimer ELDS/KKRS FokI, human codon optimised ("ELDS*/KKRS*"). All of the TALENs share the same Δ152N/+63C TALE architecture.

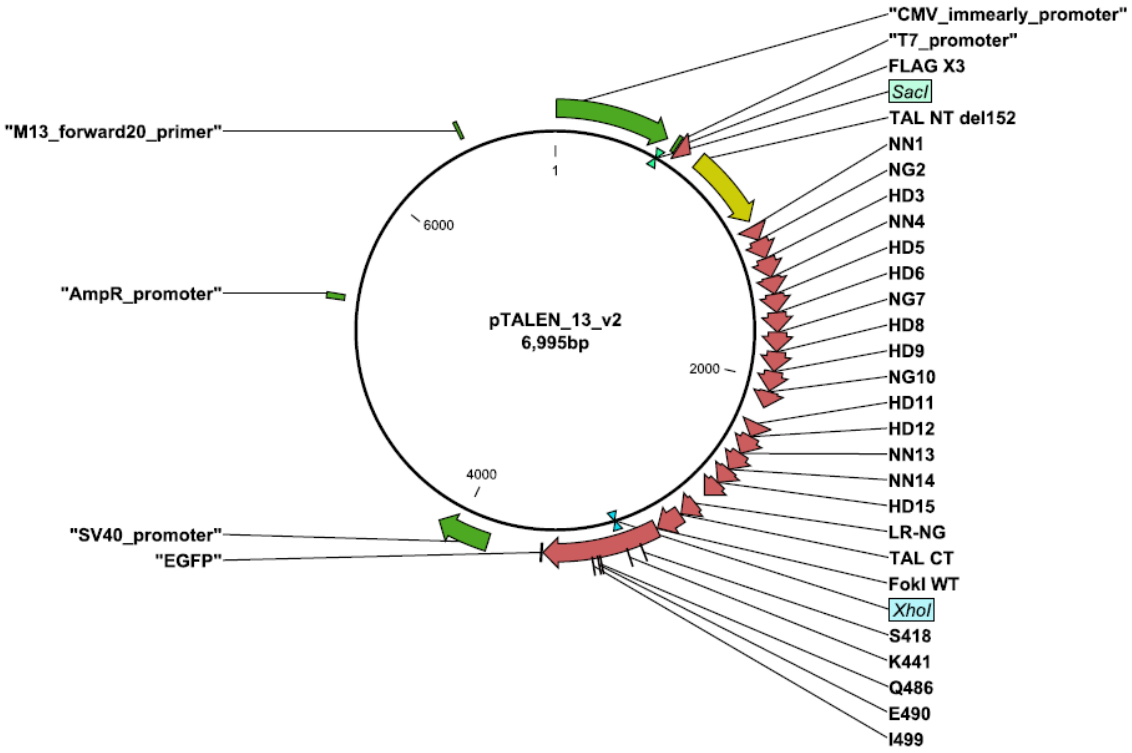

Figure 5.9 Plasmid map of pTALEN_13_v2
Scaled map showing pTALEN_13_v2, which was produced from the assembly of TAL-A13 and TAL-B13 into pTALv2-FokIWT. Restriction sites *Sac*I and *Xho*I were used to confirm the correct assembly of TALE RVDs.

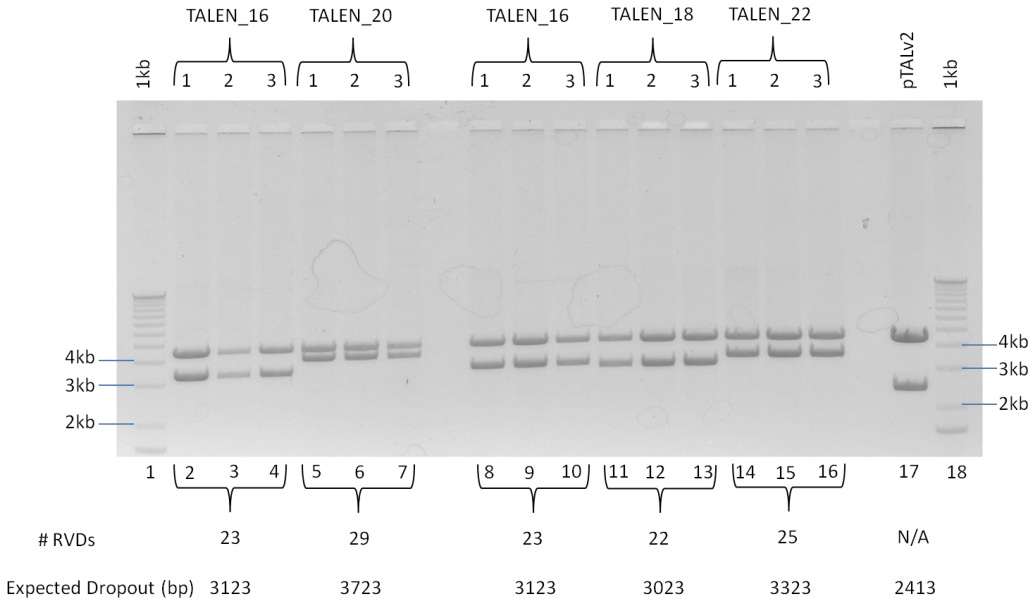

Figure 5.10 Restriction analysis of pTALEN vector assemblies
Agarose gel electrophoresis of restriction digested plasmid DNA from five different TALE assemblies into pTALv2-FokIWT. Plasmid DNA from three clones of each pTALEN assembly was digested with *Sac*I and *Xho*I. Digestion of pTALv2-FokIWT excises a 2413 bp fragment (lane 17). Digestion of pTALEN vectors yields a fragment of 823 +100 bp per TALE RVD domain. The expected dropout sizes and number of TALE RVD domains in each clone is indicated below the gel.

5.4 Design of human *VEZF1* gene targeting vector

In order to allow for designed gene and promoter mutations, conditional gene knockout and knockout of a potentially essential gene like *VEZF1*, it is desirable to undertake gene targeting via homology directed repair (HDR). HDR occurs with very low efficiency in almost all somatic human cells, but this can be stimulated up to 50,000 fold by inducing a double strand DNA break (Jasin, 1996, Elliott et al., 1998). Our first targeting strategy involved a test to see whether ZFN- or TALEN-mediated cleavage can stimulate the targeted insertion of an EGFP transgene at the human *VEZF1* gene promoter. The EGFP transgene needs to be flanked by *VEZF1* gene sequences, known as homology arms. Homology arms are recommended to be around 750 bp -- 1 kb in length and to be within 200 bp of double strand break (Davis and Stokoe, 2010, Davis and Pruett-Miller, 2010, Urnov et al., 2010).

I designed an EGFP targeting vector for the *VEZF1* promoter, which includes a puromycin resistance cassette to allow for the selection of targeted cells (Figure 5.11). The 5' homology arm consists of a 725 bp portion of the upstream *VEZF1* promoter and the 3' arm consists of a 744 bp portion of intron 1. The arms flank the core promoter, 5'UTR and the short exon 1. The proximal ends of each arm are between 70 and 230 bp from the various ZFN and TALEN cleavage sites. HDR using these homology arms should replace the *VEZF1* core promoter, TSS, ATG and exon 1 with an *EGFP-PURO* transgene (Figure 5.13). HDR-mediated insertion of the *EGFP-PURO* transgene at the *VEZF1* target locus will result in the removal of the ZFN and TALEN binding sites, so further cleavage after targeting is not possible. This is with the exception of ZFN pair 37/38, which does have a cleavage site in the homology arms. This ZFN pair were designed and constructed prior to the design of the targeting strategy, and this ZFN pair is the only one which is designed to cleave in sequences downstream of exon 1 in *VEZF1*.

The homology arms were cloned into a transgene system created by Ms. Ileana Guerrini in the West group (data not shown). This SHuTtle-LEFT-RIGHT (SHTLR) system employs multisite Gateway cloning to quickly assemble 5' and 3' homology arm sequences onto different transgenes to facilitate gene targeting and other applications (Figure 5.11). The 5' VEZF1 homology arm was obtained by genomic PCR using the primers VEZF1_5'arm_H3F and VEZF1_5'arm_AscR and cloned into pSHTLR3 via *Hin*dIII and *Asc*I to

form pSHTLR3_VEZF1_R5'arm. The 3' VEZF1 homology arm was obtained using the primers VEZF1_3'arm_H3F and VEZF1_3'arm_AscR and cloned into pSHTLR5 via *Hin*dIII and *Asc*I to form pSHTLR5_VEZF1_R3'arm. An LR Gateway reaction between the pSHTLR plasmids, the pDestR4-R3 destination vector and a pSHTLRmid plasmid containing a CMV-driven EGFP-puro transgene was performed to form pTRGT_VEZF1 (Ileana Guerrini, Figure 5.12).

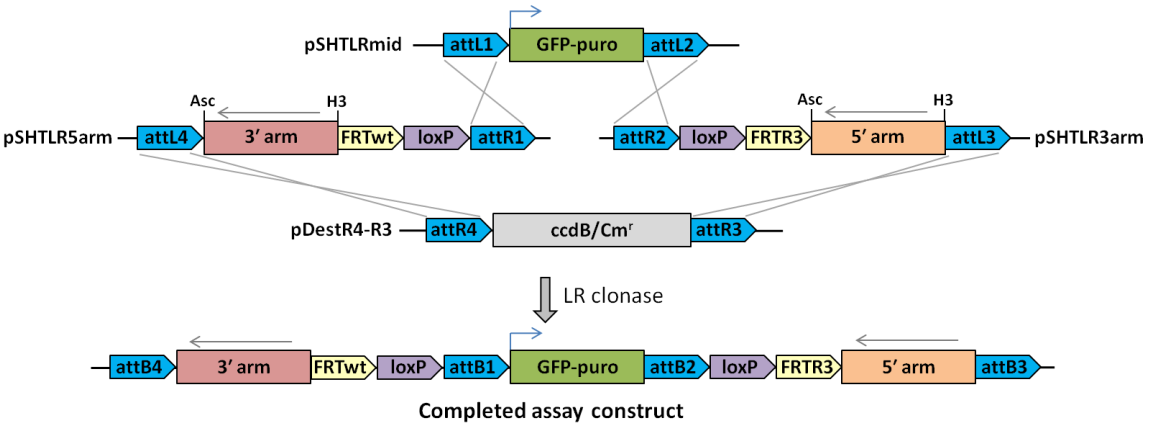

Figure 5.11 SHTLR cloning of gene targeting vectors
Schematic representation of targeting vector assembly by multisite Gateway cloning. Homology arms towards the target locus are conventionally cloned into multiple cloning sites locate in the pSHTLR5/3 plasmids. The locations of *Asc*I and *Hin*dII restriction sites used for the insertion of VEZF1 homology arms in a reverse orientation are shown. The transgene located in pSHTLRmid for this study encodes both EGFP and a puromycin resistance gene separate by a T2A self-cleavage peptide (Kim et al., 2011). LR clonase-mediated assembly results in the formation of pTRGT_VEZF1, where the GFP transgene is in the reverse orientation between the VEZF1 homology arms. The locations of recognition sequences for flp (FRT) and cre (loxP) recombinases are shown.

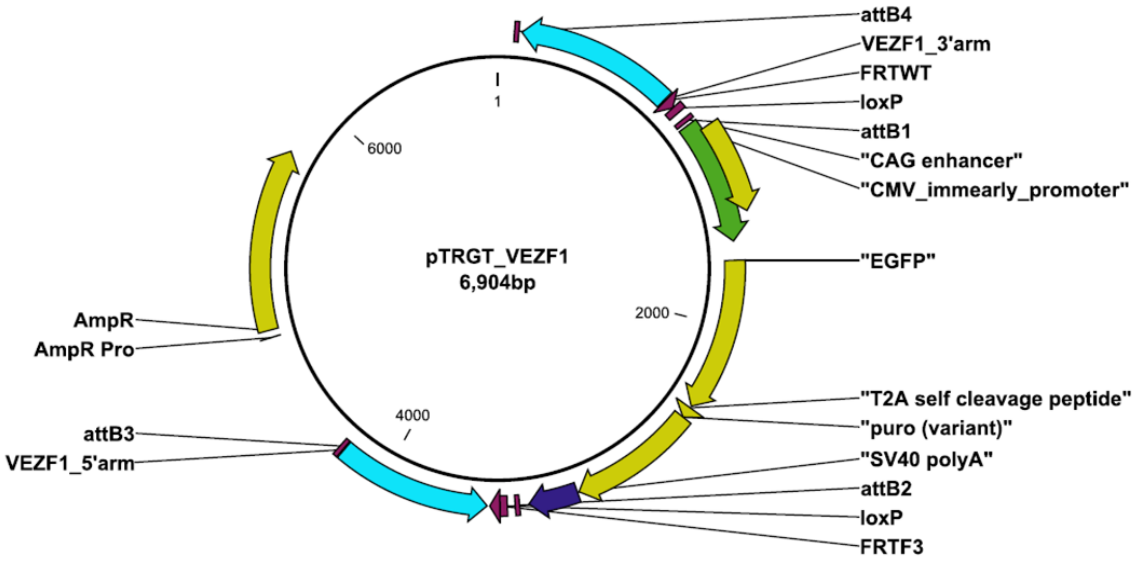

Figure 5.12 Plasmid Map of pTRGT_VEZF1
Scaled Map of pTRGT_VEZF1 showing VEZF1 homology arms flanking EFGP/puromycin transgene cassette expressed via CMV promoter. Also annotated are recognition sequences for flp (FRT) and cre (loxP) recombinases.

The VEZF1 GFP targeting vector was designed such that the *EGFP-PURO* transgene would become integrated in the reverse orientation to the *VEZF1* gene body, so that transgene transcription does not read through into the remaining *VEZF1* sequences (Figure 5.13). The transgene cassette also contains recognition sites for the site-specific recombinases Cre and Flp which enable the subsequent removal or exchange of the transgene. Cre-mediated recombination between the two loxP sites will remove the intervening *EGFP-PURO* transgene. Alternatively, Flp-mediated recombination at the FRTwt and FRTR3 sites will facilitate cassette exchange with a donor plasmid containing the same sites. The presence of these recombination sites will be key to future strategies to modify the *VEZF1* locus in a controlled manner, including conditional homozygous deletion of *VEZF1*.

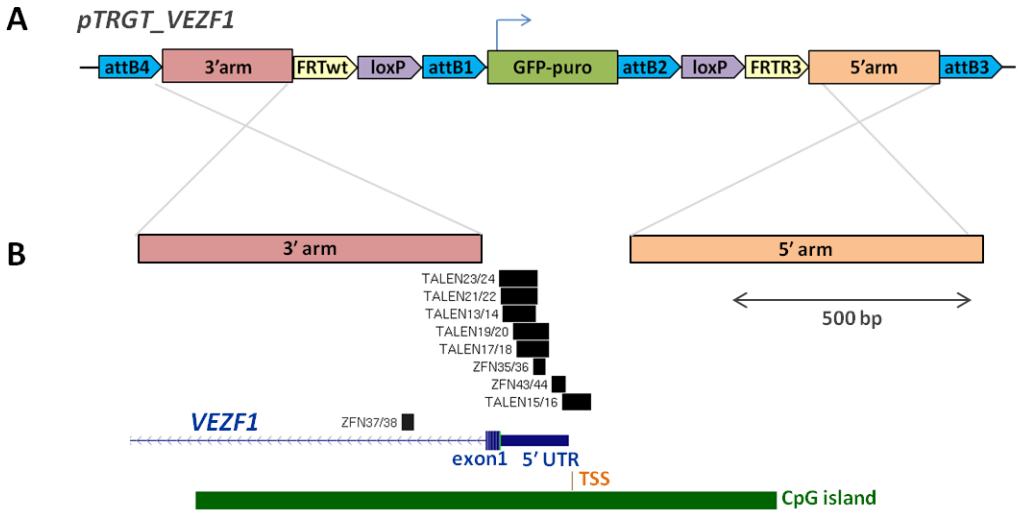

Figure 5.13 *VEZF1* Targeting Strategy
(A) Schematic cartoon depicting the pTRGT_VEZF1 vector construct which includes a CMV-driven EGFP-puro transgene flanked by recombination sites and *VEZF1* homology arms (B) Scale diagram of a 1900 bp region at the 5' end of the *VEZF1* gene promoter region based on UCSC Genome Browser alignments (hg19, chr17:56064601-56066500). The VEZF1 gene is shown in blue, which transcribes from right to left. The 5'UTR, codons of exon1, followed by intron 1 are all indicated. The extent of the CpG Island (green) and the location of the TSS (orange) are indicated below the gene. The locations of the ZFN and TALEN recognition sequences used in this study are depicted by black boxes above the gene. Above that are the locations of the sequences included as VEZF1 homology arms in pTRGT_VEZF1.

5.5 HDR-Based Targeting to Knockout the *VEZF1* gene Promoter

Circular ZFN or TALEN expression vectors were transfected into HEK293 cells with the circular targeting vector pTRGT_VEZF1 to determine whether the expression of the designed nucleases can stimulate transgene insertion at the *VEZF1* target locus. EGFP expression was measured by flow cytometry as an indicator of intracellular plasmid concentration three days post-transfection (Figure 5.14, left column). It was apparent that the transfection efficiency was greater than 75 % in all of the experiments performed. The transiently transfected cells were passaged once prior to commencing puromycin selection seven days post-transfection. The aim of selection is to enrich for cells that contain stably integrated transgenes following the dilution of non-integrated plasmids. As a control, cells which had been transfected with pTRGT_VEZF1 only were also subjected to selection to ensure that any plasmid integration was a consequence of ZFN/TALEN cleavage.

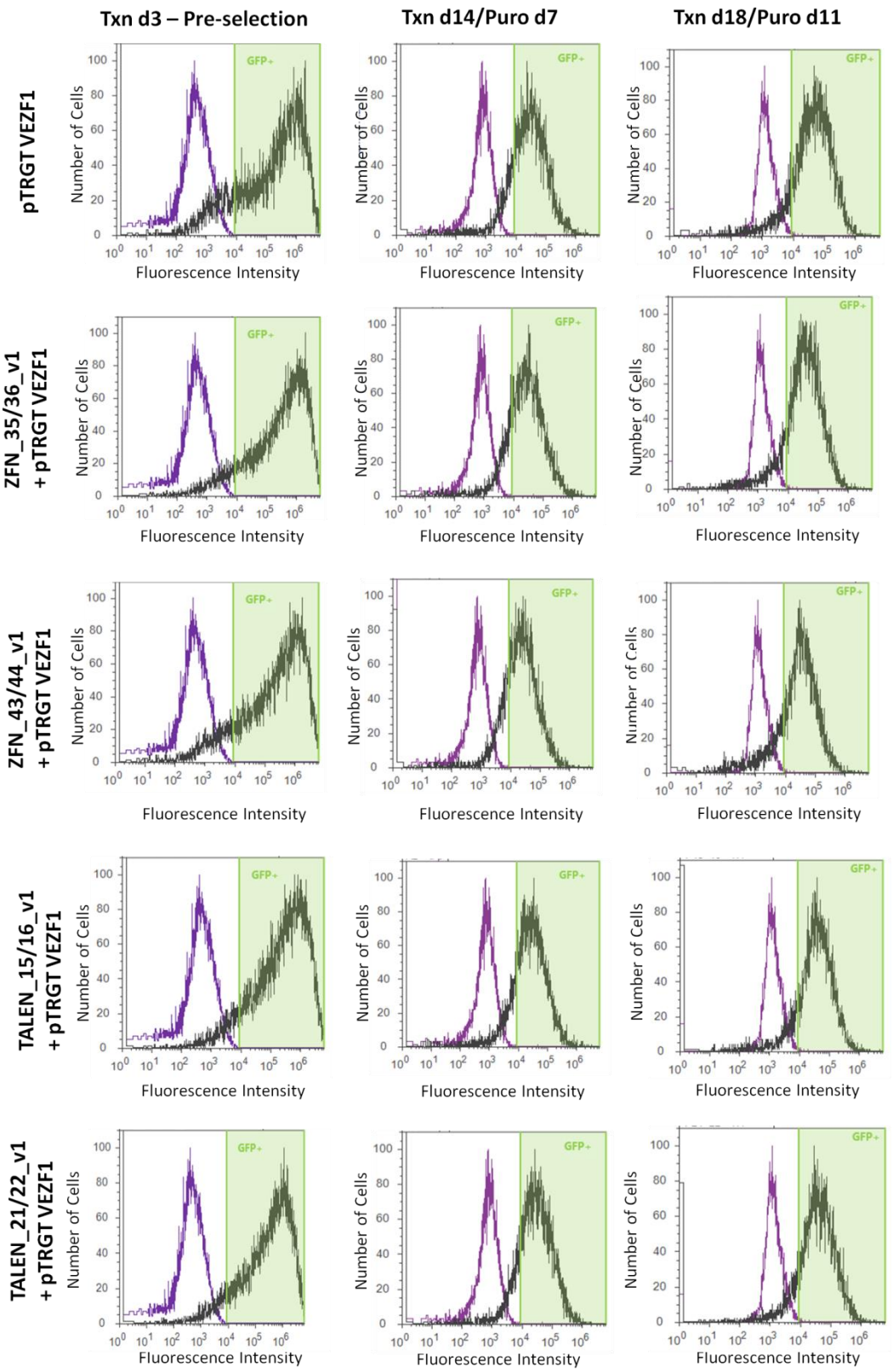

Figure 5.14 Transgene expression with and without ZFNs / TALENs
Flow cytometry analysis of EGFP fluorescence in HEK293 cells transfected with pTRGT_VEZF1 alone (top row) or pTRGT_VEZF1 with ZFN/TALEN expression vectors (lower rows). Analyses were completed 3 days post transfection (no selection), 14 days (7 days without then 7 days with selection) or 18 days (11 days of selection). The green region in each plot indicates EGFP expressing cells as the fluorescence intensity is greater than non-transfected control cells (purple).

Non-transfected HEK293 control cells were rapidly killed by puromycin selection (not shown). Unexpectedly, cells transfected with pTRGT_VEZF1 alone retained EGFP expression and grew rapidly under puromycin selection (Figure 5.14, top row). This finding indicates a substantial degree of off-target transgene integration in the conditions used. Cells that were co-transfected with nuclease expression vectors retain a similar level of EGFP expression as the pTRGT_VEZF1 only control following selection (Figure 5.14). The level of off-target integration is clearly too high to warrant further analysis of the effects of nuclease-mediated targeting in these conditions.

The control experiment where the circular pTRGT_VEZF1 plasmid is transfected alone, followed by puromycin selection, was repeated to determine whether the timing of selection is key to minimising off-target integration. The transfected cells were passaged once, two days before commencing drug selection. Five days post-transfection, the culture was divided into two populations. One culture was subjected to puromycin selection and the parallel culture was monitored for the loss of transiently transfected pTRGT_VEZF1 without selection. Flow cytometry analyses revealed that cells transfected with pTRGT_VEZF1 alone retained substantial levels of GFP expression seven days post-transfection without selection (Figure 5.15, black plots, top row). EGFP fluorescence levels only returned to non-transfected control levels between days 9 and 12 post-transfection. pTRGT_VEZF1 transfected cells subjected to puromycin selection retained high levels of EGFP expression, even after the non-selected cells had returned to WT fluorescence levels (Figure 5.15, blue plots).

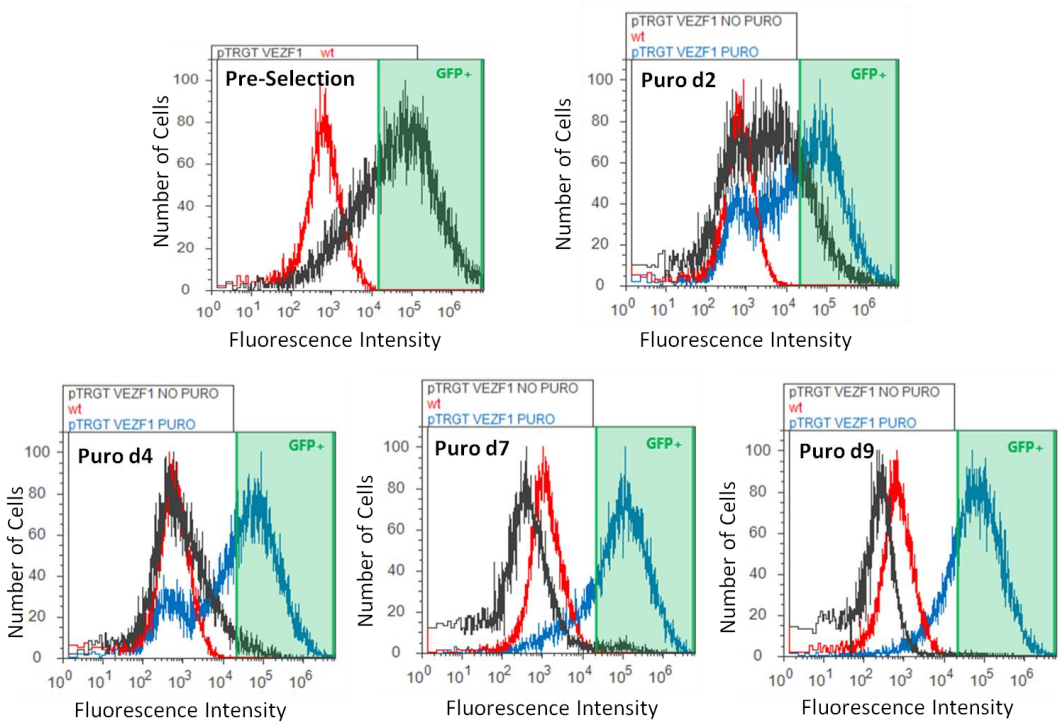

Figure 5.15 Efficient random integration of pTRGT_VEZF1 following selection
Flow cytometry analysis of EGFP fluorescence in HEK293 cells 3, 7, 9, 12 & 14 days following transfection with pTRGT_VEZF1. Parallel cultures with puromycin selection after day 5 (blue plots), or without selection (black plots), were compared to non-transfected controls (red plots).

It is clear that generation of a *VEZF1* knockout via HDR using our targeting vector is not as straightforward as initially envisaged. The issue of high level random integration that is retained following puromycin selection must be resolved before proceeding. The control experiment above demonstrated that a substantial level of pTRGT_VEZF1 plasmid is present in cells when selection was initiated 5 to 7 days post-transfection. Initiation of selection at around 10 days post-transfection may result in lower levels of random integration as any transiently transfected plasmid is lost from the population. However, a similar experiment to address this issue did not significantly reduce the level of selection for random integration of an EGFP transgene (Ileana Guerrini, data not shown). Furthermore, if the targeting of *VEZF1* results in any growth disadvantage, long periods of recovery between transfection and selection may result in the loss of targeted cells. I therefore decided that it best to take an alternative approach to studying the performance of the ZFNs and TALENs produced in this study, described in Chapter 6, until an improved strategy for gene targeting is available.

5.6 Discussion

This chapter has described the development of advanced TALEN expression vectors which were used to create three versions of TALENs that are designed to target the human *VEZF1* promoter. A gene targeting strategy was developed and experiments were performed to determine whether the TALENs and the ZFNs developed in this study can promote *VEZF1* gene targeting.

Two expression vectors were created that produce two new versions of TALENs, called v2 and v3, using the Voytas Golden Gate assembly system. These two new versions encode TALEN sequences that are codon optimised throughout, whose performance can be compared to the non-codon optimised expression vector, v1. Like v1, the v2 vector encodes a TALEN that has a wild type *Fok*I cleavage domain, whereas v3 vectors encode a number of *Fok*I mutations that have been reported to increase cleavage and restrict TALENs to function as heterodimers. The use of obligate heterodimer TALENs would be the safest way to generate *VEZF1* mutations while minimising confounding off-target effects. The assembled TALEN mammalian expression plasmids allow for the comparison of (i) differences in TALEN activity at different target sequences, (ii) differences in TALEN activity between different target sequence spacers, (iii) the effect of human codon optimisation on TALEN activity and (iv) the impact of *Fok*I mutations on TALEN activity.

The gene targeting vector pTRGT_VEZF1 was also designed and assembled, to assess the ability of ZFNs and TALENs to stimulate homology-directed double strand break repair. However, initial gene targeting experiments revealed that off-target integration of pTRGT_VEZF1 is prevalent under standard conditions and that substantial optimisations are required (Section 5.5). Delaying the point at which antibiotic selection is initiated should allow for the depletion non-integrated vector and reduce the chance of selecting for off-target integration, but poses another potential issue of population outgrowth of targeted cells. Initial investigation suggests that this optimisation alone is insufficient in abolishing random integrants (Ileana Guerrini, data not shown). Transfection of a smaller plasmid load to reduce the number of plasmid copies per cell may lessen random integration.

It is possible that off-target integration cannot be completely prevented, so it was decided to add an HSV thymidine kinase (TK) gene into the backbone of the pTRGT vector. TK expressing cells will rapidly convert the deoxy-guanosine analogue ganciclovir to ganciclovir triphosphate, which inhibits DNA replication. Cells with randomly integrated pTRGT-TK plasmids can be selected against using ganciclovir. Cells containing GFP transgenes that have integrated via HDR would lack the pTRGT-TK plasmid backbone and should survive selection. Colleagues in the West group are determining the optimal conditions for selection of cells using this method.

Once the issue of off-target integration of the targeting vector has been resolved or mitigated, a number of gene targeting steps could be undertaken to create *VEZF1* null and heterozygous cells even, if *VEZF1* is essential for growth. In the first approach, ZFN/TALEN and pTRGT_VEZF1 vectors will be transfected into cells. After a quick check that cells express GFP, clonal cell lines would be generated using puromycin selection. *VEZF1* gene integrity and VEZF1 expression would be analysed to determine whether there are lines in which targeting has been successful. At this point, it may be possible to identify lines in which homozygous *VEZF1* gene targeting has occurred, should such lines survive. These lines could be characterised at length and future rounds of re-targeting may not be required.

If it is found that all of the correctly targeted cells express VEZF1 because they are only targeted on one allele, then it is likely that VEZF1 is essential for growth. Further steps will be required to create conditional *VEZF1* homozygous null cells. The following steps rely on cells where only one *VEZF1* allele is faithfully targeted with the GFP-puro transgene. This transgene is flanked by loxP and FRT recognition sequences for the cre and flp recombinases, respectively, which will be used for cassette exchange or removal. The second allele must ideally be wild type and must retain the ZFN/TALEN target site and VEZF1 expression. The following three steps to create a conditional VEZF1 null cell line are envisaged:

1. Exchange the integrated GFP-puro transgene to restore the *VEZF1* promoter and exon 1 sequence, but leaving the flanking loxP sites. Sequences encoding RFP and a T2A self-cleavage peptide will be added immediately prior to the VEZF1 coding sequences to enable the visualisation of VEZF1 null cells in the final stage. Sequences encoding a triple FLAG epitope tag will also added to the VEZF1 N-terminus. The

ZFN/TALEN recognition sequences will be mutated. The cassette exchange will be performed by transfection of FLP recombinase expression and targeting plasmids into cells. Clonal cell lines that lack GFP expression will be selected following limiting dilution. Western blotting will be used to confirm the presence of FLAG-tagged VEZF1. This tagged allele should retain VEZF1 function in cells, but can be subsequently deleted using cre recombinase. It will no longer be a ZFN/TALEN target.

2. ZFN/TALEN and pTRGT_VEZF1 vectors will be transfected into cells selected in step 1. GFP expressing cell lines will be selected using puromycin as described above. *VEZF1* gene integrity and VEZF1 expression would be analysed to identify the lines where wild type VEZF1 has been lost and only RFP-FLAG-VEZF1 remains. These heterozygotes can be studied for any phenotype.

3. The remaining RFP-FLAG-VEZF1 allele will be deleted using cre recombinase. The transfection of a cre expression plasmid should result in loss of the *VEZF1* promoter and exon 1. The effect of VEZF1 deletion on cell growth and VEZF1 target gene regulation can be studied in the cells that lack RFP expression.

Given that the generation of a *VEZF1* knockout via nuclease-directed gene targeting requires further technical development and time was limiting, I decided to focus on examining whether the assembled nucleases are capable of inducing mutations at the *VEZF1* promoter via error-prone non-homologous end joining (NHEJ) DNA repair. I wish to determine which of the ZFNs and TALENs are most efficient. These experiments, described in Chapter 6, will feedback to future efforts to undertake precise HDR-mediated gene targeting of the *VEZF1* gene.

Chapter 6

Using Designer Nucleases to Modify the *VEZF1* Gene Promoter

6.1 Introduction

ZFNs and TALENs have been used to efficiently create gene mutations without the need for gene targeting vector strategies (Perez et al., 2008, Li et al., 2011c, Holt et al., 2010). This is because the dominant form of double strand break repair in most somatic cells is error-prone non-homologous end joining (NHEJ). The presence or absence of short insertions and deletions at the ZFN/TALEN target sites can be used to assess the performance of the designed nucleases in cells. The assay of nuclease-mediated mutations following ZFN/TALEN transfection provides information about performance in endogenous mammalian chromatin contexts, which is more relevant than exogenous bacterial or yeast reporter assays (Maeder et al., 2008, Ramirez et al., 2008). This approach will be used to compare nuclease target location, TALEN dimer spacing, *TALEN* architecture and the effect of codon optimisation. The most efficient nucleases can then be used in conjunction with an updated targeting vector to generate an HDR-mediated *VEZF1* knockout in subsequent studies (as described in Chapter 5). Given that the ZFN and TALEN targets sites lie within the *VEZF1* core promoter, it may be discovered NHEJ mutations that are sufficient to reduce or remove VEZF1 expression.

In order to accurately compare our assembled nucleases, I need to design an assay system to quantify nuclease-mediated mutations. Mutations associated with NHEJ are typically small insertions and deletions (indels) resulting from an error-prone resection and filling repair process. These mutations can be easily scored using restriction fragment length polymorphism (RFLP) assays, where the presence or absence of a restriction enzyme recognition sequence at the nuclease target site is studied (Huang et al., 2011). RFLP assays are simple to perform and interpret, but different assays must be developed for each individual target site.

A more generic assay for indels at nuclease target sites involves the use of the endonucleases *Cel*I or T7 EndonucleaseI, which cleave mismatched DNA structures (Kulinski et al., 2000, Babon et al., 2003, Oleykowski et al., 1998). Genomic PCR products of nuclease target regions that contain a mixture of wild type and mutant sequences can

be melted and reannealed to create mismatched molecules that can be visualised following cleavage by these enzymes. I intend to develop and optimise such assays to analyse the performance of ZFNs and TALENs in somatic cells.

6.2 Aims of this chapter

- Develop a reliable and quantitative assay system to test for the presence of mutations likely to be created by nucleases

- Deliver TALENs & ZFNs to cultured human cells and test their abilities to create targeted mutations at the *VEZF1* gene promoter

- Determine the nature of nuclease-mediated mutations and how they affect *VEZF1* expression

6.3 Development of an Assay to Detect NHEJ Events

6.3.1 *Cel*I & T7EI Assays

The non-sequence specific endonucleases *Cel*I (Surveyor) and T7 *endonuclease*I (T7EI) can be used to detect mismatches created following erroneous DSB repair by NHEJ (Kulinski et al., 2000, Babon et al., 2003, Oleykowski et al., 1998). The enzymes detect "bulges" in DNA molecules and digest DNA at these points. The "bulges" are created by the annealing of two strands of DNA that harbour small base pair mismatches to each other, such as the changes created following DNA repair via NHEJ. When visualised on an agarose gel, these digest products can be detected as compared to an undigested parent band which is composed of non-mismatched (WT) DNA.

6.3.1.1 Designing a Control Assay for *Cel*I & T7EI

To allow for direct comparison and optimisation of the *Cel*I and T7EI assays for NHEJ, I decided to develop a universal control assay. The Surveyor kit provides a control in the form of two plasmids named C and G, which differ from each other at only one base pair. The plasmids are individually packaged and mixed with primers for amplification of the region containing the C or G base to be used as a mismatch. The PCR products from plasmids G and C can be mixed in known ratios and the amplicons are denatured and re-

annealed to create DNA fragments containing a predicted percentage of mismatched DNA which differs at only one base pair. This mismatched DNA can then be treated with *Cel*I nuclease and analysed on an agarose gel. However, each Surveyor kit only provides enough control DNA for 10 experimental runs which, given all of the suggested optimisations, would not be adequate to thoroughly assess and compare both *Cel*I and T7EI. I have therefore devised a similar control using plasmids that are available in the lab.

Plasmids encoding the *Fok*I domains for zinc finger nuclease expression have been modified to express the Sharkey *Fok*I variant (pFOKI-EL-S, Chapter 3). This mutation is two base pair changes from the original *Fok*I plasmid (pFOKI-EL; see Figure 3.6). Primers (NHEJ CONTROL F & R) were designed to amplify the region containing these two base pair changes (Figure 6.1). By performing PCR from pFOKI-EL and pFOKI-EL-S, the amplicons generated differ from each other by only two bases. These PCR products can then be mixed in known ratios and denatured and re-annealed, as in the protocol for the *Cel*I Control Assay, to create DNA samples with two single base pair mismatches in a predicted percentage of the sample. This can be treated with both nucleases, testing a variety of conditions, and analysed on an agarose gel with the aim to determine the optimal conditions and best performing nuclease to assess ZFN and TALEN activity.

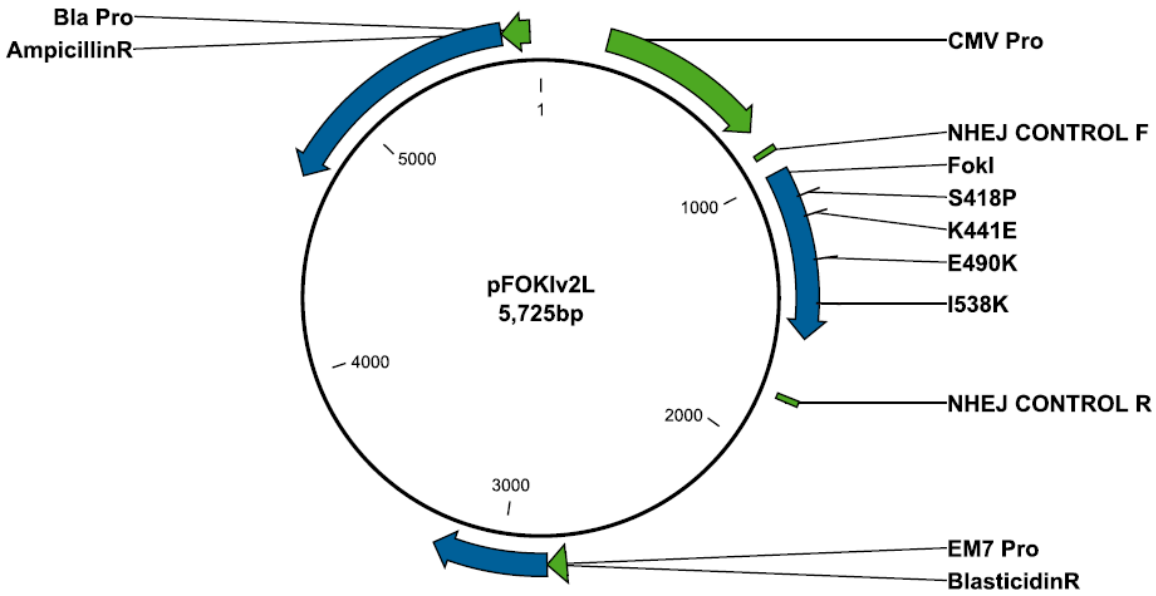

Figure 6.1 Plasmid Map of pFOKI_v2L
Scaled map of pFOKI_v2L showing the *Fok*I gene with amino acid variations encoding obligate heterodimer (E490K & I538K) and Sharkey mutations (S418P& K441E). Binding sites of primers NHEJ control F & R used for PCR amplification to create a DNA sample containing a mismatch to be used as a control for NHEJ assays.

6.3.1.2 Determining Optimal Conditions for *Cel*I & T7EI

Optimisations suggested in the manual for *Cel*I nuclease were performed using both *Cel*I and T7EI. Initially, variation in the amount of DNA and the amount of enzyme were tested using a variety of percentage mismatches in the sample, 0%, 2% and 10%. The testing of 0% mismatched DNA permits the observation of any non-specific activity of the enzyme used and may allow the determination of a false positive reporting rate. The FokI-EL and FokI-EL-S PCR products were purified via gel extraction, quantified, mixed in buffer and melted for 3 minutes at 90°C prior to annealing at 65°C, 37°C and then 22°C for 10 minutes at each temperature. PCR products were then diluted to required concentrations for digest. The suggested maximum amount of DNA to digest with *Cel*I is 400 ng and so I digested this amount, 270 ng (2/3) and 135 ng (1/3) with 5 or 10 units of each enzyme for two hours, as this is the recommended digest time for *Cel*I nuclease.

It was found that both the *Cel*I and T7EI unexpectedly digest 0% mismatch control DNA (Figure 6.2, compare lanes 2-7 with lane 20 in parts A and B). The cleavage appears to be non-specific, as the cleavage products are of random size, resulting in a smear. The 2% and 10% mismatch experiments resulted in a similar digestion profile to the 0% control (Figure 6.2, lanes 9-19 in parts A and B). The parent PCR product of 901 bp is expected to be digested into three fragments of 643 bp, 190 bp and 68bp based on the location of the mismatches in the DNA sample. It therefore appears that no mismatch-specific products are revealed by *Cel*I or T7EI endonucleases due to the high background of non-specific cleavage.

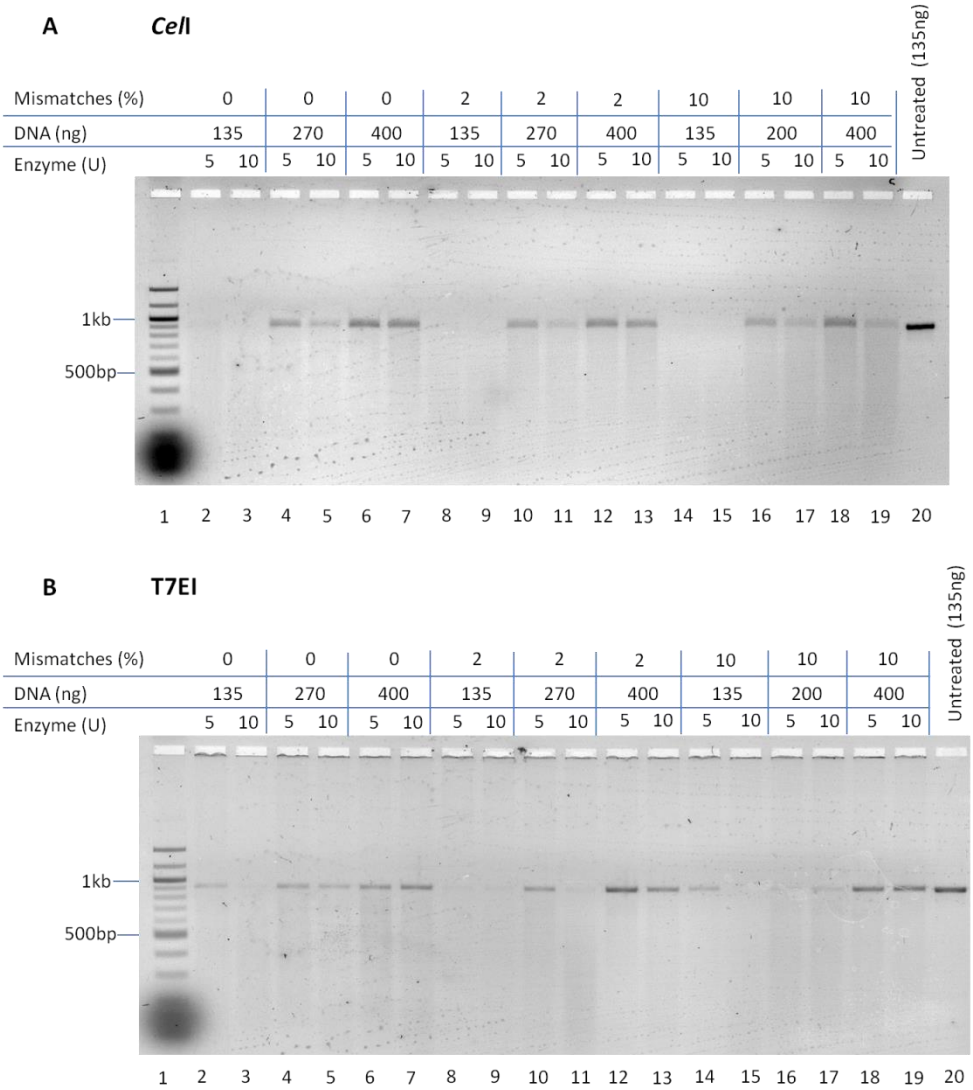

Figure 6.2 Two hour *Cel*I & T7EI digestion of control DNA
Agarose gel electrophoresis of *Fok*I control DNA treated with either *Cel*I (A) or T7EI (B). Varying amounts of PCR product DNA containing a ratio of 0, 2 or 10% mismatches after annealing were incubated with nucleases for 2 hours. The 901 bp PCR product is expected to cleave into products of 643, 190 and 68 bp when mismatches are present. Lane 20 on both gels illustrates the band intensity expected from undigested samples.

The high level of non-specific cleavage by *Cel*I and T7EI may have been due to the length of time the reaction was incubated. A timecourse experiment was carried out to determine if less time would lower the non-specific cleavage activity observed. Digestion reactions were set up as before, except for shorter incubation periods of 30, 45 or 60 minutes. Non-mismatched DNA was used to determine the time at which non-specific digest did not occur. However, prevalent non-specific cleavage was observed, despite the shorter incubation times (Figure 6.3).

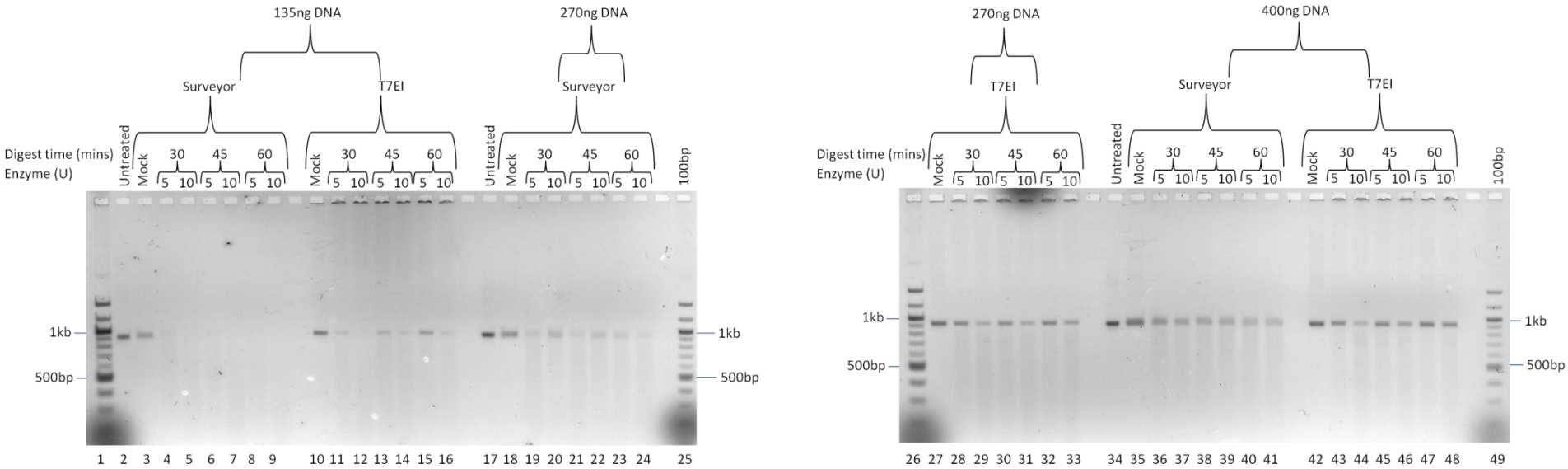

Figure 6.3 *Cel*I & T7EI digestion of 0% mismatched control DNA for 30 to 60 minutes
Agarose gel electrophoresis of *Fok*I control DNA following digestion with *Cel*I or T7EI. The amount of control DNA and incubation periods used are indicated above the gel images. The 901 bp PCR product is expected to cleave into products of 643, 190 and 68 bp when mismatches are present. Lanes 2, 17 and 34 demonstrate the band intensity of annealed PCR products, loaded neat on the gel. Lanes 3, 10, 18, 27, 35 and 43 illustrate the band intensity expected from undigested samples.

The apparent non-specific cleavage by *CelI* and T7EI may still be due to excessive incubation times. I therefore reduced the incubation times to 5 or 15 minutes with *Cel*I Figure 6.4or T7E1 to try and find a condition where 0% mismatch control DNA was not digested. Control DNA containing 0 or 30% mismatched sample was incubated with or without enzyme for 5 or 15 minutes. It was found that incubation with *Cel*I for 5 or 15 minutes resulted in little or no digestion of 0% mismatch negative control DNA (Figure 6.4 compare lanes 2, 3 with lanes 6, 7). However, none of the expected digestion products were observed in the 30% mismatch positive control DNA (Figure 6.4, compare lanes 8, 9 with lanes 4, 5). It is apparent that *Cel*I is unable to detect mismatches with desired efficiency with the control DNA and conditions used.

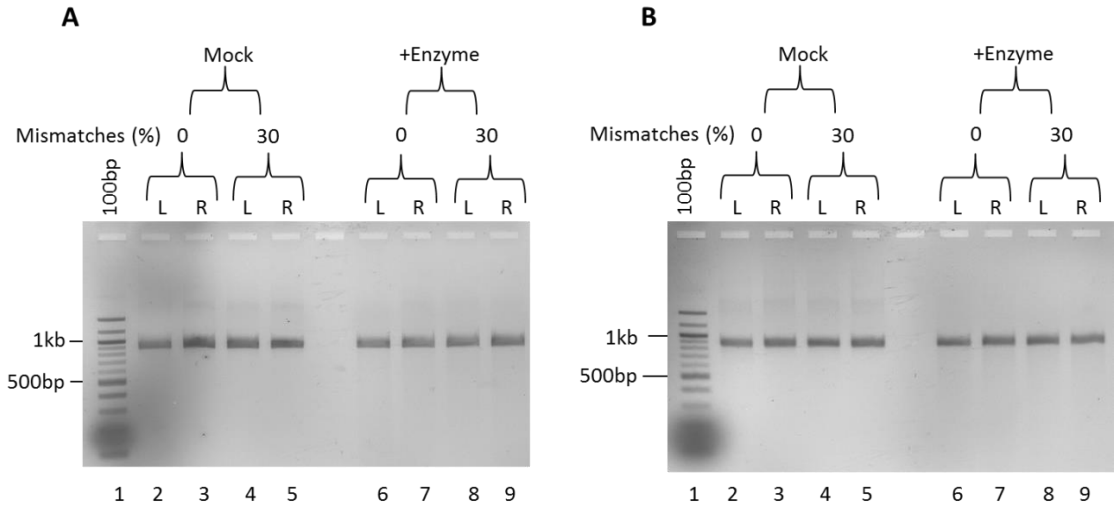

Figure 6.4 Five and fifteen minute digestion of control DNA with *Cel*I
Agarose gel electrophoresis of *Fok*I control DNA following *Cel*I digestion for (A) 5 or (B) 15 minutes. The 901 bp PCR product is expected to cleave into products of 643, 190 and 68 bp when mismatches are present. Mock experiments with no enzyme are shown in lanes 2-5.

Unlike *Cel*I, T7EI exhibited substantial non-specific cleavage activity with 15 minutes of incubation (Figure 6.5 B, lanes 6-9). This activity was minimal in 5 minute incubations (Figure 6.5 A, lanes 6-9). However, none of the expected digestion products were observed in the 30% mismatch positive control DNA (Figure 6.5 A, B, lanes 8, 9). It is apparent that like *Cel*I, T7EI is unable to detect mismatches with desired efficiency with the control DNA and conditions used.

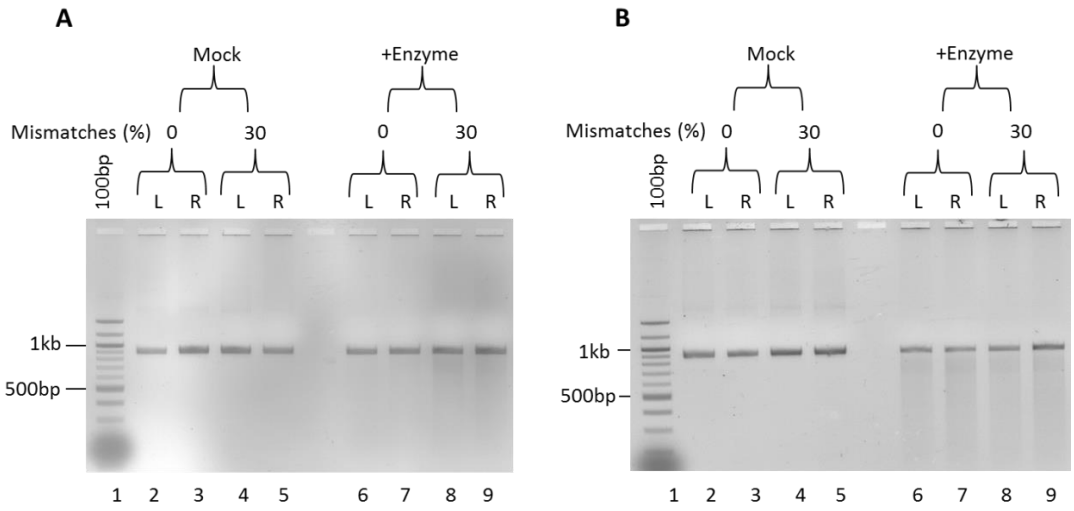

Figure 6.5 Five and fifteen minute digestion of control DNA with T7EI
Agarose gel electrophoresis of *Fok*I control DNA following T7EI digestion for (A) 5 or (B) 15 minutes. The 901 bp PCR product is expected to cleave into products of 643, 190 and 68 bp when mismatches are present. Mock experiments with no enzyme are shown in lanes 2-5.

Prevalent non-specific cleavage of 0% mismatch PCR product DNA was observed with both the *Cel*I and T7EI endonucleases. Reduction of incubation times to 15 minutes for *Cel*I and 5 minutes for T7EI removed the non-specific cleavage to acceptable levels. However, none of the expected digestion products were observed in the mismatch positive control experiments. It is possible that the mismatches for this particular control sequence are not suitable substrates for *Cel*I or T7EI, or that there are problems with the melting and re-annealing step of the assay (see Discussion). In the interests of time, I decided to discontinue using these assays as a means of assessing the level of NHEJ repair resulting from ZFN and TALEN activity. This has driven the need to design a customised assay based around the specific sequences of the ZFN and TALEN binding locations at the *VEZF1* promoter.

6.3.2 Restriction Fragment Length Polymorphism (RFLP) Analysis of TALEN-mediated mutations

I decided to employ an assay that did not rely on the melting and accurate re-annealing of PCR product DNA. Restriction fragment length polymorphism (RFLP) analysis is a simple assay where restriction digestion determines whether NHEJ has resulted in the loss of a restriction enzyme recognition sequence. The key to this assay is the identification of restriction sites that overlap the point of cleavage by the TALEN being studied. Working on the principle that TALEN-mediated DSBs repaired by NHEJ is erroneous, a restriction site present in a TALEN spacer would become mutated upon repair (Figure 6.6). Digesting TALEN-treated DNA with this restriction enzyme would result in a differential digest band pattern than would be observed from WT DNA. The ratio of the polymorphism would

directly relate to the cleavage activity of the TALEN. I designed RFLP assays to analyse NHEJ events occurring following TALEN treatment of DNA. I attempted to design similar assays for ZFN-treated samples but the restriction sites present in the spacer between our ZFN pairs were found to be highly frequent, which prevented the development of a diagnostic RFLP for these sequences (data not shown).

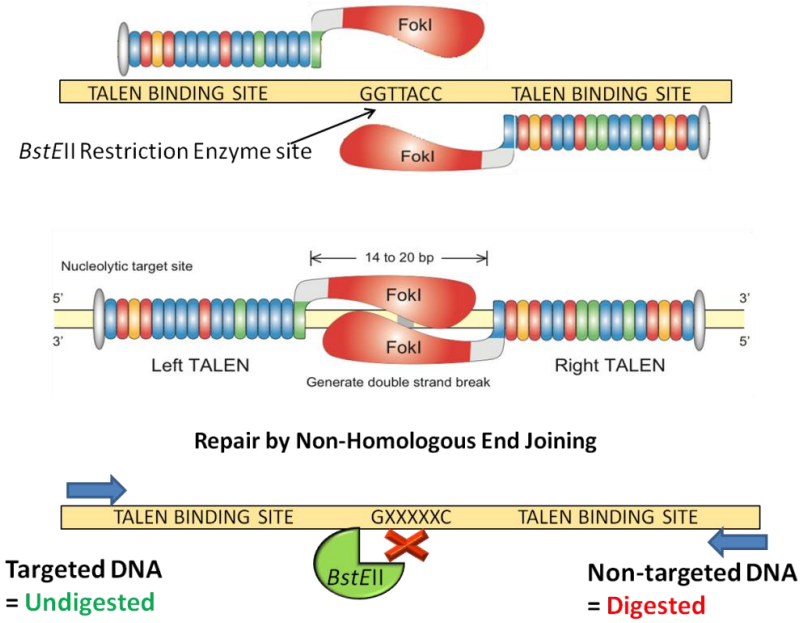

Figure 6.6 RFLP Assay Overview
Cartoon depicting the general principle of RFLP analysis. Genomic DNA is bound and cleaved by TALENs across a restriction enzyme recognition sequence; the example shown here is *Bst*EII. Repair of the TALEN-mediated DSB by NHEJ will result in mutation of the restriction enzyme recognition sequence. When incubated with the restriction enzyme, mutant DNA will not be digested and will generate an RFLP when compared to WT DNA.

The six TALEN pairs that target the human *VEZF1* promoter, described in Chapter 5, target three cleavage points. Three different RFLP assays were designed to test for TALEN activity at these target sites (Figure 6.7). Assay 1 was designed to test TALENs 13/14, 21/22 and 23/24. Assay 2 was designed for TALEN pair 15/16, and assay 3 tested TALENs 17/18 and 19/20.

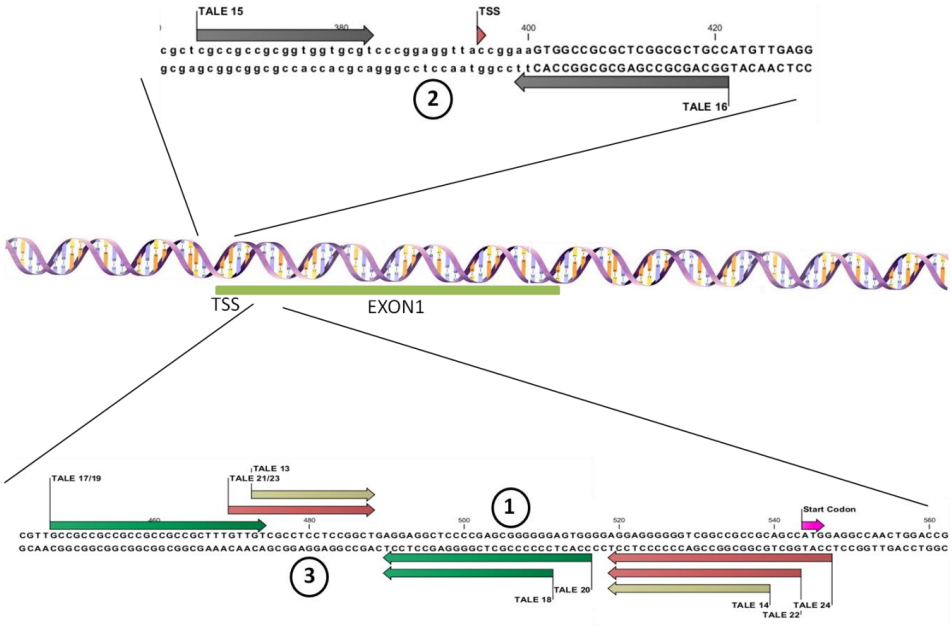

Figure 6.7 *VEZF1* Promoter Regions in RFLP Assay
Schematic overview of TALEN target sites demonstrating the requirement for three RFLP assay designs to encompass the different TALEN target regions. Restriction sites located in the spacers were selected for RFLP analyses.

6.3.2.1 RFLP analysis of *VEZF1* promoter mutagenesis directed by TALEN pairs 13/14, 21/22 & 23/24

It is desirable to have a sensitive and quantitative assay for polymorphisms so that direct comparisons can be made between efficient and inefficient TALEN designs. It was considered that the use of a DNA Bioanalyser (Agilent) would generate the most accurate data for the RFLP assay. The Bioanalyser resolves DNA molecules via electrophoresis through a matrix with high precision and quantifies different sized species to produce a quantitative electropherogram. RFLP assay 1 is designed to analyse the mutations directed by TALEN pairs 13/14, 21/22 & 23/24. The cleavage site located in the centre of the recognition sequence for these TALEN pairs overlaps a *Bsr*BI restriction enzyme recognition sequence (Figure 6.8). PCR primers were designed to generate a 1310 bp genomic PCR product for this region. This PCR product was purified and subjected to restriction digestion with the enzymes *Bsr*BI and *Bst*EII, which cleave wild type DNA into 634, 390, 172 and 114 bp fragments. TALEN-directed loss of a *Bsr*BI site would result in a mutant 286 bp cleavage product (Figure 6.8).

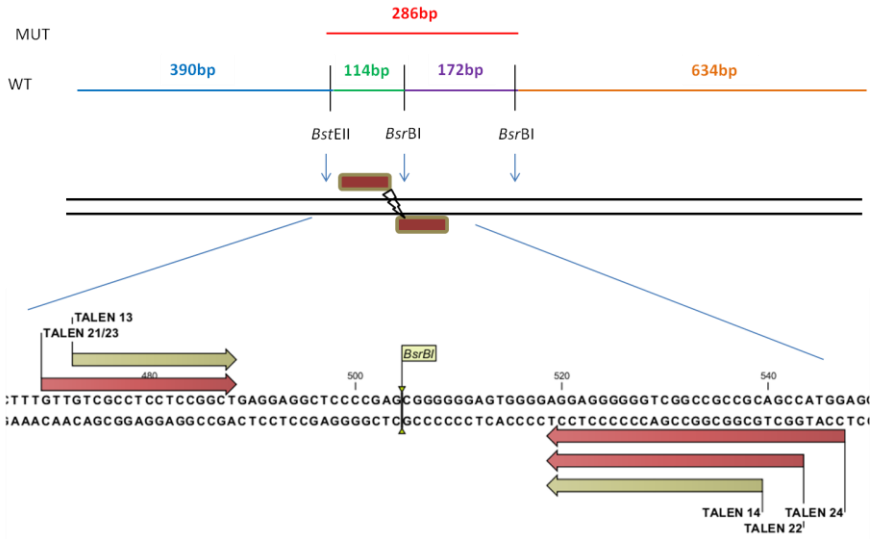

Figure 6.8 The design of RFLP Assay 1
The sequences recognised by TALEN pairs 13/14 (beige), 21/22 & 23/24 (red) are indicated by filled arrows over the DNA sequence. A *Bsr*BI restriction site overlaps the central spacer of these TALEN pairs. Above: A scale diagram depicting the cleavage pattern of a 1310 bp PCR product with *Bsr*BI and *Bst*EII.

HEK293 cells were transfected with 2 µg of each pair of expression plasmids for TALEN pairs 13/14, 21/22 and 23/24 in v1, v2 and v3 formats (9 transfections). A parallel transfection with an EGFP expression vector was monitored by FACS after 24 hours to confirm that transfection efficiency was greater than 90% (not shown). Genomic DNA was prepared from wild type and TALEN-transfected cells 72 hours after transfection. PCR was performed on the genomic DNAs using primers hVEZF1pro-374_F and hVEZF1pro+895_R. The 1310 bp PCR product was purified using a column and digested with *Bsr*BI and *Bst*EII. The product from wild type cells cleaved into the expected 634, 390, 172 and 114 bp fragments (Figure 6.9, WT). Importantly, there was no peak representing a 286 bp fragment in the WT sample, indicating complete cleavage at the *Bsr*BI site studied in the RFLP assay.

TALEN-directed mutagenesis is expected to mutate this *Bsr*BI site, which results in 286 bp fragment in the assay as *Bsr*BI would no longer be able to cleave this into 172 and 114 bp fragments. RFLP analysis of DNA from cells transfected with TALEN_23/24 FokI_v3 reveals the presence of a band matching 286 bp (Figure 6.9, bottom right). The presence of a 634 bp fragment in this same assay, and the absence of a 920 bp fragment (634 + 286) indicates that *Bsr*BI digestion was complete in this assay. Quantification of the fluorescent units (FU) indicates that the *VEZF1* promoter is mutated in at least 15% of alleles in cells transfected with TALEN_23/24 FokI_v3. The RFLP assay did not detect mutations with any of the other TALENs in this experiment (Figure 6.9).

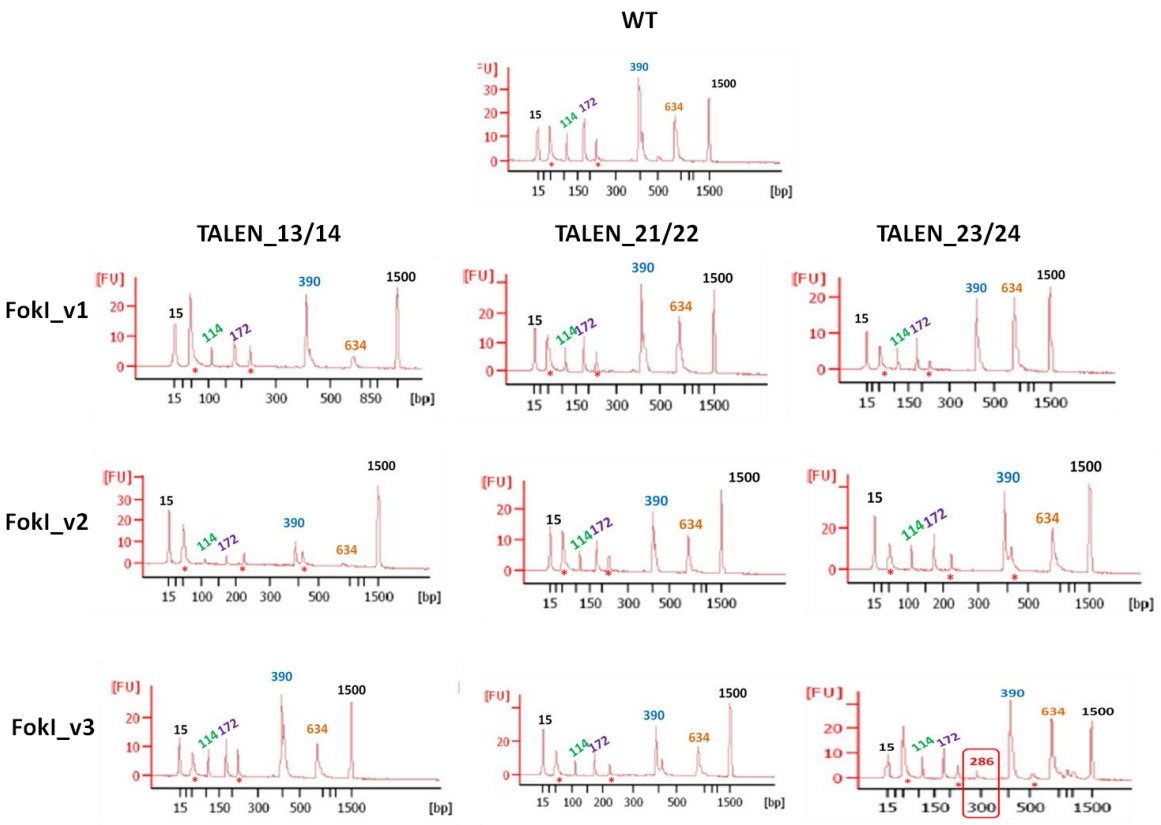

Figure 6.9 Bioanalyser RFLP analysis of *VEZF1* promoter mutagenesis directed by three versions of TALEN pairs 13/14, 21/22 & 23/24

Electropherogram plots of DNA fragments resulting from digestion of a 1310 bp PCR of the *VEZF1* promoter with *Bsr*BI and *Bst*EII. Digests were performed on PCRs of HEK293 genomic DNA from wild type (WT) cells and cells transfected with three versions of the TALEN pairs 13/14, 21/22 & 23/24. Fragment sizes are annotated on each plot, which are determined by the Bioanalyser software based on the 15 bp and 1500 bp markers (black). Unexpected peaks are indicated by red asterisks. A candidate 286 bp RFLP peak observed in TALEN_23/24_v3 is indicated by a red box.

Three DNA fragments of unexpected size were identified on the Bioanalyser from RFLP assay 1 (Figure 6.9, asterisks). These peaks were visible in all of the samples and tended to vary in intensity between replicates (not shown). It is possible that these are non-specific PCR products. The presence of these unexpected peaks makes it more difficult to infer the presence of TALEN-directed mutations. Before deciding on whether to optimise the PCR product purity further, the same PCR product digests were analysed by agarose gel electrophoresis for a second opinion.

Agarose gel electrophoresis also reveals the presence of an RFLP of expected size in PCRs from TALEN transfected cells, but there are key differences from the Bioanalyser data. A clear band consistent with the expected 286 bp RFLP fragment is found in DNA from cells transfected with TALEN_13/14 FokI_v2 and TALEN_21/22 FokI_v1 (Figure 6.10, lanes 3 and 5). This RFLP band appears to be present in DNA from the other TALEN experiments,

but this often smeared and cannot be called with confidence (Figure 6.10). It is to be expected that indels will vary in size, and thus result in smeared rather than distinct bands. The expected disappearance of the 172 and 114 bp fragments in DNA from TALEN transfected cells is another means of calling an RFLP. These bands are clearly reduced compared to WT DNA in DNA from cells transfected with TALEN_13/14 FokI_v1 and v2, TALEN_23/24 FokI_v3 (Figure 6.10, lanes 2, 3 and 10). The fact that the disappearance of the 172 and 114 bp bands and appearance of the 286 bp RFLP fragment are not entirely concordant may reflect variable NHEJ insertion/deletions (indels) between experiments. Variable indel sizes might also explain why the Bioanalyser profiles reported less polymorphism rates than agarose gel electrophoresis.

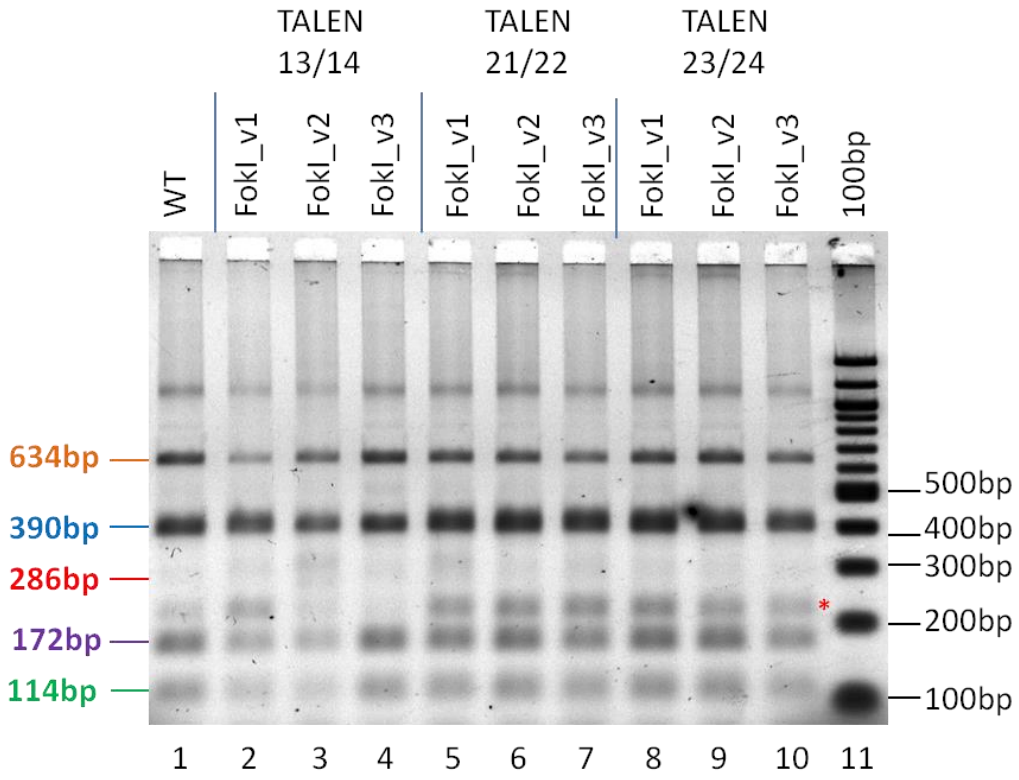

Figure 6.10 RFLP analysis of *VEZF1* promoter mutagenesis directed by three versions of TALEN pairs 13/14, 21/22 & 23/24

Agarose gel electrophoresis of DNA from cells transfected with TALEN pairs 13/14, 21/22 & 23/24 compared to wild type DNA digested with *Bsr*BI and *BstE*II. Wild type DNA is digested into 634, 390, 172 and 114 bp fragments as indicated. Mutagenesis of a *Bsr*BI site results in the loss of the 172 and 114 bp fragments and gain of a 286 bp, or similar, fragment. Non-specific PCR products are indicated by an asterisk.

6.3.2.2 RFLP analysis of *VEZF1* promoter mutagenesis directed by TALEN pair 15/16

A *Bst*EII restriction enzyme recognition site was found to overlap the cleavage site of the TALEN pair 15/16. An RFLP assay involving digestion of the 1310 bp PCR product described above with *Bst*EII and *Rsa*I will result in 732, 390 and 188 bp fragments (Figure 6.11). TALEN-directed mutagenesis of the *Bst*EII site will result in a loss of the 390 and 188 bp fragments and gain of a 578 bp fragment. It is possible that the products of mutagenesis may result in RFLP fragments that vary from 578 bp due to the variable nature of NHEJ-mediated indels, as discussed above.

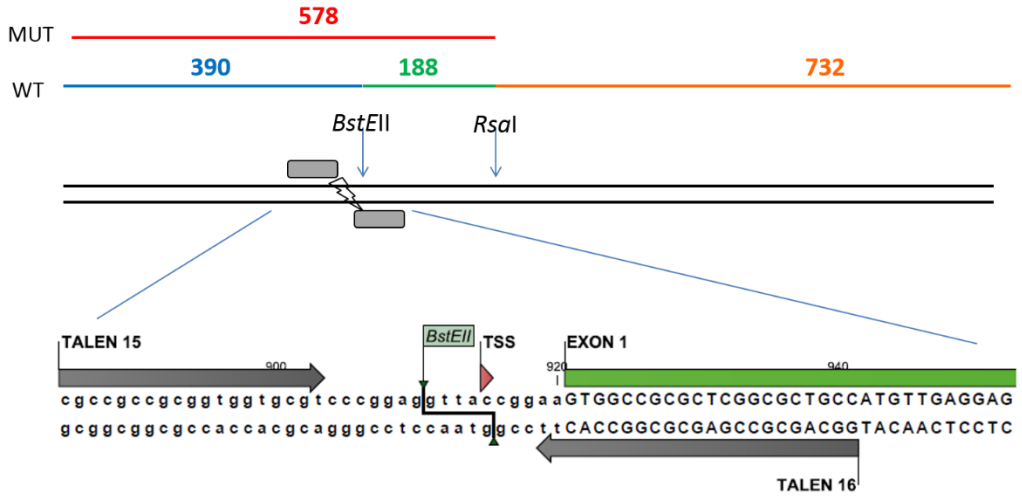

Figure 6.11 The design of RFLP Assay 2
The sequence recognised by TALEN pair 15/16 is indicated by filled grey arrows over the DNA sequence. A *Bst*EII restriction site overlaps the central spacer of these TALEN pairs. Above: A scale diagram depicting the cleavage pattern of a 1310 bp PCR product with *Rsa*I and *Bst*EII.

HEK293 cells were transfected pairs of plasmid vectors that encode the three versions of the TALEN pair 15/16. PCR DNA digested with *Rsa*I and *Bst*EII was profiled on the Agilent Bioanalyser. DNA from wild type cells was found to digest into the expected 732, 390 and 188 bp fragments (Figure 6.12, WT). Similar background peaks described above were also present in RFLP assay 2, increasing the likelihood that these were non-specific PCR products (Figure 6.12, asterisks). A very well defined peak approximating 578 bp is apparent in DNA from cells transfected with TALEN_15/16 FokI_v2 and v3 (Figure 6.12). This size matches that expected for TALEN-directed mutations which result in loss of the *Bst*EII site. In the case of TALEN_15/16 FokI_v1, however, no such mutant peak can be identified. This demonstrates that, despite the presence of non-specific background peaks, the RFLP assay can identify mutations at the anticipated site.

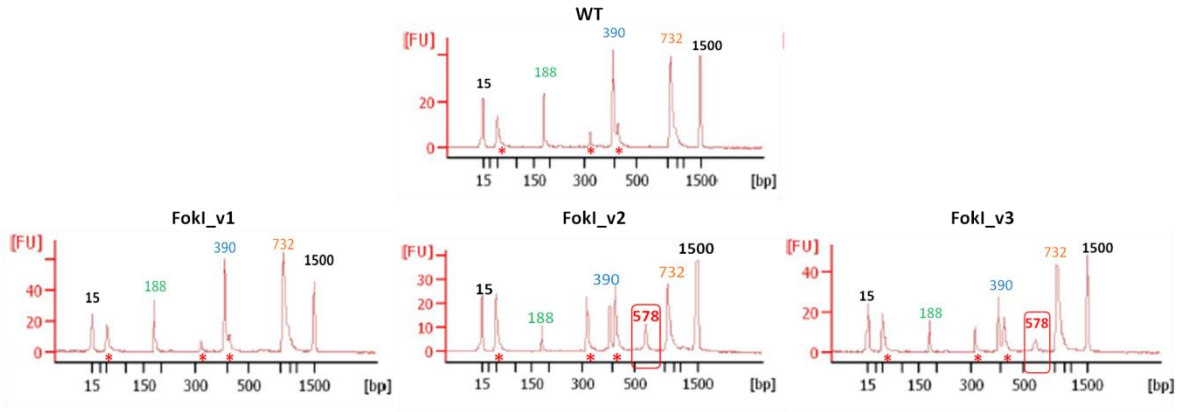

Figure 6.12 Bioanalyser RFLP analysis of *VEZF1* promoter mutagenesis directed by three versions of TALEN pair 15/16

Electropherogram plots of DNA fragments resulting from digestion of a 1310 bp PCR of the *VEZF1* promoter with *Rsa*I and *Bst*EII. Digests were performed on PCRs of HEK293 genomic DNA from wild type (WT) cells and cells transfected with three versions of the TALEN pair 15/16. Fragment sizes are annotated on each plot, which are determined by the Bioanalyser software based on the 15 bp and 1500 bp markers (black). Unexpected peaks are indicated by red asterisks. A candidate 578 bp RFLP peak observed in TALEN 15/16 v2 and v3 cells is marked by a red box.

The presence of TALEN 15/16-directed mutagenesis at the human *VEZF1* promoter was confirmed by agarose gel electrophoresis. The same *Rsa*I / *Bst*EII digested PCR products shown above were analysed again. A band approximating 578 bp is observed from the DNA of all cells transfected with the TALEN pair 15/16 (Figure 6.13, lanes 2-4). It should be noted that the 578 bp band appears to be a collection of similar sized species. The different resolution properties of agarose and the Bioanalyser matrix may explain the different levels of mutagenesis reported by the two assays from the same DNA. The presence of the ~578 bp band on agarose gels is mirrored by the reduction of the 390 and 188 bp bands (Figure 6.13, lanes 2-4). Quantification of the 188 bp band indicates that 43% and 46% of alleles are mutant within three days of transfection with the v2 and v3 15/16 TALEN pair, respectively. Analysis of the 188 band in the samples transfected with v1 of TALEN 15/16 did not reveal a significant level of mutations. This indicates that codon optimisation of TALENs offers a major advance in TALEN performance, probably due to increased TALEN protein levels. The mutation of *Fok*I to produce obligate heterodimers results in some loss of performance, but this is minimal for the 15/16 TALEN pair. This result is consistent with recent reports of combining the ELD, KKR and Sharkey mutations.

Given the high efficiency of the 15/16 v2 TALEN pair in HEK293 cells, I investigated whether this TALEN could be used to generate *VEZF1* promoter mutations in erythroid K562 cells. K562 cells are much more difficult to transfect but are an experimentally important model for studying erythroid gene regulation. A control transfection with an EGFP expression plasmid monitored by flow cytometry indicated that the transfection efficiency for K562 cells was 38 % for this experiment. Despite only a fraction of K562 cells having TALEN expression, a 578 bp RFLP band is clearly observed from the DNA of all cells transfected with the v2 TALEN pair 15/16 (Figure 6.13, lane 6). The 15/16 TALEN pair is therefore very efficient at directing *VEZF1* promoter mutagenesis in different cell types and should be useful in difficult to transfect cells like primary and stem cells.

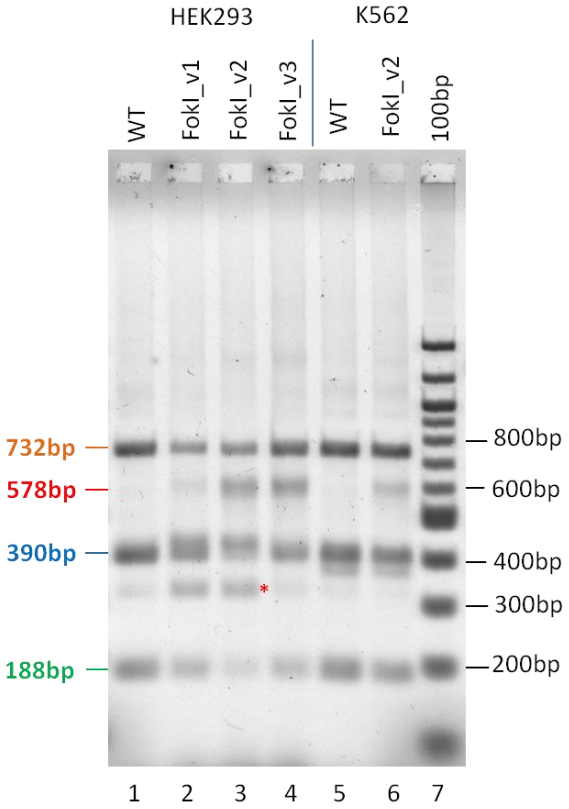

Figure 6.13 RFLP analysis of *VEZF1* promoter mutagenesis directed by three versions of the TALEN pair 15/16

Agarose gel electrophoresis of DNA from cells transfected with versions of the TALEN pair 15/16 compared to wild type DNA digested with *Rsa*I and *BstE*II. Wild type DNA is digested into 732, 390 and 188 bp fragments as indicated. Mutagenesis of a *BstE*II site results in the loss of the 390 and 188 bp fragments and gain of a 578 bp, or similar, fragment. Non-specific PCR products are indicated by an asterisk.

6.3.2.3 RFLP analysis of *VEZF1* promoter mutagenesis directed by the TALEN pairs 17/18 and 19/20

An *Msp*I restriction enzyme recognition site was found to overlap the cleavage site of the TALEN pairs 17/18 and 19/20. An RFLP assay involving digestion of the 1310 bp PCR product described above with *Msp*I and *Sac*I, results in 23 fragments (Figure 6.14). The majority of these digest products are smaller than 50 bp in size. I expect to be able to distinguish the larger 340, 125, 109 and 88 bp fragments from the smaller fragments in both the Bioanalyser and gel electrophoresis assays. TALEN-directed mutagenesis of the *Msp*I site present in the TALEN spacer should result in the loss of the 109 and 88 bp fragments and gain of a 197 bp fragment (Figure 6.14). It is possible that the products of mutagenesis may result in RFLP fragments that vary from 197 bp due to the variable nature of NHEJ-mediated indels, as discussed above.

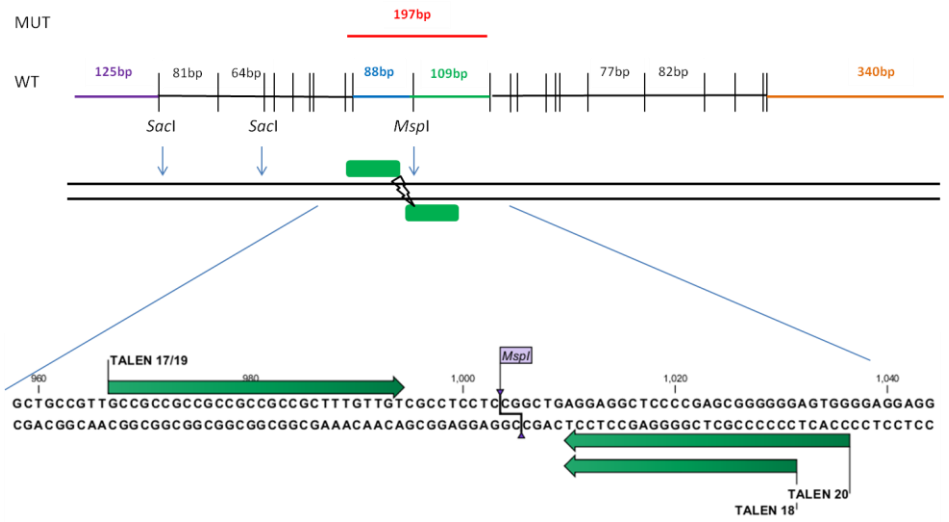

Figure 6.14 The design of RFLP Assay 3
The sequences recognised by the TALEN pairs 17/18 and 19/20 are indicated by filled green arrows over the DNA sequence. An *Msp*I restriction site overlaps the central spacer of these TALEN pairs. Above: A scale diagram depicting the cleavage pattern of a 1310 bp PCR product with *Msp*I and *Sac*I (unlabelled sites are *Msp*I).

HEK293 cells were transfected with pairs of plasmid vectors that encode the three versions of the TALEN pairs 17/18 and 19/20. PCR DNA digested with *Msp*I and *Sac*I was profiled on the Agilent Bioanalyser. DNA from wild type cells was found to digest into the expected 340, 125, 109, 88 bp and smaller fragments (Figure 6.15, WT). However, the electropherograms did not identify any peaks approximating 197 bp in DNA from cells transfected with the TALEN pairs 17/18 and 19/20 (Figure 6.15).

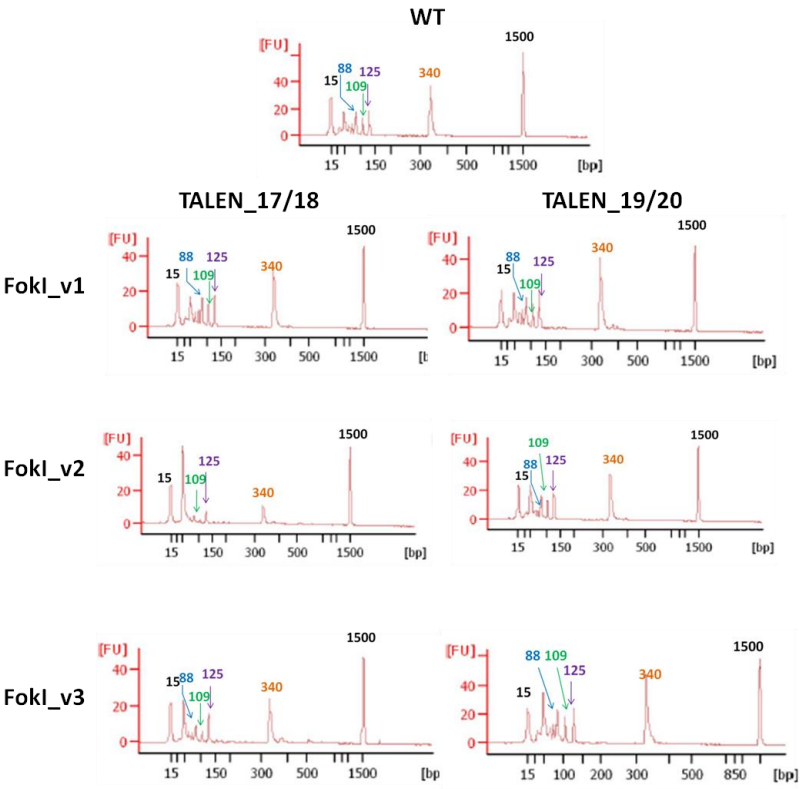

Figure 6.15 Bioanalyser RFLP analysis of *VEZF1* promoter mutagenesis directed by three versions of the TALEN pairs 17/18 and 19/20
Electropherogram plots of DNA fragments resulting from digestion of a 1310 bp PCR of the *VEZF1* promoter with *Msp*I and *Sac*I. Digests were performed on PCRs of HEK293 genomic DNA from wild type (WT) cells and cells transfected with three versions of the TALEN pairs 17/18 and 19/20. Fragment sizes are annotated on each plot, which are determined by the Bioanalyser software based on the 15 bp and 1500 bp markers (black).

Given that the Bioanalyser approach reported less mutagenesis in the RFLP assays described above, the same *Msp*I / *Sac*I digested PCR products shown above were analysed by agarose gel electrophoresis. A diffuse band approximating 197 bp was detected from the DNA of cells transfected with TALEN_17/18 FokI_v2 (Figure 6.16, lane 3). Faint smeared bands of the same size may be present in the DNA from cells transfected with the other TALENs, but mutagenesis cannot be called with any confidence in these other experiments due to background on the gel (Figure 6.16). It is not possible to accurately quantify either the 109 or 88 bp band that may be lost following mutagenesis as the gel is too crowded. It appears that the TALEN pair 17/18 results in a low degree of *VEZF1* promoter mutagenesis, but the TALEN pair 19/20 is not sufficiently active for clear detection with the RFLP assay described here. Faint banding of approximately the correct size is visible, but is not easily distinguished from background (Figure 6.16, lanes 5 – 7). The use of high percentage gels with either specialist high resolution grade agarose or polyacrylamide may be useful for future analyses with RFLP assay 3.

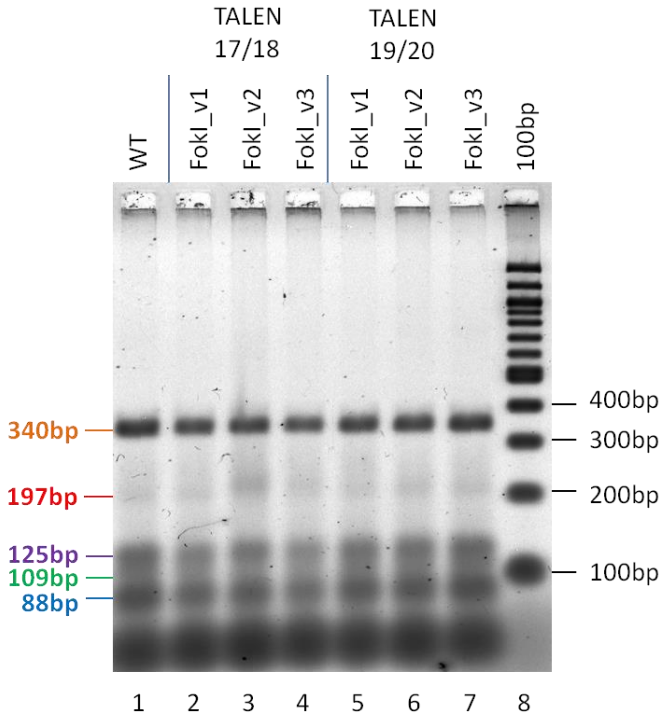

Figure 6.16 RFLP analysis of *VEZF1* promoter mutagenesis directed by three versions of the TALEN pairs 17/18 and 19/20
Agarose gel electrophoresis of DNA from cells transfected with versions of the TALEN pairs 17/18 and 19/20 compared to wild type DNA digested with *Msp*I and *Sac*I. Wild type DNA is digested into 340, 125, 109, 88 bp and smaller fragments as indicated. Mutagenesis of a *Msp*I site results in the loss of the 109 and 88 bp fragments and gain of a 197 bp, or similar, fragment.

6.4 Analysis of TALEN-directed mutations at the *VEZF1* promoter

6.4.1 Analysis of the Effect of NHEJ-induced Mutations on *VEZF1* Expression

Given that v2 and v3 of TALEN pair 15/16 directed 43 and 46% mutation of the *VEZF1* promoter, there is a possibility that these mutations alone are sufficient to result in a loss of *VEZF1* expression in HEK293 cells. This TALEN pair was originally designed to stimulate HDR-mediated removal of the *VEZF1* promoter and first exon by gene targeting (Chapter 5). However, because this TALEN pair overlaps the only annotated TSS of the *VEZF1* gene (Figure 6.11), there is a good chance that TALEN 15/16-directed mutations will reduce *VEZF1* transcription levels. Total RNA was collected from HEK293 cells seven days after transfection with plasmids encoding the v2 and v3 of TALEN pair 15/16. This was the same transfection studied by RFLP assays above. RT-PCR analysis of VEZF1 transcript levels were not reduced in these cells despite the high levels of *VEZF1* promoter mutation (Figure 6.17 A). Furthermore, western blotting analyses found no change in VEZF1 protein levels (Figure 6.17 B). It appears that in a polyclonal mixture of cells, the mutations of the *VEZF1* promoter are not sufficient to significantly reduce transcription. Alternatively, if

loss of VEZF1 expression impacts cell growth and survival, the mutant cells that were identified at day 3 may have been lost from culture by day 7 post-transfection.

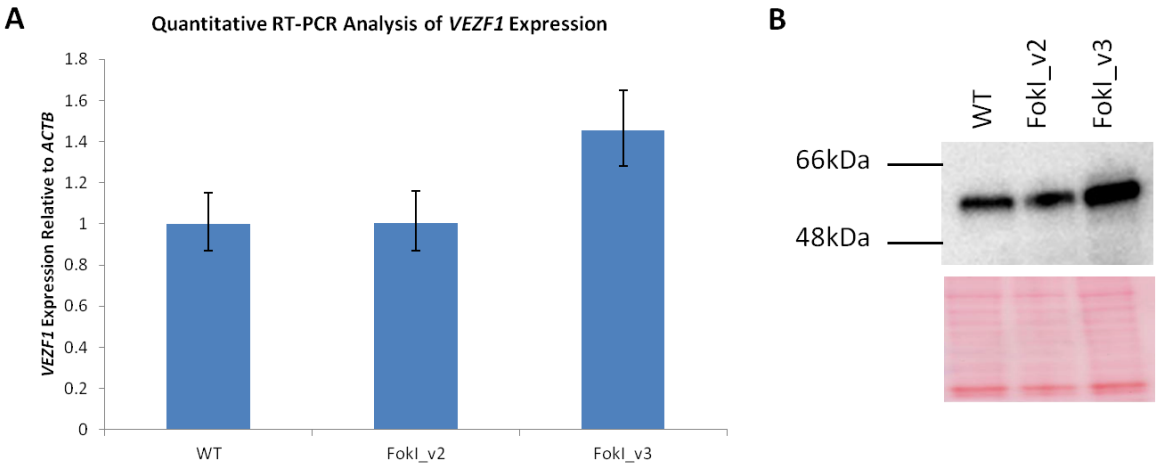

Figure 6.17 *VEZF1* Expression in HEK293 Cells following TALEN 15/16 mutagenesis
(A) Quantitative RT-PCR analysis of *VEZF1* expression levels in HEK293 cells seven days after transfection of TALEN 15/16. Expression levels are normalised relative to *ACTB* expression and *VEZF1* expression in WT cells assigned a value of 1. Error bars represent standard deviation between three technical replicates. This data is from one biological sample set (no replicates) and so no statistical analyses have been performed. (B) Western blot analysis of VEZF1 protein levels in TALEN treated HEK293 cells as compared to Ponceau loading control.

6.4.2 Determining the Nature of TALEN-directed *VEZF1* promoter mutagenesis

RFLP analysis by agarose gel electrophoresis indicated that the *VEZF1* promoter was mutated on 43% of alleles in HEK293 cells three days after transfection of TALEN 15/16 expression plasmids (Figure 6.13). However, gene expression analyses after an additional four days of culture failed to reveal any change in *VEZF1* expression (Figure 6.17). The lack of any change in *VEZF1* expression is unexpected and may be a consequence of studying a polyclonal mixture of mutant and wild type cells. It is possible to generate monoclonal cell lines following TALEN treatment, which could be screened by RFLP and RT-PCR analyses. This is a laborious undertaking, so it would be useful to know what kinds of mutations arise from TALEN-directed mutagenesis. The DNA from HEK293 cells three days after transfection of TALEN 15/16 v2 (that was previously subject to RFLP assays) was then subjected to Sanger sequencing analysis. The 1310 bp PCR product described above was inserted into the plasmid vector pGEM-T Easy via TA cloning (Figure 6.18).

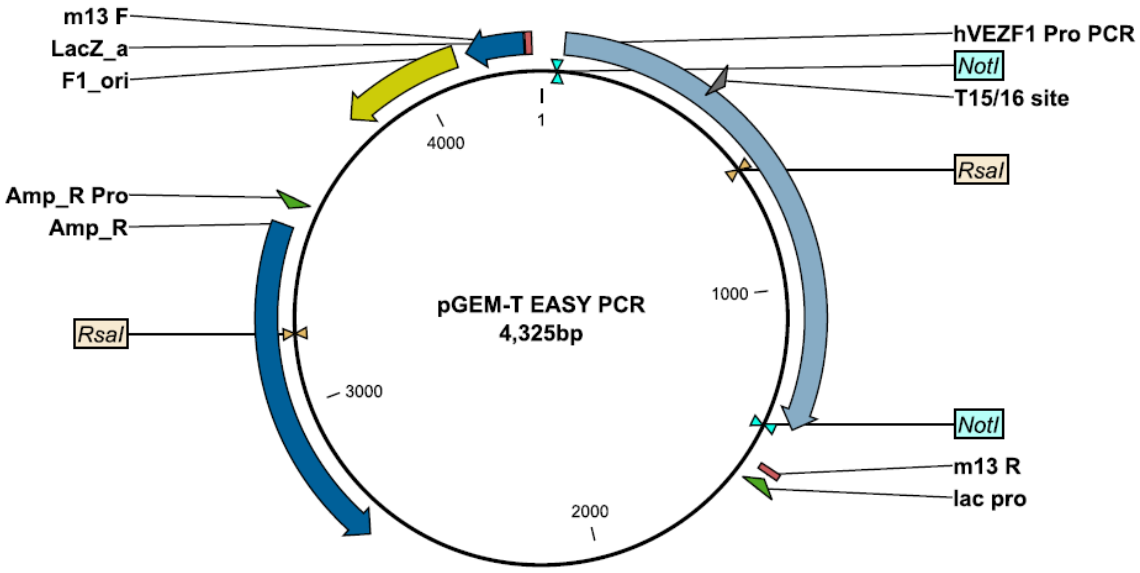

Figure 6.18 TA cloning of TALEN treated *VEZF1* promoter DNA
Scaled map of pGEM-T Easy PCR containing a 1310 bp PCR product of the *VEZF1* promoter from HEK293 cells three days after TALEN 15/16 treatment. Shown on the map are the TALEN binding location and the m13 F & R primer sites used to screen clones for correct ligation of PCR product by colony PCR. Also shown are restriction sites *Not*I & *Rsa*I employed to screen potentially positive clones by restriction digest.

Each *E.coli* transformant clone arises from a single plasmid, which represents a single allele from any given cell in the polyclonal population transfected with TALEN 15/16 v2. Plasmid DNA was prepared from 28 transformant clones derived from TALEN treated DNA and a total of 9 clones derived from wild type DNA (only 2 shown), which were screened for the presence of the cloned PCR fragment by restriction digestion (Figure 6.19). Digestion of the parent pGEM-T Easy vector with *Rsa*I and *Not*I yields 1813 and 1168 bp fragments (e.g. Figure 6.19, lanes 3, 4). Plasmids containing the *VEZF1* promoter PCR product yield additional 749 and 595bp fragments (e.g. Figure 6.19, lane 10). 24 of the plasmid clones contained TALEN treated DNA inserts. Some of these had different fragment lengths compared to wild type DNA, indicative of NHEJ indels (Figure 6.19, compare WT lane 9 with TALEN DNA in lanes 7, 18, 21 and 22).

Five clones containing WT DNA and fifteen clones containing TALEN treated DNA were sequenced using M13 primers, whose sites flank the cloned PCR product (Figure 6.18). Clones from TALEN treated DNA which generated unexpected digest patterns when screened by restriction enzyme digest were selected preferentially over those which resulted in predicted digest products (Figure 6.19, sequencing clone IDs detailed below each clone).

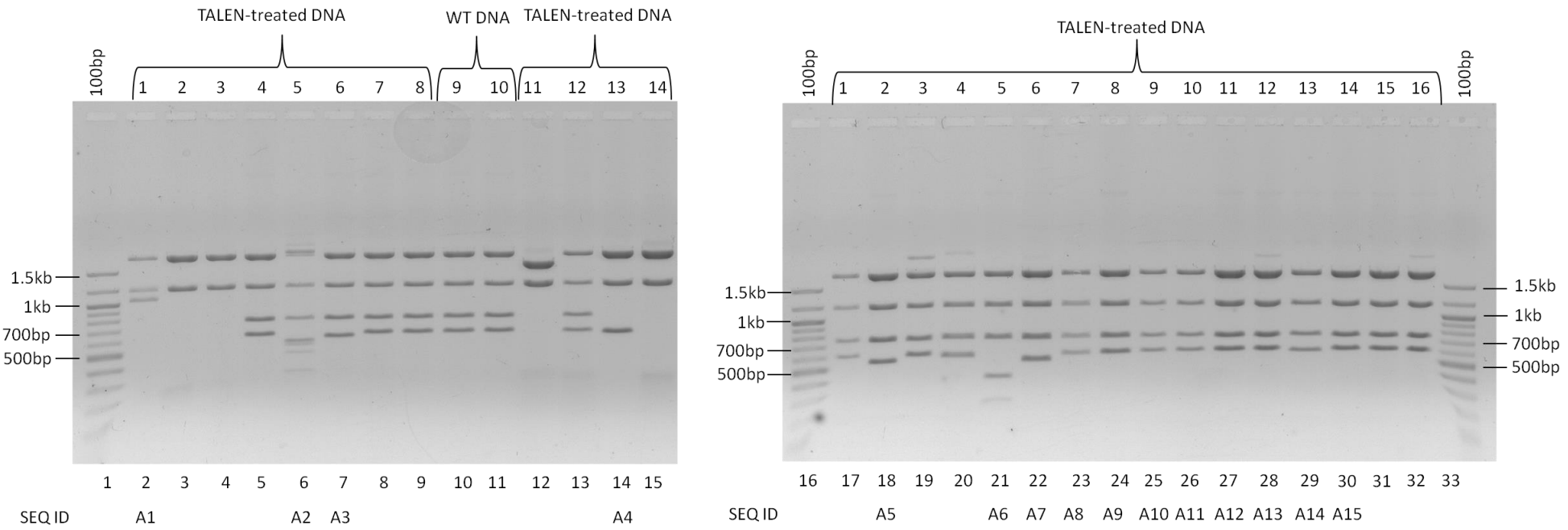

Figure 6.19 Cloning of TALEN treated *VEZF1* promoter DNA

Agarose gel electrophoresis of plasmid DNA from transformants resulting from ligation of *VEZF1* promoter PCR product into pGEM-T Easy Vector by TA-cloning. Digest of empty pGEM-T Easy vector with *Not*I & *Rsa*I generates 1813 and 1168 bp fragments and digestion of pGEM-T Easy plasmids that contain PCR product inserts results in additional products of 749 and 595 bp. Note some clones have produced unexpected digest products (lanes 2, 6, 14 & 21), these clones were preferentially picked for sequencing as they may demonstrate large insertions or deletions following TALEN-mediated DSB repair. Clones picked for sequencing were allocated a sequencing ID as denoted below each relevant lane. Lane 15 demonstrates digest of empty pGEM-T Easy vector, arising from failed ligation of TALEN –treated DNA.

The sequences from wild type clones were aligned to produce a consensus sequence for the HEK293 cells used in the West group (Figure 6.20 A, top). Five of the clones derived from TALEN treated DNA had wild type sequence, and ten were mutant. This gives a mutation rate of 66% (10 out of 15) for the clones studied, compared to the RFLP analysis value of 43% (Figure 6.13). Eight of the clones derived from TALEN treated DNA have short deletions of 5 to 62 bp, with occasional substitutions that localise to the TALEN 15/16 recognition sites (Figure 6.20 A). These mutations are characteristic of DSB repair events by NHEJ (Bibikova et al., 2002, Lloyd et al., 2005). Two of the clones of TALEN-exposed DNA exhibited more substantial deletions of 320 and 758 bp (Figure 6.20 B). These larger deletions both produced unexpected digest patterns when analysed by restriction digest (Figure 6.19, lanes 2 and 14). Furthermore, they are indicative of the types of deletions which would result in *VEZF1* knockout by NHEJ DSB repair, as in both clones the TSS and entire exon 1 of *VEZF1* are deleted. TALEN pair 15/16 can therefore be used in either the v2 or v3 formats to generate mutations of *VEZF1* that should result in the loss of expression without the need for gene targeting.

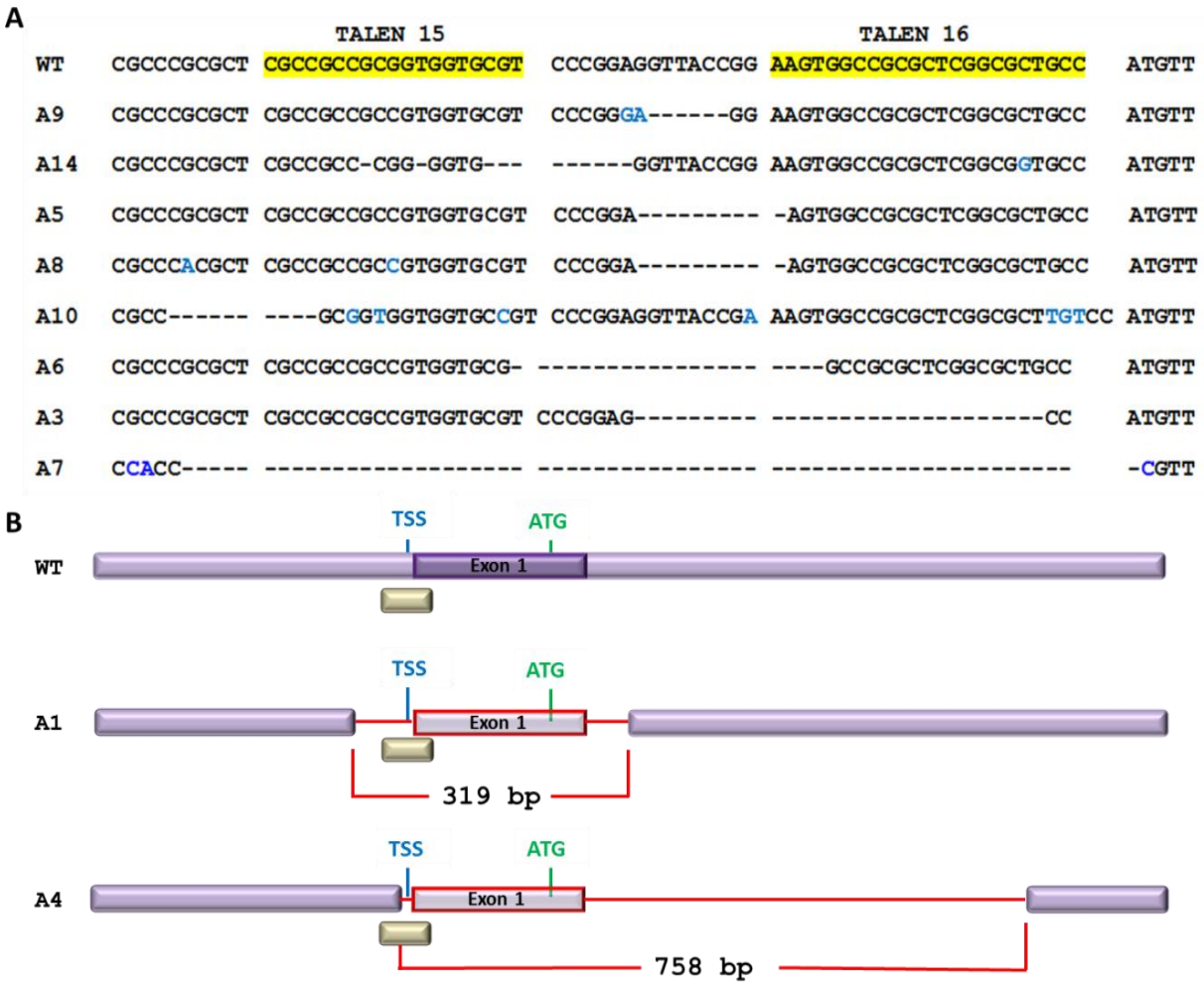

Figure 6.20 Sequence analysis of TALEN-directed *VEZF1* promoter mutagenesis
Sequencing data demonstrating the range of mutations observed at the *VEZF1* promoter three days after transfection of TALEN 15/16 v2. (A) Sequence in the region of the *VEZF1* TSS in 8 clones from TALEN treated DNA compared to wild type. Insertions (blue), substitution (red) and deletions observed (dashes). The binding sites of TALEN 15/16 are highlighted in yellow. (B) Scaled diagram showing larger deletions identified in clones 1 & 4. Purple boxes represent retained sequences and red lines represent deleted sequences. The TALEN 15/16 binding site is illustrated by beige boxes. The size of the deletion is indicated below the deleted region. The location of the TSS and ATG are represented by blue and green lines, respectively. Exon 1, illustrated by a darker purple box, was deleted in both clones.

6.5 Discussion

This chapter has described the development of reliable RFLP assays to test for ZFN and TALEN-directed mutations of the *VEZF1* gene promoter. The assays reveal that human codon optimisation considerably improves the performance of TALENs. The assays also indicate that the addition of "Sharkey" enhanced cleavage mutations allowed obligate heterodimer TALENs to perform to at a comparable level to TALENs that contain wild type *Fok*I cleavage domains. Finally, TALEN-directed mutations of the *VEZF1* gene promoter were sequenced and the effect of such mutations on VEZF1 expression was studied.

The generic mismatch cleavage assays using the *Cel*I or T7EI endonucleases were found to be challenging to optimise, with prevalent non-specific cleavage (section 6.3.1.2). Plasmids of known sequence differences were used as controls for *Cel*I and T7EI assays. The use of these controls relies upon the ability of these enzymes to cleave single base pair mismatches. Both *Cel*I and T7EI have been reported to cleave single base pair mismatches previously (Oleykowski et al., 1998, Kulinski et al., 2000, Tsuji and Niida, 2008) and this principle is the basis of the commercial control samples for the Surveyor assay that uses *Cel*I (Transgenomic). I was not able to define an incubation condition where single base matches could be identified over background non-specific cleavage by either *Cel*I or T7EI. It is possible that sequence context may be important when studying such small mismatches.

The higher than expected degree of non-specific cleavage of reannealed control DNA, irrespective of the degree of mismatch, suggests that mis-annealing may be the predominant flaw in this assay. I compared two different annealing protocols that involve either gradual or step-wise decreases in annealing temperatures after melting. Both approaches resulted in annealed products of the expected size on agarose gels, with a tighter banding observed for the stepwise procedure (data not shown), which I selected for the assays shown in this chapter. It is possible that the quality of the purified PCR product following gel extraction is not good enough for mismatch assays. Contaminants such as guanidium isothiocyanate, agarose fines or ethanol carryover from the agarose gel extraction of genomic PCR products may result in mis-annealed products that are legitimate endonuclease substrates. Future studies should avoid gel extraction completely. This approach would rely on optimised genomic PCR conditions that produce unique products that can be used directly in mismatch analyses. Mismatch analyses offer a generic solution to study the effects of ZFNs and TALENs, so further efforts to resolve the problems I have observed are worthwhile.

RFLP assays were developed to enable the analysis of TALEN-directed mutagenesis. Unfortunately, it was not possible to develop an RFLP assay for the ZFNs prepared in this study due to the paucity of suitable restriction sites in the spacers of our ZFN designs. The presence of restriction sites should be considered when designing any nuclease strategy in future. Despite the straightforward principle of RFLP assays, and careful attempts to design and perform RFLP assays, I found that it can be sometimes difficult to interpret the

digestion profiles from these assays. Low degrees of mutation in mixed populations of cells, non-specific/background PCR products and the variable size and intensity of the digestion products from mutant DNA all present challenges that need to be considered in future TALEN designs.

Primer design is obviously key in NHEJ assay design. Mispriming can lead to the production of non-specific PCR products, which interfere with interpretation. While I recommend TALEN designs be selected based on restriction enzymes present in their spacers, I recommend primers for RFLP assays be designed with expected digest patterns, from both WT and mutant samples, in mind. Inadequate primer design can result in digest patterns which can be problematic to resolve and can obscure interpretation. Primer design therefore must be carefully considered to ensure primers are specific and appropriate for RFLP patterns.

One TALEN pair was found to be very efficient at directing mutagenesis of the *VEZF1* gene promoter. Transfection of TALEN 15/16, which is designed to cleave at the transcription start site of the *VEZF1* gene promoter, resulted in up to 46% mutagenesis when the RFLP products were visualised on agarose gels. All of the TALEN pairs have demonstrated some activity in RFLP assays, although none can be said to have performed as well as TALEN 15/16. TALENs 17/18, 13/14 and 21/22 have all demonstrated clear activity in RFLP analysis visualised on agarose gels (Figure 6.16, lane 3 & Figure 6.10, lanes 3 & 5, respectively). The high activity of TALEN 15/16 could be down to genomic location, as TALEN 15/16 is the only TALEN pair to be positioned separately from the others (Figure 4.2). However, TALE domains which target the downstream genomic site have demonstrated binding as TALE Repressors (Chapter 4). The results of TALE performance are summarised in (Table 6.1).

TALE	Domain Length	Location (bp from TSS)	*VEZF1* Repression		Spacer	Total TALEN Length (domains + spacer)	TALEN performance		
			Run 1	Run 2			v1	v2	v3
13	16	+77	0.12	0.23	30	74	+	++	+
14	28	+123	1.18	1.04					
15	19	-31	0.78	0.71	15	57	++	+++	+++
16	23	+3	1.26	0.59					
17/19	28	+51	0.31	1.11	15	65	+/-	++	+/-
18	22	+94	0.20	0.93					
17/**19**	28	+51	0.31	1.11	15	72	+/-	+	+
20	29	+94	0.41	1.42					
21/23	19	+74	0.50	0.29	30	74	++	+	+
22	25	+123	0.93	0.80					
21/**23**	19	+74	0.50	0.29	30	78	+	+	+
24	29	+123	1.39	0.89					

Table 6.1 Summary of TALE performance

Performance as TALE Repressors does not necessarily translate directly into performance as TALENs, but can serve as some indication as to how TALEN pairs will perform. Take for example TALENs 21/22 and 23/24, both of these TALENs demonstrated modest activity levels, TALEN 21/22 showed a slightly improved performance over 23/24 (Figure 6.10, lane 5). When taking into account the repression mediated by TALE 22 compared to TALE 24, TALE 22 performed slightly better (Figure 4.18) and may be reflected in the performance of TALEN 21/22 versus 23/24. However, the use of TALE Repressor as predictors of TALEN activity does not hold in the case of TALEs 15 and 16, which were the greatest performing TALENs (Figure 6.13) but demonstrated only modest levels of repression (Figure 4.18).

The difficulties in using TALE Repressors as a predictor tool for TALEN activity may be due to unpredictable issues faced with reliance on both members of the TALEN pairs binding together in-situ, such as accessibility of both target sites simultaneously and ability of each partner to interact with the other in context. Experiments involving TALE Repressors do not serve to address either of these issues but can only provide information about each individual TALE domain binding in isolation and relies upon a different effector domain to report functionality. To assess the accessibility of each TALE binding site,

transcription factor ChIP data was obtained from the UCSC Genome Browser and locations of transcription factor binding was compared to TALE and ZFN recognition sites (Figure 6.21). Interestingly, the location where TALEN pair 15/16 binds has less transcription factors binding in this region, which may have enabled this TALEN pair to access its target site and resulted in the higher levels of DNA cleavage observed. Transcription factor binding at the recognition site of TALEs when used as TALE Repressors may also have inhibited Repressor functionality. However, levels of transcription factor binding at the recognition sites for TALEREP 13 and TALEREP 21/23 was similar to that at the recognition sites of other TALE Repressors used in this study. The *VEZF1* promoter remains largely uncharacterised and binding of regulatory elements has not been tested in functional assays. The true composite of transcription factors binding at this region remains to be determined and transcription factor data available from the ENCODE project is by no means an exhaustive list. However, levels of transcription factor binding at intended DNA target sites may be a parameter worth some consideration during the design process but from this data, it does not offer a definitive prediction of the functionality of one TALE domain over another.

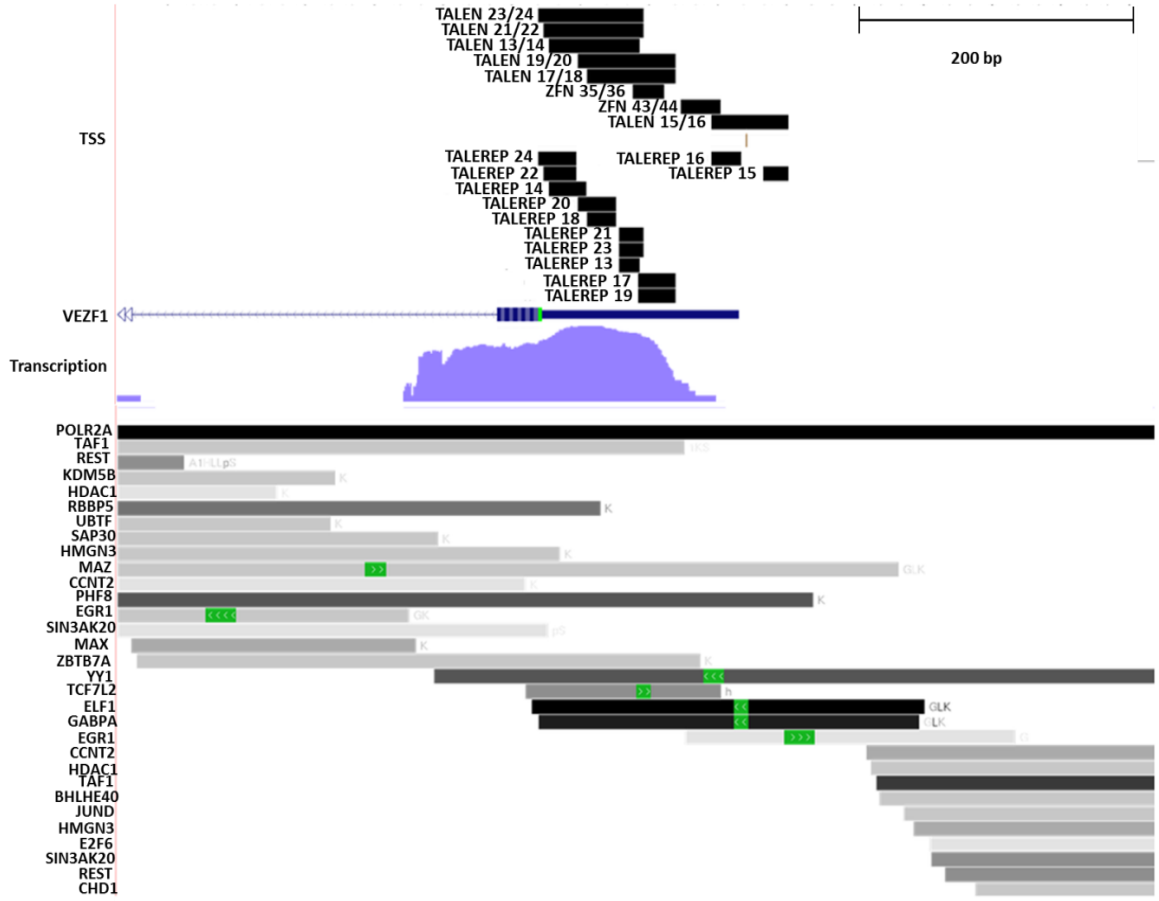

Figure 6.21 Transcription Factor Binding at the *VEZF1* Promoter
Scale diagram of a 770 bp region of at the 5' end of the *VEZF1* gene promoter region based on UCSC Genome Browser alignments (hg19, chr17:56065164-56065933). Locations of ZFN, TALEN and TALEREP recognition sequences in the *VEZF1* gene are depicted by black boxes above *VEZF1* gene, shown in blue, which transcribes from right to left. Locations of ENCODE transcription factor binding, as measured by ChIP in K562 and HEK293 cells, in this region of the *VEZF1* promoter are also shown.

One interesting parallel has been revealed in the comparison between TALENs and TALE Repressors. An inverse relationship was identified between TALE domain length and level of repression induced, and the same can be said for TALENs, when the length of both TALEN monomers, plus the spacer is added together (Figure 6.22). The target site for TALEN 15/16 is a total of 57 bp DNA and is the shortest TALEN pair in our set. TALENs 13/14 and 21/22 also performed well, and both of these TALENs bind 74 bp in total. These shorter TALE domains may have easier access to their target sites than their larger counterparts. The prospect of gaining higher activity levels from shorter domains is interesting as this would result in less labour in terms of TALE construction. Shorter TALE domains, targeting a shorter DNA sequence, may come at a cost of reduced specificity and the balance in terms of activity and specificity must be carefully managed.

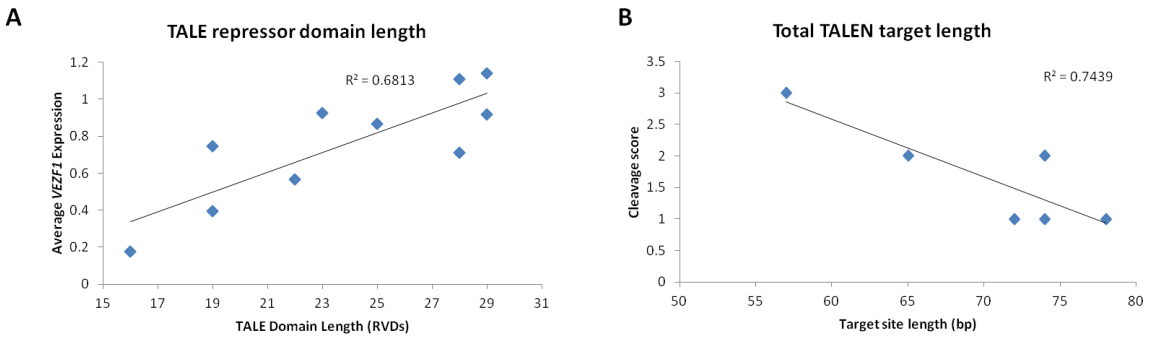

Figure 6.22 Relationship between target site length and TALE activity
Scatter plot illustrating the inverse relationship between (A) TALE repressor domain length and repression of *VEZF1* (average of two biological replicates) and (B) total TALEN target length (TALE domains + spacer) and cleavage activity of v2 TALENs, based on scoring system in Table 6.1.

TALEN pair 15/16 is centred around a 15 bp spacer, which has been demonstrated to be optimal since beginning this study (Miller et al., 2011, Lei et al., 2012). I do not see a clear relationship between spacer size and TALEN performance in our analyses. The modest activity demonstrated by most TALENs has made it difficult to attribute differences in activity to a single design characteristic, which is shared by several TALEN pairs. There is also no correlation between genomic target location and activity of both TALENs and TALEREPs (Figure 6.23). Target site accessibility was taken into account in the design of TALE domains but there does not seem to be an optimal targeting location within the VEZF1 promoter to facilitate the greatest level of gene repression or DNA cleavage.

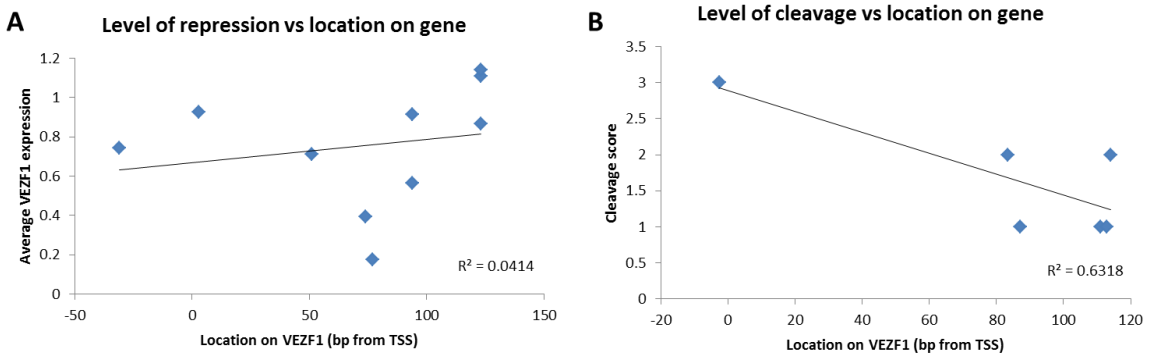

Figure 6.23 Genomic target site location does not determine TALE activity
Scatter plot demonstrating that there is no correlation between (A) TALE repressor domain target site and repression of *VEZF1* (average of two biological replicates) and (B) TALEN target site and cleavage activity of v2 TALENs, based on scoring system in Table 6.1. Location on gene is expressed as most 5' base of TALEREP binding site and as centre of TALEN target site (domains + spacer).

Codon optimisation can be seen to have generally improved activity of TALENs, compare TALEN 15/16 v1 to v2 (Figure 6.13 lane 2 and lane 3). Overall, more distinctive banding can be observed in RFLP assays of DNA treated with codon optimised TALENs, for example TALEN 13/14 v2 (Figure 6.10, lane 3) and TALEN 17/18 v2 (Figure 6.16, lane 3). Performance of TALENs in v3 architectures is similar to that of v2 (Figure 6.13, lanes 3 & 4 and Figure 6.16, lanes 6 & 7), demonstrating that obligate heterodimer *Fok*I domains, including Sharkey, have similar activity levels as WT *Fok*I.

Sequencing analysis of TALEN-mediated mutations in TALEN 15/16 treated cells has demonstrated that it is possible to utilise NHEJ as a means of generating large scale deletions. Due to its target location in the *VEZF1* promoter, there is every likelihood that the resulting mutant cells will have reduced *VEZF1* expression as compared to WT. The two large deletions observed (Figure 6.20) would be appropriate to knock out gene function, at least on one allele. However, in a polyclonal population of cells, no reduction in *VEZF1* expression was observed (Figure 6.17). This enforces the need to generate monoclonal lines from TALEN targeted cells, allowing for a large deletion to be propagated throughout a cell line. Should nuclease-mediated DSB events only occur on one allele within a cell, as is most common (Santiago et al., 2008, Kwon et al., 2013), a second round of targeting may be performed and monoclonal lines generated from re-targeted cells. Our RFLP assay design would allow for easy identification of both homo- and heterozygous targeted lines.

From the results, it has been demonstrated that at least five of our assembled TALENs to disrupt the *VEZF1* promoter are functional, and that one pair, TALEN 15/16 is capable of mediating mutations with high efficiency. The lack of effect on *VEZF1* expression from NHEJ-mediated gene disruption demonstrates the “lottery” in using NHEJ alone as a means of genome editing. The mutations created by NHEJ repair induced by TALEN 15/16 reveal there may be potential to create a *VEZF1* knockout through the generation of monoclonal lines, although this may require two rounds of TALEN targeting. Coupled with the advancements planned for the targeting vector (Chapter 5), these results demonstrate that generation of a VEZF1 gene knockout by TALEN-mediated DSBs is within grasp.

Chapter 7

Summary and Conclusions

7.1 Summary of work presented in this thesis

This study has described strategies to direct repression or mutation of *VEZF1* using custom DNA binding proteins. Previously, RNAi-mediated knockdown of VEZF1 was found to be insufficient to deplete protein binding levels sufficiently to study the effects of VEZF1 on its many gene targets (Strogantsev, 2009). The *VEZF1* gene promoter was targeted using ZFNs and TALEs fused to repressor and nuclease domains. TALE Repressors reduced *VEZF1* expression to as little as 12% of WT expression levels and nucleases disrupted the *VEZF1* gene promoter, with some TALENs resulting in over 65% mutation rates. Analysis of TALEN-mediated mutations revealed that, in some cases, the *VEZF1* promoter had been removed during DSB repair via NHEJ. This study has identified an interesting correlation between TALE domain length and level of activity, both as repressors and nucleases. The findings and reagents described here can be applied to future approaches to modify any desired gene of interest.

7.2 TALE-mediated Repression of *VEZF1*

Ten TALE proteins which target the *VEZF1* gene promoter were designed and assembled. The TALE proteins constructed were between 16 – 29 RVDs in length, in order to determine the optimal TALE design constraints. Each of the designed TALEs was assembled into a custom made expression vector, which mediated expression of our TALEs as repressor proteins utilising a SID-repressor protein. TALE-Repressors were transfected into HEK293 cells and VEZF1 repression was demonstrated by six of the ten TALE-Repressors constructed. Interestingly, I observed a correlation between TALE domain length and level of repression achieved. The longest TALE domains promoted the least repression of VEZF1, while the greatest knockdown was achieved using the shortest TALEs of our set.

This inverse correlation is not a phenomenon which has been previously reported in studies employing either TALE Repressors or TALE Activators. This observation, however, has been reported in one study using TALE Nucleases (see Section 7.3).

TALE domains constructed and fused to transcriptional repression domains, tend to be around 20 RVDs in length and achieve between 2 and 10 fold repression (Cong et al., 2012, Garg et al., 2012, Mahfouz et al., 2012). In this study, repression of down to 12% of WT levels was achieved using the shortest TALE domain constructed. This may be due to shorter TALEs perhaps finding easier access to their target sites by facing less competition with other endogenous transcription factors, or perhaps DNA-binding affinity of a TALE may decrease with TALE length. Another possibility is that the higher activity observed is due to higher expression of these shorter TALEs, which has not been assessed in this study. The utilisation of shorter TALE domains in order to gain higher levels of gene repression is a recommendation that must be considered carefully, as this cannot come at the expense of target specificity. The greatest level of repression observed in this study was mediated by a TALE which is 16 RVDs in length, which should be sufficient to target a unique DNA sequence within the human genome. The utilisation of TALE domains shorter than this may result in high levels of cytotoxicity, as has been observed with TALE nucleases of this length (Reyon et al., 2012b).

This project arose through the difficulties in creating a robust VEZF1 knockdown and TALE Repressors do offer an alternative approach to this aim. The experiment described in Chapter 4, using TALE Repressors in HEK293 cells, has since been repeated by a colleague and analysis has been performed at later time points to determine if continuation of the experiment for a longer time period would promote greater levels of knockdown. However, this has revealed that the repressive effects of TALE-Repressors are becoming silenced when left for longer periods of time. This suggests that, in order to analyse cells at their greatest level of knockdown, samples must be collected at early time points, or that selection must be administered to retain high levels of repression. Employing a combination of TALE Repressors and RNAi has been demonstrated to have a synergistic effect, resulting in almost complete target gene repression (Garg et al., 2012). Following the levels of VEZF1 gene repression achieved with TALE Repressors 13 and 21/23, work is now ongoing in the West Lab to assess the level of *VEZF1* repression attainable by combining with RNAi. This may facilitate the development of a robust and constitutive VEZF1 knockdown.

7.3 Nuclease-mediated modification of the *VEZF1* Promoter

Three pairs of ZFNs and the six pairs of TALENs were assembled to target the VEZF1 gene promoter. Originally, it was intended to create a VEZF1 knockout via HDR mediated removal of the VEZF1 promoter and insertion of a transgene. However, gene targeting experiments revealed high levels of random integration of the targeting vector, pTRGT_VEZF1. The designed nucleases were therefore used to promote NHEJ-mediated VEZF1 knockout, without the use of a targeting vector system.

In this study, it was not possible to perform analysis of nuclease-mediated NHEJ using generic mismatch assays *Cel*I and T7EI as both enzymes displayed high levels of non-specific cleavage activity. While not a widely reported issue, analysis of TALEN-induced DNA mutations has now moved largely to the use of RFLP assays and this is now incorporated as a standard consideration in TALEN design approaches, such as TALE-NT (Doyle et al., 2012).

Restriction fragment length polymorphism (RFLP) assays were developed to determine the level of nuclease-mediated NHEJ. Unfortunately, it was not possible to develop RFLP assays to determine the mutagenic activity of ZFNs used in this project, due to the nature of the restriction enzyme recognition sites found in the ZFN spacers. This demonstrates the need to consider the design of mutation detection assays at the initial TALEN and ZFN design stage. However, RFLP assays developed to asses TALEN activity have demonstrated that all TALEN pairs constructed in this project are functional, although to varying degrees. It is clear that codon optimisation improves TALEN activity and that obligate heterodimer *Fok*I domains which include the Sharkey mutations can perform at similar levels to those with the WT *Fok*I domain and this is in agreement with the findings of similar studies (Miller et al., 2011, Sanjana et al., 2012, Zhang et al., 2011).

One TALEN pair, TALEN 15/16 was very efficient at directing mutagenesis of the *VEZF1* gene promoter. This TALEN pair binds the DNA sequence at the TSS of VEZF1. Sequence analysis of targeted cells revealed large scale deletions of the VEZF1 promoter, which would be sufficient to create a VEZF1 knockout. Interestingly, TALEN pair 15/16 was the shortest TALEN pair of all those designed, binding a total of 42 bp excluding spacer. This

observation is in parallel with the observation made using TALE repressors – the shortest TALEs drive the greatest level of activity.

The observation that the greatest level of target site mutation was achieved using the shortest TALE domain, has been previously reported in another study (Reyon et al., 2012b). TALEN disruption of a GFP reporter gene was analysed 2 and 5 days post-transfection. Shorter TALENs (8 - 10 RVDs to each TALE) were found to mediate greater GFP disruption when compared to longer TALENs (14 - 19 RVDs per TALE domain) at 2 days post-transfection. However, when quantified 5 days post-transfection, GFP disruption mediated by the shorter length TALENs had significantly reduced, while mutagenicity by longer TALENs was fairly consistent. The initial high levels of GFP disruption observed when using shorter TALENs suggests that shorter TALENs can be more cytotoxic. The short TALEs used by Reyon *et al.* (2012b) are much shorter by the TALE domains used in this study while the same correlation between TALE length and activity has been observed. However, the results of both the study conducted by Reyon *et al.* (2012b) and data presented in this thesis recommend the use of TALE domains of around the same length. I would recommend each TALE domain be up to around 20 RVDs in length to reduce cytotoxicity (as observed in Reyon *et al.* (2012b)), maintain target site specificity and offer greater chance of improved functionality.

7.4 CRISPR-Cas9 Technology

During the course of this work, another gene targeting technology was developed which utilises the RNA-guided endonuclease Cas9 to create DSBs at a predefined target locus. The CRISPR-Cas9 gene targeting system offers an alternative platform for gene targeting, operating via a different mode of DNA recognition to ZFN and TALEN technologies.

The CRISPR-Cas9 system is an adaptable immune response used by many bacteria to protect from invading DNA of viruses or plasmids via RNA guided DNA cleavage (Jinek et al., 2012, Hsu et al., 2014). Short segments of foreign DNA spacers are inserted between CRISPRs (Clustered Regulatory Short Interspaced Palindromic Repeats). Once transcribed, CRISPR arrays are processed into CRISPR-RNAs (crRNAs) which harbour a unique sequence, transcribed from the invading DNA. The crRNA then hybridises with transactivating crRNA (tracrRNA), collectively forming a short guide RNA (sgRNA, Figure

7.1). The sgRNA then forms a complex with Cas9 nuclease (Deltcheva et al., 2011). The CRISPR-Cas9 complex is directed to the desired DNA target site via twenty variable nucleotides at the 5' end of the sgRNA. DNA specificity of the CRISPR-Cas9 platform is driven by these twenty nucleotides, which bind the intended DNA target site via standard Watson-Crick base pairing (Gasiunas et al., 2012, Jinek et al., 2012). The designed DNA target site must be adjacent to a short sequence called the protospacer adjacent motif (PAM) and DNA cleavage occurs 3 nucleotides 5' of this motif (Harrison et al., 2014). Cas9-mediated DSBs can then be repaired by NHEJ or HDR, performing targeted genome manipulations.

Web based design tools are now available to select target DNA sequences within a region or gene of interest. Once designed, the sgRNA sequences can then be ordered as two short oligos and cloned into an expression vector, many flavours of which are now available through www.addgene.org.

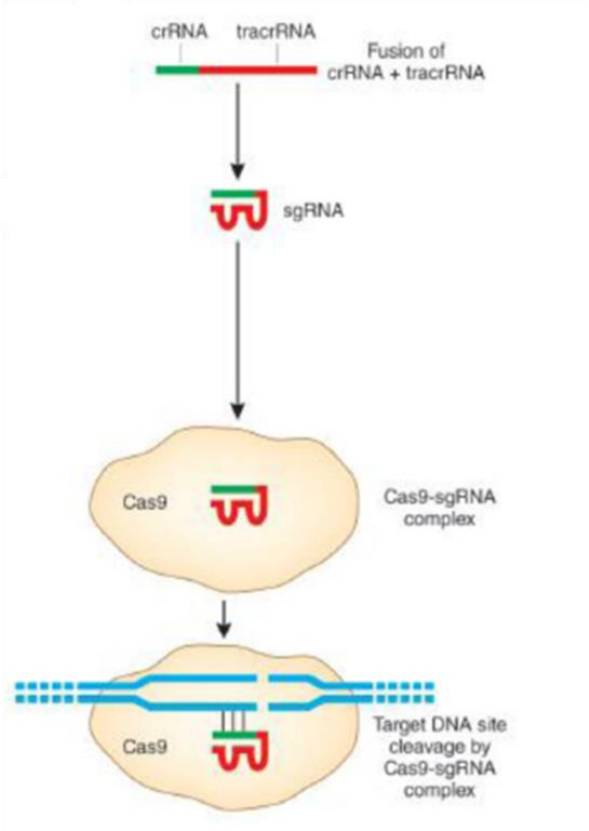

Figure 7.1 Overview of CRISPR-Cas9 System for Targeted Gene Editing
Cartoon depicting the engineered CRISPR-Cas9 system, which is composed of a fusion of crRNA and tracrRNA to form an sgRNA. The sgRNA forms a complex with Cas9 DNA nuclease. The sgRNA directs Cas9 to the target DNA cleavage site, which is complimentary to sequences in the first 20 nucleotides in the crRNA (Sander and Joung, 2014).

CRISPR-Cas9 technology has been used to successfully manipulate many genes in a variety of cell types and model organisms including cultured human cells, human pluripotent stem cells, zebrafish, fruit files and rats (reviewed in (Hsu et al., 2014). CRISPR-Cas9 technology offers the potential for higher rates of gene targeting than ZFNs and TALENs, with generally around 30% of cells in a targeted population harbouring evidence of Cas9 mediated DNA cleavage repair (Gasiunas and Siksnys, 2013).

The CRISPR-Cas9 technology offers the opportunity to introduce multiple DSBs in parallel by expression of multiple sgRNAs to various locations. This adaptation, known as multiplexing, has been used to target up to 5 genes simultaneously and offers a unique advantage of the CRISPR-Cas9 platform over ZFNs and TALENs. Multiplexing can be exploited to direct large scale targeted deletion of a desired gene or to knock out multiple genes in parallel (Cong et al., 2013, Mali et al., 2013).

Modified Cas9 endonucleases have been developed which allow the use of the CRISPR-Cas9 platform as a single stranded DNA nickase and as a fusion to a transcriptional regulators, such as activators and repressors. Cas9 nickases can be utilised to promote DNA repair via HDR rather than NHEJ, inserting targeted gene mutations (Gasiunas et al., 2012, Jinek et al., 2012). Transcriptional regulation directed by CRISPR-Cas9 has been demonstrated to enhance or repress transcription to levels of around 5-20 fold activation and around 5 - 15 fold repression of target genes in human cells. Furthermore, several studies have demonstrated a synergistic effect of when multiplexed CRISPR-transcriptional regulators are used simultaneously (Cheng et al., 2013, Gilbert et al., 2013, Maeder et al., 2013a, Perez-Pinera et al., 2013).

CRISPR-Cas9 technology is largely in its infancy and still faces similar challenges those encountered with ZFN and TALEN platforms. High levels of off target cleavage have been reported in some studies. There are ongoing studies investigating ways to improve efficiency and activity of CRISPR-Cas9 technology (Fu et al., 2013, Fu et al., 2014, Hsu et al., 2013). CRISPRs will undoubtedly benefit from some of the groundwork already covered in the development of ZFN and TALEN technologies and the advances in this platform will occur rapidly. The West Lab has now constructed CRISPRs to target *VEZF1* and is utilising knowledge gained from targeting strategy design and mutation screening in this study to develop further strategies to target *VEZF1* and other genes. Future genome modification projects conducted by the West lab will now incorporate the design of RFLP assay as part of initial gene targeting strategy design processes. From the issues faced in this study, it is evident that careful consideration must be taken in both the design of PCR primers and selection of restriction enzymes when designing RFLP mutation detection assays. For ease of analysis, it is recommended that PCR products be 1 - 1.5 kb in length and restriction enzymes chosen cleave this PCR amplicon into up to 5 fragments. Furthermore, it is useful to design the assay such that the restriction enzyme which cleaves in the TALEN spacer also cleaves at least once more in the PCR product, ensuring an internal control for enzyme functionality. Lastly, the choice of other restriction enzymes should also be carefully calculated. Both enzymes must be suitable for use in the double digest reactions and digest product resolution, including mutant fragment detection, should be considered as part of the genome modification design strategy from the first step. The implementation of design parameters recommended here have allowed CRISPR projects in the West Lab to progress much faster and have generated

results for analysis of the biological effect of VEZF1 mutation in a much shorter timescale than would have been possible without the knowledge and experience gained from genome engineering strategies described in this thesis.

7.5 Conclusions

Prior to this study, attempts at creating a VEZF1 knockdown had been insufficient for loss of function, resulting in depletion of VEZF1 protein binding at around only half of its putative target genes (Strogantsev, 2009). This project described the design, assembly and assessment of genome modification tools to target the *VEZF1* gene promoter. Engineered DNA binding modules constructed in this study have been used to successfully repress and modify the *VEZF1* promoter. Investigations carried out have enabled the development of recommended parameters to aid in the design of successful TALE domains. The knowledge and reagents described here have paved the way for future creation of a VEZF1 knockout for investigation. Furthermore, many of the vector systems constructed for this project are amenable for insertion of any designed TALE to enable future studies of any desired target gene.

The jury is, so far, still out on whether CRISPR technology will surpass TALENs as the method of choice for targeted gene modification. However, ZFNs are to date the only of these methods to have been tested in human trials. It is apparent, however, that the lessons learned in assessment of ZF and TALE technologies will spur the progress of CRISPRs and keep genome modification at the very cutting edge of science.

Appendix I

Primer Sequences

Primer Name	Sequence
pCR8_F1	TTGATGCCTGGCAGTTCCCT
pCR8_R1	CGAACCGAACAGGCTTATGT
RFP5_Nhe_Age	GATCGCTAGCACCGGTCGCCACCATGAGCG
RFP3_Bam_Not	GATCGCGGCCGCTTAGGATCCTCTGTGCCCCAGTTTGCTAG
VEZF1_5'arm_H3F	GATCAAGCTTGTTCTGGGGGTGAAGGAG
VEZF1_5'arm_AscR	GATCGGCGCGCCTCGGAGCTCAGCCAGTGC
VEZF1_3'arm_H3F	GATCAAGCTTGTACTGGGGGCCCCCCCGGG
VEZF1_3'arm_AscR	GATCGGCGCGCCGAGAGCTGGACAGAGGGAGA
FokI_Cell_Cont F	GCACCATGGGACCTAAGAAA
FokI_Cell_Cont R	GCGATGCAATTTCCTCATTT
hVEZF1pro-374 F	CTCCGCTGAGGGTCTAACAG
hVEZF1pro+895 R	CAGAGGGAGAAGGTGTCAGC
M13 F	GTAAAACGACGGCCAGT
M13 R	AACAGCTATGACCATG
β-Actin_F (qRT-PCR)	TGCTGCGCTCGTTGTTGA
β-Actin_R (qRT-PCR)	TCGCCGGGGACGATG
VEZF1_F (qRT-PCR)	AAAGGATCGCATGACCTACC
VEZF1_R (qRT-PCR)	AATGATCAGGCCTCGAGAAG

Appendix II

Sequencing Data

A. ZF35

XbaI XmaI ZF7 ZF23 ZF42 BamHI
CGACGATAAATCTAGACCCGGGGAGAAGCCGCATATTTGCCACATCCAAGGCTGTGGG
AAGGTGTACGGCCGTAGCGATGACCTGACACGTCACTTGCGCTGGCATACCGGGGAGC
GGCCATTTATGTGTACCTGGTCATACTGTGGGAAACGCTTCACACGTAGCGATGACCT
GCAACGTCACAAACGTACACACACCGGGGAGAAGAAATTTGCCTGCCCTGAGTGTCCT
AAGCGCTTCATGCGTAGCGATGACCTGACACGTCATATCAAGACCCACACCGGTGGAT
CCCAACTAGTCAAAAGTGAACTGGAGGAGAA

B. ZF36

XbaI XmaI ZF6 ZF30 ZF50 BamHI
GACGACGATAAATCTAGACCCGGGGAGAAGCCGCATATTTGCCACATCCAAGGCTGTG
GGAAGGTGTACGGCGAGCGCGGTACACTTGCACGTCACTTGCGCTGGCATACCGGGGA
GCGGCCATTTATGTGTACCTGGTCATACTGTGGGAAACGCTTCACAGATCGTAGCGCA
CTGGCACGTCACAAACGTACACACACCGGGGAGAAGAAATTTGCCTGCCCTGAGTGTC
CTAAGCGCTTCATGCAAAGCAGCGCACTGACACGTCATATCAAGACCCACACCGGTGG
ATCCCAACTAGTCAAAAGTGAACTGGAGGAGAA

C. ZF37

XbaI XmaI ZF11 ZF26 ZF42 BamHI
GGATGACGACGATAAATCTAGACCCGGGGAGAAGCCGCATATTTGCCACATCCAAGGC
TGTGGGAAGGTGTACGGCCGTAGCGATCACCTGGCACGTCACTTGCGCTGGCATACCG
GGGAGCGGCCATTTATGTGTACCTGGTCATACTGTGGGAAACGCTTCACAGATCGTAG
CCACCTCGCACGTCACAAACGTACACACACCGGGGAGAAGAAATTTGCCTGCCCTGAG
TGTCCTAAGCGCTTCATGCGTAGCGATGACCTGACACGTCATATCAAGACCCACACCG
GTGGATCCCAACTAGTCAAAAGTGAACTGGAGGAGA

D. ZF38

XbaI XmaI ZF11 ZF23 ZF36 BamHI
AAGGATGACGACGATAAATCTAGACCCGGGGAGAAGCCGCATATTTGCCACATCCAAG
GCTGTGGGAAGGTGTACGGCCGTAGCGATCACCTGGCACGTCACTTGCGCTGGCATAC
CGGGGAGCGGCCATTTATGTGTACCTGGTCATACTGTGGGAAACGCTTCACACGTAGC
GATGACCTGCAACGTCACAAACGTACACACACCGGGGAGAAGAAATTTGCCTGCCCTG
AGTGTCCTAAGCGCTTCATGCAAAGCGGGAACCTGGCACGTCATATCAAGACCCACAC
CGGTGGATCCCAACTAGTCAAAAGTGAACTGGAGGAGAA

E. ZF43

XbaI XmaI ZF98 ZF91 ZF132 ZF78 BamHI
CCTCTAGACCCGGGGAGAAGCCTTACAAATGCCCAGAATGTGGAAAGAGTTTTAGCCA
TACTGGACACCTCCTTGAACACCAGAGAACACATACCGGGGAGAAGCCTTACAAATGC
CCAGAATGTGGAAAGAGTTTTAGCAGGGCAGATAATCTCACAGAACACCAGAGAACAC
ATACCGGGGAGAAGCCTTTCCAGTGCAAGACCTGCCAGCGGAAATTCAGCCGGAGCGA
CCACCTGAAGACCCACACCCGGACACACACCGGGGAGAAGCCTTACAAATGCCCAGAA
TGTGGAAAGAGTTTTAGCTCTCCTGCTGATCTCACAAGACACCAGAGAACACATACCG
GTGGATCCTCTGCACAGTAGT

F. ZF44

XbaI XmaI ZF108 ZF139 ZF108 ZF108 BamHI
AAATCTAGACCCGGGGAGAAGCCTTACACGTGCAGCGACTGCGGCAAGGCATTTCGCG
ACAAGTCCTGCCTGAACCGCCACAGACGCACCCACACCGGGGAGAAGCCTTACGAGTG
CAACTACTGCGGCAAGACGTTCAGCGTGAGCTCCACGCTCATTCGCCACCAGCGCATC
CACACCGGGGAGAAGCCTTACACGTGCAGCGACTGCGGCAAGGCATTTCGCGACAAGT
CCTGCCTGAACCGCCACAGACGCACCCACACCGGGGAGAAGCCTTACACGTGCAGCGA
CTGCGGCAAGGCATTTCGCGACAAGTCCTGCCTGAACCGCCACAGACGCACCCACACC
GGTGGATCCCAACTAGTCAAAAG

G. FokI KK-S

XbaI HindIII BamHI FokI E490K,I538K - S418P,K441E
ACGACGATAAATCTAGATAGTAAGCTTGGATCCCAACTAGTCAAAAGTGAACTGGAGG
AGAAGAAATCTGAACTTCGTCATAAATTGAAATATGTGCCTCATGAATATATTGAATT
AATTGAAATTGCCAGAAATCCCACTCAGGATAGAATTCTTGAAATGAAGGTAATGGAA
TTTTTTATGAAAGTTTATGGATATAGAGGTGAACATTTGGGTGGATCAAGGAAACCGG
ACGGAGCAATTTATACTGTCGGATCTCCTATTGATTACGGTGTGATCGTGGATACTAA
AGCTTATAGCGGAGGTTATAATCTGCCAATTGGCCAAGCAGATGAAATGCAACGATAT
GTCAAAGAAAATCAAACACGAAACAAACATATCAACCCTAATGAATGGTGGAAAGTCT
ATCCATCTTCTGTAACGGAATTTAAGTTTTTATTTGTGAGTGGTCACTTTAAAGGAAA
CTACAAAGCTCAGCTTACACGATTAAATCATAAGACTAATTGTAATGGAGCTGTTCTT
AGTGTAGAAGAGCTTTTAATTGGTGGAGAAATGATTAAAGCCGGCACATTAACCTTAG
AGGAAGTGAGACGGAAATTTAATAACGGCGAGATAAACTTTTAAGGGCCCTTCGAAGG
TAAGCCTATCCCTAACC

H. FokI EL-S

XbaI HindIII BamHI FokI Q486E,I499L - S418P,K441E
ACGACGATAAATCTAGATAGTAAGCTTGGATCCCAACTAGTCAAAAGTGAACTGGAGG
AGAAGAAATCTGAACTTCGTCATAAATTGAAATATGTGCCTCATGAATATATTGAATT
AATTGAAATTGCCAGAAATCCCACTCAGGATAGAATTCTTGAAATGAAGGTAATGGAA
TTTTTTATGAAAGTTTATGGATATAGAGGTGAACATTTGGGTGGATCAAGGAAACCGG
ACGGAGCAATTTATACTGTCGGATCTCCTATTGATTACGGTGTGATCGTGGATACTAA
AGCTTATAGCGGAGGTTATAATCTGCCAATTGGCCAAGCAGATGAAATGGAGCGATAT
GTCAAAGAAAATCAAACACGAAACAAACATCTCAACCCTAATGAATGGTGGAAAGTCT
ATCCATCTTCTGTAACGGAATTTAAGTTTTTATTTGTGAGTGGTCACTTTAAAGGAAA
CTACAAAGCTCAGCTTACACGATTAAATCATAAGACTAATTGTAATGGAGCTGTTCTT
AGTGTAGAAGAGCTTTTAATTGGTGGAGAAATGATTAAAGCCGGCACATTAACCTTAG
AGGAAGTGAGACGGAAATTTAATAACGGCGAGATAAACTTTTAAGGGCCCTTCGAAGG
TAAGCCTATCCCTAACC

References

AITSEBAOMO, J., KINGSLEY-KALLESEN, M. L., WU, Y., QUERTERMOUS, T. & PATTERSON, C. 2001. Vezf1/DB1 is an endothelial cell-specific transcription factor that regulates expression of the endothelin-1 promoter. *J Biol Chem,* 276**,** 39197-205.

AYER, D. E., LAHERTY, C. D., LAWRENCE, Q. A., ARMSTRONG, A. P. & EISENMAN, R. N. 1996. Mad proteins contain a dominant transcription repression domain. *Mol Cell Biol,* 16**,** 5772-81.

BABON, J. J., MCKENZIE, M. & COTTON, R. G. 2003. The use of resolvases T4 endonuclease VII and T7 endonuclease I in mutation detection. *Mol Biotechnol,* 23**,** 73-81.

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

BATEMAN, A., BIRNEY, E., CERRUTI, L., DURBIN, R., ETWILLER, L., EDDY, S. R., GRIFFITHS-JONES, S., HOWE, K. L., MARSHALL, M. & SONNHAMMER, E. L. 2002. The Pfam protein families database. *Nucleic Acids Res,* 30**,** 276-80.

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

BELL, A., WEST, A. & FELSENFELD, G. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell,* 98**,** 387-96.

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

BEUMER, K. J., TRAUTMAN, J. K., CHRISTIAN, M., DAHLEM, T. J., LAKE, C. M., HAWLEY, R. S., GRUNWALD, D. J., VOYTAS, D. F. & CARROLL, D. 2013. Comparing zinc finger nucleases and transcription activator-like effector nucleases for gene targeting in Drosophila. *G3 (Bethesda),* 3**,** 1717-25.

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

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

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

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

BOGDANOVE, A. J., SCHORNACK, S. & LAHAYE, T. 2010. TAL effectors: finding plant genes for disease and defense. *Curr Opin Plant Biol,* 13**,** 394-401.

BRENNEMAN, M., GIMBLE, F. S. & WILSON, J. H. 1996. Stimulation of intrachromosomal homologous recombination in human cells by electroporation with site-specific endonucleases. *Proc Natl Acad Sci U S A,* 93**,** 3608-12.

BULTMANN, S., MORBITZER, R., SCHMIDT, C. S., THANISCH, K., SPADA, F., ELSAESSER, J., LAHAYE, T. & LEONHARDT, H. 2012. Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res,* 40**,** 5368-77.

BÜTTNER, D. & BONAS, U. 2002. NEW EMBO MEMBER'S REVIEW: Getting across—bacterial type III effector proteins on their way to the plant cell. *Embo j.* Oxford, UK.

CADE, L., REYON, D., HWANG, W. Y., TSAI, S. Q., PATEL, S., KHAYTER, C., JOUNG, J. K., SANDER, J. D., PETERSON, R. T. & YEH, J. R. 2012. Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. *Nucleic Acids Res,* 40**,** 8001-10.

CARLSON, D. F., TAN, W., LILLICO, S. G., STVERAKOVA, D., PROUDFOOT, C., CHRISTIAN, M., VOYTAS, D. F., LONG, C. R., WHITELAW, C. B. & FAHRENKRUG, S. C. 2012. Efficient TALEN-mediated gene knockout in livestock. *Proc Natl Acad Sci U S A,* 109**,** 17382-7.

CARPENTER, A. E., MEMEDULA, S., PLUTZ, M. J. & BELMONT, A. S. 2005. Common Effects of Acidic Activators on Large-Scale Chromatin Structure and Transcription. *Mol Cell Biol.*

CARROLL, D. 2011. Genome engineering with zinc-finger nucleases. *Genetics,* 188**,** 773-82.

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

CHEN, S., OIKONOMOU, G., CHIU, C. N., NILES, B. J., LIU, J., LEE, D. A., ANTOSHECHKIN, I. & PROBER, D. A. 2013. A large-scale in vivo analysis reveals that TALENs are significantly more mutagenic than ZFNs generated using context-dependent assembly. *Nucleic Acids Res,* 41**,** 2769-78.

CHENG, A. W., WANG, H., YANG, H., SHI, L., KATZ, Y., THEUNISSEN, T. W., RANGARAJAN, S., SHIVALILA, C. S., DADON, D. B. & JAENISCH, R. 2013. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res,* 23**,** 1163-71.

CHOULIKA, A., PERRIN, A., DUJON, B. & NICOLAS, J. F. 1995. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. *Mol Cell Biol,* 15**,** 1968-73.

CHRISTIAN, M., CERMAK, T., DOYLE, E. L., SCHMIDT, C., ZHANG, F., HUMMEL, A., BOGDANOVE, A. J. & VOYTAS, D. F. 2010. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics,* 186**,** 757-61.

CHUNG, J. H., WHITELEY, M. & FELSENFELD, G. 1993. A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. *Cell,* 74**,** 505-14.

CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science,* 339**,** 819-23.

CONG, L., ZHOU, R., KUO, Y. C., CUNNIFF, M. & ZHANG, F. 2012. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun,* 3**,** 968.

CONNELLY, J., BARKER, J., PRUETT-MILLER, S. & PORTEUS, M. 2010. Gene correction by homologous recombination with zinc finger nucleases in primary cells from a mouse model of a generic recessive genetic disease. *Mol Ther,* 18**,** 1103-10.

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

CROCKER, J. & STERN, D. L. 2013. TALE-mediated modulation of transcriptional enhancers in vivo. *Nat Methods,* 10**,** 762-7.

DAVIS, D. & STOKOE, D. 2010. Zinc finger nucleases as tools to understand and treat human diseases. *BMC Med,* 8**,** 42.

DAVIS, G. D. & PRUETT-MILLER, S. M. 2010. ZFN Donor Design -- Codon and Single Base Genome Editing Using Zinc Finger Nucleases. *Biowire -- A Look into the Future of Genome Editing.* Fall 2010 ed. Sigma Life Science: Sigma Aldrich.

DAWSON, J. F. 2007. Electronic Publishing as a Course Context for a Capstone Project on Protein Design. *Journal of Electronic Publishing,* 10.

DELTCHEVA, E., CHYLINSKI, K., SHARMA, C. M., GONZALES, K., CHAO, Y., PIRZADA, Z. A., ECKERT, M. R., VOGEL, J. & CHARPENTIER, E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature,* 471**,** 602-7.

DESJARLAIS J R & BERG, J. M. 1993. Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins. *Proc Natl Acad Sci U S A.,* 15**,** 2256–2260.

DICKSON, J., GOWHER, H., STROGANTSEV, R., GASZNER, M., HAIR, A., FELSENFELD, G. & WEST, A. 2010. VEZF1 elements mediate protection from DNA methylation. *PLoS Genet,* 6**,** e1000804.

DOYLE, E. L., BOOHER, N. J., STANDAGE, D. S., VOYTAS, D. F., BRENDEL, V. P., VANDYK, J. K. & BOGDANOVE, A. J. 2012. TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res,* 40**,** W117-22.

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

DREIER, B., BEERLI, R. R., SEGAL, D. J., FLIPPIN, J. D. & BARBAS, C. F., 3RD 2001. Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem,* 276**,** 29466-78.

DREIER, B., FULLER, R. P., SEGAL, D. J., LUND, C. V., BLANCAFORT, P., HUBER, A., KOKSCH, B. & BARBAS, C. F., 3RD 2005. Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem,* 280**,** 35588-97.

DREIER, B., SEGAL, D. J. & BARBAS, C. F., 3RD 2000. Insights into the molecular recognition of the 5'-GNN-3' family of DNA sequences by zinc finger domains. *J Mol Biol,* 303**,** 489-502.

DURAI, S., MANI, M., KANDAVELOU, K., WU, J., PORTEUS, M. & CHANDRASEGARAN, S. 2005. Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Res,* 33**,** 5978-90.

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

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

EMERSON, R. O. & THOMAS, J. H. 2009. Adaptive Evolution in Zinc Finger Transcription Factors. *PLoS Genet.,* 5**,** 1-12.

FAIRALL, L., SCHWABE, J. W., CHAPMAN, L., FINCH, J. T. & RHODES, D. 1993. The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature,* 366**,** 483-7.

FU, Y., FODEN, J. A., KHAYTER, C., MAEDER, M. L., REYON, D., JOUNG, J. K. & SANDER, J. D. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol,* 31**,** 822-6.

FU, Y., SANDER, J. D., REYON, D., CASCIO, V. M. & JOUNG, J. K. 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol,* 32**,** 279-84.

GAJ, T., GERSBACH, C. A. & BARBAS, C. F., 3RD 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol,* 31**,** 397-405.

GAO, X., TSANG, J. C., GABA, F., WU, D., LU, L. & LIU, P. 2014. Comparison of TALE designer transcription factors and the CRISPR/dCas9 in regulation of gene expression by targeting enhancers. *Nucleic Acids Res,* 42**,** e155.

GARG, A., LOHMUELLER, J. J., SILVER, P. A. & ARMEL, T. Z. 2012. Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Res,* 40**,** 7584-95.

GASIUNAS, G., BARRANGOU, R., HORVATH, P. & SIKSNYS, V. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A,* 109**,** E2579-86.

GASIUNAS, G. & SIKSNYS, V. 2013. RNA-dependent DNA endonuclease Cas9 of the CRISPR system: Holy Grail of genome editing? *Trends Microbiol,* 21**,** 562-7.

GEISSLER, R., SCHOLZE, H., HAHN, S., STREUBEL, J., BONAS, U., BEHRENS, S. E. & BOCH, J. 2011. Transcriptional activators of human genes with programmable DNA-specificity. *PLoS One,* 6**,** e19509.

GERALD, D., ADINI, I., SHECHTER, S., PERRUZZI, C., VARNAU, J., HOPKINS, B., KAZEROUNIAN, S., KURSCHAT, P., BLACHON, S., KHEDKAR, S., BAGCHI, M., SHERRIS, D., PRENDERGAST, G. C., KLAGSBRUN, M., STUHLMANN, H., RIGBY, A. C., NAGY, J. A. & BENJAMIN, L. E. 2013. RhoB controls coordination of adult angiogenesis and lymphangiogenesis following injury by regulating VEZF1-mediated transcription. *Nat Commun,* 4**,** 2824.

GILBERT, L. A., LARSON, M. H., MORSUT, L., LIU, Z., BRAR, G. A., TORRES, S. E., STERN-GINOSSAR, N., BRANDMAN, O., WHITEHEAD, E. H., DOUDNA, J. A., LIM, W. A., WEISSMAN, J. S. & QI, L. S. 2013. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell,* 154**,** 442-51.

GOWHER, H., STUHLMANN, H. & FELSENFELD, G. 2008. Vezf1 regulates genomic DNA methylation through its effects on expression of DNA methyltransferase Dnmt3b. *Genes Dev,* 22**,** 2075-84.

GRONER, A. C., MEYLAN, S., CIUFFI, A., ZANGGER, N., AMBROSINI, G., DENERVAUD, N., BUCHER, P. & TRONO, D. 2010. KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. *PLoS Genet,* 6**,** e1000869.

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

HALL, D. B. & STRUHL, K. 2002. The VP16 activation domain interacts with multiple transcriptional components as determined by protein-protein cross-linking in vivo. *J Biol Chem,* 277**,** 46043-50.

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

HARDY, S., LEGAGNEUX, V., AUDIC, Y. & PAILLARD, L. 2010. Reverse genetics in eukaryotes. *Biol Cell,* 102**,** 561 - 580.

HARRISON, M. M., JENKINS, B. V., O'CONNOR-GILES, K. M. & WILDONGER, J. 2014. A CRISPR view of development. *Genes Dev,* 28**,** 1859-72.

HAUSCHILD-QUINTERN, J., PETERSEN, B., COST, G. J. & NIEMANN, H. 2013. Gene knockout and knockin by zinc-finger nucleases: current status and perspectives. *Cell Mol Life Sci,* 70**,** 2969-83.

HIRAI, H., TANI, T. & KIKYO, N. 2010. Structure and functions of powerful transactivators: VP16, MyoD and FoxA. *Int J Dev Biol,* 54**,** 1589-96.

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

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

HSU, P. D., LANDER, E. S. & ZHANG, F. 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell,* 157**,** 1262-78.

HSU, P. D., SCOTT, D. A., WEINSTEIN, J. A., RAN, F. A., KONERMANN, S., AGARWALA, V., LI, Y., FINE, E. J., WU, X., SHALEM, O., CRADICK, T. J., MARRAFFINI, L. A., BAO, G. & ZHANG, F. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol,* 31**,** 827-32.

HUANG, P., XIAO, A., ZHOU, M., ZHU, Z., LIN, S. & ZHANG, B. 2011. Heritable gene targeting in zebrafish using customized TALENs. *Nat Biotechnol.* United States.

ISALAN, M. 2012. Zinc-finger nucleases: how to play two good hands. *Nat Methods,* 9**,** 32-4.

ISALAN, M., CHOO, Y. & KLUG, A. 1997. Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proc Natl Acad Sci U S A,* 94**,** 5617-21.

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

IVANOV, A. V., PENG, H., YURCHENKO, V., YAP, K. L., NEGOREV, D. G., SCHULTZ, D. C., PSULKOWSKI, E., FREDERICKS, W. J., WHITE, D. E., MAUL, G. G., SADOFSKY, M. J., ZHOU, M. M. & RAUSCHER, F. J., 3RD 2007. PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing. *Mol Cell,* 28**,** 823-37.

JACOBS, G. & MICHAELS, G. 1990. Zinc finger gene database. *New Biol,* 2**,** 583.

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

JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science,* 337**,** 816-21.

JOUNG, J. K. & SANDER, J. D. 2013. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol,* 14**,** 49-55.

KADAMB, R., MITTAL, S., BANSAL, N., BATRA, H. & SALUJA, D. 2013. Sin3: insight into its transcription regulatory functions. *Eur J Cell Biol,* 92**,** 237-46.

KAY, S., BOCH, J. & BONAS, U. 2005. Characterization of AvrBs3-like effectors from a Brassicaceae pathogen reveals virulence and avirulence activities and a protein with a novel repeat architecture. *Mol Plant Microbe Interact,* 18**,** 838-48.

KAY, S., HAHN, S., MAROIS, E., HAUSE, G. & BONAS, U. 2007. A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science,* 318**,** 648-51.

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

KIM, J. H., LEE, S. R., LI, L. H., PARK, H. J., PARK, J. H., LEE, K. Y., KIM, M. K., SHIN, B. A. & CHOI, S. Y. 2011. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One,* 6**,** e18556.

KIM, J. S., LEE, H. J. & CARROLL, D. 2010. Genome editing with modularly assembled zinc-finger nucleases. *Nat Methods.* United States.

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

KIM, Y. G. & CHANDRASEGARAN, S. 1994. Chimeric restriction endonuclease. *Proc Natl Acad Sci U S A,* 91**,** 883-7.

KOBAYASHI, N., BOYER, T. G. & BERK, A. J. 1995. A class of activation domains interacts directly with TFIIA and stimulates TFIIA-TFIID-promoter complex assembly. *Mol Cell Biol,* 15**,** 6465-73.

KOYANO-NAKAGAWA, N., NISHIDA, J., BALDWIN, D., ARAI, K. & YOKOTA, T. 1994. Molecular cloning of a novel human cDNA encoding a zinc finger protein that binds to the interleukin-3 promoter. *Mol Cell Biol,* 14**,** 5099-107.

KUHNERT, F., CAMPAGNOLO, L., XIONG, J. W., LEMONS, D., FITCH, M. J., ZOU, Z., KIOSSES, W. B., GARDNER, H. & STUHLMANN, H. 2005. Dosage-dependent requirement for mouse Vezf1 in vascular system development. *Dev Biol,* 283**,** 140-56.

KULINSKI, J., BESACK, D., OLEYKOWSKI, C. A., GODWIN, A. K. & YEUNG, A. T. 2000. CEL I enzymatic mutation detection assay. *Biotechniques,* 29**,** 44-6, 48.

KWON, D. N., LEE, K., KANG, M. J., CHOI, Y. J., PARK, C., WHYTE, J. J., BROWN, A. N., KIM, J. H., SAMUEL, M., MAO, J., PARK, K. W., MURPHY, C. N. & PRATHER, R. S. 2013. Production of biallelic CMP-Neu5Ac hydroxylase knock-out pigs. *Sci Rep,* 3**,** 1981.

LAMB, B. M., MERCER, A. C. & BARBAS, C. F., 3RD 2013. Directed evolution of the TALE N-terminal domain for recognition of all 5' bases. *Nucleic Acids Res,* 41**,** 9779-85.

LEE, H., KIM, E. & KIM, J. 2010. Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res,* 20**,** 81-9.

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

LEI, Y., GUO, X., LIU, Y., CAO, Y., DENG, Y., CHEN, X., CHENG, C. H., DAWID, I. B., CHEN, Y. & ZHAO, H. 2012. Efficient targeted gene disruption in Xenopus embryos using engineered transcription activator-like effector nucleases (TALENs). *Proc Natl Acad Sci U S A,* 109**,** 17484-9.

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

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

LI, T., HUANG, S., JIANG, W. Z., WRIGHT, D., SPALDING, M. H., WEEKS, D. P. & YANG, B. 2011b. TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res,* 39**,** 359-72.

LI, T., HUANG, S., ZHAO, X., WRIGHT, D. A., CARPENTER, S., SPALDING, M. H., WEEKS, D. P. & YANG, B. 2011c. Modularly assembled designer TAL effector nucleases for

targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res,* 39**,** 6315-25.

LIEBER, M. R. 2010. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem,* 79**,** 181-211.

LIU, J., LI, C., YU, Z., HUANG, P., WU, H., WEI, C., ZHU, N., SHEN, Y., CHEN, Y., ZHANG, B., DENG, W. M. & JIAO, R. 2012. Efficient and specific modifications of the Drosophila genome by means of an easy TALEN strategy. *J Genet Genomics,* 39**,** 209-15.

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

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

LOW, C. M. 2013. *Genomic interactions of the transcription factor VEZF1.* PhD, University of Glasgow.

LOZZIO, C. B. & LOZZIO, B. B. 1975. Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. *Blood,* 45**,** 321-34.

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

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

MAEDER, M. L., LINDER, S. J., CASCIO, V. M., FU, Y., HO, Q. H. & JOUNG, J. K. 2013a. CRISPR RNA-guided activation of endogenous human genes. *Nat Methods,* 10**,** 977-9.

MAEDER, M. L., LINDER, S. J., REYON, D., ANGSTMAN, J. F., FU, Y., SANDER, J. D. & JOUNG, J. K. 2013b. Robust, synergistic regulation of human gene expression using TALE activators. *Nat Methods,* 10**,** 243-5.

MAHFOUZ, M. M., LI, L., PIATEK, M., FANG, X., MANSOUR, H., BANGARUSAMY, D. K. & ZHU, J. K. 2012. Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein. *Plant Mol Biol,* 78**,** 311-21.

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

MALI, P., YANG, L., ESVELT, K. M., AACH, J., GUELL, M., DICARLO, J. E., NORVILLE, J. E. & CHURCH, G. M. 2013. RNA-guided human genome engineering via Cas9. *Science,* 339**,** 823-6.

MARGOLIN, J. F., FRIEDMAN, J. R., MEYER, W. K., VISSING, H., THIESEN, H. J. & RAUSCHER, F. J., 3RD 1994. Kruppel-associated boxes are potent transcriptional repression domains. *Proc Natl Acad Sci U S A,* 91**,** 4509-13.

MILLER, J., MCLACHLAN A D & KLUG, A. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO,* 4**,** 1609 - 1614.

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

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

MIYASHITA, H., KANEMURA, M., YAMAZAKI, T., ABE, M. & SATO, Y. 2004. Vascular endothelial zinc finger 1 is involved in the regulation of angiogenesis: possible contribution of stathmin/OP18 as a downstream target gene. *Arterioscler Thromb Vasc Biol,* 24**,** 878-84.

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

MORBITZER, R., ROMER, P., BOCH, J. & LAHAYE, T. 2010. Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proc Natl Acad Sci U S A,* 107**,** 21617-22.

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

MUSSOLINO, C., MORBITZER, R., LUTGE, F., DANNEMANN, N., LAHAYE, T. & CATHOMEN, T. 2011. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res,* 39**,** 9283-93.

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

PAPWORTH, M., KOLASINSKA, P. & MINCZUK, M. 2006. Designer zinc-finger proteins and their applications. *Gene,* 366**,** 27-38.

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

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

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

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

PEREZ-PINERA, P. 2013. Synergistic and tunable human gene activation by combinations of synthetic transcription factors. 10**,** 239-42.

PEREZ-PINERA, P., KOCAK, D. D., VOCKLEY, C. M., ADLER, A. F., KABADI, A. M., POLSTEIN, L. R., THAKORE, P. I., GLASS, K. A., OUSTEROUT, D. G., LEONG, K. W., GUILAK, F., CRAWFORD, G. E., REDDY, T. E. & GERSBACH, C. A. 2013. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods,* 10**,** 973-6.

PEREZ-PINERA, P., OUSTEROUT, D. G., BROWN, M. T. & GERSBACH, C. A. 2012. Gene targeting to the ROSA26 locus directed by engineered zinc finger nucleases. *Nucleic Acids Res,* 40**,** 3741-52.

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

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

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

REYON, D., KHAYTER, C., REGAN, M. R., JOUNG, J. K. & SANDER, J. D. 2012a. Engineering designer transcription activator-like effector nucleases (TALENs) by REAL or REAL-Fast assembly. *Curr Protoc Mol Biol,* Chapter 12**,** Unit 12.15.

REYON, D., TSAI, S. Q., KHAYTER, C., FODEN, J. A., SANDER, J. D. & JOUNG, J. K. 2012b. FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol,* 30**,** 460-5.

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

SANDER, J. D., CADE, L., KHAYTER, C., REYON, D., PETERSON, R. T., JOUNG, J. K. & YEH, J. R. 2011a. Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat Biotechnol.* United States.

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

SANDER, J. D. & JOUNG, J. K. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol,* 32**,** 347-55.

SANGAMO BIOSCIENCES, I. 2013. Sangamo Biosciences Presents Clinical Data From HIV Study Demonstrating Sustained Control of Viremia. *Reduction of Viral Load at or Below Limit of Detection Ongoing at 14 Weeks*.

SANJANA, N. E., CONG, L., ZHOU, Y., CUNNIFF, M. M., FENG, G. & ZHANG, F. 2012. A transcription activator-like effector toolbox for genome engineering. *Nat Protoc,* 7**,** 171-92.

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

SARGENT, R. G., BRENNEMAN, M. A. & WILSON, J. H. 1997. Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination. *Mol Cell Biol,* 17**,** 267-77.

SCHORNACK, S., MEYER, A., ROMER, P., JORDAN, T. & LAHAYE, T. 2006. Gene-for-gene-mediated recognition of nuclear-targeted AvrBs3-like bacterial effector proteins. *J Plant Physiol,* 163**,** 256-72.

SCHORNACK, S., MINSAVAGE, G. V., STALL, R. E., JONES, J. B. & LAHAYE, T. 2008. Characterization of AvrHah1, a novel AvrBs3-like effector from Xanthomonas gardneri with virulence and avirulence activity. *New Phytol,* 179**,** 546-56.

SEBASTIANO, V., MAEDER, M. L., ANGSTMAN, J. F., HADDAD, B., KHAYTER, C., YEO, D. T., GOODWIN, M. J., HAWKINS, J. S., RAMIREZ, C. L., BATISTA, L. F., ARTANDI, S. E.,

WERNIG, M. & JOUNG, J. K. 2011. In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. *Stem Cells,* 29**,** 1717-26.

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

SHIMIZU, Y., BHAKTA, M. S. & SEGAL, D. J. 2009. Restricted spacer tolerance of a zinc finger nuclease with a six amino acid linker. *Bioorg Med Chem Lett,* 19**,** 3970-2.

SHIMIZU, Y., SOLLU, C., MECKLER, J. F., ADRIAENSSENS, A., ZYKOVICH, A., CATHOMEN, T. & SEGAL, D. J. 2011. Adding fingers to an engineered zinc finger nuclease can reduce activity. *Biochemistry,* 50**,** 5033-41.

SMITH, J., BIBIKOVA, M., WHITBY, F. G., REDDY, A. R., CHANDRASEGARAN, S. & CARROLL, D. 2000. Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res,* 28**,** 3361-9.

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

STREUBEL, J., BLUCHER, C., LANDGRAF, A. & BOCH, J. 2012. TAL effector RVD specificities and efficiencies. *Nat Biotechnol.* United States.

STROGANTSEV, R. S. 2009. *Mapping and characterisation of genomic binding sites of the chromatin barrier protein VEZF1.* Thesis (Ph D), University of Glasgow.

SUN, N., LIANG, J., ABIL, Z. & ZHAO, H. 2012. Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. *Mol Biosyst,* 8**,** 1255-63.

SUN, N. & ZHAO, H. 2013. Transcription activator-like effector nucleases (TALENs): a highly efficient and versatile tool for genome editing. *Biotechnol Bioeng,* 110**,** 1811-21.

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

SZUREK, B., ROSSIER, O., HAUSE, G. & BONAS, U. 2002. Type III-dependent translocation of the Xanthomonas AvrBs3 protein into the plant cell. *Mol Microbiol,* 46**,** 13-23.

TESSON, L., USAL, C., MENORET, S., LEUNG, E., NILES, B. J., REMY, S., SANTIAGO, Y., VINCENT, A. I., MENG, X., ZHANG, L., GREGORY, P. D., ANEGON, I. & COST, G. J. 2011. Knockout rats generated by embryo microinjection of TALENs. *Nat Biotechnol.* United States.

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

TSUJI, S., FUTAKI, S. & IMANISHI, M. 2013. Creating a TALE protein with unbiased 5'-T binding. *Biochem Biophys Res Commun,* 441**,** 262-5.

TSUJI, T. & NIIDA, Y. 2008. Development of a simple and highly sensitive mutation screening system by enzyme mismatch cleavage with optimized conditions for standard laboratories. *Electrophoresis,* 29**,** 1473-83.

TUMBAR, T., SUDLOW, G. & BELMONT, A. S. 1999. Large-Scale Chromatin Unfolding and Remodeling Induced by VP16 Acidic Activation Domain. *J Cell Biol,* 145**,** 1341-54.

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

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

UTLEY, R. T., IKEDA, K., GRANT, P. A., COTE, J., STEGER, D. J., EBERHARTER, A., JOHN, S. & WORKMAN, J. L. 1998. Transcriptional activators direct histone acetyltransferase complexes to nucleosomes. *Nature,* 394**,** 498-502.

VAN DEN ACKERVEKEN, G., MAROIS, E. & BONAS, U. 1996. Recognition of the bacterial avirulence protein AvrBs3 occurs inside the host plant cell. *Cell,* 87**,** 1307-16.

VAN OEVELEN, C., WANG, J., ASP, P., YAN, Q., KAELIN, W. G., JR., KLUGER, Y. & DYNLACHT, B. D. 2008. A role for mammalian Sin3 in permanent gene silencing. *Mol Cell,* 32**,** 359-70.

VAN RENSBURG, R., BEYER, I., YAO, X. Y., WANG, H., DENISENKO, O., LI, Z. Y., RUSSELL, D. W., MILLER, D. G., GREGORY, P., HOLMES, M., BOMSZTYK, K. & LIEBER, A. 2013. Chromatin structure of two genomic sites for targeted transgene integration in induced pluripotent stem cells and hematopoietic stem cells. *Gene Ther,* 20**,** 201-14.

VOIT, R. A., HENDEL, A., PRUETT-MILLER, S. M. & PORTEUS, M. H. 2014. Nuclease-mediated gene editing by homologous recombination of the human globin locus. *Nucleic Acids Res,* 42**,** 1365-78.

WERNER, J. & GOSSEN, M. 2014. Modes of TAL effector-mediated repression. *Nucleic Acids Res,* 42**,** 13061-73.

WILSON, K. A., MCEWEN, A. E., PRUETT-MILLER, S. M., ZHANG, J., KILDEBECK, E. J. & PORTEUS, M. H. 2013. Expanding the Repertoire of Target Sites for Zinc Finger Nuclease-mediated Genome Modification. *Mol Ther Nucleic Acids,* 2**,** e88.

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

WOLFE, S. A., NEKLUDOVA, L. & PABO, C. O. 2000. DNA recognition by Cys2His2 zinc finger proteins. *Annu Rev Biophys Biomol Struct,* 29**,** 183-212.

WRIGHT, D. A., THIBODEAU-BEGANNY, S., SANDER, J. D., WINFREY, R. J., HIRSH, A. S., EICHTINGER, M., FU, F., PORTEUS, M. H., DOBBS, D., VOYTAS, D. F. & JOUNG, J. K. 2006. Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nature Protocols,* 1**,** 1637 - 1651.

XIAO, H., PEARSON, A., COULOMBE, B., TRUANT, R., ZHANG, S., REGIER, J. L., TRIEZENBERG, S. J., REINBERG, D., FLORES, O., INGLES, C. J. & ET AL. 1994. Binding of basal transcription factor TFIIH to the acidic activation domains of VP16 and p53. *Mol Cell Biol,* 14**,** 7013-24.

XIONG, J. W., LEAHY, A., LEE, H. H. & STUHLMANN, H. 1999. Vezf1: A Zn finger transcription factor restricted to endothelial cells and their precursors. *Dev Biol,* 206**,** 123-41.

YANG, L., MEI, Q., ZIELINSKA-KWIATKOWSKA, A., MATSUI, Y., BLACKBURN, M. L., BENEDETTI, D., KRUMM, A. A., TABORSKY, G. J., JR. & CHANSKY, H. A. 2003. An ERG (ets-related gene)-associated histone methyltransferase interacts with histone deacetylases 1/2 and transcription co-repressors mSin3A/B. *Biochem J,* 369**,** 651-7.

YUSUFZAI, T. M. & FELSENFELD, G. 2004. The 5'-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. *Proc Natl Acad Sci U S A,* 101**,** 8620-4.

ZHANG, F., CONG, L., LODATO, S., KOSURI, S., CHURCH, G. M. & ARLOTTA, P. 2011. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol,* 29**,** 149-53.

ZHANG, Z., XIANG, D., HERIYANTO, F., GAO, Y., QIAN, Z. & WU, W. S. 2013. Dissecting the Roles of miR-302/367 Cluster in Cellular Reprogramming Using TALE-based Repressor and TALEN. *Stem Cell Reports.*

ZHU, W., YANG, B., CHITTOOR, J. M., JOHNSON, L. B. & WHITE, F. F. 1998. AvrXa10 contains an acidic transcriptional activation domain in the functionally conserved C terminus. *Mol Plant Microbe Interact,* 11**,** 824-32.

ZOU, J., MALI, P., HUANG, X., DOWEY, S. N. & CHENG, L. 2011. Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease. *Blood,* 118**,** 4599-608.

ZOU, Z., OCAYA, P. A., SUN, H., KUHNERT, F. & STUHLMANN, H. 2010. Targeted Vezf1-null mutation impairs vascular structure formation during embryonic stem cell differentiation. *Arterioscler Thromb Vasc Biol,* 30**,** 1378-88.