Designer Nuclease-Assisted Targeting to Engineer Mammalian Genomes

DISSERTATION

zur Erlangung des akademischen Grades

Doctor rerum naturalium
(Dr. rer. nat.)

vorgelegt

der Fakultät Mathematik und Naturwissenschaften
der Technischen Universität Dresden

von

Sarah Tsurkan

geboren am 05. Juli 1985 in Deleware

Gutachter: Prof. Dr. A. Francis Stewart
Prof. Dr. Yixin Zhang

Eingereicht am (16.11.2017)

Die Dissertation wurde in der Zeit von Januar 2010 bis November 2017
am Biotechnologischen Zentrum (BIOTEC) der Technischen
Universität Dresden angefertigt.

Abstract

Designer nucleases have greatly simplified small genome modifications in many genomes. They can precisely target a specific DNA sequence within a genome and make a double stranded break (DSB). DNA repair mechanisms of the DSB lead to gene mutations or gene modification by homologous directed repair (HDR) if a repair template is exogenously supplied. Thus, small, site directed mutations are easily and quickly achieved. However, strategies that utilize designer nucleases for more complex tasks are emerging and require optimization.

To optimize CRISPR/Cas9 assisted targeting, an HPRT rescue assay was utilized to measure the relationship between targeting frequency and homology arm length in targeting constructs in mouse embryonic stem cells. The results show that different gene engineering exercises had different homology requirements. Targeted correction of a 4 bp mutation and insertion of 672 bp DNA improved steeply with homology arms up to 2 kb and 4 kb total homology, respectively, and had further increases with even longer homology arms. Deletion of a 1.5 kb stretch of DNA required homology arms of up 10 kb and supercoiled, circular DNA repair templates consistently had better targeting efficiencies than linearized templates.

The repair efficiency of oligonucleotides (ODN) was determined to be inefficient, unless high concentrations of ODNs were supplied. ODNs were only able to repair a small 4 bp mutation, correction of a larger 15 bp mutation and deletion of a 1.5 kb stretch of DNA were not possible. HDR was found to be more efficient than non-homologous end joining mediated ligation of DNA, and co-cutting of the genomic target and the donor was more efficient than insertion of a pre-linearized stretch of DNA. These results led to the conclusion that circular HDR vectors with long homology arms provide the greatest chance of achieving a complex gene engineering exercise.

The optimized conditions found in this work were then utilized to complete a complex targeting exercise: humanization of the mouse *Scn10a* gene with a hybrid mouse/human BAC targeting construct. These results show how designer nuclease-assisted targeting, along with optimized construct design, permits execution of previously impossible gene engineering exercises in mammalian genomes.

Table of Contents

Abstract iii
Table of Contents v
List of Figures ix
List of Abbreviations x

1. INTRODUCTION: 1

1.1 Gene targeting and designer nucleases 1

1.2 Scn10a 2

1.3 Humanization of the mouse genome 4

1.4 Designer nucleases 4
1.4.1 Zinc finger nucleases 5
1.4.2 TALs 6
1.4.2.1 TALENs 7
1.4.3 CRISPR/Cas9 8
1.4.4 Designer nuclease specificity and off target effects 9

1.5 DSB repair pathways 11

1.6 Recombineering 13

1.7 Gene targeting 15

1.8 Embryonic stem cells early mouse development 16

2 AIM 18

3 MATERIALS AND METHODS 19

3.1 Instruments 19

3.2 Consumables 19

3.3 Chemicals/ Reagents 19

3.4 Cell Culture Chemicals/Reagents 21

3.5 Kits and Laders 21

3.6 Enzymes 22

3.7 Buffers: 22

3.8 Synthetic Oligonucleotides 23

3.9 BACs 24

3.10 Plasmids 24

3.11 gRNA Sequences 5'-3' PAM 26

3.12 TALEN sequences: 26

3.13 ES Cell Culture 27
3.13.1 Cell Lines: 27
3.13.2 MEF Cultivation and Inactivation: 27
3.13.3 mESC Culture 27
3.13.4 mESC Passaging 28

3.14 Long Term mESC Storage 28

3.15 Cell culture reagents and media 28

3.16 mESC Transfection: 29
3.16.1 ES cell electroporation: 29
3.16.2 ES cell nucleofection: 29
3.16.3 ES cell lipofection: 30

3.17 T7E1 30

3.18 *In vitro* transcribed RNA 30

3.19 Quantitative PCR (qPCR) 30

3.20 X-Gal staining: 31

3.21 Quantitative β-Gal assay: 32

3.22 Genomic DNA extraction: 32

3.23 HPRT assay: 32

3.24 Southern assay: 33

3.25 Large scale bacterial DNA preparation: 35

3.26 Small scale bacterial DNA preparation: 36

3.27 Restriction endonuclease digest: 37

3.28 Bacterial electroporation: 37

3.29 Ethanol DNA precipitation: 37

3.30 PCR purification: 38

3.31 Gel extraction: 38

3.32 Gel electrophoresis: 38

3.33 TALEN design and assembly: 38

3.34 Plasmid Construction: 39
3.34.1 gRNA design: 39

3.34.2 gRNA expression vector construction: 40
3.34.3 TALEN expression vectors: 40
3.34.4 CMV-Rep plasmid assembly: 41
3.34.5 Subcloning *Hprt* repair vectors: 42
3.34.6 Insertion of a PGK-neo-cassette into *Hprt* exon 6: 42
3.34.7 Subcloning of Δ672 fragment: 43
3.34.8 Engineered Mutations in Targeting Vector Homology Arms: 44

3.35 Design of the *SCN10A* humanization construct 44

4 RESULTS 46

4.1 Humanization of the mouse *Scn10a* gene 46
4.1.1 Strategy to increase targeting efficiency at the Scn10a locus with TALENs 46
4.1.2 Evaluation two different TALEN pairs to test *in vivo* cleavage efficiency 47
4.1.3 Modification of the TALEN expression vector to increase *in vivo* cleavage in mESC 48
4.1.4 An *in vivo* test for TALEN cleavage 49
4.1.5 Testing for *in vivo* TALEN cleavage 50

4.2 Improving targeting efficiency via targeting construct optimization: 53
4.2.1 Hprt as a selectable locus to test targeting efficiency 53
4.2.2 Effect of homology arm length on targeting efficiency with CRISPR/Cas9 55
4.2.2.1 Introduction 55
4.2.3 Construction of targeting vectors with different homology arms 56
4.2.4 Targeting efficiency increases with total homology arm length 57
4.2.5 Use of ribonucleoprotein to induce DSB 61

4.3 ODN repair of DSB 63
4.3.1 Introduction 63
4.3.2 Targeting efficiency with ODN 64
4.3.3 Effect of asymmetrial oligos on repair efficiency 65

4.4 Effect of homology arm length on CRISPR/Cas9 mediated deletion of large stretches of DNA 66
4.4.1 Insertion of a cassette via CRISPR/Cas9 mediated HDR 66
4.4.2 Phenotypic confirmation of *Hprt* KO 67
4.4.3 Effect of gRNA position on deletion of a PGK neo cassette 68
4.4.4 Effect of homology arm length on deletion of a PGK-neo cassette 70

4.5 CRIPSR/Cas9 permits NHEJ mediated deletions and insertions 71
4.5.1 Introduction 71
4.5.2 Precise deletion of 672 bp from the Hprt gene 73
4.5.3 Design of a KiBL vector to test precise NHEJ mediated DNA insertion 73
4.5.4 NHEJ mediated ligation precisely inserts DNA 74
4.5.5 Repair of Δ672 with HDR vector templates 75

4.6 Effect of Engineered Mutations in Homology Arms on Targeting efficiency with CRISPR/Cas9 mediated HDR 76
4.6.1 Introduction 76
4.6.2 Differences between isogenic R1 DNA and RP23 BAC library at *Hprt* exon 6 77
4.6.3 Effect of engineered mutations in homology arms on targeting efficiency 77

Summary 79

4.7 Humanization of the mouse Scn10a gene using CRISPR/Cas9 80
4.7.1 Targeting strategy for the *Scn10a* gene 80
4.7.2 Loss of allele screen for humanization of the mouse genome 80
4.7.3 Generation of humanized *SCN10A* mES cells 81

5 DISCUSSION 84

5.1 Two TALENs exhibited a low cleavage efficiency 84

5.2 Optimization of construct design for designer nuclease-assisted targeting 84
5.2.1 HPRT is a selectable assay for construct design with CRISPR/Cas9 85
5.2.2 Longer HA increase targeting efficiency 85
5.2.3 Minimal homology arm length 87
5.2.4 HDR with circular vectors have higher efficiency than linear vectors 88
5.2.5 ODN repair of DSB requires large amounts of DNA 88
5.2.6 RNP has advantages over expression vectors 90
5.2.7 gRNAs have different efficiencies 91
5.2.8 Blunt end ligation is inefficient compared to HDR 91
5.2.9 Long stretches of isogenic DNA no longer necessary 92
5.2.10 Summary of optimization work 92

5.3 Optimized targeting constructs make complex targeting exercises possible 93

6 CONCLUSIONS AND OUTLOOK 94

7 REFERENCES 95

8 ACKNOWLEDGEMENTS 107
Erklärung entsprechend §5.5 der Promotionsordnung 108

List of Figures

Figure 1 Na$_v$1.8 structure 3

Figure 2 Designer Nucleases 5

Figure 3 DSB Repair Pathways 11

Figure 6 Design of *SCN10A* Targeting Construct 33

Figure 5 TALEN expression vectors 41

Figure 4 Quantitative *Hprt* rescue assay 45

Figure 7 TALEN Strategy and Validation with T7E1 assay.......................... 47

Figure 8 pCMV-Rep Vector Design 49

Figure 9 Transfection of HEK293 cells with pCMV-Rep plasmids
and different TALEN expression vectors 51

Figure 10 Design of a quantitative experiment utilizing *Hprt* 53

Figure 11 Genotypic and Phenotypic verification of *Hprt*$^-$ Cell lines 54

Figure 12 Subcloning Repair Vectors..57

Figure 13 Effect of homology arm length on targeting efficiency 58

Figure 14 ODN Repair. 65

Figure 15 Insertion of a PGK-neo cassette into the *Hprt* locus 67

Figure 16 Phenotypic confirmation of HPRT-PGK-neo clones 68

Figure 17 Deletion of a 1.5 kb stretch of DNA 69

Figure 18 Design of an experiment to test ligation efficiency 74

Figure 19 Ligation and insertion efficiency 75

Figure 20 Engineered mutations in a homology arm..................................78

Figure 21 Isogenicity does not impact targeting
efficiency when using CRISPR/Cas9 79

Figure 22 LOA Assay for Humanized *SCN10A* colonies 82

List of Abbreviations

2i	two inhibitors
6-TG	6-thioguanine
aa	amino acid
alt-NHEJ	alternative NHEJ
BAC	bacterial artificial chromosome
BSD	blasticidin
Cas9	CRISPR associated protein 9
C-NHEJ	classical NHEJ
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DNA	deoxyribonucleic acid
Dre	D6 site specific DNA recombinase
DSB	double strand break
dsDNA	double stranded DNA
E.coli	*Escherichia coli*
ES	embryonic stem
ESC	embryonic stem cell
Flp	flippase
FRT	flippase recognition target
gRNA	guide RNA
HA	homology arms
HAT	hypoxanthine, aminopterin and thymidine
HATI	histone acetyltrasferase 1
HDR	homology directed repair
HPRT	hypoxanthine-guanine phosphoribosyltransferase
HR	homologous recombination
HYG	Hygromycin B
ICM	inner cell mass
indel	insertion or deletion
KiBL	knock-in blunt ligation
KO	knockout
LacZ	structural gene for β-galactosidase
LB	lysogeny broth
LIF	leukocyte inhibiting factor
LOA	loss of allele
loxP	locus of X-over P1
MEF	mouse embryonic fibroblasts
MRRF	mitochondrial ribosome recycling factor
neo	neomycin
NHEJ	non-homologous end joining

NLS	nuclear localization signal
Nxt2	nuclear transport factor 2 like export factor 2
OD	opposite direction (vector)
ODN	oligonucleotide
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
PGK	phosphoglycerate kinase promoter
qPCR	quantative PCR
RE	restriction endonuclease
Red	lambda red system
RNA	ribonucleic acid
RNP	ribonucleoprotein
Rox	region of crossover
RPP	reduced proliferation phenotype
RT	room temperature
RVD	repeat variable diresidue
Scn10a	Sodium Voltage-Gated Channel Alpha Subunit 10
SD	same direction (vector)
SDSA	synthesis-dependent strand annealing
SNP	single nucleotide polymorphism
SSA	single strand annealing
ssDNA	single stranded DNA
ssODN	single stranded oligonucleotide
ssOR	single stranded oligo repair
T7E1	T7 Endonuclease I
TAL	transcription activator like effector
TALEN	transcription activator-like effector nucleases
TC	targeting construct
TTX	tetrodotxin
ZF	zinc finger
ZFN	zinc finger nuclease

1. Introduction:

1.1 Gene targeting and designer nucleases

In 2007, the Nobel Prize in Medicine was awarded to Mario Capecci, Martin Evans and Oliver Smithies for their work in gene-specific modification in embryonic stem cells that led to knock out mice. Mouse embryonic stem cells (mESC) can be engineered via homologous recombination with engineered exogenous deoxyribonucleic acid (DNA). Targeted mESC are then used to make transgenic mouse models. Mice are frequently used in experiments as they are the laboratory animal most closely related to humans for which gene targeting can be easily applied. As of 2016, 16,883 of the ~20,000 transcribed genes of the mouse genome have been knocked out. This enormous feat was accomplished by a consortium of scientists across the globe(1) and gives an invaluable resource for scientists interested in gene function studies. While extensive progress has been made using gene targeting strategies, this technology is still limited by the low frequency of homologous recombination, and long, laborious selection and screening strategies.(2)

So called “designer nucleases” are a valuable new tool in the toolbox for genome engineering. They induce a strand break (DSB) in the target locus and can either knock out genes through non homologous end joining (NHEJ), or increase gene targeting efficiency when a targeting construct is simultaneously provided through activation of the homologous repair machinery (HDR).(3,4) Over the last decade, the field of genetic engineering has seen technology advance from zinc finger nucleases, to transcription activator-like effector nucleases (TALENs), and finally Clustered Regularly Interspaced Short Palindromic Repeats/ CRISPR associated protein 9 (CRISPR/Cas9) and Cfp1. However, the use of nuclease-assisted targeting to achieve complex precision genome engineering still remains challenging with many unresolved issues related to experimental design.

Nuclease assisted targeting clearly has the potential to expand the area of genome engineering to encompass ambitious exercises. To exploit this potential for challenging genome engineering, such as the establishment of alleles for conditional mutagenesis, simultaneous biallelic targeting or large regional replacements, targeting constructs must be optimized. Several works have revealed that DSBs induced by nucleases can be repaired by constructs with very short homology arms (HA) (5,6) or oligonucleotides (ODNs),(7,8) however, they do not look at optimal homology arm length for the different kinds of genome engineering exercises (large insertions, large deletions, repair of small lesions, etc) nor examine how the efficiency of targeting constructs compare to ODN repair or repair by blunt end ligation. Increasing targeting efficiency not only reduces workload and costs, it also opens up avenues for clinical applications and medical research.

1.2 Scn10a

Understanding pain disorders aids in disease diagnosis and helps patient treatment by revealing new drug targets. The human gene sodium voltage-gated channel alpha subunit 10 (*SCN10A)* encodes for the tetrodotoxin (TTX) resistant sodium channel $Na_v1.8$ and is expressed primarily in small and medium diameter nocioceptive sensory neurons in the dorsal root ganglion.(9) *SCN10A* is involved in sensing pain from heat, cold, mechanical stimuli, neuropathic pain (chronic pain cause by nerve damage) and inflammation. It is a highly sought after target for pain treatment.

Mutations within *SCN10A* can lead to several disease states, with a range of disease severity depending on the mutation. Variants of *SCN10A* have been indicated in Atrial Fibrillation,(10) (11) Long QT syndrome,(12) Brugada Syndrome,(13) and in inherited neuropathies.(14) Studies of *SCN10A* are important to understand the molecular mechanism of pain, and can help identify new targets for analgesic drug developments.(15,16)

Commonly use anesthetics often block several Na_v subtypes and can cause unwanted side effects. Selectively blocking $Na_v1.8$, as it is almost exclusively expressed in nocioceptive neurons, would avoid those adverse effects while maintaining the desired anesthetic effects. $Na_v1.8$ is also a possible target for heart arrhythmia therapies.(17) Six binding locations for natural toxins and synthetic derivatives have been identified, however, there are limitations to using animal models for human diseases. Studies of various disease states that are similar in mouse and human have served to highlight the differences between the species,(18) and illustrate that while mouse may be a model, it is not a perfect match for disease states. Therefore, in order to better measure drug effects on Na$_V$1.8, a model that uses the human *SCN10A* gene would be beneficial.

The exon-intron structures of *SCN10A* are well conserved among human, rat and mouse.(19) The human *SCN10A* gene is 97 kb, has 27 coding exons and 1956 amino acids. The mouse *Scn10a* gene is 85 kb and has 27 coding exons and 1958 amino acids. All Na$_V$s have four homologous domains (D1-4) that consist of 6 transmembrane α-helical segments (S1-6) that are connected by intracellular loops (figure 1A). Both the N and C terminals are intracellular. S4 segments have positive amino acid residues that are predicted to be the voltage sensor.(20) Upon depolarization, the intracellular S4 segments move across the membrane to open the channel.(21) Elucidation of the crystal structures of Na$_V$s (22) (figure 1B) has led to the identification of the α-helical binding sites of several natural and synthetic toxins that modulate sodium channels.(16)

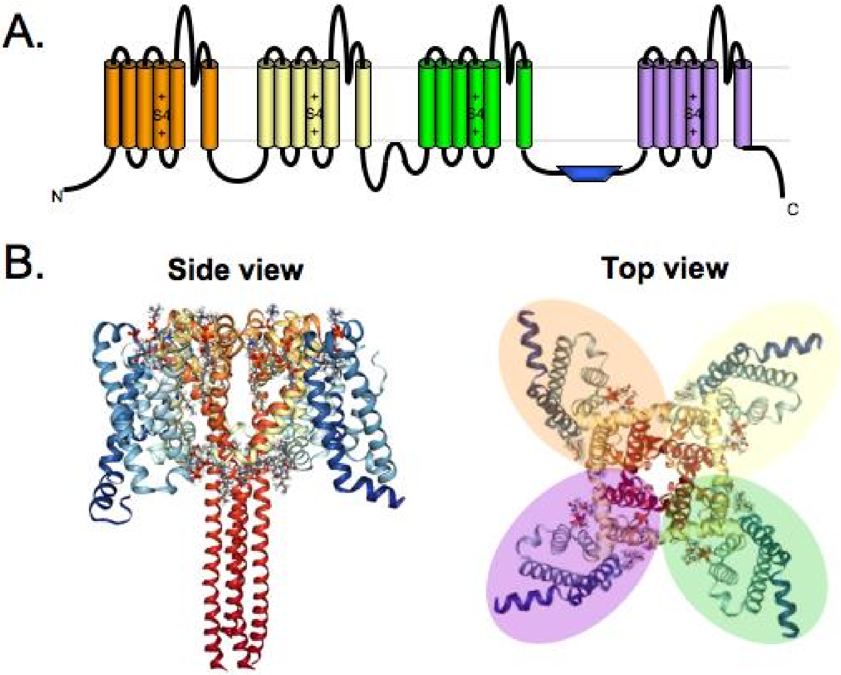


Figure 1 Na$_V$1.8 structure Na$_V$1.8 has four domains consisting of 6 transmembrane α-helical segments connected by intracellular loops. Upon depolarization, the intracellular S4 segments move across the membrane to open the channel. Na$_V$1.8 conducts Na+ ions through the pore and across the membrane during depolarization. (A) Depiction of the transmembrane segments of Na$_V$1.8 (B) Side and top views of the Na$_V$1.8 pore forming complex.(22)

1.3 Humanization of the mouse genome

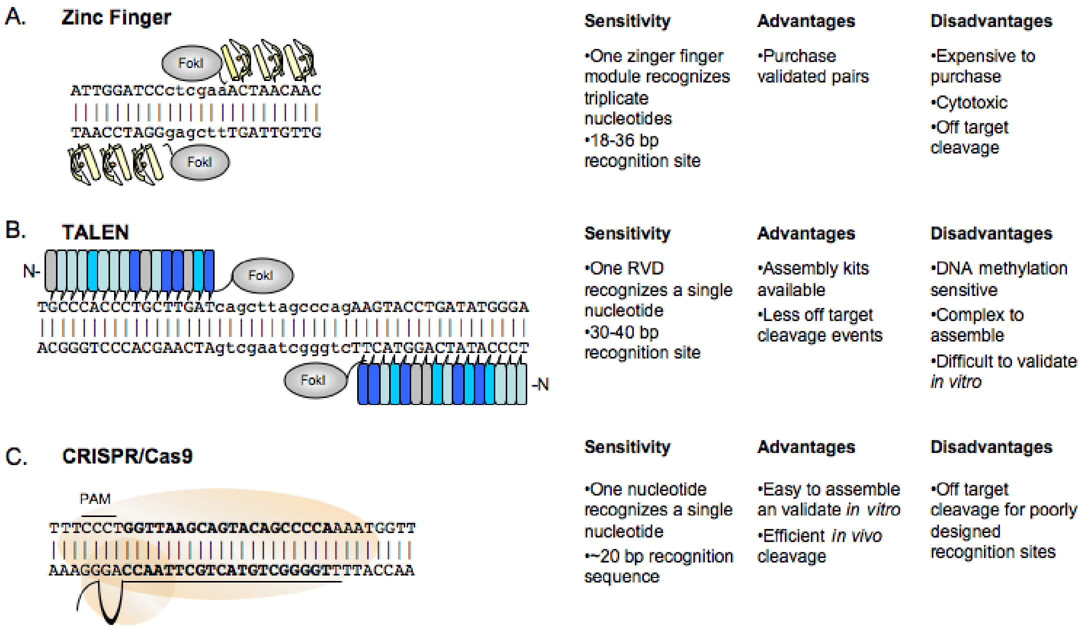
Inserting a human gene into the mouse genome, or “humanization,” is a useful technique in genome engineering for drug development and treatment. Randomly integrated small transgenes may function inefficiently due to missing *cis* acting elements(23) or be affected by the chromosomal landscape into which they are integrated.(24,25) Though large transgenes, such as BACs, can overcome some of the limitations of small transgenes, their large size makes it difficult to properly analyze copy number and to check for rearrangements.(26)

Identifying and replacing the mouse ortholog in the mouse genome avoids some of the problems that transgenes face such as position effect or copy number. However, there are differences in how orthologous mouse and human genes are regulated,(27,28) and experimental design should take that into consideration. Additionally, the large size of some genes, such as *SCN10A*, which is 85 kb in mouse and 96 kb in human, still present problems with screening for rearrangements. Methods that insert large segments of DNA often use multiple targeting steps, which is undesirable as it can decrease germline transmission.(29-31)

The use of designer nucleases to insert larger pieces of DNA presents a solution to the problem of targeting with large constructs. While targeting the mouse Mll gene with a 64 kb targeting construct with selectable markers separated by 43 kb of DNA was achievable with low efficiency (6 %),(32) nucleases have been used to homozygously replace 2.7 kb of the THY1 gene at up to 11.6% of the targeted population or heterozygously at up to 14% of the population without selection.(33) The same report found similar results when replacing 9.8 kb of the BSG gene. Baker et al.(34) used CRISPR/Cas9 to humanize the Kmt2d gene, which involved a 42 kb regional exchange and homology arms as short as 4 and 7 kb, but increased targeting efficiency when using a construct with 4 and 50 kb homology arms.

1.4 Designer nucleases

Historically, targeted gene disruption, insertion, or replacement was achieved via homologous recombination. Homologous recombination, however, has a very low efficiency that limits its utility, which meant that complex designs were difficult to complete. The discovery that DSBs increase HDR or causes knock outs by NHEJ increased the desire for efficient, specific, easy to use targeted nucleases. Designer nucleases reduce the positional effects that sometimes confound HDR, and increases HDR by several orders of magnitude. Over the past three decades, technology has been rapidly progressing in the field of designer nucleases from Zinc Finger Nucleases (ZFN), to TALENs, and finally CRISPR/Cas9.


A.
Zinc Finger
FokI
ATTGGATCCctcgaaACTAACAAC
TAACCTAGGgagctTGATTGTTG
FokI
B.
TALEN
N-
FokI
TGCCCACCCTGCTTGATcagcttagcccagAAGTACCTGATATGGGA
ACGGGTCCCACGAACTAGtcgaatcgggtcTTCATGGACTATACCCT
FokI
-N
C.
CRISPR/Cas9
PAM
TTTCCCTGGTTAAGCAGTACAGCCCCAAAATGGTT
AAAGGGACCAATTCGTCATGTCGGGGTTTTACCAA

Sensitivity	Advantages	Disadvantages
•One zinger finger module recognizes triplicate nucleotides •18-36 bp recognition site	•Purchase validated pairs	•Expensive to purchase •Cytotoxic •Off target cleavage

Sensitivity	Advantages	Disadvantages
•One RVD recognizes a single nucleotide •30-40 bp recognition site	•Assembly kits available •Less off target cleavage events	•DNA methylation sensitive •Complex to assemble •Difficult to validate *in vitro*

Sensitivity	Advantages	Disadvantages
•One nucleotide recognizes a single nucleotide •~20 bp recognition sequence	•Easy to assemble an validate *in vitro* •Efficient *in vivo* cleavage	•Off target cleavage for poorly designed recognition sites

Figure 2 Designer Nucleases (A) Zinc finger nucleases consist of a pair 3-6 zinc finger (ZF) modules fused to a FokI nuclease. (B) TALENs consist of an N-terminal region containing the type II translocation signal, 15-20 repeat variable diresidue (RVDs) conferring DNA specificity, a nuclear localization signal (NLS) and a transcriptional activation domain fused to FokI. TALENs require a 5' 'T' and must dimerize to make a DSB (C) The gRNA-Cas9 complex binds to a target sequence next to a 3 nt 'NGG" protospacer adjacent motif (PAM) where it generates a DSB 3bp upstream of the PAM.

1.4.1 Zinc finger nucleases

Zinc finger nucleases (ZFN) were the first gene editing strategy to use custom DNA endonucleases.(35) Zinc fingers are DNA binding transcription factors that are frequently found in eukaryotes. One zinc finger consists of about 30 amino acids in a ββα configuration.(36) The side chains of conserved Cys and His residues complex with a zinc ion.

Amino acids on the surface of the α-helix recognize 3-4 bp of DNA. Zinc fingers that identify all 64 of the possible nucleotide triplets have been developed, making a modular system that is useful for custom design. Pre-selected modules can be tandemly linked together to target a specific DNA sequence. Non-natural arrays that contain more than three zinc finger domains allowed synthetic ZFN construction with higher specificity, which was promoted by the discovery of a highly conserved linker sequence.(37) This permits construction of zinc fingers that recognize 9-18 bp of DNA, which can confer very high specificity.(38)

When a FokI cleavage domain is fused to the zinc finger DNA recognition module, a site directed nuclease is formed (figure 2A). FokI is a type IIS nuclease that makes cuts 9 and 13 nt downstream from the recognition site after dimerization.(39) FokI is only active upon dimerization, meaning two nuclease domains must be expressed to cause the DSB. The ZFN DNA target recognition site is two zinc finger binding sites (left and right) separated by 5-7 bp spacer that is cleaved by the FokI cleavage domain.

One major hurdle to ZFN technology is the difficulty of producing ZFNs in the lab. The intellectual property of ZFN technology is owned by Sangamo Biosciences, who has arrangement to share reagents and methods with Sigma-Aldrich, who can then construct and validate a desired ZFN pair. This comes at a cost of several thousand euros. However, a ZFN can be designed for almost any stretch of DNA. Though the specificity of ZFNs is quite high, the efficiency remains relatively low, perhaps due to zinc finger module interactions and/or chromosomal landscape.

1.4.2 TALs

Transcription activator-like (TAL) effectors were first identified in the phytopathogenic bacteria *Xanthomonas*. The effector proteins are translocated to the cytoplasm of an infected plant cell via type III secretion system where they then modulate cell function by mimicking transcription factors.(40)

TAL effectors consist of a N-terminal region which contains the type III translocation signal, a central domain of 34 amino acid (aa) tandem repeats, a nuclear localization signal, and a transcriptional activation domain(41) (figure 2B). The 34 aa repeats can be classified by their 12th and 13th amino acid residues, with each RVD recognizing a single nucleotide in the target DNA sequence.(42,43) Four repeats- HD, NG, NI, and

NN account for 75% of the total repeat usage, but over 20 repeats have been identified.(43) In naturally occurring TALs, the nucleotide 'T' is generally conserved at the 5' end of the recognition sequence (position 0) and the last repeat only contains 20aa and is considered a half repeat.(42,43) The protein sequence preceding the TAL RVD's is somewhat similar to the RVD sequence, and forms a degenerate loop similar to that found in the RVDs that may loosely interact with the 'T.'(40,44) However, work was done to evolve the N-terminal domain to allow for recognition of all the bases.(45) RVDs can range in number from 1.5 to 28.5(46) with the majority of naturally occurring TALs having between 12 and 27.(41) However Boch et al.(42) found that a minimum of 6.5 RVDs were required to induce gene expression, and RVDs with 10.5 or more repeats lead to stronger gene activation.

Analysis of the crystal structure of TAL effectors indicates that each TAL repeat comprises of a left-handed, two helix bundle: a short α-helix formed from positions 3-11 and a second, bent α-helix formed by positions 15-33.(44,47) The two helices are separated by a short RVD containing loop, which positions the RVDs close to the sense strand in the DNA major groove.(44,47) It appears that position 12 does not directly recognize DNA, but rather stabilizes the local conformation of the loop and aids position 13 to specifically recognize a DNA base.(47) The TAL repeat forms a right-handed, super-helical structure that tracks along the major groove of the DNA duplex, with the inner ridge of the super-helix having a positive electrostatic potential that allows interaction with the negatively charged phosphate of the DNA backbone.(44,47)

1.4.2.1 TALENs

TALs can be paired with a nuclease to cleave double stranded DNA at a specific site- thus introducing TALENs to the toolbox of genetic engineering. Following the design of zinc finger nucleases, a TALEN pair is designed to recognize a sequence that is separated by a spacer. The sequence specific recognition of TALs direct a FokI monomer to a sequence where the FokI dimerizes, cleaves and causes a double stranded break (DSB) within the spacer.(44)

Several TALEN assembly kits are provided at a nominal charge by depositories, which mean that TALENs can be designed and assembled in about a week. These kits come with more than 70 plasmid samples that need to be organized and stored. Assembly is a multi-step process that is relatively laborious and technically difficult. Once assembled, due to their repetitive nature, RVD order can only be checked by restriction digest. TALEN pairs are difficult to validate *in vitro*, but can be expressed and checked *in vivo*, often with a T7E1 or CelI nuclease assay. However, the efficiency of TALENs within the same work, even targeting the same locus varies from low to high, or even showing no activity.(48-51) This may be due to some undetermined property of TALEN design, or to the methylation status of the DNA target.(52,53) The ability to relatively easily design and assemble TALENs in the laboratory and reduced cytotoxicity (section 1.4.4) makes them more attractive that ZFNs, but their sensitivity to DNA methylation makes them difficult to use in cultured cells, which accumulate methylation marks with passaging.

1.4.3 CRISPR/Cas9

Bacteria have evolutionarily developed a defense mechanism against phage infection and plasmid transfer known as clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR associated protein 9 (Cas9).(54) This natural defense system has been adapted and repurposed as a RNA-guided DNA targeting genetic engineering tool.(55) (56) The *Streptocococcus pyogenes* (*S. pyogenes*) Cas9 nuclease recognizes a 20 bp DNA sequence via a single guide RNA (gRNA) (figure 2C). This results in nucleotide-nucleotide recognition, compared to the module-nucleotide recognition of TALENs and ZFNs and may give CRISPR/Cas9 a higher DNA recognition specificity.

A short, conserved 3 bp motif (NGG for *S. Pyogenes*) must be located in close proximity to the 20 bp targeted recognition site. This is known as the protospacer adjacent motif, or PAM. When the Cas9-gRNA complex binds to its target DNA sequence next to a PAM, it generates a blunt ended DSB 3 bp upstream of the PAM.(57) (55)

S. Pyogenes Cas9 is a 1,368 aa DNA endonuclease. The HNH-like nuclease domain cleaves the DNA strand complementary to the target. The RuvC-like nuclease domain cleaves the non-target strand.(58) Mutating H840A of the HNH or D10A of the RuvC domain generates a nickase, rather than a DSB.(55) Using paired nickases doubles the recognition site, thus decreasing off target effects (section 1.4.4), and makes a DSB by making off set cuts.

CRISPR/Cas9 is easily assembled in the lab, either as an expression plasmid (requires simple cloning) or produced *in vitro* and combined with Cas9 protein to form an RNP. *In vitro* production of the gRNA utilizes commercially available kits, and eliminates cloning, which makes it even more attractive. gRNA's are easily designed and verified *in vitro* via Cas9 cleavage assays, and *in vivo* with T7E1 or CelI cleavage assays. Though poorly designed gRNAs can lead to off target effects (section 1.4.4), well designed gRNAs have few off target effects, and though different gRNA's have different cleavage efficiencies,(8,33,59) most gRNAs exhibit measurable cleavage.

1.4.4 Designer nuclease specificity and off target effects

While designer nucleases efficiently cleave target DNA, they can also make unwanted cuts at genomic sites that have a high sequence similarity to the target location. In the extreme, these off target effects can cause cell toxicity, as uncontrolled DSBs can cause chromosomal rearrangements and nonsense mutations. Less drastic off target effects can cause SNPs or unintended integration of foreign DNA into the genome. Each of the three designer nuclease classes have reported off target cleavages, and have reports to reduce these effects (reviewed in(60)).

As FokI nucleases must dimerize to cleave DNA(61), the larger DNA recognition site of the pair contributes to high specificity. ZFN pairs usually recognize 18-36 bp target, and a TALEN pair recognizes a 30-40 bp target, which surpasses the complexity of the genome, meaning that well designed sites will have a single recognition site genome wide. *In silico* identification of unique target sites, however, does not translate to perfect target recognition *in vivo*. CRISPR/Cas9, in contrast, functions as a monomer, which may contribute to non-specific cleavage, especially in poorly designed recognition sequences. Conversely, however, RVD and ZF degeneracy may contribute to alternative nucleotide recognition while the nucleotide-nucleotide recognition of CRISPR/Cas9 may be less tolerant of mismatches.

ZFNs prefer guanine rich targets, which reduces the number of sequences they can target. Commercially produced ZFNs can be screened for cleavage efficiency, but come at a higher cost. ZFNs constructed from publicly available ZF resources are often cytotoxic, which may be a result of their off target effects.(62)

TALENS are generally not cytotoxic, but can still produce off target mutations.(63) There are several web-based resources that screen the target genome for potential of target locations (for common genomes such as mouse, rat or human). Kim et al. has shown that off target effects of TALENs can be prevented as long as the target sequence is at least 7 nucleotides different from any other genomic site.(64)

Cas9 requires a PAM, which for *S. pyogenes* is 5'-NGG-3', however it can cleave sites with a 5'-NAG-3' or a 5'-NGA-3', though it loses some efficiency.(65) CRISPR/Cas9 can also tolerate some mismatches in the PAM distal sequence, though mismatches in the seed region are not as well tolerated. Finally, CRISPR/Cas9 can cleave off target sites with several additional or missing nucleotides, making thousands of possible off target sites in the genome.(66)

Most publications test only for mutations at predicted off target sites, and Sanger sequencing of DNA from individual clones is the gold standard of confirming the presence or absence of off target effects. However, this method is expensive and time consuming to screening many clones. Interestingly, whole genome/exome sequencing of clonal populations of human cells that were modified with ZFNs,(67) TALENS and CRISPR/Cas9(68) reveal that mutations are almost absent in the entire genome of an individual clone, though off target mutations are detectible in a bulk population of cells. This is because the mutation frequency of on-target sites is orders of magnitude higher than mutations at off-target sites.

Nonetheless, it is important to understand off target effects, and try to prevent them. Indeed, there are several ways to decrease off target effects in designer nucleases. The easiest and most efficient way to reduce off-target effects is a well-designed recognition sequence. There are several free programs that predict off target sites, and utilizing only highly rated recognition sequence reduces the chances of off target sequence recognition. The use of paired nickases in Cas9 (thus requiring two nickases each cleaving a single strand to form a double strand break) decreases the rate of HDR, but decreases off target effects.(57,69) Cas9 ribonucleoprotein complexes rapidly cleaves target sites, and are rapidly degraded by endogenous proteases, thus reducing off target effects, without compromising on target mutation frequencies.(70,71) The final way to reduce Cas9 off target effects is to carefully control the amount of Cas9 and gRNA delivered, as high levels of Cas9 leads to off target effects.(65,72,73)

1.5 DSB repair pathways

Genome integrity is important for cell survival and proliferation. Different types of DNA lesions stall replication fork progression and can lead to replication fork collapse and therefore DSB formation. The cell has four different mechanisms to repair DSBs: classical-NHEJ (C-NHEJ), alternative NHEJ (alt-EJ), single strand annealing (SSA) and homologous recombination (HR) (figure 3). Which pathway the cell chooses depends on several interrelated factors, including cell cycle and end resectioning. End resection is diminished in non-cycling cells, and therefore NHEJ is favored over pathways that require resection.(74)

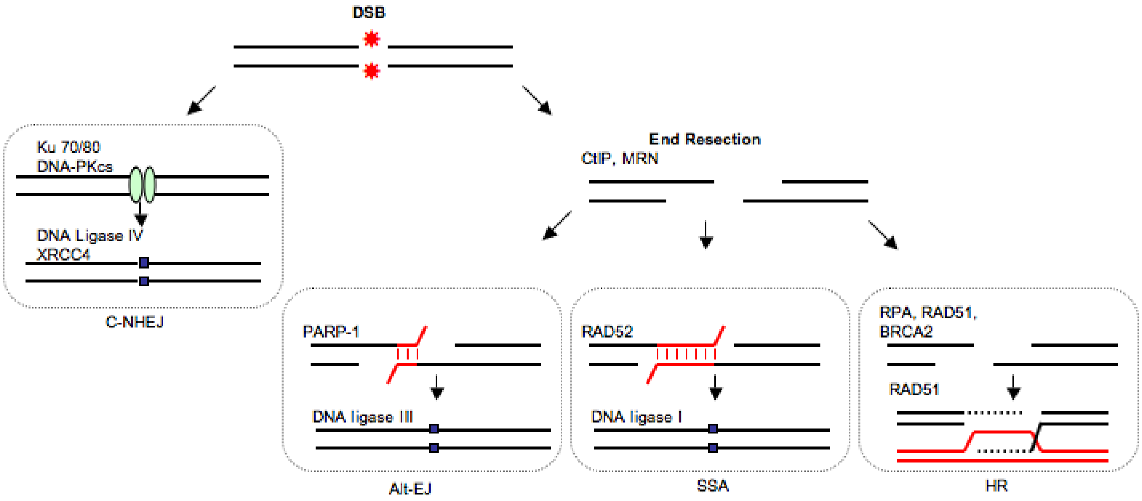


Figure 3 DSB Repair Pathways The four mechanisms to repair DSB are depicted: C-NHEJ, alt-EJ, SSA and HR. C-NHEJ results in quickly repaired DSB via ligation, and typically results in 0-4 nt loss. Alternative-NHEJ (Alt-EJ), SSA, and HR all occur after end resectioning. Alt-NHEJ uses small homologies to repair DSBs, resulting in the loss of up to 20 nt. SSA occurs after extensive end resectioning and is the result of end joining between genomic nucleotide repeats via reannealing of ssDNA covered by RPA and RAD52, resulting in the loss of large stretches of DNA. HR uses a homologous DNA for repair, and requires strand invasion mediated by RAD51 and is typically error free.

In C-NHEJ the ends are not resectioned and the DSB is repaired via blunt end ligation, independent of sequence homology. C-NHEJ has fast kinetics, and protects the genome by suppressing chromosomal translocations.(75) The C-NHEJ pathway utilizes many factors, such as Ku70/80, DNA-PKCs, and DNA ligase IV.

Once C-NHEJ is ruled out by end resectioning, a competition is held between the three remaining pathways. End resectioning generates 3' single-stranded DNA overhangs. The cell decision to resect ends likely directs the choice of pathway and repair outcome.(74) Alt-EJ, SSA, or HR can repair the resected junction. The end resection begins with 'end clipping' by the MRE11 nuclease and CtlP, and relatively few base pairs (ie up to 20 bp in mammalian cells) are removed, thus producing a small overhang that can be repaired by alt-EJ, which uses microhomology pairing for the repair. The evidence of the alternative end joining pathway has been presented in recent years, as cells deficient in C-NHEJ could still repair DSBs by end joining.(76)

'Extensive resection' is the second phase of end resectioning and involves helicases and exonucleases (like BLM, CtlP, EXO1 and WRN) producing long stretches of ssDNA. Once the long stretches of ssDNA are produced, the cell is committed to HR or SSA.(74,77) RAD51 and RPA compete to bind to the ssDNA to prevent secondary structure formation, DNA end degradation, and prevent spontaneous annealing between microhomologies(78) (thus regulating alt-EJ).(79)

HR uses a sister chromatid or homologous DNA for repair, and requires strand invasion mediated by RAD51, and is typically error free. Once the ends are resected, a D-loop is formed after homology search and strand invasion. In the following steps, RAD51 is removed from dsDNA, strand extension, junction resolution and chromatin reassembly occurs.(80) This is a complex process that is mediated by several checkpoints and many different factors. This naturally occurring cellular process is exploited during gene targeting. DNA can be inserted, deleted, or replaced when homologous stretches of DNA surrounding the change are supplied.

SSA does not require a donor sequence and does not require strand invasion, and occurs independently from RAD51. Rather, SSA repairs via resected ends that anneal to exposed complementary sequences. SSA consists of end joining between genomic nucleotide repeats via reannealing of ssDNA covered by RPA and RAD52. This ultimately results in a loss of DNA, as one copy of the repeat and the sequence between repeats are deleted in the repair product.(81)

The cell cycle regulates the competition between C-NHEJ and end resectioning dependent pathways. C-NHEJ occurs mostly in G0/G1 and G2, but can occur throughout the cell cycle.(82,83) In G1, BRCA1 recruitment is inhibited and DNA end resection is blocked, thus promoting C-NHEJ. However, in S/G2, where DNA replication is highest, cyclin dependent kinases (such as ATM) phosphorylate substrates that favor the three resection depend pathways: HR, alt-EJ, or SSA.(83)

Each of the four types of repair have typical outcomes for DSB repair. This is especially important to recognize when using designer nucleases, as the DSBs they cause can lead to different outcomes. For instance, the FokI nuclease (used with ZFNs and TALENs) leave a 4 nt 5’ overhang. This means that end resectioning must occur, and HR, or one of the error prone end resection pathways must repair the lesion. Cas9, however, makes a straight DSB that can be repaired through blunt end ligation (C-NHEJ) as it competes with the end resectioning pathways. Typically, C-NHEJ is either perfectly repaired through blunt end ligation, or results in small 1-4 nt deletions. HR results in either accurate repair, or loss of heterozygosity. SSA results in large deletions, and alt-EJ produces chromosomal rearrangements, and small insertions or deletions.(80)

Synchronizing cell culture before introducing a designer nuclease can help direct repair to a certain pathway, as different pathways are chosen at different parts of the cell cycle.(83) Therefore, if one desires HDR, introduction of a repair template and nuclease and synchronization of the cell cycle to S/G2 will increase the chance of HR. Alternatively, if a knockout phenotype via small indels is desired, a nuclease that produces overhangs is beneficial. Finally, inhibiting or enhancing modulators of any of the pathways can sway the cell’s decision for or against that pathway.(84)

1.6 Recombineering

Recombineering is a method of *in vivo* DNA engineering that utilizes phage-derived recombination systems in *E. coli* and other bacteria. The first uses of phage derived recombination systems utilized the lambda phage(85) and the RecE/RecT system of the Rac prophage.(86) The term “recombineering” comes from **recombin**ation-mediated genetic engin**eering**.(87,88) Plasmids, BACs and the *E. coli* genome have all been modified using recombineering, without the need for traditional *in vitro* cloning methods such as restriction digestions and ligation reactions that rely on the location and sequence of restriction endonucleases.(89) Recombineering is advantageous in that it overcomes the limitations of cloning, as it is independent of restriction sites. The location of the recombineering reaction is defined by the sequence of user defined homology arms that flank a recombineering cassette. A typical homology arm is 35-50 bp long, as this is enough to identify a unique locus in even the largest mammalian genome. Recombineering is based on constitutive or inducible expression of the *red αβγ* genes of the Red system or the *recET* genes of the RecE/RecT system together with *red γ* in *E. coli*.(90) Expression of the *red γ* protein increases the stability of transformed linear DNA substrates by inhibiting the RecBCD nuclease. This means that the very long homology arms (of about 1000 bp) that was used in initial experiments(85) are unnecessary and that shorter homology arms, which can be introduced during PCR via chemically synthesized olignucleotides that contain both the ~50 bp homology arm and a ~21 bp PCR primer can be easily used for DNA engineering.(86) RecE and Redα are 5′->3′ exonucleases, and RecT and Redβ are DNA annealing proteins. DSB repair is initiated by RecE or Redα. First RecE or Redα degrades the DNA in a 5’-3’ direction, starting from the DSB, thereby creating a 3’ ssDNA overhang. Then, RecT or Redβ binds to the ssDNA forming a recombinogenic proteonucleic filament which is used in recombination, either by single strand annealing or by strand invasion.(91,92)

Plasmid based recombineering expression systems have the advantage of being easily transferred into the desired *E. coli* strain. The recombineering genes are typically expressed in *trans* from a low to medium copy number, temperature-sensitive pSC101 plasmid under the control of the L-arabinose inducible pBAD promoter.(93) This allows for temporal control of gene expression in the cell of choice by a plasmid that can be eliminated from the host by a temperature shift.

Recombineering can be used for many applications, among them: a) utilizing gap repair to subclone sections of DNA from larger DNA molecules (ex BAC),(94,95) b) insertion, deletion or replacement of a selectable fragment,(86,94) c) single-stranded oligonucleotides for site-directed mutagenesis (ssOR) for single point mutations and small insertions and deletions(96-98) d) high throughput pipelines for rapid generation of modified BACs based on sequential recombineering.(99-101)

1.7 Gene targeting

The use of recombineering greatly simplifies the design and assembly of the constructs that are required for gene targeting. There are many possibilities in the design of targeting vectors that can be used for a variety of purposes. For gene targeting without a nuclease, DNA must be linear, with long stretches of isogenic DNA (3-5 kb HA on each side), and require a strategy to detect targeting must (common strategies are Southern assay, long range PCR and loss of allele assays). Using selectable markers (such as drug resistance genes) enriches targeting events, and additional sequences, such as site-specific recombinases can be used to make knock out or conditional alleles. Recombineering easily allows vector assembly by sublcloning from bacterial artificial chromosomes (BAC) libraries, integrating cassettes, and integrating additional DNA sequences such as RE sites or site specific recombinase recognition sequences.

mES cells provide an excellent and versatile model for the study of mammalian biology. mESC can be derived from early mouse embryos,(102,103) can be modified with exogenous DNA(104) and have the ability to reconstitute fertile animals from cultured cells, i.e., to colonize all of the tissues of the embryo, including the germ line. Mouse embryonic stem (ES) cells have a higher frequency of HDR than other cultured cells- most likely due to their rapid growth and resultantly high DNA replication, which means that S phase constitutes a greater proportion of the cell cycle. Though gene targeting is also possible in ES cells from other species such as human and rat;(105,106) the legal restrictions placed on human ESC manipulation, the high efficiency of mouse gene targeting, and the relative low cost of mouse maintenance mean that mouse is one of the most used mammalian models.

There are many different ways to engineer the mouse genome, and transgenic mice have opened the door for major advancements in biomedical research. A common application of gene targeting is the generation of knockout (KO) mice, in which a gene is disrupted resulting in a non-functional protein. If this “loss of function” causes a phenotypic change, this can be studied and the function of this gene and the protein it codes for can be identified. In addition to being used for animal studies, mES are a model for pluripotency and stem cell issues.(107) mESC can also be differentiated in to other cell types(108-111) which allows for the study of lineage commitment and differentiation.(112)

1.8 Embryonic stem cells early mouse development

Mammalian development begins after fertilization, with the single cell zygote. The zygote undergoes a series of cleavage divisions leading to the formation of the morula, which consists of cells called blastomeres. Further divisions and morphogenic changes result in the early blastocyst. The blastocyst consists of the inner cell mass (ICM) and the surrounding trophectoderm cells which forms the extra embryonic tissues. While the zygote and blastomeres are totipotent, the ICM consists of pluripotent cells that will form the embryo proper. In late blastocyst formation, the ICM forms a monolayer of primitive endoderm that encloses the remaining pluripotent epiblast cells and the embryo is ready to implant into the uterus. The next step is gastrulation, in which the epiblast specializes into the three germ layers (endoderm, ectoderm, and mesoderm) and the germ cell lineage.(113) Fetal somatic stem cells, which are multipotent, further differentiate into the cells of each organ as the fetus grows and develops. A small percentage of fetal stem cells remain multipotent after fetal development and form adult stem cell pools, which are required for homeostasis or regeneration. Most adult tissues maintain adult somatic stem cells, which divide asymmetrically giving rise to one stem cell and one committed progenitor. The progenitor can then differentiate into the specialized cells of the tissue.

The ICM of the blastocyst can be extracted and cultured *in vitro*. In order to be classified as ES cells, they must meet several criteria: they must be undifferentiated and able to divide, they must be able to develop into the endoderm, ectoderm and mesoderm, and they must have the ability to self-renew. In culture, ES cells can be maintained in an undifferentiated state when grown on mouse MEFs in medium containing serum.(102,103) ES cells can be weaned from MEF feeders and cultured in culture medium supplemented with LIF.(114) Serum can be replaced in growth medium by bone morphogenetic protein 4 (BMP4). Cells maintained with LIF and serum, however, vacillate between a naïve and “primed” state, and have mosaic transcription factor expression levels.(115) Supplementing culture media with two inhibitors (2i- PD0325901 and CHIR99021) inhibits Mitogen-activated protein kinase (MEK) and Glycogen syntase kinase 3 (GSK3)(116) and returns ES cells to their “ground state” of pluripotency.(117)

Alternatively, adding growth factors to culture medium can promote ES cell differentiation into specific cell types.(107,118) This allows for the study of large, homogenous cell populations at different developmental stages. The molecular study of genes, the proteins they code for, and their interactions with other proteins, DNA, and RNA has been greatly advanced by the development of immunoprecipitation,(119) mass spectrometry(120) and chromatin immunoprecipitation (ChIP/ChIP Seq) and RNA profiling has greatly improved our understanding of what is occurring in cells at a molecular level both in the ES stage and upon differentiation.

2 AIM

Nuclease assisted targeting has drastically changed the outlook of gene therapy over the last decade. Designer nucleases like CRISPR/Cas9 have greatly simplified small genome modifications in many genomes. DNA repair mechanisms of nuclease induced DSB leads to gene mutations or gene replace by homologous directed repair (HDR) if a repair template is exogenously supplied. Thus, small, site directed mutations are easily and quickly achieved. However, strategies that utilize designer nucleases for more complex tasks are emerging and require optimization.

Several works have revealed that DSBs induced by nucleases can be repaired by constructs with very short arms(5,6) or even oligonucleotides (ODNs) (7,8), however, they do not look at optimal homology arm length for the different kinds of kinds of genome engineering exercises (large insertions, large deletions, repair of small lesions, etc) nor examine how the efficiency of targeting constructs with smaller arms compare to ODN repair or repair via NHEJ mediated ligation of microhomologies or blunt ends. Increasing targeting efficiency not only reduces workload and costs, it also opens up avenues for clinical applications and medical research.

Therefore, in this work, I aim to clarify optimal construct design for designer nuclease assisted targeting by evaluating the targeting efficiency of different constructs. I will examine the impact of homology arm length on targeting efficiency in different gene engineering exercises and evaluate the requirements for long stretches of isogenic DNA in the homology arms when utilizing nuclease assisted targeting. I will compare the efficiency of HDR to repair techniques that utilize minimal or no homology such as ODN repair and NHEJ mediated ligation, and utilize this information to form a recommendation for optimal vector design for nuclease assisted targeting in mammalian genomes. I will then use these optimized vector design to attempt an ambitious targeting exercise: humanization of the mouse *Scn10a* gene.

3 Materials and Methods

3.1 Instruments

CASY cell counter (model TTC)	Roche
CK 40	Olympus
CKX 41	Olympus
Electroporator 2510	Eppendorf
G:Box gel documentation system	Syngene
Genepulser Xcell	Biorad
Mx3000P / Mx3005P multiplex PCR instr.	Agilent
NanoDrop ND-1000 spectrophotometer	PeQLab
Nucleofector II device	Lonza
Phosphoimaging plates BAS cassette 2340	Fujifilm
Phosphoimaging scanner FLA 3000	Fujifilm
Stereomicroscope with camera	Nikon SMZ 1500
Thermocycler ep Gradient S	Eppendorf
Thermomixer compact	Eppendorf
Ultrospec 2100 Pro	Amersham Biosciences
UV Stratalinker 2400	Stratagene
Vortexer VortexGenie2	Scientific Industries
Water bath 1003	GFL m.b.H.

3.2 Consumables

10 cm tissue culture dishes	Nunc	168381
6 - and 24 - well plates	Nunc	140675, 142475
96 - well plate (flat bottom)	Greiner Bio-One	4410193
Blotting paper	Bio-Rad laboratories	170-3967
Cryo tube vials	Nunc	366656
Microlance needles (23G)	Beckton Dickinson	300800
Electroporation cuvettes (0.4 cm)	VWR	732-1137
Electroporation cuvettes (0.1 cm)	VWR	732-2267
Nylon Transfer Membrane: Biodyne B 0,45 µm	PALL Life Sciences	60200
Quick Spin G-50	Sigma	11273973001
Maxi-500 Columns	Qiagen	10063

3.3 Chemicals/ Reagents

Acetic Acid	VWR	20104
agarose	Serva	11404

Ampicillin sodium salt	Sigma	A9518
QC Wash Buffer	Qiagen	1015371
Bovine serum albumin (BSA)	Sigma Aldrich	A7906
Chloromphenical	Sigma Aldrich	23275
Coomassie Brilliant Blue G20	Merck Millipore	115.444.025
Dimethylformamide (DMF)	Sigma Aldrich	D4551
di-Sodium hydrogen phosphate dihydrate ($Na_2HPO_4 * 2 H_2O$)	Merck Millipore	106580
Dithiothreitol (DTT)	Sigma Aldrich	D9779
Ethanol	VWR	20821
Ethidium Bromide	Sigma Aldrich	E8751
Ethylenediaminetetraacetic acid (EDTA)	Merck Millipore	108418
Formaldehyde	Merck Millipore	104003
Glutaraldehyde	Sigma Aldrich	G6257
Glycerol	Merck Millipore	104092
Hydrochloric acid	VWR	20252
Hygromycin-B	Roth	1287.2
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Sigma	N-1127
Isopropanol	VWR	20842330
L arabinose	Sigma	A3256
Agar, Bacteriological	Amresco	232-658-1
Magnesium chloride ($MgCl_2$)	Sigma Aldrich	208337
Methanol	VWR	20847
Kanamycin sulfate	Sigma	K1377
N-Lauroyl sarcosine sodium salt	Sigma Aldrich	L9150
ortho-Nitrophenyl-β-galactoside (ONPG)	Roth	CN22.1
P1 Resuspension Buffer	Qiagen	1014854
P2 Lysis Buffer	Qiagen	1014939
P3 Neutralization Buffer	Qiagen	1014955
Phenylmethylsulonyl fluoride(PMSF)	Sigma	P-7626
Potassium chloride (KCl)	Merck Millipore	529552
Potassium dihydrogen phosphate (KH_2PO_4)	Merck Millipore	104873
Potassium-ferricyanide (K-Ferricyanide) Sigma Aldrich	Sigma Aldrich	P8131
Potassium-ferrocyanide (K-Ferrocyanide)	Sigma Aldrich	P9387
QBT Equilibration Buffer	Qiagen	1015291
QF Elution Buffer	Qiagen	1025572
Sodium chloride (NaCl)	VWR	27810
Trisodium Citrate dihydrate	Merck Millipore	1.064.481.000
Sodium dihydrogen phosphate ($NaH_2PO_4 * H_2O$)	Merck Millipore	1.063.461.000

Sodium dodecyl sulfate (SDS) (20% Solution)	Fisher Scientific	BP1311-1
Sodium hydroxide (NaOH)	Merck Millipore	1.064.981.000
TBE buffer (10x)	AppliChem	A0972
Tetracyclin Hydrochloride	Sigma	T-3383
Tris	Carl Roth	AE15
X-galactosidase (X-gal)	Fisher Scientific	10365410

3.4 Cell Culture Chemicals/Reagents

2-Mercaptoethanol (β-Mercaptoethanol)	Sigma Aldrich	M6250
6TG	Sigma	A 4660
Chicken serum	Invitrogen	16110082
CT99021	ABCR	AB 253776
D-glucose	Merck Millipore	108342
Dimethylsulfoxide (DMSO)	Sigma Aldrich	D2650
DMEM + GlutaMAX	Invitrogen	61965-026
Fetal calf serum (FCS)	PAA	A15-101
Gelatin	Sigma Aldrich	G2500
Geneticin (G418 sulfate)	Gibco/Invitrogen	11811-064
HAT	Invitrogen	21060- 017
HEPES	Sigma Aldrich	54457
L-glutamine	Invitrogen	25030-024
LIF	Obtained by cultivation of human embryonic kidney cells (293)	
Lipofectamine Ltx	Lipofectamine LTX and Plus reagent	Invitrogen
Mitomycin C	Sigma Aldrich	M0503
Non-essential amino acids (NEAA)	Invitrogen	11140-035
PD0325901	ABCR	391210-10-9
Penicillin/Streptomycin	Invitrogen	15140-122
Sodium pyruvate	Invitrogen	11360-039
Trypsin 2.5% (10x)	Invitrogen	15090-046

3.5 Kits and Laders

1 kb DNA ladder	NEB	N3232S
100 bp DNA ladder	NEB	N3231
Amaxa Mouse ES Cell Nucleofector Kit	Lonza	VPH 1001
Deoxynucleotide mix (dNTPs)	NEB	N0447S
High Prime Kit	Roche	11 585 592 001
Invisorb Spin DNA Extraction Kit	Stratec	1020110300

Mega Shortscript T7	Ambion	AM1354
MSB Spin PCRapace	Stratec	1020220400
NEB Buffers	NEB	Provided with Enzyme
Nucleobond Bac100	Macherey-nagel	740579
Plasmid Safe DNAse	Epicentre	E3110K
Voytas TALEN	Gift from Voytas Lab	

3.6 Enzymes

BsaBI	NEB	R0537S
BsaI	NEB	R3535S
BsmI	NEB	R0134S
BssSI	NEB	R0680S
Bstz171	NEB	R3594S
Cas9-NLS Wt protein	P.E.P MPI-CBG	
EcoNI	NEB	R0521S
EcoRI	NEB	R3101S
Esp3.I	Thermo Scientific	ER0452
GoTaq qPCR Master Mix	Promega	A6001
NdeI	NEB	R0111S
Proteinase K	Fisher Scientific	BP1700-100
SexAI	NEB	R0605S
T4 Ligase	NEB	M0202S
Taq DNA polymerase and buffer	5 prime	2200010

3.7 Buffers:

Phosphate Buffered Saline (1x PBS, pH 7.4)

The media supply center of the Bioinnovation Center, TU-Dresden provided PBS without magnesium and calcium.

	Final Concentration
NaCl	137.0 mM
KCl	2.7 mM
Na_2HPO_4	10.0 mM
KH_2PO_4	

Lysis Buffer for ES cells

	Final Concentration
1M Tris, PH 7.5	5mM
0.5M EDTA	5mM
1M NaCl	5mM
N-Lauroylsarcosine	17mM
Proteinase K (added immediately prior to use)	1 mg/ml

Reagents were dissolved in sterile H_2O with a final volume of 0.5L and autoclaved. Proteinase K was added directly before use.

LB medium

The media supply center of the Bioinnovation Center, TU-Dresden provided PBS without magnesium and calcium.

	Final Concentration
Tryptone	10 g/L
Yeast Extract	5 g/L
NaCl	5 g/L

3.8 Synthetic Oligonucleotides

qPCR Primers

Scn10a Mouse End F	GAACCAATAGCAACCACCCT
Scn10a Mouse End R	GCGTTGCAACATGTAGTTCC
Scn10a Human Start F	ACCAAGAAGAGAAGCCTCGG
Scn10a Human Start R	ACTAGTAGCTCTTACCCGGTG
Scn10a Human End F	AGCCAACTTTGTCACGTCTC
Scn10a Human End R	GGATGACTCTGAAGAGCGTT
MRRF F	ATCGACAGAGGGCTTGCATC
MRRF R	GCAGCTGTACACTGAAAACCA
HATI F	CCCATCATGCATGTAAACGGT
HATI R	GAGTTTGCCGTGAGTTCAGC

HPRT Repair ODN

	GTGGTGGTTTTATTTACCATTAAATGTCTCTTTTC TTTTTAAAAGGATATAATTGACACTGGTAAAACA ATGCAAACTTTGCTTTCCCTGGTTAAGCAGTACA
Asymmetric Oligo Sense	GCCCCAAAATGGTTAAGGTTGCAA

Asymmetric Oligo Antisense	TTGCAACCTTAACCATTTTGGGGCTGTACTGCTTAACCAGGGAAAGCAAAGTTTGCATTGTTTTACCAGTGTCAATTATATCCTTTTAAAAAGAAAAGAGACATTTAATGGTAAATAAAACCACCAC
HPRT-ss50bp-Sense	AATGCAAACTTTGCTTTCCCTGGTTAAGCAGTACAGCCCCAAAATGGTTA
HPRT-ss50bp-Antisense	TAACCATTTTGGGGCTGTACTGCTTAACCAGGGAAAGCAAAGTTTGCATT
HPRT-ss100bp-Sense	AGGATATAATTGACACTGGTAAAACAATGCAAACTTTGCTTTCCCTGGTTAAGCAGTACAGCCCCAAAATGGTTAAGGTTGCAAGGTATGTATGCCACTT
HPRT-ss100bp-Antisense	AAGTGGCATACATACCTTGCAACCTTAACCATTTTGGGGCTGTACTGCTTAACCAGGGAAAGCAAAGTTTGCATTGTTTTACCAGTGTCAATTATATCCT

3.9 BACs

mScn10a BAC RP24-303H7

hSCN10A BAC RP11-1400P20

Hprt BAC RP23-13N1

P1-amp mou-hum scn10a (Shenbiao Hu)

3.10 Plasmids

HPRT repair vectors

pBR322-amp-HPRT-200bp	This Work
pBR322-amp-HPRT-500bp	This Work
pBR322-amp-HPRT-1kb	This Work
pBR322-amp-HPRT-2kb	This Work
pBR322-amp-HPRT-5kb	This Work
pBR322-amp-HPRT-10kb	This Work
pBR322-amp-HPRT-1kb-200bp	This Work
pBR322-amp-HPRT-1kb-500bp	This Work
pBR322-amp-HPRT-1kb-2kb	This Work
pBR322-amp-HPRT-1kb-5kb	This Work
pBR322-amp-HPRT-1kb-10kb	This Work
pBR322-amp-Hprt-PGK-neo-5kb	This Work
pBR322-amp-HPRT-1kb-isogenic	This Work

HA With Engineered Mutations

pBR322-amp-Hprt-(-5bp)	This Work
pBR322-amp-Hprt-1kb-BssSI	This Work

pBR322-amp-Hprt-1kb-BssSI-ndeI-(-883)	This Work
pBR322-amp-Hprt-1kb-BssSI-ndeI-(-701)	This Work
pBR322-amp-Hprt-1kb-BssSI-ndeI-(-464)	This Work
pBR322-amp-Hprt-1kb-BssSI-ndeI-(-181)	This Work
pBR322-amp-Hprt-1kb-BssSI-(-5bp)-(-883)	This Work
pBR322-amp-Hprt-1kb-BssSI-(-5bp)-(-701)	This Work
pBR322-amp-Hprt-1kb-BssSI-(-5bp)-(-464)	This Work
pBR322-amp-Hprt-1kb-BssSI-(-5bp)-(-181)	This Work

TALEN expression vectors

pCAG-T7-TALE-FokI-ELD destination	Addgene 40131
pCAG-T7-TALE-FokI-KKR destination	Addgene 40132
pCAG-T7-TALE-FokI-ELD Start 1	This Work
pCAG-T7-TALE-FokI-KKR Start 2	This Work
pCAG-T7-TALE-FokI-ELD Middle 1	This Work
pCAG-T7-TALE-FokI-KKR Middle 2	This Work
pCDNA3.1(-)-pTAL3	This Work
pCDNA3.1(-)-pTAL3 neo del ccdb-hyg-17aa	This Work

Rep Vectors

pCMV-RabCht-Rep	Kuhn Lab
pCMV-ccdb-hyg-Rep	This Work
pCMV-Scn10a Start-Rep	This Work
pCMV-Scn10a Middle-Rep	This Work

gRNA expression plasmids

pBR322-U6-cm-ccdb-tracrRNAnew-amp	Stewart Lab
pBR322-U6-Hprt 6-1-tracrRNAnew-amp	This Work
pBR322-U6-Hprt 67-tracrRNAnew-amp	This Work
pBR322-U6-Hprt 76-tracrRNAnew-amp	This Work
pBR322-U6-Hprt 67/76-4-tracrRNAnew-amp	This Work
pBR322-U6-neo5-tracrRNAnew-amp	This Work
pBR322-U6-PGK1-tracrRNAnew-amp	This Work
pBR322-U6-PGK8-tracrRNAnew-amp	This Work
pBR322-U6-Hprt C15-tracrRNAnew-amp	This Work
pBR322-U6-Hprt C18-tracrRNAnew-amp	This Work
pBR322-U6-Scn10a 3-tracrRNAnew-amp	This Work
pBR322-U6-Scn10a 7-tracrRNAnew-amp	This Work

Cas9 Expression Vector

pBR322-CAGGS-NLS-FLAG-linker-hCas9-IRES-puro	Stewart Lab

Ligation vectors

pBR322-amp-ligation vector SD	This Work
pBR322-amp-ligation vector OD	This Work
p15a-hyg-ccdb-cm	Stewart Lab
p15a-cm-HPRT-EcoRV	Stewart Lab/Jun Fu
pR6K-PGK-EM7-neo	Stewart Lab

3.11 gRNA Sequences 5’-3’ PAM

Hprt 6-1 TGGGGCTGTACTGCTTAACC AGG

Hprt C15 TTTGCATTACAGCCCCAAAA TGG

Hprt C18 TTTGGGGCTGTACTGTACCA GGG

Hprt 67 CTGAGGTAGTTCTAGATCCA AGG

Hprt 76 CCTCTTAGGAGTCTAAAGTA GGG

Hprt 67/76-4 CTAGATTTTAGACTCCTAAG AGG

Neo-5 TGAGAGCTCCCTGGCGAATT CGG

Pgk-1 ATGCGCTTTAGCAGCCCCGC TGG

Pgk-8 TAGGGGAGGCGCTTTTCCCA AGG

Scn10a g3 GGTAGTTCCCACGGACCCAA AGG

Scn10a g7 ACGGTTCACTCCAGAGTCGC TGG

3.12 TALEN sequences:

Scn10a Start TALEN 21

Tal 1 Length: 29 GGTTATTTTCCTCCCAGAAGAGTGTAAAT
1.1: NN_1 NG_2 NG_3 NI_4 NG_5 NG_6 NG_7 NG_8 HD_9 HD_10 pFUSA_30A
1.2: NG_1 HD_2 HD_3 HD_4 NI_5 NN_6 NI_7 NI_8 NN_9 NI_10 pFUSA_30B
1.3: NN_1 NG_2 NN_3 NG_4 NI_5 NI_6 NI_7 pFUS_B7
LR: pLR_NG

Tal 2 Length: 28 AGATGGAGTTCCCCTTTGGGTCCGTGGG
2.1: HD_1 HD_2 HD_3 NI_4 HD_5 NN_6 NN_7 NI_8 HD_9 HD_10 pFUSA_30A
2.2: HD_1 NI_2 NI_3 NI_4 NN_5 NN_6 NN_7 NN_8 NI_9 NI_10 pFUSA_30B
2.3: HD_1 NG_2 HD_3 HD_4 NI_5 NG_6 HD_7 pFUS_B7
LR: pLR_NG

Spacer Length: 21
Spacer Sequence: CCTTCCCCAAGAAGAATGAGA NN

Scn10a Middle TALEN 13

Tal 1 Length: 16 GCCCACCCTGCTTGAT

```
1.1: NN_1 HD_2 HD_3 HD_4 NI_5 HD_6 HD_7 HD_8 NG_9 NN_10 pFUS_A
1.2: HD_1 NG_2 NG_3 NN_4 NI_5 pFUS_B5
LR: pLR_NG

Tal 2 Length: 16 AAGTACCTGATATGGG
1.1: HD_1 HD_2 HD_3 NI_4 NG_5 NI_6 NG_7 HD_8 NI_9 NN_10p FUS_A
1.2: NN_1 NG_2 NI_3 HD_4 NG_5 pFUS_B5
LR: pLR_NG
```

Spacer Length: 13
Spacer Sequence: CAGCTTAGCCCAG

3.13 ES Cell Culture

3.13.1 Cell Lines:

Mouse Embryonic Fibroblasts (MEF or Feeders): isolated from 13.5dpc embryos

E14Tg2a: mESC with a 129 male genetic background. Feeder independent

R1: mESC with a 129 male genetic background. Feeder dependent, but weaned for several experiments

HEK 293: Human embryonic kidney cells

3.13.2 MEF Cultivation and Inactivation:

MEFs from E13.5 embryos were expanded in DMEM with 4.5 g/l glucose supplemented with 10 % heat-inactivated FCS, 2mM L-glutamine, and 1 % P/S. Confluent cell layers were inactivated by mitomycin C (10 mg/ml) to prevent further cell divisions, and flash frozen at -80C before long term storage in liquid nitrogen. MEFs were thawed and plated at 80 % confluency 24 hours before seeding mESC.

3.13.3 mESC Culture

Feeder- independent ES cell lines were kept in an undifferentiated state by culturing them in standard ES medium. ES cells were kept at 37°C and 5 % CO2 and passaged every 48-72 hours to avoid confluent growth and differentiation.

Feeder-dependent ES cell lines were kept in an undifferentiated state in culture at 37°C and 5 % CO2. Mitomycin C inactivated MEFs were plated on gelatinized cell culture dishes for 24 hours in MEF medium. After 24 hours, ES cells were plated on the feeders in mESC medium. The cells were passaged every 48-72 hours to avoid confluency and differentiation.

3.13.4 mESC Passaging

ES cells were passaged by trypsinization (Trypsin 2x + G) for 5 minutes at 37°C. Trypsin was neutralized by addition of standard mESC medium and the cells clumps were manually dispersed by pipetting to form a single cell suspension. The cells were then distributed to new dishes at an appropriate concentration.

3.14 Long Term mESC Storage

Cell lines were flash frozen at -80°C and transferred to liquid nitrogen for long term storage in culture medium containing 15 % FCS and 10 % DMSO.

3.15 Cell culture reagents and media

0.1 % Trypsin (for mESC)

	Final Concentration
PBS	1x
2.5 % Trypsin	0.1 %
Chicken Serum	1 %
EDTA	0.2 mg/ml
D-glucose	1 mg/ml

mES Medium (FCS+LIF)

	Final Concentration
DMEM+ GlutaMax	
FCS	10 %
2-mercaptoethanol	100 mM
L-glutamine	2 mM
100 mM Sodium Pyruvate	1 mM
100x non-essential amino acids	1x
Stewart lab recombinant LIF	1 ml/500 ml
Additional if 2i conditions	
PD-0325901	1 µM
CT-99021	3 µM

Beta-mercaptoethanol Solution

7 μl of β-mercaptoethanol was added to 10 ml of PBS and filter sterilized. Aliquots were stored at -20°C.

3.16 mESC Transfection:

3.16.1 ES cell electroporation:

Actively growing mESC were rinsed with 1x PBS, trypsinized, collected with mESC medium in a 15 ml Falcon tube, pelleted by centrifugation, and washed once with PBS. Unless otherwise indicated, $5x10^6$ cells were resuspended in 800 μl of cold PBS mixed with DNA. The electroporation was performed in a chilled 4 mm electroporation cuvette with a Gene Pulser X-Cell (BioRad) using the exponential settings (250 V, 500 μF). The cuvettes were left for 5 min at RT after electroporation and the contents were transferred to a 15 ml Falcon tube with 5 mL pre-warmed mES medium. Cell clumps were distributed by manual pipetting, and the cell suspension was distributed to gelatinized 10 cm culture plates containing 9 ml pre-warmed medium. After 24 hours, fresh ES medium supplemented with appropriate selection was added and the selection medium was changed frequently until colonies were visible for picking (using a microscope) or counting after staining (by eye).

3.16.2 ES cell nucleofection:

Actively growing mESC were rinsed with 1x PBS, trypsinized, collected with mESC medium in a 15 ml Falcon tube, pelleted by centrifugation, and washed once with PBS. $2x10^6$ cells were resuspended in 100 μl of Nucleofection mix (2:9 mix of supplement 1 with mES Nucleofector solution (Lonza) and RNP. The RNP was formed by mixing 7.5 μg Cas9 protein (MPI-CBG) and 7.5 μg *in vitro* transcribed mRNA in H_2O for 10 min at RT. The cell-RNP suspension was transferred to nucleofection cuvettes (Lonza) and the nucleofection was carried out using a Nucleofector II device, program A-013 as recommended by the manufacturer. The contents of the cuvettes were immediately transferred to a 15 ml Falcon tube with pre-warmed mES medium. Cell clumps were distributed by manual pipetting, and the cell suspension was distributed to gelatinized 10 cm culture plates containing 9 ml pre-warmed medium. After 24 hours, fresh ES medium supplemented with appropriate selection was added and the selection medium was changed frequently until colonies were visible for counting after staining (by eye).

3.16.3 ES cell lipofection:

An appropriate amount of circular plasmid or BAC DNA was lipofected into $2.5x10^5$ mES cells seeded in 6-well plates 24 h before. The DNA was added into 300 µl Opti-MEM with Glutamax and 2.5 µl of Plus Reagent was added and incubated 5 min at RT. 6.25 µl of Lipofectamine LTX was added and the solution was carefully mixed by inversion. After 30 min RT incubation, the lipofection mix was added dropwise to the cells in 3 ml of fresh mES medium and incubated overnight.

3.17 T7E1

200 ng of column purified PCR product was denatured and re-annealed in a PCR cycler. A 20 µl reaction with 1x NEB buffer 2 and 1µl T7E1 enzyme was incubated for 1 hour at 37°C and the digest was run on a 1 % agarose gel to check for cleavage products.

3.18 *In vitro* transcribed RNA

In vitro transcribed RNA was made using the manufacturer's suggestions for the Mega Shortscript T7 kit. PCR primers were designed with a T7 promoter and a 'guard' integrated into the forward primer. A PCR was performed with a high fidelity polymerase, the product was extracted from a gel, and 1 µg of the DNA was used for the transcription reaction (using the buffer, dNTPs and T7 enzyme from the kit). The reaction was left to run overnight at RT. 1 µl of Turbo DNAse was added and the reaction was incubated at 37ºC for 15 minutes. The RNA was then purified twice with phenol/chloroform and precipitated with sodium acetate and EtOH. The RNA pellet was dissolved in H_2O and stored at -80ºC.

3.19 Quantitative PCR (qPCR)

Quantative PCR amplifications were completed in a Mx3000P or Mx3005P multiplex PCR instrument with the following reaction mixtures and thermal profile.

	20 µl	Final conc.
ddH_2O	7,4 µl	-
2x GoTaq qPCR Master Mix	10 µl	1x
10 µM Primer 1	0.8 µl	100 nM
10 µM Primer 2	0.8 µl	100 nM
DNA	1 µl	10 ng/ 20µl

		Temperature	Time
Initial denaturation		95ºC	5 s
40 Cycles	Denature	95ºC	10 s
	Anneal	62ºC	20 s
	Elongation	72ºC	30 s
Dissociation curve		95ºC	1 m
		62ºC	30 s
		0.5ºC steps 62ºC - 95ºC	30 s

The analysis of qPCRs was performed in triplicate, and the cycle threshold (Ct) values were normalized against the Ct of an internal control (MRRF or Nxt2). Fold change was calculated using the ΔΔCt method(121):

$$2^{-\Delta\Delta Ct}$$

$$\Delta\Delta Ct = (Ct_{gene} - \text{avg. } Ct_{housekeeping})_{sample} - (\text{avg. } Ct_{gene} - \text{avg. } Ct_{housekeeping})_{control}$$

3.20 X-Gal staining:

6-well plates with attached cells were washed once with 1x PBS, then 2 ml fixative solution was added to each well, followed by a two-minute incubation at room temperature. After aspiration of the fixative solution, the wells were washed three times with 1x PBS and incubated overnight at 37°C with 2 ml staining solution. The next day, the wells were washed once with 1x PBS and blue staining was documented with a stereomicroscope.

	Fixative Solution		Staining Solution
675 µl	Formaldehyde [37 %]	25 µl	$MgCl_2$ [1 M]
50 µl	Glutaraldehyde [25 %]	1.25 ml	Potassium Ferricyanide [50 mM]
11.78 ml	PBS	1.25 ml	Potassium Ferrocyanide [50 mM]
12.5 ml		9.9 ml	PBS
		62.5 µl	X-Gal [200 mg/ml] in DMF*
		12.5 ml	

*heat staining solution to 50°C before adding X-Gal to prevent crystallization

3.21 Quantitative β-Gal assay:

Cells were washed with PBS and harvested by scraping in 1x PBS, transferred to 1.5 ml Eppendorf tubes and centrifuged at 4,000 rpm for four minutes at 4°C. The liquid supernatant was aspirated and the cell pellets were resuspended in 100 µl Extraction Buffer and disrupted by three cycles of flash freezing in liquid nitrogen and thawing in a warm water bath. The samples were centrifuged at 12,000 rpm for 15 minutes at 4°C to remove cell debris. The protein concentration was measured on the nanodrop, and the protein was diluted to 40 µg. The protein extracts were mixed with 800 µl of Solution I followed by 200 µl of Solution II. The samples were thoroughly mixed by vortexing and incubated in the dark at 37°C for 1 hour (until the yellow color is visible). The reaction was stopped by the addition of 500 µl [1 M] $NaCO_3$. The mixture was transferred into plastic cuvettes for measurement of absorbance at 420 nm using a UV/Visible Spectrophotometer.

Extraction Buffer	Solution I	Solution II
0.25 M Tris-HCl (pH 7.8)	1 M $Na_2HPO_4*2H_2O$	1 M $Na_2HPO_4*2H_2O$
100 mM DTT	1 M $NaH_2PO_4*2H_2O$	1 M $NaH_2PO_4*2H_2O$
100 mM PMSF	2 M KCl	2 mg/ml ONPG
(Adjust pH to 7.0)	1 M $MgCl_2$	

3.22 Genomic DNA extraction:

ES cells were grown at high density in a 24 well tissue culture plate for DNA extraction. Cells were washed with PBS and incubated with 200 µl of lysis buffer at 55 °C overnight in a humid box. Genomic DNA was precipitated from the lysis mix ethanol precipitation with 400 µl of NaCl and 100% EtOH. The DNA was precipitated in the plate according to the protocol described in "Genomic DNA Microextraction: A Method to Screen Numerous Samples"(122).

3.23 HPRT assay:

mESC were grown on gelatinized cell culture dishes. The medium was discarded, and the plates were washed with PBS. The mESC colonies were fixed in 4 mL of 80 % EtOH for 10 minutes and then stained in 4 ml Coomassie blue staining solution (500 ml: 362.5 ml H_2O, 100 ml MeOH, 37.5 ml acetic acid, 0.25 g Coomassie G250) for 10 minutes. The staining solution was decanted (and saved to be reused) and the culture dishes washed with H_2O and air-dried. The stained colonies were manually counted.

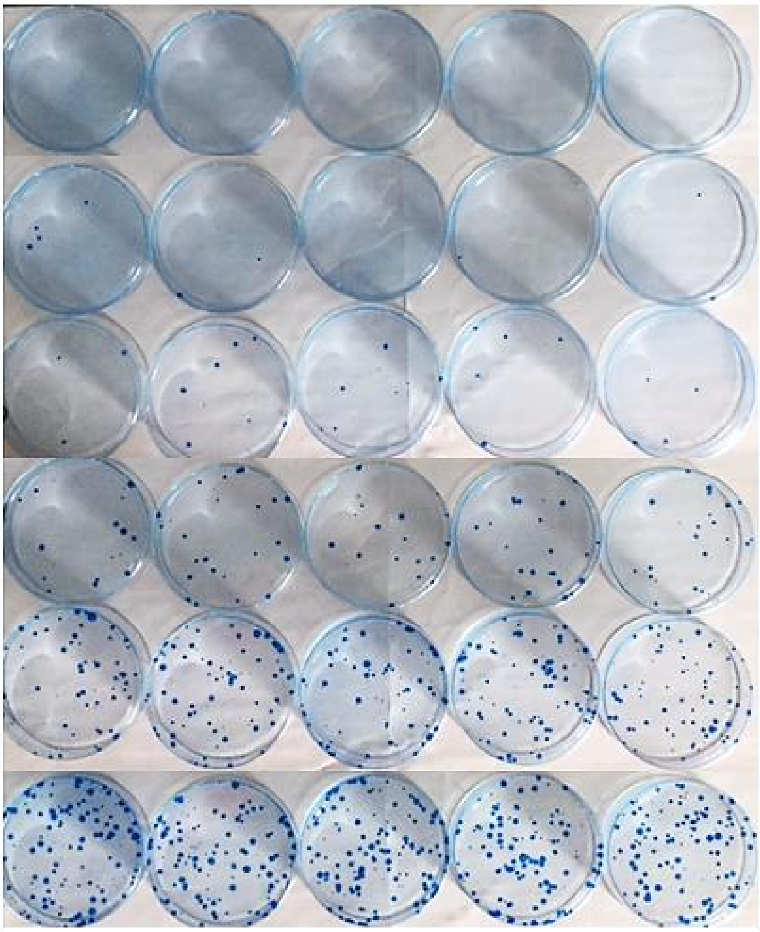


Figure 4 Quantitative *Hprt* rescue assay Cells with a damaged *Hprt* were transfected with repair constructs or ODN and Cas9/gRNA. After 24 hours, HAT supplemented media was applied to the plates to select for cells with a repaired *Hprt*. After 10 days, the colonies were fixed and counted to provide a quantitative evaluation of nuclease assisted targeting efficiency.

3.24 Southern assay:

Genomic DNA (5 µ-10 µg per ES cell clone) was digested with NdeI restriction enzyme and the digested DNA bands were separated by electrophoresis through a 0.8 % agarose gel in 1x TBE buffer. Gels were run at low voltage to allow for proper band separation. Following electrophoresis gels were denatured twice with denaturing buffer for 15 minutes and once with neutralizing solution for 20 minutes. The DNA was transferred via capillary action to a nylon membrane by blotting with 20x SSC overnight. After blotting, the membrane was rinsed with 6x SSC and the DNA was fixed to the membrane by UV crosslinking.

Denaturation Buffer for Southern blotting

	Final Concentration
NaCl	1.5 M
NaOH	0.5 M

Reagents were dissolved in sterile H_2O with a final volume of 1 L.

Neutralization Buffer for Southern blotting

	Final Concentration
Tris	1 M
NaCl	1.5 M

*The Tris and NaCl were dissolved 800 ml of H_2O, the pH value was adjusted to pH 7.4 with HCl and the final volume was adjusted to 1 L with H_2O.

20 x SSC

	Final Concentration
NaCl	3 M
Sodium Citrate	0.3 M

Reagents were dissolved in 800 ml H_2O, pH adjusted with HCl to pH 7.0, and the final volume was adjusted to one liter and sterilized by autoclaving.

Random prime DNA labeling was achieved using the HighPrime Kit. Template DNA (50 – 100 ng) from a gel extracted PstI fragment of the neo gene was denatured in 11 µl distilled water for 10 minutes at 99°C and chilled on ice immediately to prevent renaturation. HighPrime (4 µl) and radioactive [α32P] dCTP (5 µl) were added and samples were incubated at 37°C for 1 hour. Unincorporated dNTPs were removed by filtration with Sephadex G50 Quick Spin Columns. Radiolabelled DNA probes were loaded onto Quick Spin Columns and centrifuged for 4 minutes at 1,000 x g. Purified radiolabelled DNA probes were used immediately.

Nylon membranes with fixed target DNA were rinsed with 6x SSC and rolled into hybridization bottles. Prehybridization solution for southern blots was pre-warmed to

37°C before use. 8 ml CG-prehybridization solution was added to each hybridization tube and membranes were prehybridized for 1 hour at 63°C. Radio-labeled DNA probes were heat-denatured at 99°C for 10 minutes and chilled rapidly on ice. Denatured probes were added to the prehybridization solution and membranes were hybridized at 63 °C overnight. After hybridization, membranes were washed 2x for 20 minutes with CG washing buffer at 65°C to remove residual unbound probe and probes that were not specifically bound to the membrane. The washed membranes were covered with plastic wrap and exposed to X-ray films overnight before detection with the FLA3000.

CG Wash Solution

	Final Concentration
Na_2HPO_4* $2H_2O$	20 mM
20 % SDS	1 %
0.5 M EDTA pH 8.0	1 mM

Reagents were dissolved in sterile H_2O with a final volume of 1 L.

CG Hybridization Solution

Stock solution Final concentration Amount for 0.5 liter

	Final Concentration
Na_2HPO_4* $2H_2O$, pH 7.2	0.25 M
20 % SDS	7 %
10 % Bovine serum Albumin (BSA)	1 %
EDTA, pH 8.0	1 mM
H_2O	

Reagents were mixed, filtered with a 0.33 µM filter, aliquoted, and stored at -20°C

3.25 Large scale bacterial DNA preparation:

Large-scale extractions of plasmid DNA were made using Qiagen reagents from 250-500 ml overnight cultures in LB supplemented with appropriate antibiotics following the manufactures directions. Briefly, the cells were pelleted and resuspended in 10 ml of Qiagen buffer P1, then mixed with 10 ml of Qiagen Buffer P2 and incubated at room temperature for 5 minutes to lyse the cells. 10 ml of Qiagen Buffer P3 was added to the mixture, and the mixture was incubated for 20 minutes on ice. The sample was briefly centrifuged to clear the lysate, then filtered through a Whatman filter into a Qiagen 500 column that was equilibrated with Qiagen Buffer QBC. The column was washed twice with 30 ml of Qiagen Buffer QC and eluted with 15 ml of Qiagen Buffer QF. 10.5 ml of isopropanol was added, the solution was vigorously mixed, and the sample was centrifuged for 1 hour at 9,000 rpm. The DNA pellet was washed with 70% ethanol, air dried and resuspended in water to an appropriate concentration.

Large-scale extractions of BAC DNA were made using the MN BAC 100 kit from 500 ml overnight cultures in LB supplemented with appropriate antibiotics following the manufactures directions. Briefly, the cells were pelleted and resuspended in buffer 20 ml of S1, then mixed with 20 ml of buffer S2 and incubated at room temperature for 5 minutes to lyse the cells. 20 ml of buffer S3 was added to the mixture, and the mixture was incubated for 20 minutes on ice. The sample was briefly centrifuged to clear the lysate, then filtered through a Whatman filter into a BAC 100 column that was equilibrated with buffer NBT. The column was washed twice with 30 ml of buffer N3 and eluted with 15 ml of pre warmed buffer N5. 10.5 ml of isopropanol was added, the solution was vigorously mixed, and the sample was centrifuged for 1 hour at 9,000 rpm. The DNA pellet was washed with 70 % ethanol, air dried and resuspended in water to an appropriate concentration.

3.26 Small scale bacterial DNA preparation:

Plasmid and BAC DNA was extracted by alkaline lysis mini prep. In general, a 2 mL culture of LB supplemented with appropriate antibiotics was inoculated with a bacterial culture and left to grow overnight at an appropriate temperature (30 or 37°C) in a thermoshaker at 950 rpm. The next day, the culture was centrifuged at 13,200 rpm for 1 min to pellet the bacteria. 200 μL of Qiagen buffer P1 was added and the pellet resuspended. Then 200 μL of Qiagen buffer P2 was added and the mixture was gently mixed to allow lysis of the cells. 200 μL of Qiagen buffer P3 was added to the mixture and gently mixed to neutralize the solution. The samples were then centrifuged for 10-20 min to pellet the lysed cell components. The DNA containing supernatant was then transferred to a new microcentrifuge tube containing 500 μL isoproponal, and the samples were vigorously mixed. The samples were then centrifuged at 13,200 rpm for 20 minutes to pellet the DNA. The supernatant was decanted, and the DNA pellet gently washed with 500 µL 70 % EtOH. Finally, the pellet was then left to dry before resuspension in water.

3.27 Restriction endonuclease digest:

Restriction Endonuclease digestion of double stranded DNA was performed according to the instructions of the manufacturer using the provided buffer.

3.28 Bacterial electroporation:

DNA was transfected into bacterial cells via electroporation. A 1 mL culture was grown to confluency (usually overnight) in LB supplemented with the appropriate antibiotics at the appropriate temperature (30 or 37°C) in a thermoshaker at 950 rpm. From the confluent culture, 30 µL was used to inoculate tubes with a punctured lid containing 1.4 mL fresh LB supplemented with the appropriate antibiotics. After 2 hours growth at 30°C shaking at 950 rpm, the cultures were induced with arabinose (if applicable) and transferred to 37°C shaking at 950 rpm for 45 min. The cultures were then centrifuged at 4°C at 9,000 rpm for 30 s, the supernatant decanted, and the cell pellet resuspended in 1mL cold water. The samples were then centrifuged a second time (4°C,10,000 rpm, 30 sec) the supernatant decanted, and the pellets resuspended in 1 mL cold water. The samples were centrifuged a final time (11,000 rpm, 4°C, 30 sec) and all but ~30 µL was decanted from the tubes. The pellet was resuspended in this ~30 µL, mixed with DNA and transferred to a chilled 1 mm electroporation cuvette. The cuvette was immediately pulsed in a Biorad electoporator at 1250 V aiming for a time constant between 3.0 and 6.0 ms.

3.29 Ethanol DNA precipitation:

To precipitate (and therefore purify and concentrate the DNA), 1/10 of the sample volume of buffer P3 (or 3M NaAc) and 3x the volume of 100% ice cold EtOH were added to the DNA solution. The sample was thoroughly and gently mixed by inversion, and the tube was incubated at -20°C for 1 hour or -80°C for 5 minutes. The samples were then centrifuged at 20°C at maximum speed and the pellet washed in 70% EtOH and air dried. The DNA was then dissolved in water to an appropriate concentration.

3.30 PCR purification:

PCR products were purified of primers, dNTPs, and PCR reagents using the QIAquick PCR purification kit (Qiagen) according to the manufactures instructions. Briefly, 5 volumes of binding buffer were mixed with the PCR reaction mixture. The mixture was applied to the QIAquick spin column, centrifuged, and the flow-through discarded. 10-25 µl of water was added to the column, incubated for 5 minutes at RT and then centrifuged. The eluate was collected in a 1.5 ml micocentrifuge tube.

3.31 Gel extraction:

DNA fragments were separated by electrophoresis on agarose gels in 1x TBE. The desired band was visualized via excited blue light on the Dark reader and excised. DNA was extracted using the QIAquick gel extraction kit (Qiagen) according to the manufactures instructions.

3.32 Gel electrophoresis:

DNA was electrophoresed in TBE agarose gels with 0.8 % to 2 % agarose depending on the size of the desired fragment. Ethidium bromide was added to the gels at 100 ng/ml to visualize the fragments, and the fragments were photographed using the syngene G:Box GeneSnap gel documentation system

3.33 TALEN design and assembly:

TALENs target sites were identified using the TALE-NT TALEN Targeter tool(123) by identifying the ideal region of TALEN cleavage and selecting the surrounding sequence. The sequence was submitted to the TALEN Targeter, which locates all possible TALEN target sites in the submitted sequence. From the list of identified TALEN targets, one TALEN pair was selected based on its proximity to the ideal cleavage site, it’s spacer length, and TAL length (with shorter, easier-to-assemble TAL effectors preferred with a minimum length of 16 RVDs. Some submitted sequences returned many possible target sites, while others had very few. The chosen target sites were then compared against the mouse genome for potential off-target effects using the TAEN-NT tool.

Our lab used the kit designed and graciously provided by the Voytas lab.(123) TAL effector constructs were assembled as described using Golden Gate cloning. The Golden Gate method utilizes unique 4 bp overhangs (sticky ends) that are left by Type IIS restriction endonucleases. Briefly, RVD’s are assembled in two steps 1) assembly of the RVDs into intermediary arrays and 2) joining of the intermediary arrays into a final TAL effector expression vector.

First, the module plasmids that contain RVD's 1-10 plus intermediate plasmid pFUS_A and RVDs 11-11+n with the appropriate pFUS_B are digested and ligated in the same reaction mixture by adding both BsaI and T4 ligase in the T4 reaction buffer and incubating in a thermocycler for 10 cycles of 5 min at 37°C and 10 min at 16°C. The enzymes are inactivated for 5 min at 50°C followed by 5 min at 80°C. To prevent linear, incomplete DNA fragments from recombining or self-circularizing, the mixture is treated with Plasmid Safe DNase for 1 hour at 37°C. The reaction mixture is then dialyzed to remove buffer salts, electroporated into frozen competent GB05 or GB06 cells, and plated on LB-agar containing 60 μg/mL spectinomycin, with Xgal and IPTG for blue/white screening. Several white colonies were picked, cultured, and checked by restriction digest and the junctions were sequenced to show correct assembly of the RVDs.

In the second step, the intermediate, assembled RVD arrays, the appropriate last repeat (LR plasmid) and the expression vector are similarly digested and ligated, this time using the enzyme Esp3.I. The mixture was then dialyzed and electroporated into GB05 or GB06 cells and plated on LB supplemented with 100 μg/mL ampicillin. Depending on the expression vector, the LB was additionally supplemented with X-gal and IPTG (for blue white screening) or with ampicillin alone (for vectors where ccdb was replaced with RVDs). The resulting colonies are checked by restriction digest for correct assembly of the RVDs and sequence verified when possible.

3.34 Plasmid Construction:

3.34.1 gRNA design:

gRNA recognition sequences were identified by the search engine provided by the Zhang Lab.(65) A DNA sequence was submitted to the CRISPR search engine to look for potential gRNA recognition sequences. The search engine scans the submitted sequence for possible CRISPR guides (20 nucleotides followed by a PAM sequence: NGG) and scans for potential off target matches throughout the selected genome. The program then rates the identified gRNA sequences based on their potential off target effects. Only gRNAs with a high score (so low potential off target effects) were chosen for the experiments presented in this work.

3.34.2 gRNA expression vector construction:

gRNA expression vectors were assembled via recombineering. pBR322-U6-cm-ccdB-tracrRNAnew-amp was digested with BstZ171 and NheI and the digest was run on a 0.8 % agarose gel. The 3019 bp band was extracted from the gel and used for all further reactions. The fragment of pBR322-U6-cm-ccdB-tracrRNAnew-amp was co-electroporated with a synthesized oligo that contained left and right HA to the vector and the 20 bp gRNA recognition site into competent bacterial cells that carried a psc101-ETgA plasmid. The cells were then plated on LB agar plates with ampicillin. The next day colonies were picked, and the plasmid DNA was extracted by alkaline lysis mini-prep and the insertion of the linear oligonucleotide was confirmed by restriction digest. Correct colonies were then confirmed by sequence analysis to ensure correct integration and the absence of mutations from the oligo.

3.34.3 TALEN expression vectors:

In order to shorten the linker, and decrease background from cloning, a ccdb-hyg cassette recombineering cassette was designed. Two synthetic ODNs (pTAL3-ccdb-F and pTAL3-ccdb-R) were purchased that had 50bp HA to the vector, an Esp3.1 site, and ~21bp PCR primers that amplified the ccdb-hyg portion of the p15A-hyg-ccdb plasmid. The pCDNA3.1(-)-pTAL3Δneo vector was linearized, and the purified PCR product and the linearized vector were co-electroporated into GB05-dir GyrA462 cells to produce pTAL3-ccdb-hyg-17aa. The cells were plated on LB-agar with Hygromycin selection and single colonies were picked after 24 hours. Positive colonies were verified first by restriction digest, then by sequence confirmation of the junctions to confirm correct integration of the cassette and the absence of mutation from the ODNs.

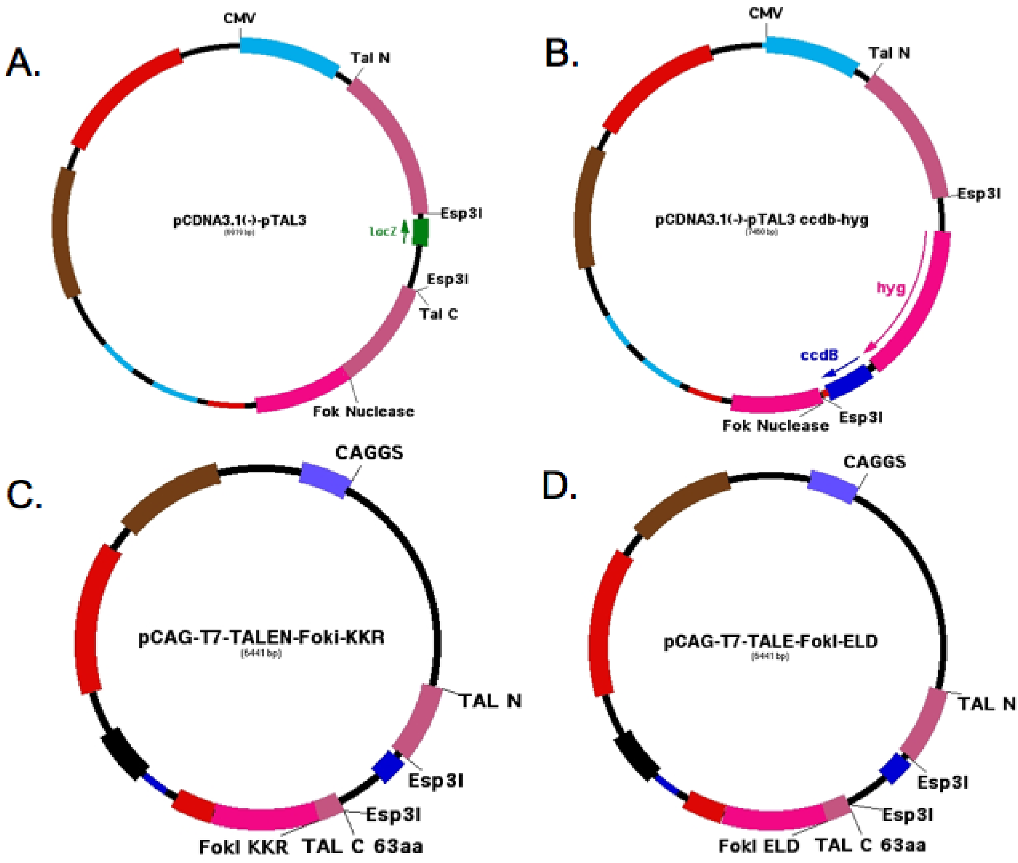

A.
CMV
Tal N
pCDNA3.1(-)-pTAL3
lacZ
Esp3I
Esp3I
Tal C
Fok Nuclease
B.
CMV
Tal N
Esp3I
pCDNA3.1(-)-pTAL3 ccdb-hyg
hyg
ccdB
Esp3I
Fok Nuclease
C.
CAGGS
pCAG-T7-TALEN-Foki-KKR
(6441 bp)
TAL N
Esp3I
Esp3I
FokI KKR
TAL C 63aa
D.
CAGGS
pCAG-T7-TALE-FokI-ELD
(6441 bp)
TAL N
Esp3I
Esp3I
FokI ELD
TAL C 63aa

Figure 5 TALEN expression vectors The different TALEN vectors vary in their TAL C linker length, FokI nuclease sequence, and eukaryotic promoter. A) pCDNA3.1(-)-pTAL3 has a 236 aa linker and a CMV promoter. RVD insertions can be screened for with blue/white colonies. B) pCDNA3.1(-)-pTAL3-ccdb-hyg has a shortened linker (17 aa) and RVD insertion has minimal background because of the ccdb selection. It has a CMV promoter and is derived from pCDNA3.1(-)-pTAL3. C and D) pCag-t7-TALE-FokI-ELD/KKR destination vectors are obligate heterodimers. RVD insertion is screened for with blue/while colonies. The constructs have a 63 aa linker and a CAGGS promoter.

3.34.4 CMV-Rep plasmid assembly:

Two ODNs (LacZ ccdb F and LacZ ccdb R) were purchased that added 2x BstBI RE sites to PCR primers for the ccdb-hyg cassette from a p15A-ccdb-hyg plasmid template. The LacZ ccdb F primer additionally contained NheI and NaeI RE sites and the Lac ccdbR primer additionally contained AscI and KpnI RE sites to the plasmid could be used for additional downstream experiments. The PCR produced a 1633 bp product. This product, and the CMV-Rab-Rep plasmid from the Kuhn lab were digested with BstBI and the digestion products were purified by column and ligated overnight at 16°C with T4 DNA ligase to produce a CMV-Rep-ccdb-hyg plasmid. The ligation product was electroporated into electrocompetent GB-dir GyrA462 cells and the cells were plated on LB agar plates with hygromycin and ampicillin. Single colonies containing the correct plasmid were verified by restriction digest and sequence confirmed for absence of mutations from the ODN. This plasmid could then be used to insert the recognition sequence of other TALEN pairs and used for *in vivo* screening of TALEN cleavage efficiency.

ODNs were then purchased that had the Scn10a Start and Scn10a Middle TALEN recognition sequences flanked by 2x NaeI and 2x KpnI RE sites. The Scn10a-Start-Rep-Ins and the Scn10a-Middle-Rep-Ins ODNs and the CMV-Rep-ccdb-hyg plasmid were digested with NaeI and KpnI and the digestion products were column purified. The purified products were ligated using T4 Ligase overnight at 16°C and the ligation products were electroporated into GB05 cells to produce CMV-Scn10a Start-Rep and CMV-Scn10a Middle-Rep. Positive colonies were screened by restriction digest and the absence of mutations was confirmed by sequencing.

3.34.5 Subcloning *Hprt* repair vectors:

psc101-γβαA was electroporated into electrocompetent bacterial cells containing the RP23-13N1 BAC (CHORI). The cells with the BAC + psc101-γβαA were plated on LB agar plates with chloromphenical and tetracycline. The plates were grown overnight at 30°C. Recombineering oligos that contained a 50 bp HA to the BAC sequence and 21 bp PCR primers were used to amplify a pBR322-amp cassette from a pBR322-amp-ccdb-hyg plasmid. The subcloning back-bone PCR was electroporated into electrocompetent BAC + psc101-γβαA that had the recombineering proteins produced after induction with L-arabinose. After recovery, the cells were plated on LB agar plates with ampicillin. Single colonies were grown, the plasmid DNA extracted by alkaline lysis mini prep and subclones of *Hprt* were confirmed by restriction digest and sequence analysis of the junctions to ensure no mutations were inserted from the ODNs.

3.34.6 Insertion of a PGK-neo-cassette into *Hprt* exon 6:

ODN (Quad HPRT Neo R and Quad HPRT PGK F) were purchased that contained 50 bp HA to the RP23-13N1 and PCR primers to the pR6K-PGK-em7-neo plasmid. The HA to the BAC were designed such that 43 bp of *Hprt* exon 6 and was deleted upon insertion of the PGK-neo cassette which also caused a frameshift to ensure a KO phenotype. A 1604 bp PCR product was amplified and column purified. Recombineering proficient bacterial cells with the *Hprt* containing RP23-13N1 BAC +psc101-γβαA were electroporated with a purified PCR product to produce RP23-13N1-PGK-neo cells. The cells were plated on LB agar plates with ampicillin and kanamycin. Single colonies were grown, the plasmid DNA extracted by alkaline lysis mini prep and the insertion of the PGK neo cassette was confirmed by restriction digest and sequence analysis of the junctions to ensure no mutations were inserted from the ODNs.

The RP23-13N1-PGK-neo cells were then made electrocompetent and electroporated with psc101-γβαA. The cells were plated on LB agar plates with chloromphenical and tetracycline and incubated at 30°C. The cells were then grown and made recombineering proficient and electroporated with a back-bone PCR (Hprt-5kb-a and Hprt-5kb-b amplified pBR322-amp) to subclone a 10 kb segment of the *Hprt* gene (pBR322-amp-Hprt-Pgk-neo-5kb). The cells were plated on LB agar with ampicillin and kanamycin. Single colonies were grown, the plasmid DNA extracted by alkaline lysis mini prep and subcloning of the PGK neo cassette and surrounding DNA was confirmed by restriction digest and sequence analysis of the junctions to ensure no mutations were inserted from the ODNs.

3.34.7 Subcloning of Δ672 fragment:

The 672bp fragment of *Hprt* that was removed via CRISPR/Cas9 was subcloned from the *Hprt* containing RP23-13N1 BAC. ODNs (Ligation Vector OD F/R and Ligation Vector SD F/R) were designed that contained 50 bp HA to the BAC, the HPRT 67/76 gRNA 4 recognition site in either the same direction (SD) or opposite (OD) direction as the gRNA genomic recognition site, and 21 bp PCR primers that amplified the pBR322-amp of a pBR322-amp-ccdb-hyg plasmid. Recombineering proficient bacterial cells with the *Hprt* containing RP23-13N1 BAC +psc101-γβαA were electroporated with the purified PCR products to produce pBR322-Hprt-OD-Δ672-OD and pBR322-Hprt-SD-Δ672-SD. The cells were plated on LB agar plates with ampicillin. Single colonies were grown, the plasmid DNA extracted by alkaline lysis mini prep and the insertion of the OD-Δ672-OD and SD-Δ672-SD was confirmed by restriction digest and sequence analysis of the junctions to ensure no mutations were inserted from the ODNs.

3.34.8 Engineered Mutations in Targeting Vector Homology Arms:

A BssSI recognition site was inserted in *Hprt* exon 6 for simplified screening of recombination events. The substituted base pairs make a silent mutation. pBR322-amp-HPRT-1kb was linearized with BsmI. GB05+ psc101ETγA cells were made recombineering proficient by expression of the recombineering genes and the linearized pBR322-amp-HPRT-1kb gel extracted fragment was co-electroporated with purchased ODNs HPRT-LLHR BssSI-1n and HPRT-LLHR BssS1-2. The oligos contain 100mers that contain 50 bp HA to the pBR322-amp-HPRT-1kb plasmid and 50 bp homology to each other, which includes the BssSI silent mutations. The cells were plated on LB agar plates with ampicillin to select for the pBR322-amp-Hprt-1kb-BssSI cells. Single colonies were isolated, the plasmid DNA extracted by alkaline lysis mini prep and the insertion of the BssSI was confirmed by restriction digest and sequence analysis of the junctions to ensure no mutations were inserted from the ODNs.

pBR322-amp-Hprt-1kb-BssSI cells were linearized with either BsmI, BstZ171, SexAI, EcoNI, or BsaBI. GB05+ psc101ETγA cells were made recombineering proficient by expression of the recombineering genes and the linearized pBR322-amp-HPRT-1kb-BssSI gel extracted fragment was co-electroporated with the appropriate LLHR oligo. The oligos either inserted a 6 bp NdeI or deleted 5 bp at the BsmI site (+61), the Bstz171 site (-181), the SexA1 site (-464), the EcoNI site (-701), or the BsaBI (-883) site. The ODNs contained 50 bp HA each side of the engineered mutation (100mers for the deletion ODNs, 106mers for the NdeI ODNs). The cells were plated on LB agar plates with ampicillin to select for the plasmids with engineered mutations. Single colonies were isolated, the plasmid DNA extracted by alkaline lysis mini prep and the insertion of the BssSI was confirmed by restriction digest and sequence analysis of insertion/deletion event.

3.35 Design of the *SCN10A* humanization construct

Shengbiao Hu, a former student in the Stewart lab developed an ambitious targeting exercise to humanize the mouse *Scn10a* gene. The *SCN10A* humanization construct is a complex design that includes three cassettes with multiple site-specific recombinase recognition sites. The construct was assembled via recombineering. The first cassette (Figure 6A) consists of a FRT flanked splice acceptor and lacZneo cassette and a loxP site inserted into the first intron. This would allow for screening of the targeting event by lacZ staining. However, the lacZ cassette is driven by the endogenous promoter, and it was later determined that *Scn10a* is not expressed in mESC. Therefore, the LacZ gene and G418 selection cannot be used in mES cells. The LacZneo cassette can be deleted upon expression of the Flp recombinase leaving a LoxP site in the genome. A second cassette (Figure 6B) consisting of Rox flanked PGK-BSD and a loxP site was inserted after the second exon. As blasticidin (BSD) is driven by its own promoter, there is no need for endogenous expression of *Scn10a* to drive the BSD resistance. The PGK-BSD cassette can be deleted upon expression of the Dre recombinase, leaving a loxP site in the genome. The final cassette (Figure 6C) is a F3 flanked hygromycin gene driven by a ubiquitin promoter and loxP site, which was inserted before the stop codon. Upon Flp expression, the F3 sites will recombine, thus deleting the hyg cassette and the LoxP site. Subsequent Cre expression will result in a KO phenotype while leaving the gene body intact. Alternatively, if the Cre recombinase is expressed before expression of the Flp recombinase, the LoxP sites will recombine, thus deleting the entire *Scn10a* gene body.

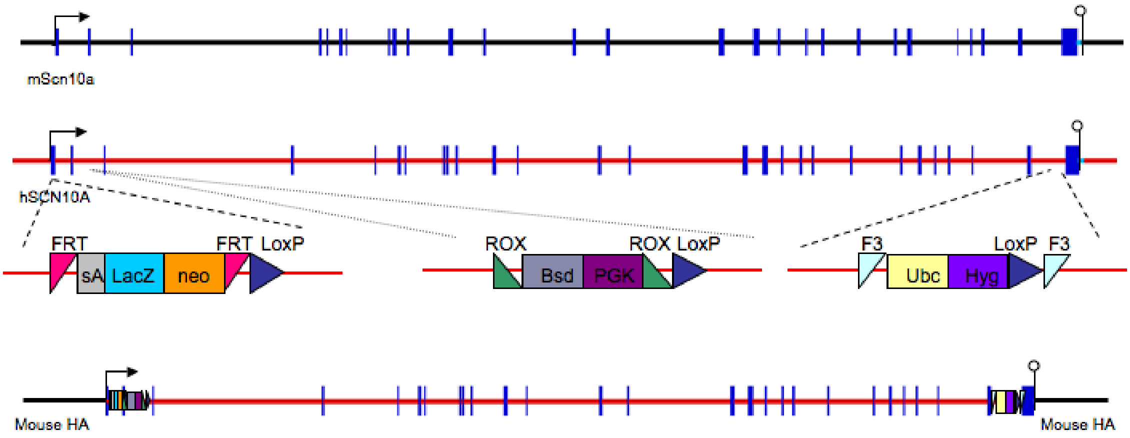


Figure 6: Design of *SCN10A* Targeting Construct A) *mScn10a* gene is 85 kb and has 28 exons. B) *hSCN10A* is 96 kb and has 27 exons. Three cassettes were inserted into a BAC carrying the human *SCN10A* gene using standard recombineering methods C) Recombineering added 8 kb and 7 kb mouse HA to the human *SCN10A* gene resulting in a humanization targeting construct.

4 Results

4.1 Humanization of the mouse *Scn10a* gene

As TALENs were one of the first designer nucleases that could be easily assembled in the laboratory, our lab was excited to utilize this emergent technology to increase targeting efficiency in our own projects. Designer nucleases opened doors for experiments that were too ambitious without them, specifically the newly emergent TALENs as they were an affordable solution to ZFNs (CRISPR/Cas9 systems for gene targeting had not yet been published). An ambitious targeting exercise to humanize the mouse *Scn10a* gene, in which the 85 kb of the mouse genomic *Scn10a* gene would be replaced with a 96 kb human *SCN10A* gene (modified with three selection cassettes (section 3.35)) flanked by 7 kb and 8 kb homology arms was not possible, perhaps due to the large size of the construct. Several attempts were made to target the mouse genome in E14Tg2a mES cells using this construct, but were never successful. To overcome this challenge, TALENs offered an exciting method to increase targeting efficiency at the Scn10a locus.

4.1.1 Strategy to increase targeting efficiency at the Scn10a locus with TALENs

TALEN recognition sequences were identified at three separate locations in the mouse genome that did not recognize the TC or the human *SCN10A* gene. These TALENs recognized sequences present at the 5'(start), Middle, and 3' (end) of the mouse *Scn10a* gene (figure 7A). TALENs were assembled using the Golden Gate system from the Voytas lab(123) (section 3.33). The TALEN Start pairs, Middle pairs and the *SCN10A* targeting vector were then co-lipofected into E14Tg2a mES cells. To select for targeted cells, after 24 hours, medium supplemented with BSD was applied to the cells for 5 days, and then medium supplemented with HYG was put on the cells for an additional 5 days. After 10 days, there were no colonies on the plates. In an extended effort to overcome this issue, this experiment was repeated several times, with varying amounts of TC and TALEN constructs, and no targeted colonies were ever obtained. This indicated that either the Scn10a TALENs had poor cleavage efficiency, or there was a problem with the TC design. In an attempt to tackle these questions systematically, a more in depth examination of TALEN expression was necessary.

4.1.2 Evaluation two different TALEN pairs to test *in vivo* cleavage efficiency

In order to validate the TALENs, an assay was performed to check for DNA cleavage in an *in vivo* experiment. Ten µg of each TALEN Scn10a Start and Scn10a Middle Pairs (5 µg left and 5 µg right TALEN) were electroporated into 5x10^6 E14Tg2a cells which were plated into a 96 well plate, giving small bulk populations. After 30 hours, DNA was extracted from the wells and checked for cleavage with the T7E1 assay. Primer pairs (Scn10a Start -300F/ +200R and Scn10a Middle -300F/+200R) that amplified ~500 bp around the cut site were used to amplify genomic DNA extracted from the E14Tg2a cells. The PCR reaction was column filtered to remove primers and salts, and 200ng of each sample was denatured and re-annealed (figure 7B) in a PCR cycler, digested with T7E1 nuclease, and the digestion products were run on a 1 % gel. Positive and negative controls provided in the Surveyor nuclease kit were used to verify T7E1 cleavage, and undigested PCR product was loaded to show the original PCR band size (figure 7C). T7E1 cleavage of the mismatches produced by heteroduplex formation during reannealing would result in two bands, one at 200 bp and one at 300 bp. No cleavage product was observed for the bulk population of cells (figure 7C), which indicates that the cleavage efficiency of the TALENs was less than ~5 %, which is the detection efficiency of the T7E1 assay.(124) As the two different TALEN pairs that were tested exhibited low *in vivo* cleavage efficiency, the next step was to try to improve TALEN efficiency in our culture conditions.

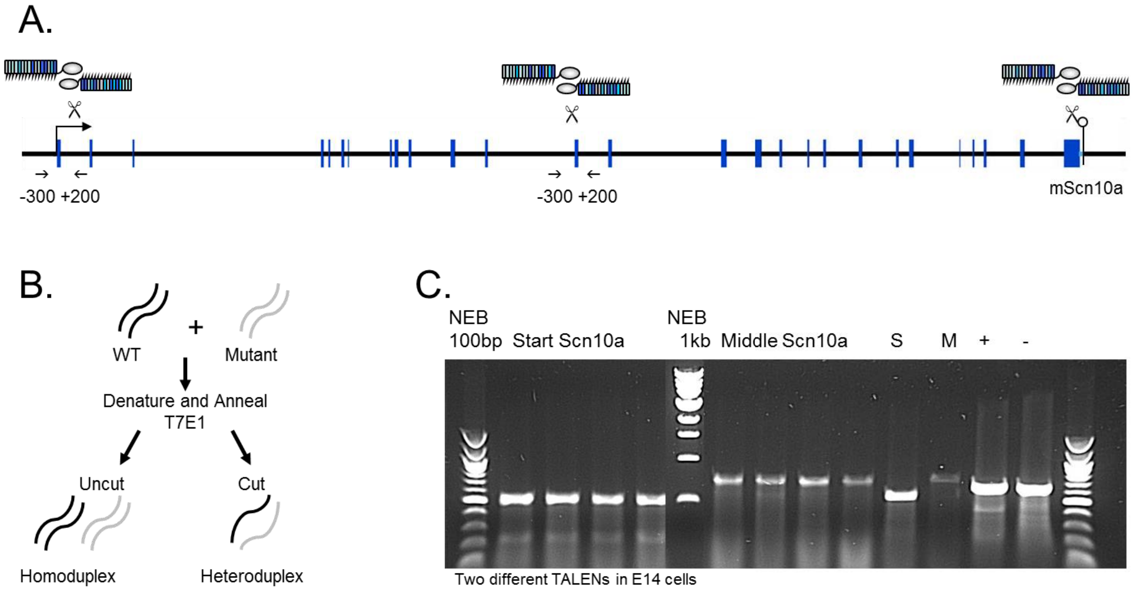


Figure 7: TALEN Strategy and Validation with T7E1 assay A) TALEN design locations and primer pair location for amplification of genomic DNA B) Amplified genomic DNA is denatured and

reannealed. The T7E1 nuclease recognizes and cleaves non-perfectly matched heteroduplex DNA. C) T7E1 assay of amplified genomic DNA of E14tg2a cells transfected with different TALEN pairs.

4.1.3 Modification of the TALEN expression vector to increase *in vivo* cleavage in mESC

Mussolino et al.(63) indicated that the linker between the TAL RVDs and the FokI nuclease could affect the cleavage efficiency of the TALENs by FokI positioning, and suggested that a shortened C-terminal linker increased genome editing activity while reducing toxicity. The linker in the Voytas kit is 236 aa, which could be easily shortened to the suggested 17 aa, as the first 17 aa of the linker in the Voytas kit are supplied by the plasmid encoding the last repeat. This made shortening the linker very simple by removing the rest of the linker in the destination vector (pCDNA3.1(-)-pTAL3). In addition to shortening the linker, cloning background was reduced by inserting a ccdb-hyg cassette into the final destination vector. CcdB is a toxin that poisons the gyrase-DNA complex, which blocks polymerase activity and leads to DSB. Cells that have a mutation in the gene encoding the A subunit of topoisomerase II (gyrA462) are resistant to the ccdb toxin.(125) Primers were designed that amplified a ccdb-hygromycin cassette (see section 3.43.3) and added Esp3.I restriction endonuclease sites and homology arm the pCDNA3.1(-)-pTAL3 final destination vector. Because of the integrated Esp3.I sites, the new destination plasmid with the short linker can be used in conjunction with the Voytas kit assembled RVDs.

The RVDs of the Start and Middle TALENs were then inserted into the new TALEN expression vector with the shortened linker (pcDNA3.1 (-)-pTAL3-ccdb-hyg) using ligation of the cleaved Esp3.1 restriction endonuclease sites. Correct integration of the RVDs was confirmed by restriction digest and the junctions were sequence confirmed. The entire length of RVDs could not be sequence confirmed due to the repetitive nature of the RVD sequence.

Variations of the FokI domain were reported to increase the specificity and activity of cleavage in ZFN,(126) and were an interesting possibility to increase TALEN cleavage, as well. The ELD/KKR mutations (ELD:KKR denotes Q486E,I499L,N496D and E490K,I538K,H537R) in FokI make it an obligate heterodimer, which reduces off-target effects. TALEN expression vectors were obtained from Addgene that could be utilized with the Voytas Goldengate kit (Addgene 40132 and 40131). The RVDs for the Scn10a Start and Scn10a Middle TALEN pairs were ligated into the ELD/KKR vectors using Esp3.1. Correct cloning events were confirmed by restriction digest and the junctions were sequence confirmed.

4.1.4 An *in vivo* test for TALEN cleavage

A TALEN pair with established target cleavage *in vivo* via TALEN mRNA microinjection into the embryo(127) was provided to us by a colleague for a control along with system they used to confirm TALEN cleavage *in vivo*. The pCMV-Rab-Rep plasmid is an unpublished construct that consists of the first 135 amino acids of the LacZ gene followed by a stop codon and the recognition site for the Rab^{cht} TALEN pair followed by the last 1,022 aa of the LacZ gene (figure 8). This means that there is a repeat of 132 aa of the LacZ gene separated by the TALEN recognition site. Expression of the TALEN pair that recognizes the sequence that separates the LacZ gene causes a DNA DSB that is then repaired by the cell machinery. The 100 % homology between the replicated portions of LacZ form a perfect repair template to form a fully functioning LacZ gene that can be visualized by X-gal staining (figure 8).

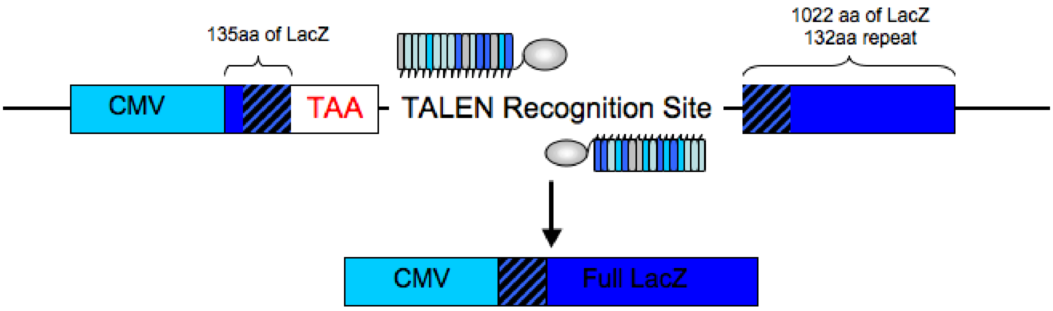


Figure 8 pCMV-Rep Vector Design The Rep plasmids consist of the first 135 amino acids of the LacZ gene followed by a stop codon and the recognition site for the a TALEN pair followed by the last 1,022 aa of the LacZ gene. Expression of the TALEN pair that recognizes the sequence interrupting the LacZ gene causes a DNA DSB that is then repaired by the cell machinery. The homology between the replicated portions of LacZ forms a repair template to form a fully functioning LacZ gene.

In order to use the pCMV-Rep vectors to test the efficiency of our Scn10a TALENs, the Rab^{cht} TALEN recognition site was deleted from the pCMV-Rab-Rep plasmid (see section 3.34.4). Both the pCMV-Rep plasmid and purchased oligos with the Scn10a Start and Scn10a Middle recognition site flanked by NaeI and KpnI RE sites were digested with NaeI and KpnI enzymes. Ligation of the digested products formed a transiently expressed plasmid that could test for *in vivo* cleavage of our Scn10a TALENs. The modified pCMV-Rep-ccdb-hyg plasmid can be modified to test any TALEN pair *in vivo*. With the inserted Scn10a Start and Middle TALEN recognition sites, the next step was to test the new TALEN expression plasmids for increased TALEN cleavage.

4.1.5 Testing for *in vivo* TALEN cleavage

Many of the published works that utilized TALENs worked with HEK293 cells. HEK293 cells are a human embryonic kidney cell line that was transferred with sheared human adenovirus and adenovirus DNA. The resulting cell line is hypotriploid with a modal chromosome number of 64. HEK293 cells are feederless and easy to transfect, making them a commonly used cell line for experiments.

TALEN pairs and their corresponding pCMV-Rep vectors were lipofected into 2.5×10^5 HEK293 cells to compare the cleavage efficiencies of the long linker supplied by the Voytas lab, the 17aa linker, and the ELD/KKR expression vectors. Cells were transfected with PBS as a negative control, a βactin-LacZ expression vector as a transfection and staining control, and the pCMV-Rab-Rep plasmid with and without the Rab^{cht} TALEN pair as a positive control for TALEN cleavage. Cells were also transfected with the pCMV-Scn10a-Start-Rep plasmid without TALENs and with the Scn10a Start RVDs in the long linker, 17aa, and ELD/KKR vectors and the pCMV-Scn10a-Middle-Rep plasmid without TALENs, and with the Scn10a Middle RVDs in the long linker, 17aa and ELD/KKR vectors. After 72 hours, the cells were fixed and stained to test for LacZ expression.

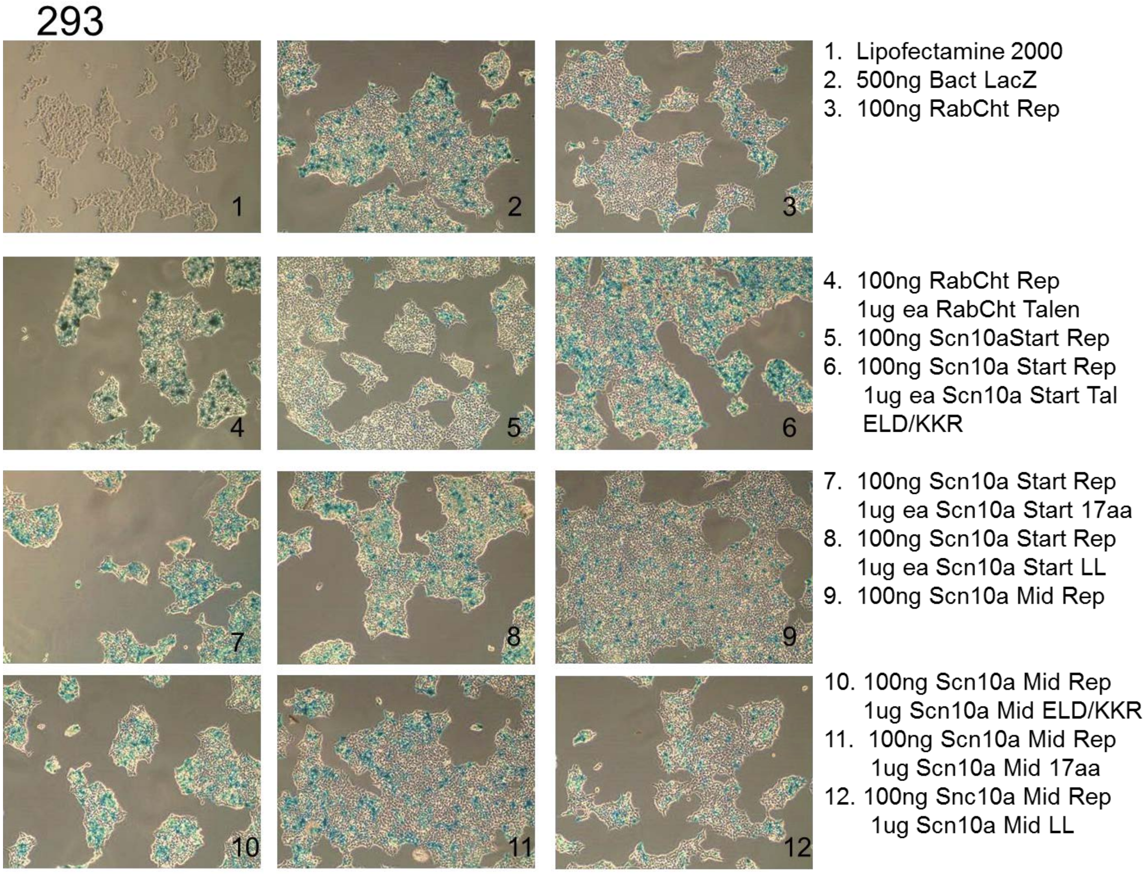

293
1
2
3
4
5
6
7
8
9
10
11
12
1. Lipofectamine 2000
2. 500ng Bact LacZ
3. 100ng RabCht Rep
4. 100ng RabCht Rep
1ug ea RabCht Talen
5. 100ng Scn10aStart Rep
6. 100ng Scn10a Start Rep
1ug ea Scn10a Start Tal
ELD/KKR
7. 100ng Scn10a Start Rep
1ug ea Scn10a Start 17aa
8. 100ng Scn10a Start Rep
1ug ea Scn10a Start LL
9. 100ng Scn10a Mid Rep
10. 100ng Scn10a Mid Rep
1ug Scn10a Mid ELD/KKR
11. 100ng Scn10a Mid Rep
1ug Scn10a Mid 17aa
12. 100ng Snc10a Mid Rep
1ug Scn10a Mid LL

Figure 9 Transfection of HEK293 cells with pCMV-Rep plasmids and different TALEN expression vectors HEK293 cells were transfected with PBS, a βactin-LacZ expression vector, the pCMV-Rab-Rep plasmid with and without the Rab^{cht} TALEN pair, the pCMV-Scn10a-Start-Rep plasmid without TALENs and with the Scn10a Start RVDs in the long linker, 17aa, and ELD/KKR vectors and the pCMV-Scn10a-Middle-Rep plasmid without TALENs, and with the Scn10a Middle RVDs in the long linker, 17aa and ELD/KKR vectors.

As expected, there were no blue stained cells in the negative control (figure 9-1). There were blue staining cells in the cells transfected with a βactin-LacZ expressing plasmid (figure 9-2) indicating efficient DNA transfection and cell staining. Surprisingly, there were some blue staining cells from all conditions transfected with the LacZ rep constructs (figure 9-3, 9-5, 9-9). This most likely comes from recombination of the LacZ repeated portions of the LacZ gene in the absence of DSB. The Rab^{cht} TALEN pair shows an increase over the pCMV-Rab-rep alone (compare figure 9-3 to 9-4). The pCMV-Scn10a-Start-rep also shows some background staining (figure 9-5). However, the Scn10a Start RVDs in the ELD/KKR, 17aa and original TALEN expression vectors showed an increase in staining. An increase in blue staining cells was not as evident with the Scn10a Middle TALENs, in any of the expression vectors (compare figure 9-9 to 9-10, 9-11, 9-12).

As the staining results looked promising, the experiment was repeated in HEK293 cells using a quantifiable β-Gal assay. As there was no visible increase in staining with the Scn10a Middle TALEN pairs, those vectors were excluded from this experiment. As before, $2.5\text{x}10^5$ HEK293 cells were lipofected with PBS as a negative control, a βactin-LacZ expression vector as a transfection and staining control, and the pCMV-Rab-Rep plasmid with and without the Rab^{cht} TALEN pair as a positive control for TALEN cleavage. Cells were also transfected with the pCMV-Scn10a-Start-Rep plasmid without TALENs and with the Scn10a Start RVDs in the long linker, 17aa, and ELD/KKR vectors. After 72 hours, the cells were collected, pelleted and the quantifiable β-Gal assay was performed. Expression of the repaired LacZ gene was tested for by X-Gal staining by measuring absorbance at 420nm. Again, as expected, there was no background from the cells or Lipofectamine, and a plasmid expressing LacZ showed a sharp increase in LacZ expression. The Rab^{cht} TALEN pair showed an increase in LacZ expression over the pCMV-Rab-Rep plasmid alone ($p = 0.05$ in a paired T-test). The results with the SCN10a Start TALEN pairs were less impressive. Neither the ELD/KKR, the 17aa expression vector, nor the original expression vector showed a significant increase compared to the pCMV-Scn10a-Rep plasmid background (17aa linker $p = 0.1$ in a paired T-Test, ELD/KKR and Long linker $p > 0.5$).

This was a turning point for the project. It was obvious that the designed TALENs had poor cleavage efficiency *in vivo*. As the goal was to humanize the *Scn10a* gene, a new strategy needed to be designed: either re-design the TALENs, or switch to a different designer nuclease. During the last decade designer nucleases have made enormous strides, and a new emergent technology, CRISPR/Cas9 for genome engineering, had been successfully used in our lab. CRISPR/Cas9 design and gRNA assembly has the advantage of being less laborious than TALENs, and had been successfully used in our hands to modify mES cells. Therefore, a new experimental design was developed with CRISPR/Cas9.

Simultaneously, the question still about optimized construct design still remained. The large TC for *SCN10A* was not designed with nuclease assisted targeting in mind. It has long HA, flanked by FseI sites to linearize the construct (two requirements for traditional gene targeting). The question remained as to how designer nuclease assisted targeting in mammalian genomes effected vector design. Therefore, a new experiment was planned that could easily test different vector design parameters and their effect on targeting efficiency when using a designer nuclease.

4.2 Improving targeting efficiency via targeting construct optimization:

4.2.1 Hprt as a selectable locus to test targeting efficiency

HPRT codes for the HGPRT protein; a transferase that has an important role in the purine salvage pathway. It catalyzes the conversion of hypoxanthine to inosine monophosphate and guanine to guanosine monophosphate. While purines can be synthesized *de novo* by the cell, this expends energy. By utilizing the salvage pathway, the cell economizes energy expenditure. HPRT allows for both positive and negative selection by using medium supplemented with hypoxanthine, aminopterin and thymidine (HAT) which permits the growth of *Hprt*$^+$ cells, but not *Hprt*$^-$ cells, while medium supplemented with 6 thioguanine (6-TG) kills cells that are *Hprt*$^+$ and allows *Hprt*$^-$ cells to grow.(128) As *Hprt* is monoallelic in male cells (*Hprt* is located on the X chromosome) this is an ideal system to test the efficiency of targeting in cells in different experiments (figure 10).

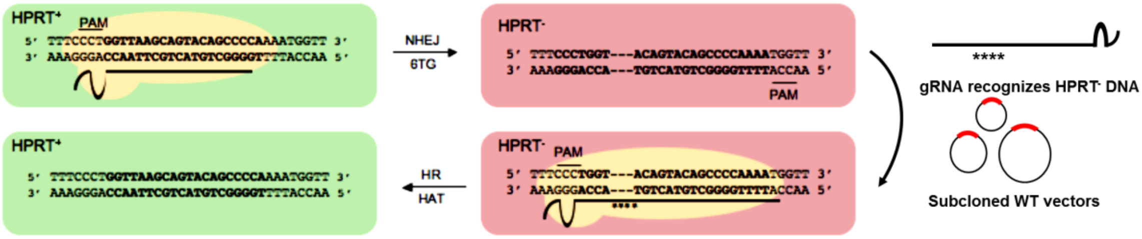
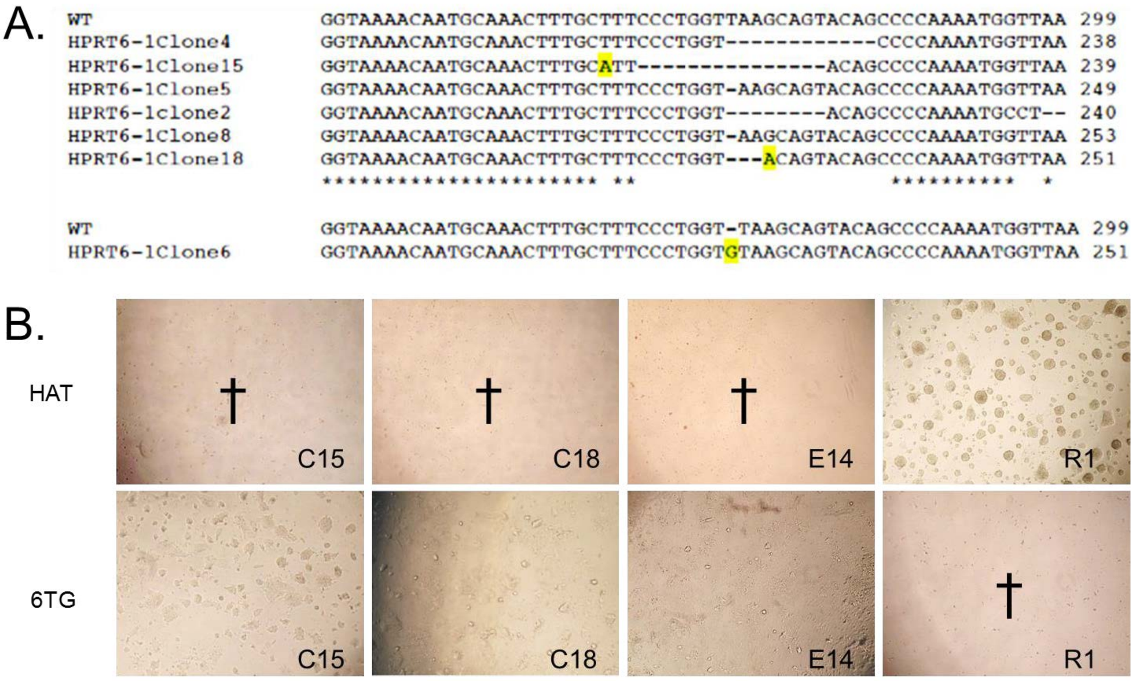

HPRT+
PAM
5' TTTCCCTGGTTAAGCAGTACAGCCCCAAAATGGTT 3'
3' AAAGGGACCAATTCGTCATGTCGGGGTTTTACCAA 5'
NHEJ
6TG
HPRT-
5' TTTCCCTGGT---ACAGTACAGCCCCAAAATGGTT 3'
3' AAAGGGACCA---TGTCATGTCGGGGTTTTACCAA 5'
PAM
****
gRNA recognizes HPRT- DNA
Subcloned WT vectors
HPRT-
PAM
5' TTTCCCTGGT---ACAGTACAGCCCCAAAATGGTT 3'
3' AAAGGGACCA---TGTCATGTCGGGGTTTTACCAA 5'
****
HR
HAT
HPRT+
5' TTTCCCTGGTTAAGCAGTACAGCCCCAAAATGGTT 3'
3' AAAGGGACCAATTCGTCATGTCGGGGTTTTACCAA 5'

Figure 10: Design of a quantitative experiment utilizing *Hprt* Wt R1 cells were electroporated with Cas9 and gRNA expression plasmids that recognized *Hprt* exon 6. NHEJ was allowed to occur, and cells with a damaged *Hprt* were selected for with 6-TG. A new gRNA expression plasmid recognizing the mutation (and not wt) was designed. The mutated exon 6 was repaired via HDR with donor DNA.

In order to knock out *Hprt* in R1 m ES cells, exon 6 of the wild type gene was damaged with Cas9. Exon 6 was selected as its structure is vital to the catalytic pocket of HPRT.(129) Wt R1 cells were electroporated with Cas9 and gRNA expression plasmids that recognized exon 6 (gRNA HPRT 6-1). NHEJ was allowed to occur, and cells with a damaged *Hprt* were selected for with 6-TG (figure 10). Multiple colonies were expanded and the NHEJ mutation was identified via sequencing (figure 11A). Mostly deletions, but also some substitutions, and a single integration were observed. The PCR primers used to sequence this region were placed 500 bp apart, so very large mutations will not be identified using this method. Two clones (c15 and c18) were chosen for phenotypic verification. Indeed, these clones are resistant to 6-TG and susceptible to HAT (figure 11B). The wt R1 cells were used as a control, and are HAT resistant and 6-TG susceptible. E14tg2a cells were used as an additional control. E14tg2a cells are missing part of the *Hprt* promoter and exons 1 and 2, and are therefore susceptible to HAT and resistant to 6-TG. C18 was chosen for experiments, as the 4 bp mutation resembles a commonly desired mutation (i.e. deletion or exchange of an amino acid). A new gRNA expression plasmid recognizing the c18 mutation was designed. This gRNA (gRNA C18) encompasses the 4 bp mutation from c18, and can no longer recognize and cleave wt DNA.



```
WT                      GGTAAAACAATGCAAACTTTGCTTTCCCTGGTTAAGCAGTACAGCCCCAAAATGGTTAA 299
HPRT6-1Clone4           GGTAAAACAATGCAAACTTTGCTTTCCCTGGT------------CCCCAAAATGGTTAA 238
HPRT6-1Clone15          GGTAAAACAATGCAAACTTTGCATT---------------ACAGCCCCAAAATGGTTAA 239
HPRT6-1Clone5           GGTAAAACAATGCAAACTTTGCTTTCCCTGGT-AAGCAGTACAGCCCCAAAATGGTTAA 249
HPRT6-1clone2           GGTAAAACAATGCAAACTTTGCTTTCCCTGGT--------ACAGCCCCAAAATGCCT-- 240
HPRT6-1Clone8           GGTAAAACAATGCAAACTTTGCTTTCCCTGGT-AAGCAGTACAGCCCCAAAATGGTTAA 253
HPRT6-1Clone18          GGTAAAACAATGCAAACTTTGCTTTCCCTGGT---ACAGTACAGCCCCAAAATGGTTAA 251
                        ********************** **                   **********  *

WT                      GGTAAAACAATGCAAACTTTGCTTTCCCTGGT-TAAGCAGTACAGCCCCAAAATGGTTAA 299
HPRT6-1Clone6           GGTAAAACAATGCAAACTTTGCTTTCCCTGGTGTAAGCAGTACAGCCCCAAAATGGTTAA 251
```

Figure 11 Genotypic and Phenotypic verification of *Hprt⁻* Cell lines (A) The sequence of seven *Hprt⁻* clones that have CRISPR/Cas9 induced damage were sequenced. All seven clones had indels around the cut site of the Cas9 induced DSB. (B) Two clones (C15) and (C18) were chosen for phenotypic confirmation. These clones are 6-TG resistant and HAT sensitive, and can therefore be used in selectable rescue experiments for repair of the *Hprt* gene.

4.2.2 Effect of homology arm length on targeting efficiency with CRISPR/Cas9

4.2.2.1 Introduction

Homologous recombination is a routinely used technique in gene targeting, however, though this technique has been in use for several decades, it has a relatively low efficiency. Early works focused on understanding the mechanism of HR, and trying to increase its efficiency. One parameter that was clearly shown to improve efficiency was using long stretches of isogenic DNA in the targeting vectors.

Thomas and Capecchi described an increase in targeting efficiency with increasing HA length.(104) This study only compared three HA totals (4 kb, 5.4 kb and 9.1 kb) that suggested a non-linear relationship between HA length and homologous recombination at the *Hprt* locus. The dependence of targeting efficiency on HA length was also noted at the μ gene of hybridoma cells,(130) 2 different locations in the *Hprt* gene,(131) and the Chinese Hamster *Aprt* locus.(132) Hasty et al. examined this relationship further, using a variety of targeting vectors between 1.3 and 6.8 kb at the *Hprt* locus.(133) They, too, observed a relationship between homology arm length and targeting efficiency, along with several additional observations. They noted that total homology less than 1.7 kb was insufficient to generate targeted events, and an increase of homology to 1.9 kb increased targeting events 5x.

Hasty et al.(133) also examined asymmetric constructs (constructs with HA of differing sizes) and noted that an arm with 472 bp was just as efficient as a 1.2 kb arm in forming a double crossover. This indicates that it is total homology that is limiting, rather than a single homology arm. These findings were also observed by Scheerer and Adair(132) at the Chinese Hamster *Aprt* locus; when using asymmetric homology arms, they noted that one arm less than 500 bp made less effective substrates for targeting than substrates with a similar total HA length. They noted no effect on targeting frequency when the short arm was at the 5’ or the 3’ end of the construct. However, several reports have indicated that when a single homology arm was reduced to below 1 kb, the recombination does not have the same level of fidelity.(131,133,134)

The optimal amount of homology is also unclear. Adding homology longer than 6 kb did not increase targeting efficiency in *Hprt,*(133) though Deng et al. showed a strong dependence on homology that saturated at 14 kb, rather than 6 kb in *Hprt.*(131) Similarly, Lu et al. did not observe an increase in targeting efficiency with arms longer than 8 kb (up to 100 kb) at the *β-globin* locus of ES cells.(135) The optimal amount of homology may depend on chromosomal landscape, isogenicity, or other, undefined parameters.

The presence of additional 3 kb of nonhomologous DNA at the ends of targeting vectors does not effect targeting efficiency at the *Hprt* locus,(131) nor did ~2 kb non homologous DNA at the ends of targeting vectors effect targeting efficiency at the μ locus.(130) This knowledge led to the use of negative selection genes (such as HSV-TK) to select against unintended integration events.(136) Additionally, the size of the insert seems to have no impact on targeting frequency, as inserts ranging from 8 bp to 12 kb in the *Hprt* locus all had similar rates of targeting.(137)

The use of designer nucleases has abolished the need for long homology arms for HR to occur.(5,6) Indeed, recent publications have evaluated the requirements of homology with designer nucleases. Byrne et al.(33) utilized varying arm lengths from 94 bp to 5 kb, including asymmetric constructs. They determined that flanking homology arms up to 2 kb resulted in the optimal targeting efficiency, and additional homology up to 5 kb reduced gene targeting efficiency. However, the authors used a standard amount of DNA per construct (not equimolar amounts).

This leaves several unanswered questions about the optimal length for HA when using nuclease assisted targeting. The selectable HPRT assay (section 4.2.1) provides an eloquent solution to quantifiably answer this question.

4.2.3 Construction of targeting vectors with different homology arms

In order to examine how homology requirements are impacted by designer nucleases, a series of repair vectors of DNA with varying homology arm length were cloned from the BAC RP23-13N1 using standard recombineering techniques.(86) The homology arms varied in length from 200 bp each side (i.e. 400 bp total homology) to 10 kb each side (ie 20 kb total homology) (figure 12). Additional constructs that had a standard 1 kb arm and one arm with varying lengths were cloned. A PvuI and an AgeI site was integrated in all constructs between the backbone and the homology arms to allow linearization of the construct.

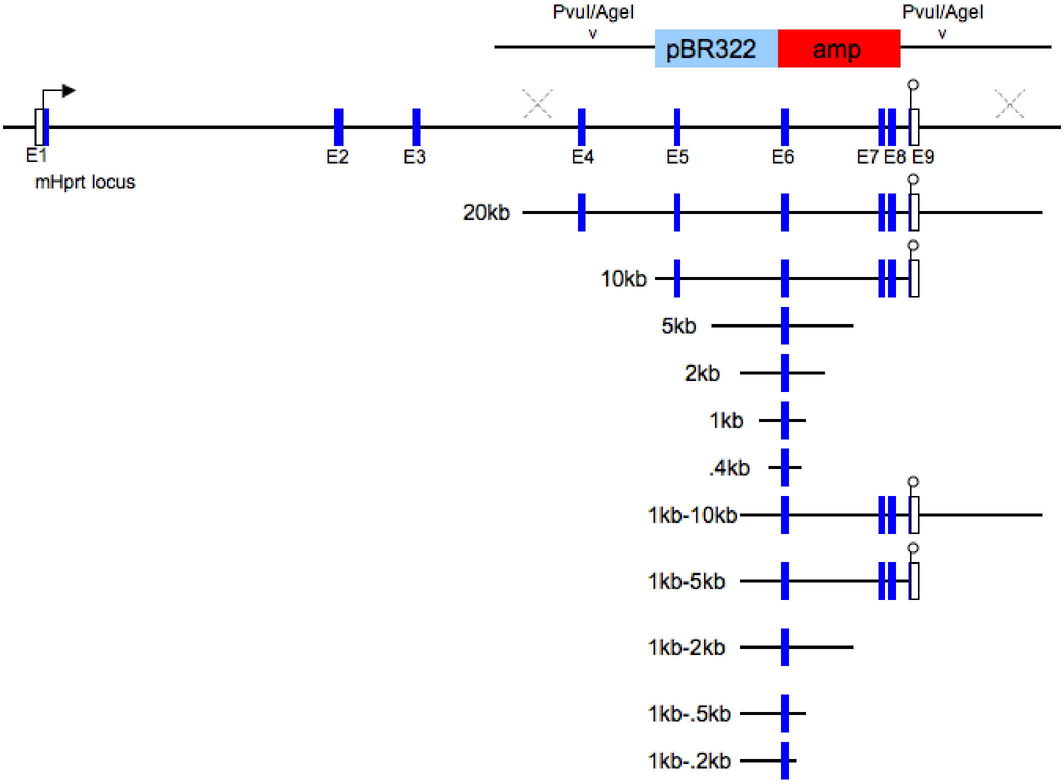


Figure 12 Subcloning Repair Vectors A) A series of repair vectors with homology arm length varying from 200 bp to 10 kb on each side were subcloned from a BAC carrying the *Hprt* gene. Some constructs were symmetrical (same amount of homology on both sides) and some asymmetrical (one standard 1 kb arm and one arm of varying length).

4.2.4 Targeting efficiency increases with total homology arm length

In order to test the effect of homology arm length on Cas9 induced DSB repair, standard amounts of DNA were used in all experiments. Repair vector amounts equimolar to 10 µg of a 4 kb total homology repair vector were used. This amount was chosen, as when a titration with increasing concentration of repair vector was performed, at 10 µg, the system was not yet saturated (figure 13A).

For plasmid electroporation experiments, a great excess of gRNA expression plasmid (15 µg) compared to Cas9 expression plasmid (5 µg) was used to limit off target effects that could be caused by excess Cas9 protein in cells.(65,72,73) Each electroporation was completed at least twice.

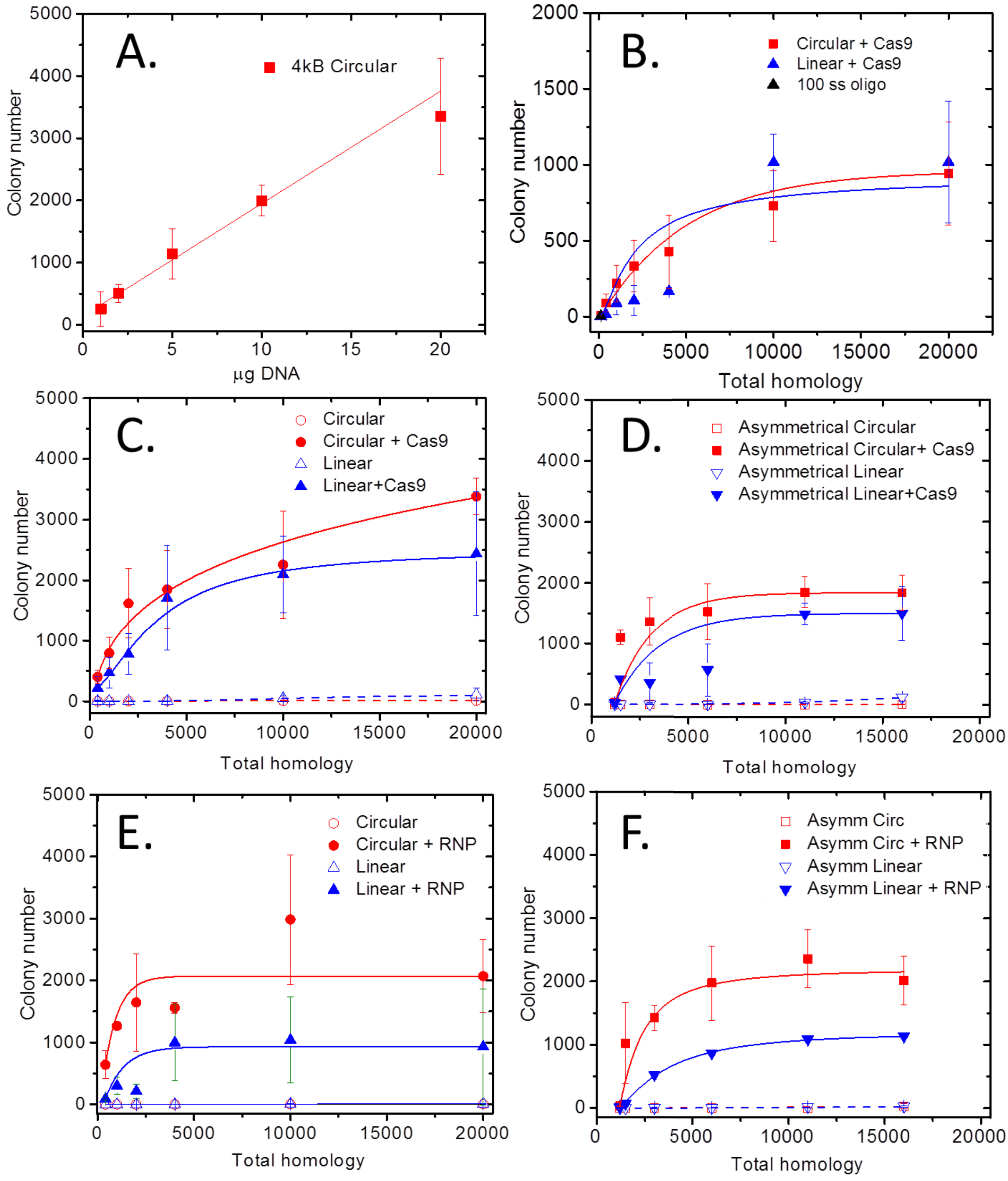


Figure 13 Effect of homology arm length on targeting efficiency A) Titration of different concentrations of a 4 kb repair vector. At 10 µg of DNA/ 5x10^{6} cells, the system is not saturated, and this amount was chosen for calculations of equimolar concentrations of the other repair constructs. B) Repair of *Hprt*$^{-}$ c15 cells with circular and linear repair constructs with varying HAs and a 100 bp single stranded oligo with gRNA and Cas9 expression plasmids. C) Repair of *Hprt*$^{-}$ c18 cells with circular and linear repair constructs with varying HAs with and without gRNA and Cas9 expression plasmids. D) Repair of *Hprt*$^{-}$ c18 cells with asymmetrical circular and linear repair constructs with one

standard 1 kb HA and one arm of varying length with and without gRNA and Cas9 expression plasmids. E) Repair of *Hprt*$^{-}$ c18 cells with circular and linear repair constructs with varying HA with and without RNP. F) Repair of *Hprt*$^{-}$ c18 cells with asymmetrical circular and linear repair constructs with one standard 1 kb HA and one arm of varying length with and without RNP.

A set of experiments was completed to compare several parameters of vector design for nuclease assisted targeting. First, the series of subcloned wt DNA with varying HA lengths with and without gRNA and Cas9 expression vectors were electroporated into 5×10^6 c18 cells. After electroporation, the cells were plated on gelatinized 10 cm culture plates and after 24 hours, 1x HAT supplemented media as added to the plates to select for *Hprt*$^{+}$ cells. After 10 days, the colonies were stained and counted.

As a control, C18 cells were electroporated with gRNA and Cas9 expression vectors alone. This gives us the background number of colonies that are produced by NHEJ at the DSB in mutated exon 6 in C18 cells. An average of 61 colonies were obtained. Nine of these colonies were sequenced to identify the NHEJ events. In 8/9 colonies, the three deleted nucleotides were replaced with random nucleotides. In 1/9 colonies, there were more significant deletions that led to a Hprt^{+} phenotype.

A steep increase in targeting efficiency with homology arms up to 2 kb in length was observed, with a more gradual increase up to 20 kb. This is in contrast to previous results that indicate homology arms more than 14 kb (without nuclease assisted targeting) do not increase targeting efficiency(131,133,135) and in contrast to reports that indicate that HA more than 2 kb (with nuclease assisted targeting) decrease targeting efficiency.(33) Additionally, targeting efficiency with and without the Cas9 expression plasmid was evaluated, and found that targeting increased dramatically when Cas9 was expressed. A maximum of 350-fold difference was observed when Cas9 was used. Finally, the targeting efficiency of linear vectors and circular vectors was compared. There was a slight increase in targeting with supercoiled, circular donor vectors for HDR when a DSB was induced. There were no colonies obtained with circular vectors without Cas9, as targeting without a nuclease requires a linear construct. When using a nuclease, the linearization and purification step can be eliminated, which simplifies the targeting protocol.

In order to ensure that the results obtained in this experiment were not exclusive to this C18, the experiment was repeated with another clone. Clone 15 has a 15 bp deletion and a single bp substitution (figure 11A). It was derived at the same time as C18, and is also 6-TG resistant and HAT sensitive (figure 11B). A gRNA that recognizes the C15 genome, but not the WT genome, was cloned into the gRNA expression vector. In contrast to the C18 cells, when C15 was electroporated with its gRNA and Cas9 alone, no HPRT$^+$ clones were made. This is likely because the 15 bp deletion is too large for the cell to repair and produce a functional HPRT protein. When the cells were electroporated with the gRNA and Cas9 plasmids along with 2 µg of 100 bp single stranded oligo, an average of 4 colonies were obtained. This shows that while ssOR is useful for small mutagenesis (section 4.3), its utility is greatly reduced for experiments that require larger mutations. The overall efficiency of HPRT repair is less with C15 (compare 13B to 13C). This is likely partially due to differences in the gRNA efficiency (section 4.5.3 for further discussion), but may also be due to the size of the lesion. The small 4 bp mutation can be corrected by c-NHEJ, but the larger mutation of C15 precludes the use of c-NHEJ. There was also a slight shift in the curve, with a sharp increase in efficiency up to 4 kb HA length vs the 2 kb HA in C18, which is especially noticeable in linear constructs. However, these results confirm that longer HAs increase targeting efficiency, that circular constructs are more efficient than linear constructs, and that while 2 kb total homology is sufficient for targeting using CRISPR/Cas9, an increase in HA length to 4 kb would be optimal.

Next, targeting efficiency of asymmetrical constructs that had one standard 1 kb arm and a second arm of varying length (figure 13 D) was determined. As previously, $5x10^6$ C18 cells were electroporated with equimolar amounts of the both circular and linear repair vectors, 15 µg of C18 gRNA expression vector and compared targeting efficiency with, and without, 5 µg of Cas9 expression vector. Each electroporation was completed at least twice, and the colony numbers were counted after the HPRT assay. Again, a large increase in targeting efficiency with CRISPR/Cas9 was observed. Also, a steep increase in targeting as total homology increased was again observed. The difference between linear and circular construct targeting efficiency was more apparent in this experiment, with supercoiled, circular donors having a higher targeting efficiency. Constructs with one homology arm shorter than 200 bp show a minimal increase in targeting above NHEJ background (figure 13D), even though symmetrical constructs with the same total homology had an increase in targeting above background. Once the HA was increase to 500 bp, an average of 1097 colonies for circular and 415 colonies for linear constructs was obtained. This indicates that that there is a requirement for each homology arm, and that it is somewhere between 200 and 500 bp. This observation is in accordance with results of targeting without a nuclease induced DSB.(132,133) Though colonies can be obtained with homology arms as short as 200 bp each side (figure 13C), targeting efficiency is markedly decreased.

4.2.5 Use of ribonucleoprotein to induce DSB

Cas9 protein can either be produced by the cell when an expression plasmid is transfected into the cell, or purified Cas9 protein can be combined with an *in vitro* transcribed gRNA to form a RNP complex that can be transfected into the cell. To determine the effect on targeting efficiency, the two different transfection methods: co-electroporation of the donor, Cas9, and gRNA expression plasmids, and co-nucleofection of the preformed RNP and donor DNA, were compared. The use of an RNP has several advantages over gRNA and Cas9 plasmid expression vectors. Firstly, RNPs have the possibility of being “cloning-free” by using an *in vitro* transcribed gRNA and commercially available Cas9 protein. Secondly, there have been several reports of expression plasmid DNA integration at Cas9 induced DSB both on target and off target(134,138) when small homologies between the plasmid and the genome lead to insertion of expression vector plasmid DNA. Finally, utilizing the RNP delivers a complete package of gRNA and Cas9, and does not require the cell to synthesize these components, thus ensuring temporal coordination of the reagents. Compared to expression plasmids, RNP editing is very fast.(70) Indels are detectible very shortly after electroporation, and reaches a plateau after 24 hours. Cas9 is rapidly removed from the cells, and is mostly absent within 24 hours. In contrast, plasmids delivered by electroporation take longer to begin editing the genome, and persist up to 72 hours in the cell.(139) Since on-target cleavage reaches its maximum while Cas9 protein is still being expressed, the only remaining effect would be off target cleavages. Indeed, Cas9 RNP exhibits a lower level of off target mutagenesis.(139)

The C18 gRNA that recognized the 4 nt mutation in *Hprt*$^{-}$ C18 cells was *in vitro* transcribed and Cas9 protein was purchased from MPI-CBG. The RNP was preformed at RT for 10 minutes, and the RNP was mixed with the donor vector in Lonza ES Cell nucleofection solution supplemented with S1 and electroporated into C18 cells using a Lonza nucleofector. For this experiment, 2×10^6 C18 cells were used, which is 2.5x less cells than electroporations using the Biorad electroporator, because of the limitations of the nucleofector solution and cuvettes (as suggested by the manufacturer). The cells were plated immediately after electroporation onto gelatinized 10 cm plates, and after 24 hours medium supplemented with 1x HAT was added to the plates to select for cells with a repaired *Hprt*. Each nucleofection was repeated at least twice.

An electroporation of the RNP without repair template to determine the number of background colonies repaired by NHEJ resulted in an average of 116 colonies (more than the 61 colonies from the plasmid electroporation control) indicating that even with 2.5x less cells, cleavage was more efficient with the RNP than the plasmid system. As this experiment utilized 2.5x less cells, our initial thought was to use 2.5x less DNA, however, this produced so many colonies it was difficult to count using the stained colony HPRT assay. Therefore, the concentration of repair vector DNA was decreased eight fold to obtain a countable number of colonies.

The RNP was then co-electroporated with equimolar amounts of the series of repair vectors used in figure 13 C/D to test the effect of HA length on HDR when a DSB is caused by the RNP. As seen in figure 13 E/F, even with 2.5x less cells and 8x less repair vector, the RNP produced similar number of repaired colonies with up to 10 kb total homology. The largest, 20 kb total homology showed a decrease in targeting, which may be due to the tendency of nucleofections to break up large constructs.(140) The same tendency was seen with the larger asymmetrical constructs (figure 13F). It appears that constructs over ~15 kb have less targeting efficiency when nucleofection is used, and this decrease is more prominent with circular vectors. This is in contrast to electroporation, which did not see a decrease in targeting efficiency after ~15 kb (figure 13 C-F) (though no sharp increase is observed, either). This experiment once again verifies that targeting efficiency of DSB repair induced by Cas9 RNP (figure 13E) increases with homology arm length, that circular constructs are more efficient than linear constructs, and that there is a steep increase in targeting efficiency with homology arm length up 2 kb then a leveling off of efficiency increase. The asymmetrical (figure 13F) constructs once again show that it is overall homology that has the biggest impact on repair efficiency, as long as there is one homology arm longer than ~500 bp.

While HDR is an effective means for gene targeting, it has several pitfalls. Vector design and assembly is a laborious process. The requirement for long HA is diminished with nuclease assisted targeting, meaning HA can be added via PCR, or even come from commercially purchased oligos. Therefore, the next section of this work examines the efficiency of repairing DSB with commercially purchased oligonucleotides.

4.3 ODN repair of DSB

4.3.1 Introduction

Modification of the genome with oligonucleotides is possible without nuclease induced DSBs with efficiencies up to 2 %,(141,142) which is attractive in gene therapy applications, as ODNs do not require multiple targeting steps like selection marker removal. Additionally, in contrast to most HDR donors, ssODNs can be commercially purchased, and require no vector assembly. However, the targeting efficiency of ssODNs is quite low, though it can be increased by slowing replication fork progression,(143) or by using phosphorothioate bonds to prevent nuclease degradation.(144,145)

The main limitation of targeting with ssODNs is decreased survival of modified cells,(8,145,146) especially when using protected oligos.(7,145) Gene editing by ssODNs requires a large amount of molecules to be introduced into cells, and can cause a reduced proliferation phenotype (RPP).(147,148) The mechanism of RPP is still not clear, however stalled replication forks have been suggested as the main cause.(149,150) However, when ssODNs are used in conjunction with nuclease induced DSBs, the amount of ssODN molecules that need to be introduced into the cell are reduced,(148) thus opening the doors for ssODN repair as an attractive alternative to HDR.

Much work has gone into optimizing ODN donor design by examining the repair efficiency of single stranded vs double stranded ODNs, strand bias, length and symmetry around cut site. Earlier works had indicated that without nuclease induced DSBs, oligos must be single stranded in order to be incorporated,(7) however, with induced DSBs, dsODNs can be inserted, but at a lower frequency than ssODNs.(146,151,152)

Several reports have indicated a strand bias when using ssODN repair,(7,8,149,152,153) yet others have not seen a strand bias.(151,152,154) Reasons for strand bias have been attributed to titration of the ssODN by the gRNA(153) and to the dynamics of Cas9 protein interaction with DNA (Cas9 asymmetrically releases Cas9 from the cleaved, PAM distal, non target strand first).(151) Richardson et al. optimized ssODN design by designing ssODNs that are complementary to the nontarget strand and overlap the Cas9 cut site with 36 bp on the PAM distal side and 91bp on the PAM proximal side (151), and Liang et al. also observed an increase when using asymmetric oligos,(146) though they did not observe a strand bias and therefore suggested a mechanism based on 5’ resectioning. The “optimal” length (80-90 bp) of donor ssODNs might agree with the later logic. In contrast to HDR donors in which longer is better, ssODN repair is hindered by longer ssODN donors.(8,146) This may be due resectioning of the cut template resulting in overhangs of ~30-40 bp(155) that form more stable annealing with shorter ODNs, though other factors such as secondary structures may also be at play.

The *Hprt* system introduced earlier in this work (section 4.2.1) can also be used to examine the use of ssODN repair with nuclease induced DSBs. Therefore, the next part of this work focuses on repair of the C18 *Hprt*$^-$ cells with ODNs.

4.3.2 Targeting efficiency with ODN

As constructs with as little as 200 bp HA (400 bp total homology) were still able to repair a 4 nt defect in *Hprt*, the repair efficiency of ssODNs was the next logical step. An equimolar amount (calculated according to the repair constructs in section 4.2.4) of a 50 bp ssODN, 100 bp ssODN, ds50 bp annealed ODNs, and ds100 bp annealed ODNs was electroporated with 15 µg C18 gRNA and 5 µg Cas9 expression vectors into 5x10^6 C18 *Hprt*$^-$ cells. After 24 hours, the medium was replaced with 1X HAT to select for *Hprt*$^+$ cells. After 10 days, the colonies were fixed, stained and counted. To determine if a strand bias occurs at this location, both a sense and antisense ssODN were commercially purchased. These oligos were annealed to form the dsODNs. When an equimolar amount of ODN was used, there were no colonies obtained above the background number of colonies from NHEJ, indicating that the reduced homology in ODNs up to 100 bp is not enough for efficient gene repair. Therefore, increasing amounts of ODN were electroporated into the C18 cells (figure 14A). Here a linear increase in repair efficiency with increasing amounts of DNA was observed. The 50nt ss and dsODN showed the least amount of repair. ds100nt ODN and ss100nt sense showed similar levels of repair and 100nt antisense oligo showed the highest level of repair, though this difference was not statistically significant at all data points on the plot (figure 14A).

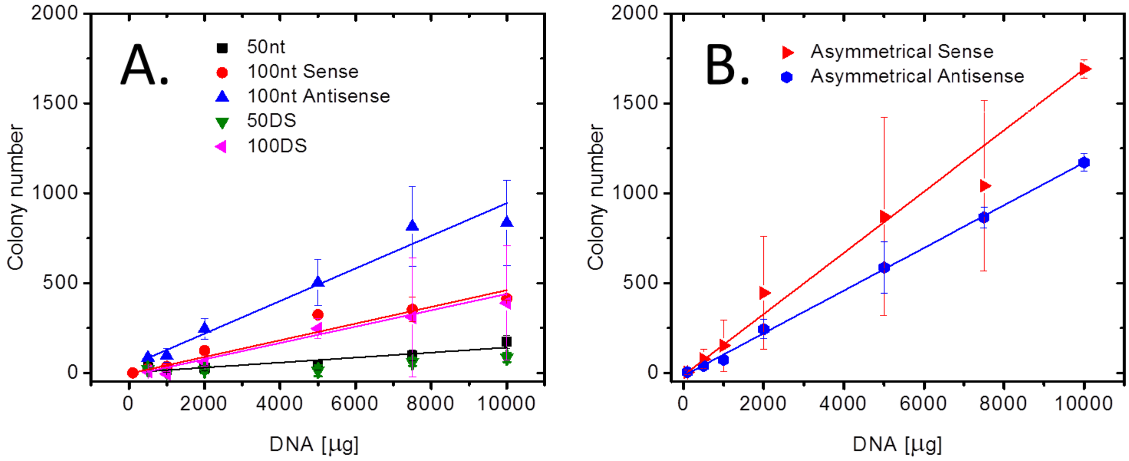


Figure 14 ODN Repair A) Repair of C18 *Hprt*$^-$ cells with 50nt ssODN, sense and antisense 100nt ssODN, ds50nt ODNs, and ds100nt ODNs. B) Repair of C18 *Hprt*$^-$ cells with sense and antisense asymmetric oligos.

4.3.3 Effect of asymmetrical oligos on repair efficiency

As others had noted an increase in repair efficiency when using asymmetric oligos,(146,151) it was interesting to see if this would also be the case in C18 cells. Asymmetrical ssODN with 36 bp on the PAM distal side and 91 bp on the PAM proximal side(151) were purchased and increasing concentrations were electroporated into 5x10^6 C18 cells with 15 µg C18 gRNA and 5 µg Cas9 expression plasmids. A linear increase in repair was observed with increasing amounts of ssODN. There was an increase in repair efficiency compared to symmetric ODNs (compare figure 14B to

14A). There was a slight difference between sense and antisense asymmetric ssODNs (figure 14B), though the sense symmetrical ODN repaired better whereas the antisense asymmetric oligo performed better. The increase in repair of sense vs antisense asymmetrical oligo is not significant at all data points in the plot, and may be due to experimental differences.

There are several advantages to targeting with ODN- the largest being affordable, commercial availability of ODNs that do not require vector assembly. However, there are several pitfalls that make ODNs impractical for gene therapy. Though optimization of ODN design may decrease some of these problems, the large amount of ODN required and resulting RPP will require more problem solving before they are a commonly used solution for gene therapy.

Additionally, ODNs can only correct small mutations. When 2 μg of 100 bp ssODN was used to correct the 4 bp mutation in C18 cells, and average of 244 colonies were obtained (figure 14A). However, when 2 μg of 100 bp ssODN were used to correct the 15 bp mutation in C15 cells, an average of 4 colonies were obtained (figure 13B). As HDR showed clear advantages over repair with short homologies, the next step was to examine the HDR donor requirements for more even complex targeting exercises.

4.4 Effect of homology arm length on CRISPR/Cas9 mediated deletion of large stretches of DNA

4.4.1 Insertion of a cassette via CRISPR/Cas9 mediated HDR

While gene knock-outs can be achieved with a single amino acid change, some experiments might require more extensive deletions or substitutions of DNA. In order to observe the effect of nucleases on a large deletion of DNA, a PGK-neo-pA (1528 bp) cassette was inserted into exon 6 of *Hprt* (figure 15A). The insertion of this cassette deleted 43 bp of *Hprt* to ensure a KO phenotype of *Hprt*. The same gRNA (6-1) that was used to cleave exon 6 of WT R1 DNA (section 4.2.1) was used to increase targeting of the cassette. A circular PBR322-amp-HPRT-PGK-neo vector with 10 kb total homology, and 15 μg RNA 6-1 and 5 μg Cas9 expression plasmids were electroporated into 5x10^6 R1 cells. After 24 hours, G418 selection was applied to the cells. After 7 days, 6-TG was applied to the cells for 4 days and clones were picked and expanded.

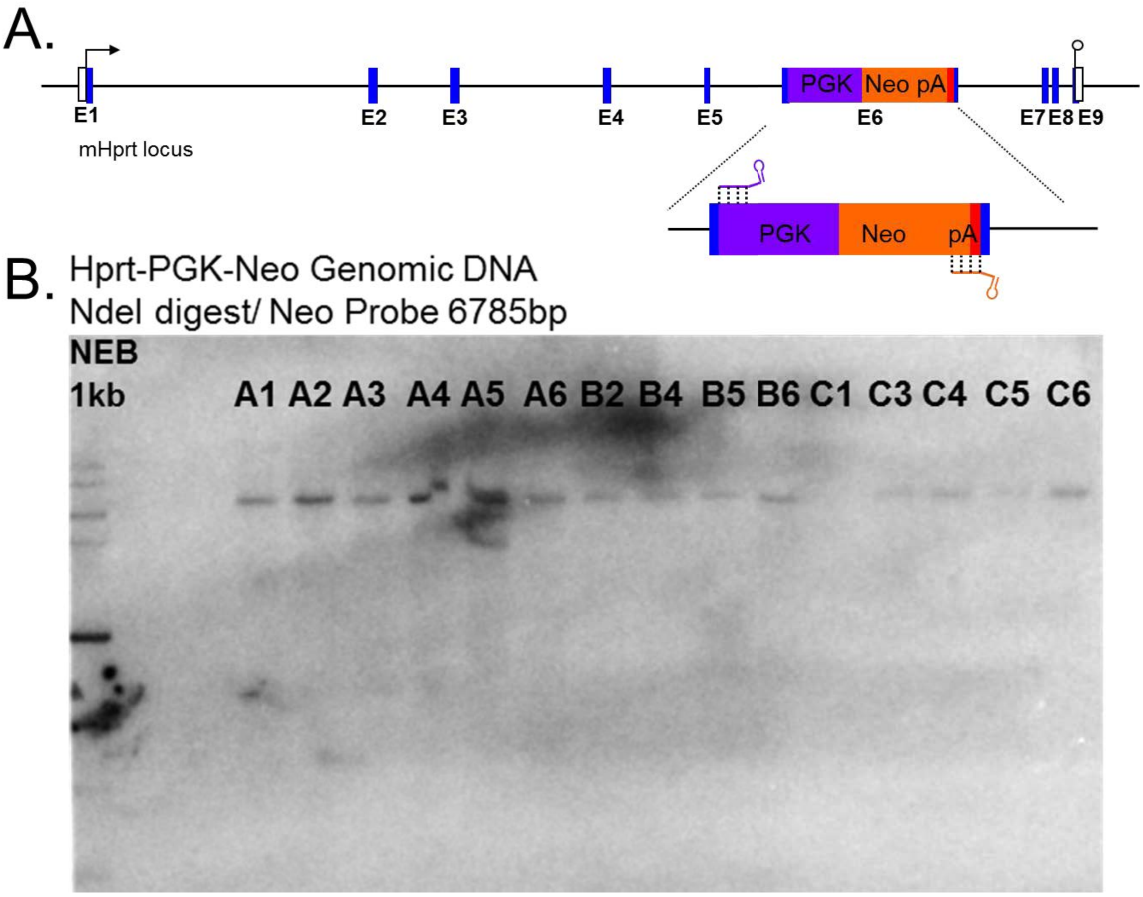

A.
E1
E2
E3
E4
E5
E6
E7
E8
E9
mHprt locus
PGK
Neo
pA
B.
Hprt-PGK-Neo Genomic DNA
Ndel digest/ Neo Probe 6785bp
NEB
1kb
A1 A2 A3 A4 A5 A6 B2 B4 B5 B6 C1 C3 C4 C5 C6

Figure 15 Insertion of a PGK-neo cassette into the *Hprt* locus A) Design of the PGK-neo cassette inserted into exon 6 of *Hprt* B) Southern assay to confirm single integration of the circular PGK-neo cassette into R1 cells. An internal neo probe was used on NdeI digested genomic DNA, producing a single band at 6785 bp in targeted cells.

Because the neomycin resistance gene has its own promoter, multiple or random integrations of the cassette into the genome were possible. A southern assay was performed with an internal neo probe to ensure single integration of the PGK-neo cassette. All clones contained a single integration of the cassette, with the anticipated band size (figure 15B).

4.4.2 Phenotypic confirmation of *Hprt* KO

From these clones, A1 was chosen for further experiments. The phenotype of clone A1 was confirmed to be G418 resistant, HAT sensitive and 6-TG resistant (figure 16). R1 was used as a wt control, and is G418 sensitive, 6-TG sensitive, and HAT resistant. The PGK-neo cassette is 1528 bp, and gRNAs were designed to target the

PGK (5’ end of the cassette) and the neo/ poly A (3’ end of the cassette). The phenotype of these cells confirms they are suitable for rescue experiments for repair of the *Hprt* gene.

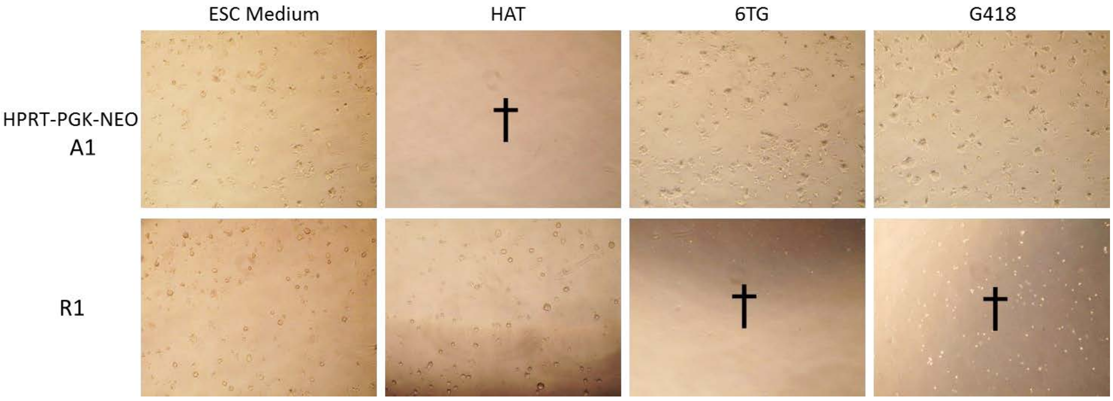


Figure 16 Phenotypic confirmation of HPRT-PGK-neo clones HPRT-PGK-neo A1 is G418 and 6-TG resistant and HAT sensitive, and can therefore be used in rescue experiments for repair of the *Hprt* gene.

4.4.3 Effect of gRNA position on deletion of a PGK neo cassette

The Hprt-PGK-neo targeted cells allow examination of several interesting questions pertaining to deletion of larger sizes of DNA. While there was a steep increase in targeting efficiency with up to 2 kb arms in C18 cells (figure 13 C-F) this was not the case for C15 cells, which have a larger mutation (figure 13B). In C15 cells, there was a sharp increase up to 4 kb, which shows better efficiency with longer HAs. HPRT-PGK-neo cells allow for quantification of the effect of HA length on deletion of a 1.5 kb PGK-neo cassette. Additionally, the length of the cassette allows for examination of the utility of using multiple gRNAs vs a single gRNA (figure 15A).

The efficiency of the gRNAs were tested with the 4 kb circular repair vector previously used (section 4.2.4), by electroporating $5x10^6$ HPRT-PGK-neo A1 cells with 15 µg gRNA expression plasmid, 5 µg Cas9 expression plasmid and the 4 kb total homology repair vector. Each electroporation was repeated at least twice.

When HPRT-PGK-neo cells were electroporated with the expression plasmids for the gRNA and Cas9 without any repair vector, there were no colonies obtained from the control (gRNA+ Cas9 alone) (figure 17A). For simple deletions, a strategy that uses ligation of the DSBs formed by two gRNAs is possible (section 4.6), however in this experiment, when two gRNAs were used together to excise the PGK-neo cassette, no colonies were obtained, as the 43 bp that were deleted upon insertion of the cassette is too large to be repaired by the cell and form a functional HPRT protein.

An average of 355 colonies were obtained with neo-gRNA-5 that recognizes the neo (3') end of the cassette (figure 17A). Two gRNAs for the PGK (5') end of the cassette were designed and tested, as Pgk-gRNA1 had significantly lower repair efficiencies (41 colonies). The second gRNA designed to recognize the 5' of the cassette (Pgk-gRNA 8) exhibited a much better cleavage efficiency and produced an average of 484 colonies (figure 17A). The difference in gRNA cleavage efficiency has been previously observed, and may be the result of the different melting temperatures of the gRNA sequence, or the chromatin environment(8,33,59) though other unidentified factors are likely to also play.

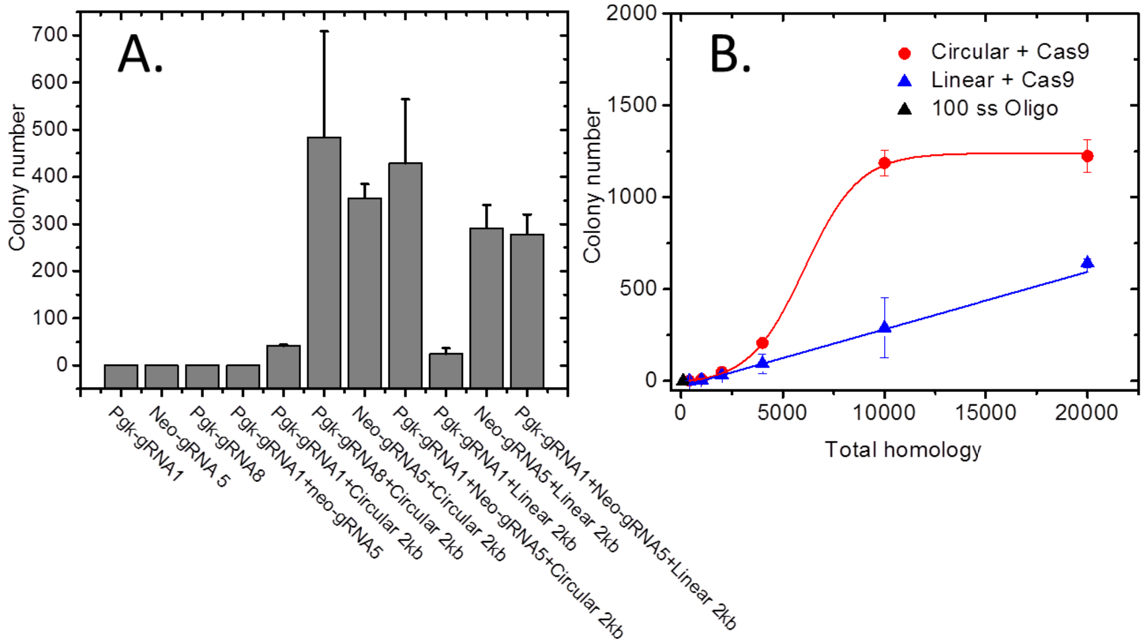


Figure 17 Deletion of a 1.5 kb stretch of DNA A. Test of single and double gRNA efficiency with a 4 kb circular HDR vector B. Deletion of a 1.5 kb PGK-neo cassette by HDR using linear and circular repair constructs with varying HA lengths.

Two gRNAs co-electroporated with the circular 4 kb repair vector to remove the PGK-neo cassette was not more efficient than using a single gRNA at either end of the cassette (figure 17A). The efficiency of a single gRNA over two gRNAs has also previously been observed at two genes in human iPSCs. (33)

4.4.4 Effect of homology arm length on deletion of a PGK-neo cassette

As no increase in targeting efficiency was observed with two gRNAs, further experiments were completed using the range of repair vectors previously used (section 4.2.3) and a single gRNA that recognized the poly A of the neo gene. A 100 bp oligo was not able to remove the cassette, even at high concentrations (2 µg). In fact, a high efficiency of cassette deletion was not observed until 10 kb total homology was used (figure 17B). This is in contrast to the 4 nt repair, which saw a steep increase up to 2 kb total homology, which is even more severe than the shift to 4 kb observed with C15 cells (compare figure 17B to 13B). Additionally, there was an overall decrease in efficiency (comparing repaired colony number) between deletion of a 1.5 kb cassette and a 4 nt repair. While some of this difference may be due to the difference in gRNA efficiency, as a gRNA that efficiently cleaved genomic DNA was used, this decrease in targeting efficiency is likely due to the size of the deletion.

Once again, an increase in targeting efficiency with longer homology arms was observed, and, there was a difference in the HDR efficiency of circular vs linear constructs. The circular constructs were much more efficient at removing the cassette and still showed a non-linear increase in efficiency as homology length increased. The linear constructs showed a more linear increase and a much lower overall efficiency compared to circular constructs.

Our results have shown the utility of HDR for small repair (section 4.2) and larger deletions (section 4.4). Other common gene targeting experiments require insertion of large stretches of DNA. The DSB that are induced by designer nucleases offers two interesting opportunities for inserting DNA: both by HDR and NHEJ. The next section of this work examines the efficiency of DNA insertion via these two DNA repair pathways.

4.5 CRIPSR/Cas9 permits NHEJ mediated deletions and insertions

4.5.1 Introduction

Several experiments with ZFNs and TALENs utilized the overhangs produced by the FokI nuclease to repair DSB with foreign DNA. Orlando et al.(6) utilized double stranded oligonucleotides with overhangs that matched the overhang produced by a ZFN to integrate the oligo into the genome. This process utilizes the microhomology between the oligo and the overhang produced by the ZFN to insert the foreign DNA via NHEJ. While overhangs that were homologous exhibited a higher level of insertion efficiency, sequence that was flanked by non homologous overhangs was also inserted. Perfect ligation events occurred more often with protected oligos, though non protected oligos were also perfectly inserted, though less efficiently. In a follow-up paper, Cristea et al.(156) utilized both TALENs and ZFNs to insert larger (~3 kb) transgenes utilizing NHEJ capture. In this set of experiments, the authors used a transgene donor with the same TALEN or ZFN recognition sites as the genomic recognition sites. The authors postulated that concurrent transgene cleavage *in vivo* using the same nuclease that cuts the genomic target site would increase targeting efficiency by decreasing transgene degradation. Indeed, the transgene was inserted only when simultaneously cut by the nuclease. The insertions were almost never perfect (small indels at either the 5’ or 3’ junction, most likely caused by reformation of the cleavage site and sequential cleavage by the nuclease) and sometimes resulted in multiple integrations of the transgene, either at additional genomic locations or in tandem at the target site. When the authors pre-linearized the transgene, they noted less efficient targeting efficiency due to exonucleic degradation in the cell.

In a similar set of experiments, Maresca et al.(157) also utilized ZFN and TALEN induced overhangs to integrate large stretches of DNA into the genome. Called ObLiGaRe, this method solves the problem of sequential recognition site reconstitution and recleavage by alternating the orientation of the recognition sequences, thus locking the ligation product in a palindrome of half recognition sequences. This method sometimes resulted in perfect ligation sequences, but also resulted in indels and multiple, sequential integrations of the transgene.

Nakade et al.(158) reported a system called PITCh which utilized three gRNA sequences and two different target recognition sites on the donor vector to integrate the donor DNA. While this proved the possibility of perfect ligation products from Cas9 ligation, the system was unnecessarily complex.

Bachu et al.(159) examined Knock in Blunt-end Ligation (KiBL) of ~5 kb in HEK 293 and CHO K1 cells. They observed that circular constructs integrated better than linear, and that cell type had an impact on both KIBL efficiency (0.17% in HEK293 cells and 0.45% in CHO cells, with selection), as well as the fidelity of KIBL (HEK293 cells showed indels in 32/32 sequenced ligation events while CHO KI showed perfect ligation in 4/4 clones at the HPRT locus and a mixture of perfect ligations and indels at another location in the genome).(159)

Geisinger et al.(160) also observed that deletions from two gRNAs depends on the gRNA itself (different gRNAs have different frequencies), and the orientations of the PAMs, but concluded that the majority of deletions were precisely repaired at several loci in the genome and with varying deletion sizes. They then used KiBL with PCR amplicons (without homology arms) to replace the sequence between two gRNAs precisely in both an immortalized human cell line and JF10 hiPSCs. Their findings indicated that the blunt ends generated by Cas9 cleavage can be used to knock in PCR cassettes in a homology independent manner, though with less efficiency than HDR repair (up to 4 percent without selection and up to 22 % with selection). Precise cassette junction integration only occurred in JF10 cells, and rarely occurred in HEK293 cells (as seen by Bachu et al.(159)).

Cas9 makes a blunt end DSBs that are substrates for error free repair by c-NHEJ, thus reconstituting the gRNA recognition site that can be re-cleaved by Cas9. This cycle of cleavage, repair, and re-cleavage is broken only when the exonuclease processing of c-NHEJ or alt-NHEJ prevent cleavage by the nuclease. The heterogenous nature of the indels that are generated, combined with the fact that not all indels produce knock-out phenotypes, means that screening and confirmation is required for downstream applications.

One drawback of KiBL is the theoretical caveat that only 50% of integrations will have the desired orientation. The *Hprt* experimental design used in this work excludes this possibility, as the *Hprt* gene function must be restored by a correctly integrated cassette. However, a screening process must be designed to screen for this possibility if the integration event is not selectable.

The rescue *Hprt* assay gives the perfect opportunity to test the efficiency of Cas9 induced blunt end ligation for insertion of DNA and to compare this to the efficiency of insertion of DNA via HDR.

4.5.2 Precise deletion of 672 bp from the Hprt gene

In this experiment 672 bp of genomic HPRT DNA in R1 cells, including exon 6, were deleted by electroporating two gRNA expression plasmids (Hprt 67 and Hprt 76) and a Cas9 expression plasmid into $5\text{x}10^6$ R1 mESC and plating them on gelatinized 10 cm dishes in a low density in ES medium. After seven days, the media was supplemented 6-TG selection to isolate *Hprt*$^-$ cells. The gRNAs, and therefore the NHEJ ligation product, are located within introns (figure 18). Several colonies were expanded and sequenced. Δ672 clones 6 and 10 were precisely repaired, with an exact 672 bp deletion. These clones were 6-TG resistant and HAT sensitive, thus making them suitable for selectable *Hprt* experiments.

4.5.3 Design of a KiBL vector to test precise NHEJ mediated DNA insertion

A new gRNA that identified the junction of the repaired cells was designed. A repair construct that consisted of the deleted 672 bp flanked by the same gRNA as the junction of the Δ672 cells was made using subcloning. The gRNA was introduced via the recombineering oligos in either the same direction (SD) as the genomic gRNA recognition site or in the opposite direction (OD) as the genomic gRNA recognition site (figure 18). When the gRNAs are in the same direction, the gRNA recognition site will be reconstituted by perfect ligation, and can be continually cleaved and ligated until indels form and stop the gRNA recognition. When the gRNA site is in the opposite direction, the gRNA will cleave both the genomic DNA and the vector, but the ligation product between the vector and the genomic DNA will no longer be recognized by the gRNA. Additionally, a PvuI site was inserted between the gRNA and the Δ672 bp in the ligation vector. This allowed us to linearize the construct *in* *vitro* and compare ligation of the 672 bp vs ligation of the *in vivo* linearized ligation vector.

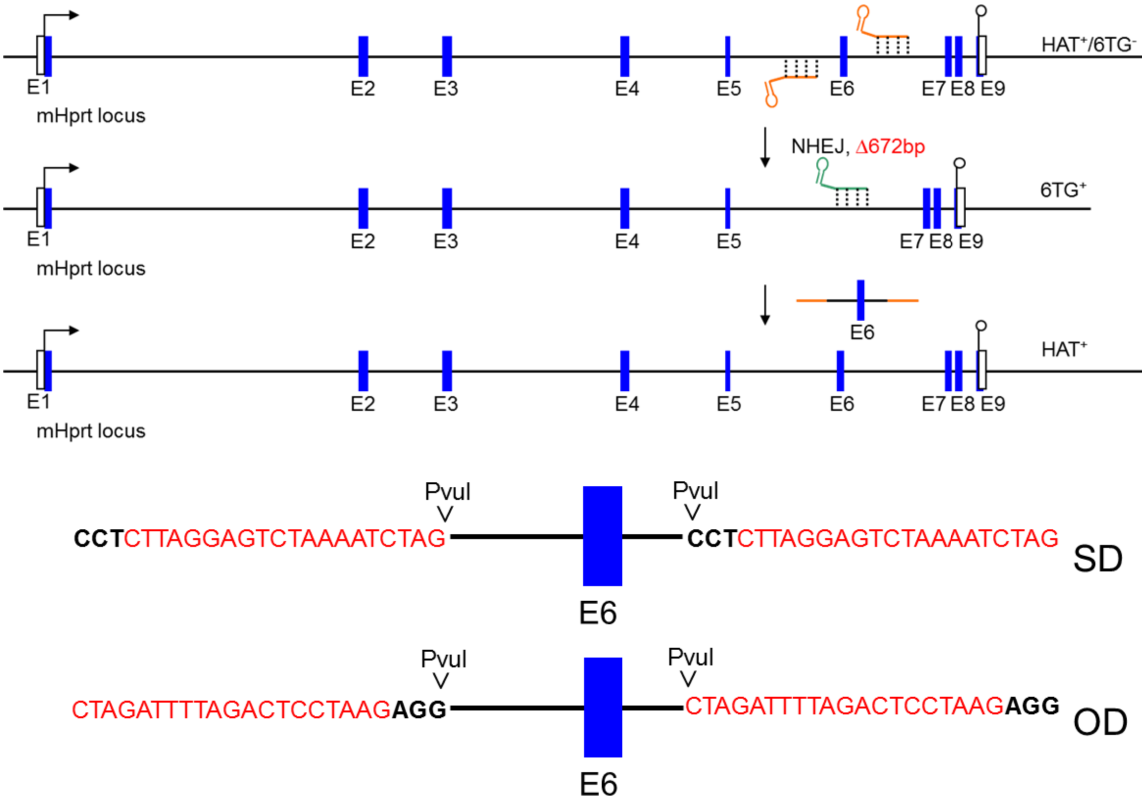

HAT+/6TG-
E1
E2
E3
E4
E5
E6
E7 E8 E9
mHprt locus
NHEJ, Δ672bp
6TG+
E1
E2
E3
E4
E5
E7 E8 E9
mHprt locus
E6
HAT+
E1
E2
E3
E4
E5
E6
E7 E8 E9
mHprt locus
PvuI
PvuI
CCTCTTAGGAGTCTAAAATCTAG
CCTCTTAGGAGTCTAAAATCTAG
SD
E6
PvuI
PvuI
CTAGATTTTAGACTCCTAAGAGG
CTAGATTTTAGACTCCTAAGAGG
OD
E6

Figure 18. Design of an experiment to test ligation efficiency Two gRNAs (gRNA 67 and gRNA76) were designed that recognize sequence flanking exon 6. Expression plasmids for these two gRNAs and Cas9 were electroporated in R1 cells. NHEJ was allowed to occur and *Hprt*$^-$ cells were selected for with 6-TG. Perfect ligation (with no indels) occurred in two colonies. A new gRNA (67/76 gRNA 4) was designed that recognized the unique junction caused by NHEJ. This gRNA sequence was added to a repair vector that replaces the 672 deleted bp in the same direction (SD) and in the opposite direction (OD) as the gRNA site in the genome of Δ672 C6 and C10 cells.

4.5.4 NHEJ mediated ligation precisely inserts DNA

Following the previous protocols, 5×10^6 Δ672 C10 and C6 cells were electroporated with 15 µg of the 67/76 gRNA 4 expression vector and 5 µg of the Cas9 expression vector. When C6 and C10 were electroporated with expression plasmids for the gRNA and Cas9 (with no donor template), there were no colonies obtained, as the Δ672 (which removed all of exon 6) cannot be repaired by the cell without a template (figure 19A). Next, *in vitro* linearized vector and circular vectors SD and OD that would be linearized *in vivo* were electroporated into the cells along with expression plasmids for Cas9 and the gRNA. The *in vitro* linearized vector integrated with a very low efficiency (average of 5 colonies) while the *in vivo* linearized vector integrated much more efficiently. OD versions of the construct were anticipated to have a higher ligation efficiency, as the SD gRNA would be reconstituted and re-cleaved, allowing for deletions or inversions of the intervening DNA. However, this was not the case, and the SD *in vivo* ligation construct had higher levels of ligation (figure 19 A). The efficiency of *in vitro* linearized integration is quite low, and though *in vivo* linearized ligation is much higher, it is still less efficient than HDR (Compare 19A to 13C-F).

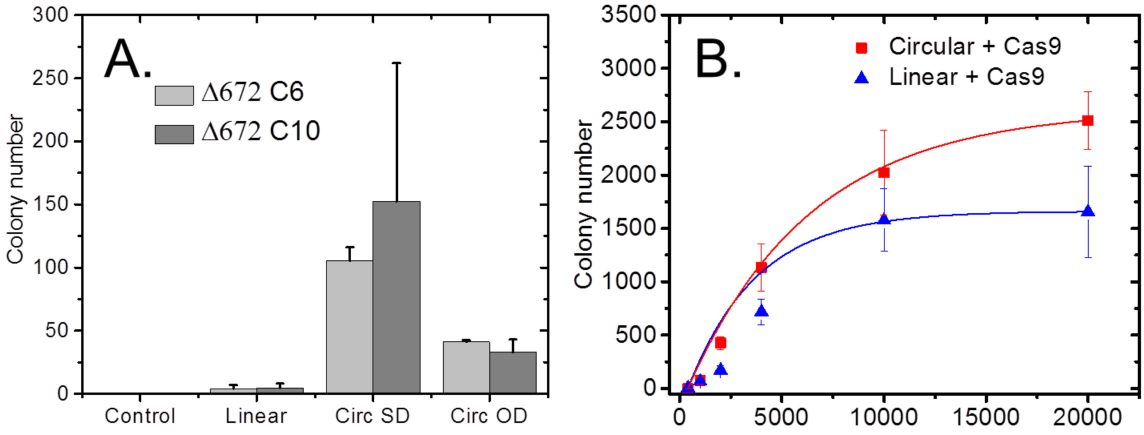


Figure 19. **Ligation and insertion efficiency** A) Insertion of Δ672 with *in vivo* and *in vitro* linearized repair constructs. The *in vivo* circularized vectors have the gRNA flanking the insert in both the same direction (SD) and opposite direction (OD) as the genomic gRNA recognition site. B) Repair of Δ672 C10 cells with a series of HDR vectors (section 4.2.3) with varying HA length.

4.5.5 Repair of Δ672 with HDR vector templates

Another common gene editing exercise is insertion of DNA sequence. In order to evaluate the efficiency of insertion of a deleted stretch of 672 bp, the same HDR repair vectors with varying arm lengths (section 4.2.3) were used. 5x10^6 Δ672 C10 were electroporated with equimolar amounts of the HDR vectors, 15 µg Hprt 67/76 gRNA 4 and 5 µg Cas9 expression vector. The cells were plated on gelatinized 10 cm culture dishes, and after 24 hours, the medium was supplemented with 1x HAT to select for *Hprt*$^+$ cells. After 9 days the colonies were fixed, stained, and counted. There were no positive colonies for 400 bp circular repair construct, as the HA do not extend past the deleted section. There were an average of 5 colonies obtained for the linear 400 bp constructs. This comes from ligation of the linearized cassette, and occurs with a partial deletion of the genomic region, as the deleted 672 bp is larger than the 400 bp HDR vector.

All of the HDR vectors have a decreased overall HA length of 672 bp. As such, especially for the smaller HDR constructs, a lower level of gene repair was expected. Indeed, this was the case with a slight decrease in the number of corrected colonies compared to HDR of the 4 bp mutation in C18 cells (compare figure 19B to 13C). Once again, here there was in increase in targeting efficiency with HA length and circular vectors repaired more efficiently than linear. There was a sharp increase in efficiency up to 4 kb, with a more gradual increase after that. Even with the longer HA, there is a slight decrease in targeting efficiency compared to repair of a 4 bp mutation (compare figure 19B to 13C). However, insertion of a 672 bp section of DNA was more efficient that deletion of a 1.5 kb fragment of DNA (compare Figure 19B to 17B).

This work has so far clarified several requirements for optimization of HDR vectors for several different types of gene engineering exercises with designer nucleases. One additional factor that can be easily examined with the selectable *Hprt* experiment is the requirement for isogenic DNA. In the next section, the requirements for long stretches of isogenic DNA when using designer nucleases in HDR are examined.

4.6 Effect of Engineered Mutations in Homology Arms on Targeting efficiency with CRISPR/Cas9 mediated HDR

4.6.1 Introduction

The use of isogenic DNA, or rather DNA that lacks significant mismatches greatly impacts targeting efficiency. Letsou and Liskay showed that gene conversion was sensitive to a single mismatch within 1 kb,(161) and others showed that the use of isogenic DNA increases targeting frequency 10-20 fold at the retinoblastoma locus(162) and 25 fold at the creatine kinase m gene.(163) At the *Hprt* locus, a 4-5x increase in targeting efficiency was found when using isogenic DNA,(131) though the number and location of the mismatches were not identified, this data indicates a requirement of a minimal stretch of DNA with perfect homology to the target locus. Reports using BACs as targeting constructs indicated that as homology arm length increases, the requirement for isogenicity decreases,(164) and as repair of DSB requires less homology, determining the requirements for isogenicity when using CRISPR/Cas9 would be valuable for optimizing HA length.

4.6.2 Differences between isogenic R1 DNA and RP23 BAC library at *Hprt* exon 6

For previous experiments (section 4.2.1), a 4 nt mutation in C18 *Hprt*$^-$ R1 cells were repaired with vectors that were subcloned from a commercially purchased BAC. The RP23-13N1 BAC was derived from pooled tissues from three C57BL/6J female mice. R1 cells, in which these experiments are performed, have a 129 genetic background. It was unclear if these two genetic backgrounds had identical sequences at the *Hprt* locus- and if there were mutations, if they effected targeting efficiency when using CRISPR/Cas9. In order to better study this, genomic DNA from R1 cells was extracted, digested with EcoRV, and used to subclone the 2 kb total homology vector. The sequence of the HA of the isogenic vector and the BAC vector were compared. A single C>G SNP was found, however, when the targeting efficiency of an equimolar amount of the isogenic construct was compared to the efficiency of the BAC derived vector with the same HAs, there was no significant difference in targeting efficiency found (figure 21A/B).

4.6.3 Effect of engineered mutations in homology arms on targeting efficiency

As a single SNP is quite small compared to the 2 kb HA, it was postulated that perhaps a larger mutation would have a greater effect. Mutations were engineered within the homology arms to determine the effect of non-isogenic DNA on targeting efficiency when using CRISPR/Cas9. In this experiment, silent mutations were first inserted into exon6 of *Hprt* using linear plus linear recombineering (section 3.34.8). These two mutations maintain the amino acid sequence of the gene while inserting a BssSI restriction endonuclease recognition site into the genome to allow for screening of the cells (figure 20). Then 5 bp were deleted from the 3' homology arm where a single cutting BsmI restriction endonuclease site allowed easy deletion of 5 bp via linear plus linear recombination. This location was verified to lack a splice acceptor or branching location.

As before, 5x10^6 C18 cells were electroporated with 5 µg Cas9 expression vector, 15 µg C18 gRNA expression vector and the same molar amount as the 2 kb total HA constructs so the targeting efficiency could be compared. There was no decrease in targeting efficiency observed with the 5 bp deletion. The lack of decrease in targeting efficiency was surprising, but it was postulated that as the mutation was only 65 bp away from the 4 nt mutation in exon 6 (figure 20), the repair machinery treated the mutations as a single lesion. This led to the idea that mutations at different locations in the HA would have an increased/decreased effect on targeting efficiency, as it would decrease the length of homology that could be used during homology search. Therefore, a new design was established with engineered mutations along the 5' homology arm that would put the lesions at varying distances from exon 6 (figure 20).

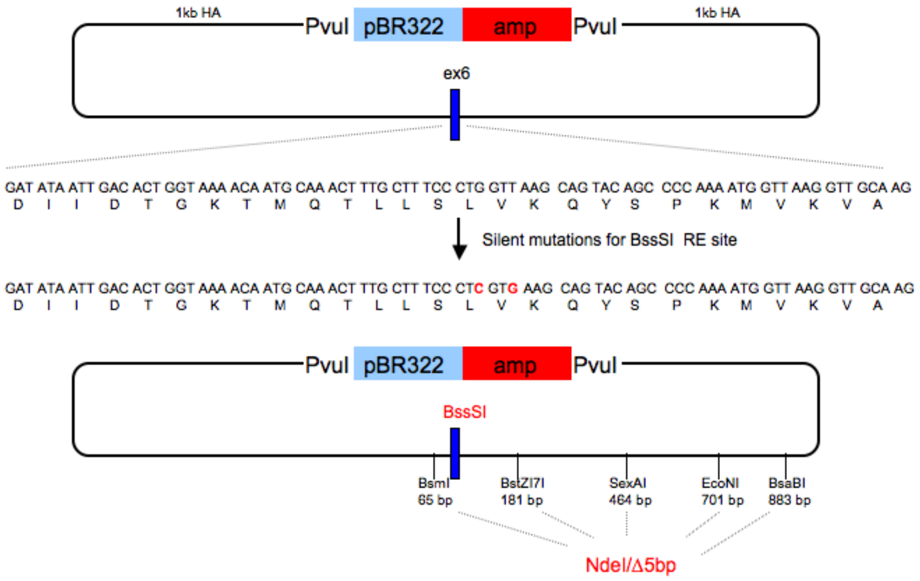

1kb HA
PvuI
pBR322
amp
PvuI
1kb HA
ex6
GAT ATA ATT GAC ACT GGT AAA ACA ATG CAA ACT TTG CTT TCC CTG GTT AAG CAG TAC AGC CCC AAA ATG GTT AAG GTT GCA AG
D I I D T G K T M Q T L L S L V K Q Y S P K M V K V A
Silent mutations for BssSI RE site
GAT ATA ATT GAC ACT GGT AAA ACA ATG CAA ACT TTG CTT TCC CTC GTG AAG CAG TAC AGC CCC AAA ATG GTT AAG GTT GCA AG
D I I D T G K T M Q T L L S L V K Q Y S P K M V K V A
PvuI
pBR322
amp
PvuI
BssSI
BsmI 65 bp
BstZ17I 181 bp
SexAI 464 bp
EcoNI 701 bp
BsaBI 883 bp
NdeI/Δ5bp

Figure 20 Engineered mutations in a homology arm Two silent mutations were introduced into exon 6 to make a BssSI restriction site. Single cutting restriction sites were identified in the homology arms of the 2 kb HDR construct. At each of these locations, either 5 bp were deleted, or a 6 bp NdeI site was inserted. This interrupts the homology at varying distances from exon 6.

Four additional single cutting restriction endonuclease sites were identified along the 5' homology arm. Linear plus linear recombineering was used to either delete 5 bp or insert a 6 bp NdeI restriction endonuclease recognition site at each of these locations (figure 20). The 5 bp deletion or NdeI insertions were confirmed by restriction endonuclease digestion and sequencing of the vector. 5x10^6 C18 cells were electroporated with 5 µg Cas9 expression vector, 15 µg C18 gRNA expression vector and the same molar amount as the 2 kb total HA constructs so the targeting efficiency of vectors with engineered mutations could be evaluated. Surprisingly, no significant increase or decrease in targeting efficiency was observed with any of the mutations, no matter the location on the HA (figure 21). This indicates that the homology requirements for designer nuclease assisted repair are less than that provided in these constructs.

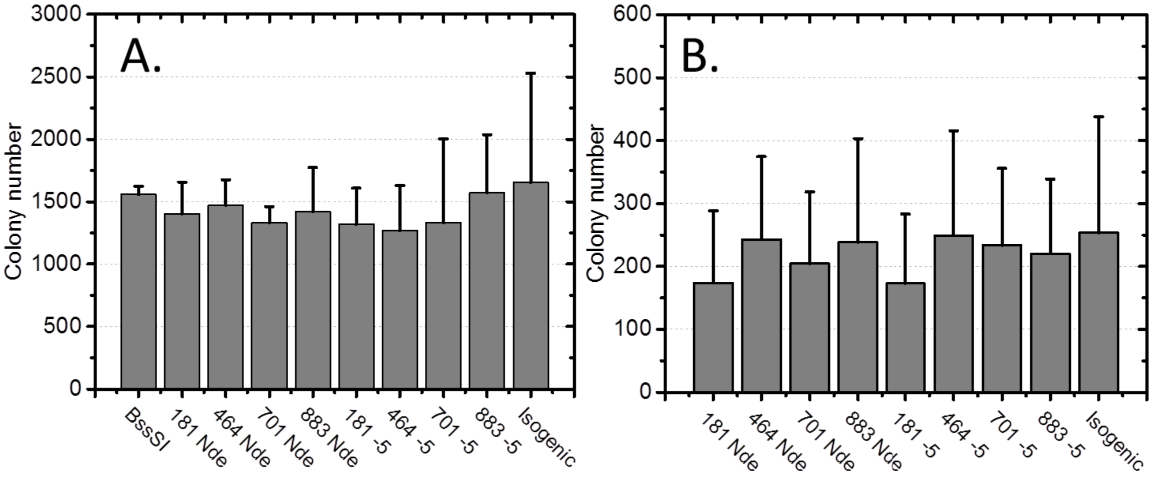


Figure 21 Isogenicity does not impact targeting efficiency when using CRISPR/Cas9 Engineered lesions along the 3' HA of a 2k total homology construct do not impact targeting efficiency, no matter their location on the homology arm. Neither a 6 bp insertion nor a 5 bp deletion have a significant impact.

Summary

The results so far of this work have established that while arms as short as 400 bp total HA are sufficient for HDR, for small mutations, there is an advantage to using longer HA, as HA up to 2 kb total homology drastically increase targeting efficiency. When larger, more complex exercises are required, even longer HA, up to 10 kb total homology drastically increase targeting efficiency. For both small and large exercises, increased homology beyond 4 kb provides further, smaller increases in targeting efficiency. Circular constructs perform better than linear constructs in all HR targeting exercises, and there is no need to clone isogenic DNA from the genome. With the optimized TC parameters in mind, humanization of the *Scn10a* gene was attempted again.

4.7 Humanization of the mouse Scn10a gene using CRISPR/Cas9

4.7.1 Targeting strategy for the *Scn10a* gene

Previous works have examined the use of two nuclease recognition sites to evaluate if dsDNA breaks at both ends of the gene replacement improved targeting efficiency.(33) In this experiment they replaced a 2.7 kb *THY1* human gene with its mouse counterpart in iPCS without selection. When using two nuclease recognition sites, they observed gene excision and inversions in up to 20.5 % of the population, while single or double gene replacements occurred in only 2 % or 8 % of the population respectively. While the use of antibiotic screening can ensure targeting of one allele, the other allele must be screened for deletions. Additionally, depending on the design of the experiment, the cells must be screened for inversions. As the authors found that targeting with a single gRNA produced single gene replacements in up to 15 % of the population, and double targeting in up to 11 % of the cell population, they recommended using a single gRNA targeting strategy.(33) However, several publications have shown that deletion frequency declines with increasing deletion size,(165-167) and Byrne et al.(33) noted that the efficiency of deletions relied more heavily on the efficiency of the specific gRNA pairs. Therefore, a strategy with two gRNAs that cleaved at the 5’ end and the 3’ end of *mScn10a* was utilized, as the gRNAs are 85 kb apart.

4.7.2 Loss of allele screen for humanization of the mouse genome

In order to quickly screen colonies for humanization of the *Scn10a* gene, a loss of allele (LOA) assay was utilized. qPCR primers were designed using the NCBI primer designer tool that amplified 70-100 bp of DNA from the mouse 5’ *Scn10a* gene body, the mouse 3’ gene body, the human 5’ *SCN10A* gene body or the human 3’ gene body. These primers were tested on genomic R1 DNA and genomic H7 DNA to ensure they were specific, both in band size and to the species they were designed for.

The fold change of expression was calculated relative to a wild type control using the ΔΔCt method.(168) In this method, the cycle threshold (Ct) values need to be normalized against the average Ct of an internal control. However, a control that worked for both human and mouse DNA was first needed.

Bejerano et al.(169) published a list of 481 different DNA segments that are more than 200 bp that have 100 % homology between human, mouse, and rat DNA. These areas of ultrahomology provided an excellent location to design qPCR primers that gave not just similar sized products, but completely identical PCR products that could be used as an internal control for both human and mouse genomic DNA. From the list provided, any regions from the X chromosome (1N) were excluded and actively transcribed regions that had ~50 % GC content were focused on. From this list, several genomic regions identified in which specific qPCR primers within the ultrahomologous regions could be designed using the NCBI primer design tool. Sequences located within the *HAT1* gene (copy number =2N) and the *MRRF* gene (copy number =2N) were chosen and qPCR primers were ordered for these regions. The primers were tested in triplicate with both H7 and R1 genomic DNA and the results were compared to a previously used internal control Nxt2 (34) (copy number =2N mouse). The primer pairs correctly recognized both mouse and human DNA and produced a single band. Using the ΔΔCt method, both HAT1 and MRRF primers identified 2N in R1 and H7 DNA. As expected, the Nxt2 primer pair only identified 2N in R1 DNA, and did not produce a product for H7 DNA.

4.7.3 Generation of humanized *SCN10A* mES cells

The TC carrying the human *SCN10A* gene is very large (137,763 bp). As shown earlier (section 4.2.4), very large constructs break up upon electroporation. Therefore, the TC was lipofected, along with the gRNA expression vectors for the 5' and 3' ends of the mouse *Scn10a* gene and Cas9 expression vectors into feederless R1 cells. After 24 hours, the medium was supplemented with BSD and three days later changed to HYG selection. Although several colonies were resistant to BSD and HYG, only one colony had a deleted mouse *Scn10a*. This clone (G2) however, also exhibited multiple integrations of the BAC throughout the genome (figure 22). This low efficiency can be explained by the low transfection efficiency of three different expression vectors and a TC into the same cell via lipofection.

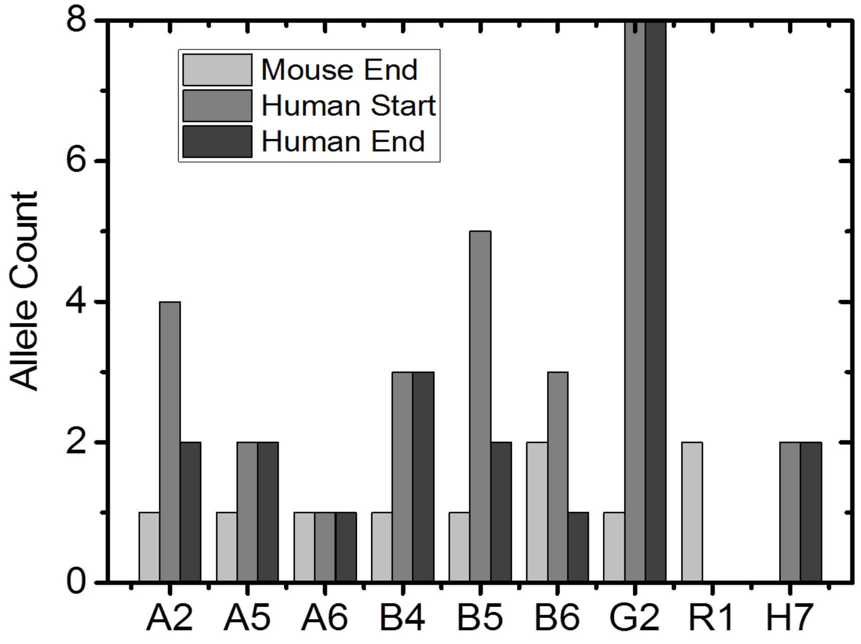


Figure 22 LOA Assay for Humanized *SCN10A* colonies Allele counts for R1 cells transfected with the *SCN10A* humanization construct and expression vectors for gRNA and Cas9. Primer pairs were designed for qPCR that amplify either the mouse or human genomic region.

As good results were previously obtained when electroporating R1-derived cells with multiple gRNA and Cas9 expression vectors (section 4.2), the protocol was modified. First, the SCN10A TC was prepared for lipofection, and applied to an empty 6 well culture dish. Next 1x10^6 feederless R1 cells were electroporated with 3 µg of each gRNA expression vector and 1 µg of Cas9 expression vector, using the standard electroporation protocol. After the standard 5 minute recovery, the cells were divided to 3 wells of the 6 well dish (~333.000 cells /well) that contained the TC treated with lipofectamine. After 24 hours, the medium was supplemented with BSD for 3 days, and then changed to medium supplemented with Hyg. After 10 days, colonies were picked, expanded and screened by qPCR for *mScn10a* deletion and *hSCN10A* insertion.

This protocol proved more successful, with 6 colonies showing a deletion of the *mScn10a*. However, only one colony, A6, showed a deletion of the *mScn10a* gene and a single insertion of the *hSCN10A* gene. The rest of the colonies have multiple integrations, and sometimes partial integrations, of the TC. Partial integrations could come from TC that are broken from DNA extraction, or they could be a result of crossovers between the high levels of homology between the mouse and human gene.

This could be avoided by multiple targeting steps (as in section 4.5), but multiple targeting steps can reduce germline transmission. (29-31)

Additionally, LOA allele screening in itself is not definitive test for correct gene targeting. These clones must be further screened for correct integration by FISH or southern analysis. However, this work shows that it is possible to complete very complex targeting exercises when using designer nucleases when using an optimized TC design.

5 Discussion

Designer nucleases are rapidly changing how gene targeting is approached. This has led to several questions about optimal target construct design when using designer nucleases. This work has helped to clarify the optimum targeting vector design for several different kinds of complex targeting exercises, and then utilized this optimal design to complete a very complex targeting exercise: humanization of the *SCN10A* gene in ES mouse cells.

5.1 Two TALENs exhibited a low cleavage efficiency

This work evaluated two different TALENs that recognized the *Scn10a* gene. These TALENs both exhibited undetectable levels of cleavage with the T7E1 assay (figure 7). Attempts to increase cleavage efficiency by optimizing the linker and using obligate heterodimeric TALEN expression vectors did not appreciably raise cleavage efficiency of the Scn10a TALENs (section 4.1.1.6). There are several possible reasons that TALEN cleavage was inefficient. Firstly, it is possible that the TALENs in this work were confounded by chromosomal position or some other unknown factor. The variability of TALEN binding efficiency has been noted in several different works.(48-51) It is possible that designing and testing additional TALENs would solve the problem.

However, the overall low efficiency of TALENs, especially in cultured ES cells has been noted in other works.(8) The low efficiency of TALENs in cultured cells may be due to the sensitivity of TALENs to cytosine methylation.(52,53) This sensitivity might be overcome by utilizing special TALEN assembly kits,(53) freshly extracted mESC,(170) or by injecting TALEN mRNA into blastocysts(127) (which have reduced methylation(171,172)). The relatively low efficiency of TALENs in cultured mESC, and their comparatively complex assembly led us to change strategies and utilize CRISPR/Cas9 to induce DSB.

5.2 Optimization of construct design for designer nuclease-assisted targeting

Nuclease assisted targeting to achieve complex genome engineering has several unresolved issues related to experimental design. DSB repair differs mechanistically from conventional targeting, as it is promoted by the ends of the linear DNA targeting construct. In order to clarify targeting vector requirements for nuclease assisted targeting, an experiment was designed that could easily test different design parameters and their effect on targeting efficiency when using a designer nuclease.

5.2.1 HPRT is a selectable assay for construct design with CRISPR/Cas9

HPRT is a selectable assay using medium supplemented with HAT, which permits the growth of *Hprt*$^+$ cells, but not *Hprt*$^-$ cells, and medium supplemented with 6-TG, which kills cells that are *Hprt*$^+$ and allows *Hprt*$^-$ cells to grow.(128) As *Hprt* is located on the X chromosome, it is monoallelic in male R1 mES cells and can be used to test the efficiency of targeting with different targeting constructs. R1 cells can be weaned from feeders, and grown on gelatinized plates to form colonies that are easily visible to the human eye. They can be fixed and stained to make counting easier, and *Hprt* rescue makes an excellent quantifiable experiment.

This work presented several systems in which *Hprt* was interrupted. First, cleavage by a single gRNA (section 4.2.1) resulted in a 4 bp mutation from NHEJ. This small deletion could be repaired by HDR and various ODNs. *Hprt* was also interrupted by two gRNAs, which deleted a 672 bp stretch of DNA. This deletion was repaired via HDR and blunt end ligation (section 4.5.4). Additionally, a PGK-neo cassette was inserted into *Hprt* and this cassette was subsequently removed via HDR (section 4.4.4). These mutated *Hprt* alleles and their quantifiable repair via ODN, HDR, NHEJ, and ligation indicate the versatility of the HPRT assay.

5.2.2 Longer HA increase targeting efficiency

Homologous recombination is a routinely used technique in gene targeting, however, it has a relatively low efficiency. One parameter that was clearly shown to improve targeting efficiency was using long stretches of isogenic DNA in the targeting vectors. While longer arms are better, the dependence on homology seemed to saturate somewhere between 6 and 14 kb,(131,133,135) with the optimal amount of homology likely depending on chromosomal landscape, isogenicity, or other, undefined parameters.

Recent advances in designer nuclease-assisted targeting have led to unanswered questions about the optimal length for HA to repair induced DSB. The selectable HPRT assay (section 4.2.1) provides a solution to quantifiably answer this question. In order to systematically study the effect of HA length on targeting efficiency, a series of repair vectors with varying HA length were cloned from the RP23-13N1 BAC. The homology arms varied in length from 200 bp each side (i.e. 400 bp total homology) to 10 kb each side (i.e. 20 kb total homology) (figure 12). Additional constructs that had a standard 1 kb arm and one arm with varying lengths were cloned. These repair vectors were transfected into the cell with and without expression plasmids for a gRNA and Cas9. In all cases, targeting increased dramatically when Cas9 was expressed. A maximum of 350-fold difference was observed when Cas9 was used.

When a small, 4 bp lesion in *Hprt* was repaired with the series of HDR vectors, a steep increase in targeting efficiency with homology arms up to 2 kb in length was observed, with a more gradual increase up to 20 kb (figure 13C). When a, larger, 15 bp lesion was corrected with the repair vectors, there was also a slight shift in the curve, with the sharp increase in efficiency up to the 4 kb HA length vs the 2 kb HA in C18 cells, which is especially noticeable in linear constructs (compare figure 13C to 13B). The overall decrease in targeting efficiency in C15 cells compared to C18 cells can be attributed to the efficiency of the gRNA and the size of the lesion.

Overall, deletion of a stretch of 1.5 kb DNA was less efficient than repair of a 4 or 15 bp mutation (section 4.5.5). High efficiency of cassette deletion was not observed until 10 kb total homology was used (figure 17B), which is much different than the repair of a 4 bp mutation, which had efficient targeting with 2 kb total homology (compare figure 17B to 15C). While some of the decrease in targeting efficiency may be due to differences in gRNA efficiency, the shift in the curve indicates that the decrease in targeting efficiency is likely due to the size of the deletion. While these results need to be repeated, perhaps with a different chromosomal location, these results indicate that deletions are much less efficient than small corrections or insertions.

When a deleted 672 bp stretch of DNA was reinserted by the same series of repair vectors used in section 4.2.1, there is a decrease in targeting efficiency, likely due do the decreased overall HA length (from the deleted 672 bp) (compare figure 19B to 13C). Once again, there is a shift in the curve, with an increase in targeting up to 4 kb, and a more gradual increase after that. Though insertion of 672 bp is less efficient than correction of a 4 bp mutation, it is comparatively more efficient than deletion of a 1.5 kb fragment of DNA (compare Figure 19B to 17A).

In none of these cases were longer homology arms found to have a negative effect on targeting efficiency. This indicates that while short arms can be used, there are benefits to having longer homology arms. This work has shown this to be especially true for different targeting exercises, with deletions requiring longer HAs, and insertions and correction of larger lesions exhibiting greater targeting efficiencies with 4 kb total HA. Smaller lesions could be efficiently repaired with 2 kb total HA as previously reported.(33) Therefore, there is not a single “optimal” homology arm length when it comes to nuclease assisted HDR. The optimal length depends on the experiment, and there are no drawbacks to having longer HAs.

Another potentially important observation is that all curves that measure the effect of homology arm length are bimodal (except for linear repair vectors for large deletions), showing a steep increase followed by further moderate increases with longer HA. This may be because stability of the synapse requires 1-2 kb homology on either side, and that longer targeting constructs increase the probability that targeting constructs will find their matching sequences in the genome. The idea that a minimal amount of homology is required for synapse stability is further supported by the results obtained from asymmetrical construct repair and repair constructs with engineered mutations, as targeting efficiency does not decrease when mutations are introduced to the homology arm (section 4.6).

5.2.3 Minimal homology arm length

Asymmetrical constructs with one homology arm shorter than 500 bp showed a minimal increase in targeting above NHEJ background (figure 13D). This observation is in accordance with results of targeting without a nuclease induced DSB.(132,133) Once the asymmetrical HA was increased to 500 bp, an average of 1097 colonies for circular and 415 colonies for linear constructs was obtained. Symmetrical constructs showed a similar pattern, with constructs that had 200 bp each side showing minimal increases over background NHEJ, and constructs with 500 bp each side (1kb total homology) showing a greater increase (figure 13D). Other works have identified that HA as small as 50-100 bp can repair a DSB caused by a nuclease, (5,6) and this work has not identified a true "minimal" homology arm requirement. In fact, though equimolar amounts of a 50 bp dsODN did not repair the 4 bp mutation in C18 cells, when the concentration of dsODN was increased, even this short oligo with 25 bp homology each side of the mutation was able to repair *Hprt* (figure 14A). It is likely that the true minimal HA length for nuclease assisted targeting varies with chromosomal landscape, targeting exercise, and nuclease efficiency. Indeed, nuclease assisted targeting can even use microhomologies(6,156-158) or no homology(159,160) to ligate in pieces of DNA (section 4.5). What is clear, however, is that smaller homologies are not optimal for any of the targeting exercises examined in this work, with longer HA always being more efficient.

5.2.4 HDR with circular vectors have higher efficiency than linear vectors

In all experiments that compared HDR with circular and linear vectors (vectors linearized with RE before transfection into the cell) the circular constructs have a better targeting efficiency. This is especially true for the more complex targeting exercises (deletion of 1.5 kb and insertion of a stretch of DNA, but is also noticeable in the correction of the smaller, 4 bp mutations). This may be due to increased transfection efficiency of supercoiled, circular DNA. However, it may also be a result of degradation of the linear pieces of DNA by nucleases. The differences in the shapes of the curves when comparing circular and linear HDR constructs indicates nuclease degradation may be the cause (figures 13,17,19).

5.2.5 ODN repair of DSB requires large amounts of DNA

There are several advantages to gene repair with ODN- the largest being the affordable, commercial availability of customized ODNs. However, there are several pitfalls that make ODNs impractical for gene therapy. Though optimization of ODN design may decrease some of these problems, further optimization will be required before they are a commonly used solution for gene therapy.

When an equimolar amount of ODN (calculated according to the HDR vectors in 4.2.4) was used to repair the 4 bp mutation in C18 cells, there were no colonies obtained above the background number of colonies from NHEJ, indicating that the reduced homology of 50 to 100 bp is not enough for efficient gene repair. Increasing the concentrations of ODNs increased repair efficiency in all cases (figure 14). Even at a vast molar excess (i.e. 10 μg of ODN), the efficiency achieved does not match the targeting construct with 1 kb homology arms. Though a point of cell death caused by excess ODN was not reached with this titration, this would eventually occur, as excess ODN would eventually cause a toxic RPP.(147,148)

There was no strand bias observed in this work. While there is evidence both for and against strand bias when using ODN to repair DSB, this work does not support strand bias related to the orientation of the ODN, and indicates that the ODN can be designed in either orientation with no impact on repair efficiency.

As others had noted an increase in repair efficiency when using asymmetric oligos,(146,151) the efficiency of repair of a 4 bp mutation by both asymmetric and symmetric oligos was compared (figure A and B). There was an increase in targeting efficiency observed when asymmetric oligos were used. While the reason asymmetric oligos are more efficient at gene targeting is not completely clear, it may be due to 5' resectioning,(146) gRNA titration of oligo, (153) or Cas9 release mechanisms.(151) As no strand bias was observed, the work supports a hypothesis of 5' resectioning.

This work did not screen for random, unintended ODN integration. ODN's have been reported to clog replication forks and cause replication fork collapse (143) and double strand breaks. This work and others (6,156-160) show that exogenous DNA can be ligated into DSB. In fact, this principle is so well known that it is one of the methods of testing for off target effects.(173) When using ODNs to make small changes to DNA, a strategy should be identified that can check for off target integrations of the ODN. As higher concentrations of ODN are more likely to stall replication forks and have unintended integrations, it is important to use the smallest concentration of ODN possible.

ODNs can only correct small mutations. When 2 µg of 100 bp ssODN was used to correct the 4 bp mutation in C18 cells, an average of 244 colonies were obtained (figure 14A). However, when 2 µg of 100 bp ssODN was used to correct the 15 bp mutation in C15 cells, an average of 4 colonies were obtained (section 4.2.4). An additional drawback of ODN repair is that it requires extensive screening for correction events. This is not the case for *Hprt* as this gene is selectable, but for other small mutations a screening strategy must be designed and many colonies must be screened for correct integration while excluding off target integrations. Though ODNs have some advantages (easily commercially available), their usefulness in gene repair is somewhat diminished by the consequences of ODNs (only correct small mutations, off target integrations, and RPP). In many cases, an easily assembled HDR vector would be a more effective solution.

5.2.6 RNP has advantages over expression vectors

Cas9 protein can either be produced by the cell when an expression plasmid is transfected into the cell, or delivered to the cell as an RNP package. The use of an RNP has several advantages over gRNA and Cas9 plasmid expression vectors: RNPs can be "cloning-free," they avoid the problem of expression plasmid DNA integration at Cas9 induced DSB,(134,138) RNP editing is very fast,(70) and RNP is quickly degraded by the cell which means that RNPs exhibit a lower level of off target mutagenesis.(139) In this work, the RNPs were more effective at HDR of DSB (section 4.2.5) than expression plasmids for gRNA and Cas9. The RNP required 8x less DNA (for 2.5x less cells) to achieve the same level of targeting compared to electroporation methods. This may be due to transfection efficiency of the RNP and repair vectors over the Cas9 and gRNA expression vectors and repair vectors. Less DNA however reduces the chances of off target integrations in the cell, making RNP delivery Cas9 an excellent method for inducing DSB.

However, utilizing RNPs comes with several downsides. RNPs are most often transfected into cells via nucleofection. Nucleofections are known to break up large constructs,(140) which was also observed in this work (figure 13E). As seen earlier (section 5.2.2), different targeting exercises have different optimal homologies, with some exercises (like deletions) requiring larger total HA lengths. This means that co-delivery of large HDR vectors and RNP is not optimal. RNPs can be lipofected into cells, however, the transfection efficiency is lower, and varies greatly between cell lines.(174) Additionally, the nucleofection kits for purchase (Lonza Mouse ES Cell Kit was used in this work), are expensive compared to standard electroporation.

This means that for experiments there must be a cost/benefit analysis that will be different for each design. RNP cleavage to promote HDR that uses small HA would be beneficial, as it has less off target effects and requires less vector DNA. However, larger HDR constructs that use larger HA would still require another mode of transfection (such as lipofection or electroporation) and therefore preclude the use of RNPs.

5.2.7 gRNAs have different efficiencies

The difference in gRNA cleavage efficiency has been previously observed, and may be the result of the different melting temperatures of the gRNA sequence, or the chromatin environment(8,33,59) though other unidentified factors likely also play a role. This work also observed a difference in gRNA efficiency (section 4.4.3), as two gRNAs in the same region had very different targeting efficiencies with a 4 kb HDR construct. This result suggests that it would be time saving to design and test several gRNAs simultaneously whenever possible, to increase the likelihood of choosing a gRNA with a good cleavage efficiency.

5.2.8 Blunt end ligation is inefficient compared to HDR

Several works have investigated the possibility of using NHEJ to ligate DNA into DSB.(6,156-160) As this work showed that with high concentrations of ODN, even 50 bp total homology was sufficient to correct a 4 bp mutation (section 4.3.2), the next step was to evaluate the efficiency of integration of constructs with no homology. Therefore, a new experiment at the *Hprt* locus was designed. A stretch of DNA was deleted by CRISPR/Cas9 mediated DSB induced NHEJ. A new gRNA sequence was identified at the NHEJ junction. The deleted stretch of DNA was inserted into a vector flanked by the new gRNA recognition sites and PvuI sites. The PvuI sites permit *in vitro* linearization of the construct while the gRNA sites permit *in vivo* linearization when Cas9 and the gRNA expression vector are also expressed.

The *in vitro* linearized vector integrated with a very low efficiency (average of 5 colonies) while the *in vivo* linearized vector integrated much more efficiently. The co-cutting of donor vector and target sequence has been noted before. (156,157) The gRNAs were inserted into the donor vector either in the same direction (SD) or opposite direction (OD) as the target vector. OD versions of the construct were anticipated to have a higher ligation efficiency, as the SD gRNA would be reconstituted and re-cleaved, allowing for deletions or inversions of the intervening DNA. However, this was not the case, and the SD *in vivo* ligation construct had higher levels of ligation (figure 19A). Though the gRNA sequences are recognized by the same gRNA/Cas9 complex, the Cas9 binds to opposite strands of the DNA. Others have noted a strand bias in ODN repair due to the strand that Cas9 binds(154), and this may also be the case for blunt end ligation.

When a series of HDR vectors with varying arm lengths was used to re-insert the excised DNA, the efficiency of insertion was much higher than *in vivo* linearized ligation (compare figure 19A to 19B). It was therefore concluded that HDR is a more efficient means of inserting DNA than blunt end ligation.

5.2.9 Long stretches of isogenic DNA no longer necessary

The use of isogenic DNA, or rather DNA that lacks significant mismatches greatly impacts targeting efficiency. Letsou and Liskay showed that gene conversion was sensitive to a single mismatch within 1 kb,(161) and others showed that the use of isogenic DNA increases targeting frequency up to 25 fold.(131,162,163) To determine the effect of mismatches within a homology arm on nuclease assisted targeting efficiency, an experiment was designed with engineered mutations along the 5' homology arm that would put lesions at varying distances from *Hprt* exon 6 (figure 20). Surprisingly, no significant increase or decrease in targeting efficiency was observed with any of the mutations, no matter the location on the HA (figure 21).

Other works have identified that conversion tracts when a DSB is repaired is often less than 58 bp,(175) and in cases where the conversion tracts are longer, extensive end resectioning of the DSB has occurred. Further screening of the colonies obtained in this experiment can show if this work also finds such short conversion tracts. It is clear, however that along with reduced HA requirements, the requirement for perfect HA is also no longer necessary with nuclease induced DSB.

5.2.10 Summary of optimization work

This work has so far clarified several requirements for optimized HDR vectors for several different types of gene engineering exercises with designer nucleases. The results show that while arms as short as 400 bp total HA are sufficient for HDR, for small mutations, there is an advantage to using longer HA, as HA up to 2 kb total homology drastically increase targeting efficiency. When larger, more complex exercises are required, even longer HA, up to 10 kb total homology drastically increase targeting efficiency. For both insertions and small mutations, increased homology beyond 4 kb provides further increases in targeting efficiency and for deletions homology greater than 10 kb total homology shows the best efficiency. Circular constructs perform better than linear constructs in all HR targeting exercises, and there is no need to clone DNA from the genome, DNA with minimal differences from a BAC library is sufficient. ODN repair and NHEJ ligation can incorporate DNA, but are significantly less efficient than HDR.

5.3 Optimized targeting constructs make complex targeting exercises possible

With the clarified requirements for HDR vectors, a very ambitious targeting exercise was attempted: humanization of the *SCN10A* gene. As circular vectors have higher targeting efficiency in all targeting exercise in this work, a circular construct was utilized. The construct was lipofected, and not nucleofected as nucleofection breaks up large constructs. The long HA as designed in the original construct were not reduced, as long HA improve targeting efficiency. The long HA made screening for targeting via southern assay difficult, so a LOA assay was designed to test for integration. In order to utilize the ΔΔCt method method(168) (in which the fold change of expression is calculated relative to a wild type control), an internal control for both human and mouse needed to be designed to normalize the Ct values. From a list of ultrahomologous DNA segments,(169) qPCR primers were designed that gave not just similar sized products, but completely identical PCR products that could be used as an internal control for both human and mouse genomic DNA.

By electroporating the gRNA and Cas9 expression vectors into R1 cells, and plating them on lipofectamine treated TC DNA, one colony, was obtained that showed a deletion of the *mScn10a* gene and a single insertion of the *hSCN10A* gene. The rest of the colonies had multiple or partial integrations, of the TC. Partial integrations could come from fractured TC, or they could be a result of crossovers between the high levels of homology between the mouse and human gene. While these clones must be further screened for correct integration by FISH or southern (as LOA allele screening in itself is not definitive test for correct gene targeting), this work shows that it is possible to complete very complex nuclease assisted targeting exercises when using an optimized TC design.

6 Conclusions and Outlook

Nuclease assisted targeting has drastically changed the outlook of gene therapy over the last decade. Designer nucleases have reduced the requirements for targeting vectors and have opened the doors for more complex targeting exercises such as simultaneous biallelic gene targeting or large regional replacements. Nuclease assisted targeting is mechanistically different from conventional targeting, and thus the requirements for targeted repair are different. The aim of this work was to clarify optimal construct design for designer nuclease assisted targeting and compare efficiency of HDR to repair with constructs that utilize little or no homology.

While conventional targeting requires linear constructs with long homology arms containing long stretches of isogenic DNA, these requirements are all diminished when using designer nuclease assisted targeting. This work has compared homology arm requirements of several gene engineering exercises: correction of a small mutation, insertion of DNA, and deletion of DNA. The results show that circular constructs, with long HA have the best targeting efficiency. While repair via ligation and ODNs are possible, it is comparatively inefficient, and has the potential for off target integrations.

Future work will focus on confirming off target integrations of ssODNs. While it is postulated that ODNs can be integrated off target at DSB or at junctions via SDSA, this is yet to be confirmed. This is a vital piece of the story, and will greatly impact the usefulness of ODN for gene therapy. More data points on the deletion curve will also clarify the curve shape. There is no current explanation as to why deletion of DNA with linear constructs is the only curve that is not bimodal. Repeating this experiment, perhaps at another locus will verify that this is not an experimental inaccuracy. Finally, sequencing of the repair events with engineered mutations will confirm that while longer arms stabilize the synapse, the actual gene conversion tract is much shorter.

7 References

1. Skarnes, W.C., Rosen, B., West, A.P., Koutsourakis, M., Bushell, W., Iyer, V., Mujica, A.O., Thomas, M., Harrow, J., Cox, T. *et al.* (2011) A conditional knockout resource for the genome-wide study of mouse gene function. *Nature*, **474**, 337-342.
2. Capecchi, M.R. (2005) Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nature reviews*, **6**, 507-512.
3. Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. and Gregory, P.D. (2010) Genome editing with engineered zinc finger nucleases. *Nature reviews*, **11**, 636-646.
4. Handel, E.M. and Cathomen, T. (2011) Zinc-finger nuclease based genome surgery: it's all about specificity. *Current gene therapy*, **11**, 28-37.
5. Phillips, J.E. and Calos, M.P. (1999) Effects of homology length and donor vector arrangement on the efficiency of double-strand break-mediated recombination in human cells. *Somatic cell and molecular genetics*, **25**, 91-100.
6. Orlando, S.J., Santiago, Y., DeKelver, R.C., Freyvert, Y., Boydston, E.A., Moehle, E.A., Choi, V.M., Gopalan, S.M., Lou, J.F., Li, J. *et al.* (2010) Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. *Nucleic acids research*, **38**, e152.
7. Radecke, S., Radecke, F., Peter, I. and Schwarz, K. (2006) Physical incorporation of a single-stranded oligodeoxynucleotide during targeted repair of a human chromosomal locus. *The journal of gene medicine*, **8**, 217-228.
8. Yang, L., Guell, M., Byrne, S., Yang, J.L., De Los Angeles, A., Mali, P., Aach, J., Kim-Kiselak, C., Briggs, A.W., Rios, X. *et al.* (2013) Optimization of scarless human stem cell genome editing. *Nucleic acids research*, **41**, 9049-9061.
9. Akopian, A.N., Souslova, V., England, S., Okuse, K., Ogata, N., Ure, J., Smith, A., Kerr, B.J., McMahon, S.B., Boyce, S. *et al.* (1999) The tetrodotoxin-resistant sodium channel SNS has a specialized function in pain pathways. *Nature neuroscience*, **2**, 541-548.
10. Delaney, J.T., Muhammad, R., Shi, Y., Schildcrout, J.S., Blair, M., Short, L., Roden, D.M. and Darbar, D. (2013) Common SCN10A variants modulate PR interval and heart rate response during atrial fibrillation. *Europace*, **16**, 485-490.
11. Jabbari, J., Olesen, M.S., Yuan, L., Nielsen, J.B., Liang, B., Macri, V., Christophersen, I.E., Nielsen, N., Sajadieh, A., Ellinor, P.T. *et al.* (2015) Common and rare variants in SCN10A modulate the risk of atrial fibrillation. *Circulation*, **8**, 64-73.
12. Abou Ziki, M.D., Seidelmann, S.B., Smith, E., Atteya, G., Jiang, Y., Fernandes, R.G., Marieb, M.A., Akar, J.G. and Mani, A. (2017) Deleterious protein-altering mutations in the SCN10A voltage-gated sodium channel gene are associated with prolonged QT. *Clinical genetics*.
13. Hu, D., Barajas-Martinez, H., Pfeiffer, R., Dezi, F., Pfeiffer, J., Buch, T., Betzenhauser, M.J., Belardinelli, L., Kahlig, K.M., Rajamani, S. *et al.* (2014) Mutations in SCN10A are responsible for a large fraction of cases of Brugada syndrome. *Journal of the American College of Cardiology*, **64**, 66-79.

14. Faber, C.G., Lauria, G., Merkies, I.S., Cheng, X., Han, C., Ahn, H.S., Persson, A.K., Hoeijmakers, J.G., Gerrits, M.M., Pierro, T. *et al.* (2012) Gain-of-function Nav1.8 mutations in painful neuropathy. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 19444-19449.
15. Bagal, S.K., Bungay, P.J., Denton, S.M., Gibson, K.R., Glossop, M.S., Hay, T.L., Kemp, M.I., Lane, C.A., Lewis, M.L., Maw, G.N. *et al.* (2015) Discovery and Optimization of Selective Nav1.8 Modulator Series That Demonstrate Efficacy in Preclinical Models of Pain. *ACS medicinal chemistry letters*, **6**, 650-654.
16. Bagal, S.K., Chapman, M.L., Marron, B.E., Prime, R., Storer, R.I. and Swain, N.A. (2014) Recent progress in sodium channel modulators for pain. *Bioorganic & medicinal chemistry letters*, **24**, 3690-3699.
17. Chen, X., Yu, L., Shi, S., Jiang, H., Huang, C., Desai, M., Li, Y., Barajas-Martinez, H. and Hu, D. (2016) Neuronal Nav1.8 Channels as a Novel Therapeutic Target of Acute Atrial Fibrillation Prevention. *Journal of the American Heart Association*, **5**.
18. Amisten, S., Atanes, P., Hawkes, R., Ruz-Maldonado, I., Liu, B., Parandeh, F., Zhao, M., Huang, G.C., Salehi, A. and Persaud, S.J. (2017) A comparative analysis of human and mouse islet G-protein coupled receptor expression. *Scientific reports*, **7**, 46600.
19. Souslova, V.A., Fox, M., Wood, J.N. and Akopian, A.N. (1997) Cloning and characterization of a mouse sensory neuron tetrodotoxin-resistant voltage-gated sodium channel gene, Scn10a. *Genomics*, **41**, 201-209.
20. Stuhmer, W., Conti, F., Suzuki, H., Wang, X.D., Noda, M., Yahagi, N., Kubo, H. and Numa, S. (1989) Structural parts involved in activation and inactivation of the sodium channel. *Nature*, **339**, 597-603.
21. Yang, N., George, A.L., Jr. and Horn, R. (1996) Molecular basis of charge movement in voltage-gated sodium channels. *Neuron*, **16**, 113-122.
22. Payandeh, J., Scheuer, T., Zheng, N. and Catterall, W.A. (2011) The crystal structure of a voltage-gated sodium channel. *Nature*, **475**, 353-358.
23. Kleinjan, D.A. and van Heyningen, V. (2005) Long-range control of gene expression: emerging mechanisms and disruption in disease. *American journal of human genetics*, **76**, 8-32.
24. Alami, R., Greally, J.M., Tanimoto, K., Hwang, S., Feng, Y.Q., Engel, J.D., Fiering, S. and Bouhassira, E.E. (2000) Beta-globin YAC transgenes exhibit uniform expression levels but position effect variegation in mice. *Human molecular genetics*, **9**, 631-636.
25. Kaufman, R.M., Pham, C.T. and Ley, T.J. (1999) Transgenic analysis of a 100-kb human beta-globin cluster-containing DNA fragment propagated as a bacterial artificial chromosome. *Blood*, **94**, 3178-3184.
26. Peterson, K.R., Navas, P.A., Li, Q. and Stamatoyannopoulos, G. (1998) LCR-dependent gene expression in beta-globin YAC transgenics: detailed structural studies validate functional analysis even in the presence of fragmented YACs. *Human molecular genetics*, **7**, 2079-2088.
27. Anguita, E., Sharpe, J.A., Sloane-Stanley, J.A., Tufarelli, C., Higgs, D.R. and Wood, W.G. (2002) Deletion of the mouse alpha-globin regulatory element (HS -26) has an unexpectedly mild phenotype. *Blood*, **100**, 3450-3456.
28. Garrick, M.D., Kuo, H.C., Vargas, F., Singleton, S., Zhao, L., Smith, J.J., Paradkar, P., Roth, J.A. and Garrick, L.M. (2006) Comparison of mammalian

cell lines expressing distinct isoforms of divalent metal transporter 1 in a tetracycline-regulated fashion. *The Biochemical journal*, **398**, 539-546.

29. Baer, A. and Bode, J. (2001) Coping with kinetic and thermodynamic barriers: RMCE, an efficient strategy for the targeted integration of transgenes. *Current opinion in biotechnology*, **12**, 473-480.
30. Wallace, H.A., Marques-Kranc, F., Richardson, M., Luna-Crespo, F., Sharpe, J.A., Hughes, J., Wood, W.G., Higgs, D.R. and Smith, A.J. (2007) Manipulating the mouse genome to engineer precise functional syntenic replacements with human sequence. *Cell*, **128**, 197-209.
31. Yu, Y. and Bradley, A. (2001) Engineering chromosomal rearrangements in mice. *Nature reviews*, **2**, 780-790.
32. Testa, G., Zhang, Y., Vintersten, K., Benes, V., Pijnappel, W.W., Chambers, I., Smith, A.J., Smith, A.G. and Stewart, A.F. (2003) Engineering the mouse genome with bacterial artificial chromosomes to create multipurpose alleles. *Nature biotechnology*, **21**, 443-447.
33. Byrne, S.M., Ortiz, L., Mali, P., Aach, J. and Church, G.M. (2015) Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. *Nucleic acids research*, **43**, e21.
34. Baker, O., Tsurkan, S., Fu, J., Klink, B., Rump, A., Obst, M., Kranz, A., Schrock, E., Anastassiadis, K. and Stewart, A.F. (2017) The contribution of homology arms to nuclease-assisted genome engineering. *Nucleic acids research*, **45**, 8105-8115.
35. Beerli, R.R., Segal, D.J., Dreier, B. and Barbas, C.F., 3rd. (1998) Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 14628-14633.
36. Beerli, R.R. and Barbas, C.F., 3rd. (2002) Engineering polydactyl zinc-finger transcription factors. *Nature biotechnology*, **20**, 135-141.
37. Liu, Q., Segal, D.J., Ghiara, J.B. and Barbas, C.F., 3rd. (1997) Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 5525-5530.
38. Bhakta, M.S., Henry, I.M., Ousterout, D.G., Das, K.T., Lockwood, S.H., Meckler, J.F., Wallen, M.C., Zykovich, A., Yu, Y., Leo, H. *et al.* (2013) Highly active zinc-finger nucleases by extended modular assembly. *Genome research*, **23**, 530-538.
39. Pingoud, A. and Jeltsch, A. (2001) Structure and function of type II restriction endonucleases. *Nucleic acids research*, **29**, 3705-3727.
40. Bogdanove, A.J., Schornack, S. and Lahaye, T. (2010) TAL effectors: finding plant genes for disease and defense. *Current opinion in plant biology*, **13**, 394-401.
41. Boch, J. and Bonas, U. (2010) Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annual review of phytopathology*, **48**, 419-436.
42. Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509-1512.
43. Moscou, M.J. and Bogdanove, A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 1501.

44. Mak, A.N., Bradley, P., Cernadas, R.A., Bogdanove, A.J. and Stoddard, B.L. (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science*, **335**, 716-719.
45. Lamb, B.M., Mercer, A.C. and Barbas, C.F., 3rd. (2013) Directed evolution of the TALE N-terminal domain for recognition of all 5' bases. *Nucleic acids research*.
46. White, F.F. and Yang, B. (2009) Host and pathogen factors controlling the rice-Xanthomonas oryzae interaction. *Plant physiology*, **150**, 1677-1686.
47. Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.K., Shi, Y. and Yan, N. (2012) Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science*, **335**, 720-723.
48. Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D. and Joung, J.K. (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nature biotechnology*, **30**, 460-465.
49. Cade, L., Reyon, D., Hwang, W.Y., Tsai, S.Q., Patel, S., Khayter, C., Joung, J.K., Sander, J.D., Peterson, R.T. and Yeh, J.R. (2012) Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. *Nucleic acids research*, **40**, 8001-8010.
50. Moore, F.E., Reyon, D., Sander, J.D., Martinez, S.A., Blackburn, J.S., Khayter, C., Ramirez, C.L., Joung, J.K. and Langenau, D.M. (2012) Improved somatic mutagenesis in zebrafish using transcription activator-like effector nucleases (TALENs). *PloS one*, **7**, e37877.
51. Wood, A.J., Lo, T.W., Zeitler, B., Pickle, C.S., Ralston, E.J., Lee, A.H., Amora, R., Miller, J.C., Leung, E., Meng, X. *et al.* (2011) Targeted genome editing across species using ZFNs and TALENs. *Science*, **333**, 307.
52. Bultmann, S., Morbitzer, R., Schmidt, C.S., Thanisch, K., Spada, F., Elsaesser, J., Lahaye, T. and Leonhardt, H. (2012) Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic acids research*, **40**, 5368-5377.
53. Valton, J., Dupuy, A., Daboussi, F., Thomas, S., Marechal, A., Macmaster, R., Melliand, K., Juillerat, A. and Duchateau, P. (2012) Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. *The Journal of biological chemistry*, **287**, 38427-38432.
54. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709-1712.
55. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816-821.
56. Gasiunas, G., Barrangou, R., Horvath, P. and Siksnys, V. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, E2579-2586.
57. Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L. and Church, G.M. (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature biotechnology*, **31**, 833-838.
58. Chen, H., Choi, J. and Bailey, S. (2014) Cut site selection by the two nuclease domains of the Cas9 RNA-guided endonuclease. *The Journal of biological chemistry*, **289**, 13284-13294.

59. Wang, T., Wei, J.J., Sabatini, D.M. and Lander, E.S. (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, **343**, 80-84.
60. Guha, T.K., Wai, A. and Hausner, G. (2017) Programmable Genome Editing Tools and their Regulation for Efficient Genome Engineering. *Computational and structural biotechnology journal*, **15**, 146-160.
61. Bitinaite, J., Wah, D.A., Aggarwal, A.K. and Schildkraut, I. (1998) FokI dimerization is required for DNA cleavage. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 10570-10575.
62. Kim, H.J., Lee, H.J., Kim, H., Cho, S.W. and Kim, J.S. (2009) Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome research*, **19**, 1279-1288.
63. Mussolino, C., Morbitzer, R., Lutge, F., Dannemann, N., Lahaye, T. and Cathomen, T. (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic acids research*, **39**, 9283-9293.
64. Kim, Y., Kweon, J., Kim, A., Chon, J.K., Yoo, J.Y., Kim, H.J., Kim, S., Lee, C., Jeong, E., Chung, E. *et al.* (2013) A library of TAL effector nucleases spanning the human genome. *Nature biotechnology*, **31**, 251-258.
65. 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. *et al.* (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology*, **31**, 827-832.
66. Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K. and Sander, J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology*, **31**, 822-826.
67. Yusa, K., Rashid, S.T., Strick-Marchand, H., Varela, I., Liu, P.Q., Paschon, D.E., Miranda, E., Ordonez, A., Hannan, N.R., Rouhani, F.J. *et al.* (2011) Targeted gene correction of alpha1-antitrypsin deficiency in induced pluripotent stem cells. *Nature*, **478**, 391-394.
68. Smith, C., Gore, A., Yan, W., Abalde-Atristain, L., Li, Z., He, C., Wang, Y., Brodsky, R.A., Zhang, K., Cheng, L. *et al.* (2014) Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell stem cell*, **15**, 12-13.
69. Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y. *et al.* (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, **154**, 1380-1389.
70. Kim, S., Kim, D., Cho, S.W., Kim, J. and Kim, J.S. (2014) Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome research*, **24**, 1012-1019.
71. Ramakrishna, S., Kwaku Dad, A.B., Beloor, J., Gopalappa, R., Lee, S.K. and Kim, H. (2014) Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. *Genome research*, **24**, 1020-1027.
72. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. and Adli, M. (2014) Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature biotechnology*, **32**, 677-683.
73. Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A. and Liu, D.R. (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature biotechnology*, **31**, 839-843.

74. Symington, L.S. and Gautier, J. (2011) Double-strand break end resection and repair pathway choice. *Annual review of genetics*, **45**, 247-271.
75. Difilippantonio, M.J., Zhu, J., Chen, H.T., Meffre, E., Nussenzweig, M.C., Max, E.E., Ried, T. and Nussenzweig, A. (2000) DNA repair protein Ku80 suppresses chromosomal aberrations and malignant transformation. *Nature*, **404**, 510-514.
76. Deriano, L. and Roth, D.B. (2013) Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. *Annual review of genetics*, **47**, 433-455.
77. Sturzenegger, A., Burdova, K., Kanagaraj, R., Levikova, M., Pinto, C., Cejka, P. and Janscak, P. (2014) DNA2 cooperates with the WRN and BLM RecQ helicases to mediate long-range DNA end resection in human cells. *The Journal of biological chemistry*, **289**, 27314-27326.
78. Wang, X. and Haber, J.E. (2004) Role of Saccharomyces single-stranded DNA-binding protein RPA in the strand invasion step of double-strand break repair. *PLoS biology*, **2**, E21.
79. Deng, S.K., Gibb, B., de Almeida, M.J., Greene, E.C. and Symington, L.S. (2014) RPA antagonizes microhomology-mediated repair of DNA double-strand breaks. *Nature structural & molecular biology*, **21**, 405-412.
80. Ceccaldi, R., Rondinelli, B. and D'Andrea, A.D. (2016) Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends in cell biology*, **26**, 52-64.
81. Heyer, W.D., Ehmsen, K.T. and Liu, J. (2010) Regulation of homologous recombination in eukaryotes. *Annual review of genetics*, **44**, 113-139.
82. Chiruvella, K.K., Liang, Z. and Wilson, T.E. (2013) Repair of double-strand breaks by end joining. *Cold Spring Harbor perspectives in biology*, **5**, a012757.
83. Karanam, K., Kafri, R., Loewer, A. and Lahav, G. (2012) Quantitative live cell imaging reveals a gradual shift between DNA repair mechanisms and a maximal use of HR in mid S phase. *Molecular cell*, **47**, 320-329.
84. Shibata, A. (2017) Regulation of repair pathway choice at two-ended DNA double-strand breaks. *Mutation research*, **803-805**, 51-55.
85. Murphy, K.C. (1998) Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli. *Journal of bacteriology*, **180**, 2063-2071.
86. Zhang, Y., Buchholz, F., Muyrers, J.P. and Stewart, A.F. (1998) A new logic for DNA engineering using recombination in Escherichia coli. *Nature genetics*, **20**, 123-128.
87. Copeland, N.G., Jenkins, N.A. and Court, D.L. (2001) Recombineering: a powerful new tool for mouse functional genomics. *Nature reviews*, **2**, 769-779.
88. Court, D.L., Sawitzke, J.A. and Thomason, L.C. (2002) Genetic engineering using homologous recombination. *Annual review of genetics*, **36**, 361-388.
89. Cohen, S.N., Chang, A.C., Boyer, H.W. and Helling, R.B. (1973) Construction of biologically functional bacterial plasmids in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, **70**, 3240-3244.
90. Fu, J., Bian, X., Hu, S., Wang, H., Huang, F., Seibert, P.M., Plaza, A., Xia, L., Muller, R., Stewart, A.F. *et al.* (2012) Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nature biotechnology*, **30**, 440-446.

91. Maresca, M., Erler, A., Fu, J., Friedrich, A., Zhang, Y. and Stewart, A.F. (2010) Single-stranded heteroduplex intermediates in lambda Red homologous recombination. *BMC molecular biology*, **11**, 54.
92. Ander, M., Subramaniam, S., Fahmy, K., Stewart, A.F. and Schaffer, E. (2015) A Single-Strand Annealing Protein Clamps DNA to Detect and Secure Homology. *PLoS biology*, **13**, e1002213.
93. Guzman, L.M., Belin, D., Carson, M.J. and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of bacteriology*, **177**, 4121-4130.
94. Muyrers, J.P., Zhang, Y., Testa, G. and Stewart, A.F. (1999) Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic acids research*, **27**, 1555-1557.
95. Zhang, Y., Muyrers, J.P., Testa, G. and Stewart, A.F. (2000) DNA cloning by homologous recombination in Escherichia coli. *Nature biotechnology*, **18**, 1314-1317.
96. Ellis, H.M., Yu, D., DiTizio, T. and Court, D.L. (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 6742-6746.
97. Swaminathan, S., Ellis, H.M., Waters, L.S., Yu, D., Lee, E.C., Court, D.L. and Sharan, S.K. (2001) Rapid engineering of bacterial artificial chromosomes using oligonucleotides. *Genesis*, **29**, 14-21.
98. Zhang, Y., Muyrers, J.P., Rientjes, J. and Stewart, A.F. (2003) Phage annealing proteins promote oligonucleotide-directed mutagenesis in Escherichia coli and mouse ES cells. *BMC molecular biology*, **4**, 1.
99. Poser, I., Sarov, M., Hutchins, J.R., Heriche, J.K., Toyoda, Y., Pozniakovsky, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A.W. *et al.* (2008) BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nature methods*, **5**, 409-415.
100. Sarov, M., Schneider, S., Pozniakovski, A., Roguev, A., Ernst, S., Zhang, Y., Hyman, A.A. and Stewart, A.F. (2006) A recombineering pipeline for functional genomics applied to Caenorhabditis elegans. *Nature methods*, **3**, 839-844.
101. Fu, J., Teucher, M., Anastassiadis, K., Skarnes, W. and Stewart, A.F. (2010) A recombineering pipeline to make conditional targeting constructs. *Methods in enzymology*, **477**, 125-144.
102. Evans, M.J. and Kaufman, M.H. (1981) Establishment in culture of pluripotential cells from mouse embryos. *Nature*, **292**, 154-156.
103. Martin, G.R. (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, **78**, 7634-7638.
104. Thomas, K.R. and Capecchi, M.R. (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell*, **51**, 503-512.
105. Geurts, A.M., Cost, G.J., Freyvert, Y., Zeitler, B., Miller, J.C., Choi, V.M., Jenkins, S.S., Wood, A., Cui, X., Meng, X. *et al.* (2009) Knockout rats via embryo microinjection of zinc-finger nucleases. *Science*, **325**, 433.
106. Lombardo, A., Genovese, P., Beausejour, C.M., Colleoni, S., Lee, Y.L., Kim, K.A., Ando, D., Urnov, F.D., Galli, C., Gregory, P.D. *et al.* (2007) Gene

editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nature biotechnology*, **25**, 1298-1306.

107. Keller, G. (2005) Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes & development*, **19**, 1129-1155.

108. Stavridis, M.P. and Smith, A.G. (2003) Neural differentiation of mouse embryonic stem cells. *Biochemical Society transactions*, **31**, 45-49.

109. Boheler, K.R., Czyz, J., Tweedie, D., Yang, H.T., Anisimov, S.V. and Wobus, A.M. (2002) Differentiation of pluripotent embryonic stem cells into cardiomyocytes. *Circulation research*, **91**, 189-201.

110. Wei, H., Juhasz, O., Li, J., Tarasova, Y.S. and Boheler, K.R. (2005) Embryonic stem cells and cardiomyocyte differentiation: phenotypic and molecular analyses. *Journal of cellular and molecular medicine*, **9**, 804-817.

111. Niwa, H. (2010) Mouse ES cell culture system as a model of development. *Development, growth & differentiation*, **52**, 275-283.

112. Wobus, A.M. and Boheler, K.R. (2005) Embryonic stem cells: prospects for developmental biology and cell therapy. *Physiological reviews*, **85**, 635-678.

113. Takaoka, K. and Hamada, H. (2012) Cell fate decisions and axis determination in the early mouse embryo. *Development (Cambridge, England)*, **139**, 3-14.

114. Smith, A.G., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M. and Rogers, D. (1988) Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature*, **336**, 688-690.

115. Chambers, S.M., Boles, N.C., Lin, K.Y., Tierney, M.P., Bowman, T.V., Bradfute, S.B., Chen, A.J., Merchant, A.A., Sirin, O., Weksberg, D.C. *et al.* (2007) Hematopoietic fingerprints: an expression database of stem cells and their progeny. *Cell stem cell*, **1**, 578-591.

116. Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P. and Smith, A. (2008) The ground state of embryonic stem cell self-renewal. *Nature*, **453**, 519-523.

117. Nichols, J. and Smith, A. (2009) Naive and primed pluripotent states. *Cell stem cell*, **4**, 487-492.

118. Smith, A.G. (2001) Embryo-derived stem cells: of mice and men. *Annual review of cell and developmental biology*, **17**, 435-462.

119. Medeiros, R.B., Papenfuss, K.J., Hoium, B., Coley, K., Jadrich, J., Goh, S.K., Elayaperumal, A., Herrera, J.E., Resnik, E. and Ni, H.T. (2009) Novel sequential ChIP and simplified basic ChIP protocols for promoter co-occupancy and target gene identification in human embryonic stem cells. *BMC biotechnology*, **9**, 59.

120. Mann, M. (2006) Functional and quantitative proteomics using SILAC. *Nat Rev Mol Cell Biol*, **7**, 952-958.

121. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif*, **25**, 402-408.

122. Ramirez-Solis, R., Rivera-Perez, J., Wallace, J.D., Wims, M., Zheng, H. and Bradley, A. (1992) Genomic DNA microextraction: a method to screen numerous samples. *Analytical biochemistry*, **201**, 331-335.

123. Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J. and Voytas, D.F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic acids research*, **39**, e82.

124. Vouillot, L., Thelie, A. and Pollet, N. (2015) Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3 (Bethesda, Md*, **5**, 407-415.

125. Bernard, P. and Couturier, M. (1992) Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *Journal of molecular biology*, **226**, 735-745.

126. 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. and Holmes, M.C. (2011) Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nature methods*, **8**, 74-79.

127. Wefers, B., Meyer, M., Ortiz, O., Hrabe de Angelis, M., Hansen, J., Wurst, W. and Kuhn, R. (2013) Direct production of mouse disease models by embryo microinjection of TALENs and oligodeoxynucleotides. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 3782-3787.

128. Melton, D.W., Konecki, D.S., Brennand, J. and Caskey, C.T. (1984) Structure, expression, and mutation of the hypoxanthine phosphoribosyltransferase gene. *Proceedings of the National Academy of Sciences of the United States of America*, **81**, 2147-2151.

129. Keough, D.T., Brereton, I.M., de Jersey, J. and Guddat, L.W. (2005) The crystal structure of free human hypoxanthine-guanine phosphoribosyltransferase reveals extensive conformational plasticity throughout the catalytic cycle. *Journal of molecular biology*, **351**, 170-181.

130. Shulman, M.J., Nissen, L. and Collins, C. (1990) Homologous recombination in hybridoma cells: dependence on time and fragment length. *Molecular and cellular biology*, **10**, 4466-4472.

131. Deng, C. and Capecchi, M.R. (1992) Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Molecular and cellular biology*, **12**, 3365-3371.

132. Scheerer, J.B. and Adair, G.M. (1994) Homology dependence of targeted recombination at the Chinese hamster APRT locus. *Molecular and cellular biology*, **14**, 6663-6673.

133. Hasty, P., Rivera-Perez, J. and Bradley, A. (1991) The length of homology required for gene targeting in embryonic stem cells. *Molecular and cellular biology*, **11**, 5586-5591.

134. Petrezselyova, S., Kinsky, S., Truban, D., Sedlacek, R., Burtscher, I. and Lickert, H. (2015) Homology arms of targeting vectors for gene insertions and CRISPR/Cas9 technology: size does not matter; quality control of targeted clones does. *Cellular & molecular biology letters*, **20**, 773-787.

135. Lu, Z.H., Books, J.T., Kaufman, R.M. and Ley, T.J. (2003) Long targeting arms do not increase the efficiency of homologous recombination in the beta-globin locus of murine embryonic stem cells. *Blood*, **102**, 1531-1533.

136. Mansour, S.L., Thomas, K.R. and Capecchi, M.R. (1988) Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature*, **336**, 348-352.

137. Mansour, S.L., Thomas, K.R., Deng, C.X. and Capecchi, M.R. (1990) Introduction of a lacZ reporter gene into the mouse int-2 locus by homologous recombination. *Proceedings of the National Academy of Sciences of the United States of America*, **87**, 7688-7692.

138. Gabriel, R., Lombardo, A., Arens, A., Miller, J.C., Genovese, P., Kaeppel, C., Nowrouzi, A., Bartholomae, C.C., Wang, J., Friedman, G. *et al.* (2011) An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nature biotechnology*, **29**, 816-823.
139. Lin, S., Staahl, B.T., Alla, R.K. and Doudna, J.A. (2014) Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife*, **3**, e04766.
140. Rostovskaya, M., Fu, J., Obst, M., Baer, I., Weidlich, S., Wang, H., Smith, A.J., Anastassiadis, K. and Stewart, A.F. (2012) Transposon-mediated BAC transgenesis in human ES cells. *Nucleic acids research*, **40**, e150.
141. Aarts, M. and te Riele, H. (2010) Progress and prospects: oligonucleotide-directed gene modification in mouse embryonic stem cells: a route to therapeutic application. *Gene therapy*, **18**, 213-219.
142. Olsen, P.A., Randol, M., Luna, L., Brown, T. and Krauss, S. (2005) Genomic sequence correction by single-stranded DNA oligonucleotides: role of DNA synthesis and chemical modifications of the oligonucleotide ends. *The journal of gene medicine*, **7**, 1534-1544.
143. Wu, X.S., Xin, L., Yin, W.X., Shang, X.Y., Lu, L., Watt, R.M., Cheah, K.S., Huang, J.D., Liu, D.P. and Liang, C.C. (2005) Increased efficiency of oligonucleotide-mediated gene repair through slowing replication fork progression. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 2508-2513.
144. Papaioannou, I., Disterer, P. and Owen, J.S. (2009) Use of internally nuclease-protected single-strand DNA oligonucleotides and silencing of the mismatch repair protein, MSH2, enhances the replication of corrected cells following gene editing. *The journal of gene medicine*, **11**, 267-274.
145. Rios, X., Briggs, A.W., Christodoulou, D., Gorham, J.M., Seidman, J.G. and Church, G.M. (2012) Stable gene targeting in human cells using single-strand oligonucleotides with modified bases. *PloS one*, **7**, e36697.
146. Liang, X., Potter, J., Kumar, S., Ravinder, N. and Chesnut, J.D. (2016) Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *Journal of biotechnology*, **241**, 136-146.
147. Bonner, M., Strouse, B., Applegate, M., Livingston, P. and Kmiec, E.B. (2012) DNA damage response pathway and replication fork stress during oligonucleotide directed gene editing. *Molecular therapy*, **1**, e18.
148. Strouse, B., Bialk, P., Niamat, R.A., Rivera-Torres, N. and Kmiec, E.B. (2014) Combinatorial gene editing in mammalian cells using ssODNs and TALENs. *Scientific reports*, **4**, 3791.
149. Olsen, P.A., Randol, M. and Krauss, S. (2005) Implications of cell cycle progression on functional sequence correction by short single-stranded DNA oligonucleotides. *Gene therapy*, **12**, 546-551.
150. Engstrom, J.U., Suzuki, T. and Kmiec, E.B. (2009) Regulation of targeted gene repair by intrinsic cellular processes. *Bioessays*, **31**, 159-168.
151. Richardson, C.D., Ray, G.J., DeWitt, M.A., Curie, G.L. and Corn, J.E. (2016) Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nature biotechnology*, **34**, 339-344.
152. Gravells, P., Ahrabi, S., Vangala, R.K., Tomita, K., Brash, J.T., Brustle, L.A., Chung, C., Hong, J.M., Kaloudi, A., Humphrey, T.C. *et al.* (2015) Use of the

HPRT gene to study nuclease-induced DNA double-strand break repair. *Human molecular genetics*, **24**, 7097-7110.

153. Bialk, P., Rivera-Torres, N., Strouse, B. and Kmiec, E.B. (2015) Regulation of Gene Editing Activity Directed by Single-Stranded Oligonucleotides and CRISPR/Cas9 Systems. *PloS one*, **10**, e0129308.
154. Kan, Y., Ruis, B., Takasugi, T. and Hendrickson, E.A. (2017) Mechanisms of precise genome editing using oligonucleotide donors. *Genome research*, **27**, 1099-1111.
155. Zierhut, C. and Diffley, J.F. (2008) Break dosage, cell cycle stage and DNA replication influence DNA double strand break response. *The EMBO journal*, **27**, 1875-1885.
156. Cristea, S., Freyvert, Y., Santiago, Y., Holmes, M.C., Urnov, F.D., Gregory, P.D. and Cost, G.J. (2013) In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. *Biotechnology and bioengineering*, **110**, 871-880.
157. Maresca, M., Lin, V.G., Guo, N. and Yang, Y. (2013) Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. *Genome research*, **23**, 539-546.
158. Nakade, S., Tsubota, T., Sakane, Y., Kume, S., Sakamoto, N., Obara, M., Daimon, T., Sezutsu, H., Yamamoto, T., Sakuma, T. *et al.* (2014) Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nature communications*, **5**, 5560.
159. Bachu, R., Bergareche, I. and Chasin, L.A. (2015) CRISPR-Cas targeted plasmid integration into mammalian cells via non-homologous end joining. *Biotechnology and bioengineering*, **112**, 2154-2162.
160. Geisinger, J.M., Turan, S., Hernandez, S., Spector, L.P. and Calos, M.P. (2016) In vivo blunt-end cloning through CRISPR/Cas9-facilitated non-homologous end-joining. *Nucleic acids research*, **44**, e76.
161. Letsou, A. and Liskay, R.M. (1987) Effect of the molecular nature of mutation on the efficiency of intrachromosomal gene conversion in mouse cells. *Genetics*, **117**, 759-769.
162. te Riele, H., Maandag, E.R. and Berns, A. (1992) Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 5128-5132.
163. van Deursen, J. and Wieringa, B. (1992) Targeting of the creatine kinase M gene in embryonic stem cells using isogenic and nonisogenic vectors. *Nucleic acids research*, **20**, 3815-3820.
164. Valenzuela, D.M., Murphy, A.J., Frendewey, D., Gale, N.W., Economides, A.N., Auerbach, W., Poueymirou, W.T., Adams, N.C., Rojas, J., Yasenchak, J. *et al.* (2003) High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nature biotechnology*, **21**, 652-659.
165. Chen, F., Pruett-Miller, S.M., Huang, Y., Gjoka, M., Duda, K., Taunton, J., Collingwood, T.N., Frodin, M. and Davis, G.D. (2011) High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nature methods*, **8**, 753-755.
166. Canver, M.C., Bauer, D.E., Dass, A., Yien, Y.Y., Chung, J., Masuda, T., Maeda, T., Paw, B.H. and Orkin, S.H. (2014) Characterization of genomic

deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *The Journal of biological chemistry*, **289**, 21312-21324.

167. Lee, H.J., Kim, E. and Kim, J.S. (2010) Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome research*, **20**, 81-89.
168. Schmittgen, T.D. and Livak, K.J. (2008) Analyzing real-time PCR data by the comparative C(T) method. *Nature protocols*, **3**, 1101-1108.
169. Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S. and Haussler, D. (2004) Ultraconserved elements in the human genome. *Science*, **304**, 1321-1325.
170. Ochiai, H., Sugawara, T., Sakuma, T. and Yamamoto, T. (2014) Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells. *Scientific reports*, **4**, 7125.
171. Okamoto, Y., Yoshida, N., Suzuki, T., Shimozawa, N., Asami, M., Matsuda, T., Kojima, N., Perry, A.C. and Takada, T. (2016) DNA methylation dynamics in mouse preimplantation embryos revealed by mass spectrometry. *Scientific reports*, **6**, 19134.
172. Inoue, A. and Zhang, Y. (2011) Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science*, **334**, 194.
173. Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P. *et al.* (2015) GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature biotechnology*, **33**, 187-197.
174. Liang, X., Potter, J., Kumar, S., Zou, Y., Quintanilla, R., Sridharan, M., Carte, J., Chen, W., Roark, N., Ranganathan, S. *et al.* (2015) Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *Journal of biotechnology*, **208**, 44-53.
175. Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J.A. and Jasin, M. (1998) Gene conversion tracts from double-strand break repair in mammalian cells. *Molecular and cellular biology*, **18**, 93-101.

8 Acknowledgements

I would like to thank Francis Stewart for patiently supporting my thesis work and career development over the last years. His guidance and scientific input have been invaluable.

Thanks to all current and former members of the Stewart and Anastassiadis Labs for the excellent work atmosphere- for both the scientific support and comic relief they provide. I would like to especially thank the outstanding technical assistants in our lab: Madeleine, Susan, Mandy, Isabel, Steffi, and Jing have always been willing to share their secret methods and lunchtime laughs.

I am grateful to Jun Fu, Frank Groß and Rupam Choudhury for their invigorating scientific discussions and input.

I would like to thank members of my committee, especially Francis and Yixin Zhang for reviewing this thesis.

I would like to thank my husband, Mikhail, for his unwavering belief in me, and his unending support of my aspirations. Without his support, this work would not have been possible.

Erklärung entsprechend §5.5 der Promotionsordnung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Püfungsbehörde vorgelegt. Die Dissertation wurde von Prof. Dr. A. Francis Stewart am Biotechnologischen Zentrum (BIOTEC) der TU Dresden betreut und im Zeitraum vom Januar 2010 bis November 2017 verfasst. Ich erkläre hiermit, dass keine früheren erfolglosen Promotionsverfahren stattgefunden haben. Ich erkenne die Promotionsordnung der Fakultät für Mathematik und Naturwissenschaften an der Technischen Universität Dresden an.

Dresden, 16. November 2017 Sarah Dorothea Tsurkan