Development and Applications of CRISPR/Cas9 Genome Editing Technology

Fatwa Adikusuma

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

Department of Molecular and Cellular Biology
School of Biological Science
University of Adelaide, Australia

October 2017

Contents

This page intentionally left blank

Abstract

The ability to edit the genome of organism can positively impact biomedical research. Using genome editing we can dissect the function of gene(s) and their regulatory elements and, more importantly, facilitate modelling of human genetic diseases to understand their pathology and develop treatments. The recently developed CRISPR/Cas9 genome editing platform has become rapidly and widely used in biomedical research due to its ease of use, highly efficiency and low cost. The research conducted during my PhD attempted to further develop this technology and apply it to a range of biological questions that are of interest to the Thomas laboratory.

I showed that this technology can be applied to generate a gene-swap mouse model to study functional redundancy between closely related transcription factors in SOXB1 family, SOX2 and SOX3. By swapping *Sox3* with *Sox2* in vivo we showed that the presence of Sox2 in the absence of Sox3 rescues *Sox3*-null phenotypes. This finding provides strong and direct evidence that they are functionally redundant. I also develop CRISPR/Cas9-based strategies to allow targeted elimination of an entire chromosome. These strategies termed centromere removal and chromosome shredding could facilitate efficient chromosome deletion as shown by successful elimination of the Y chromosome in mouse ES cells and zygotes.

We also contribute to the development of CRISPR/Cas9 toolbox by generating plasmids that can express dual gRNAs and other required components such as the Cas9 or Cas9-nickase as well as selection markers within a single plasmid. Interestingly, our vector design allows facile generation of two unique guides in a simple one-step reaction, rendering these plasmids user-friendly for researchers requiring simultaneous expression of two gRNAs. Targeting two sites can be achieved with very high efficiency using these plasmids, which will be made freely available to all researchers via the Addgene plasmid

repository. Lastly, my research show that large deletions are frequently generated as the repair outcome of CRISPR/Cas9-mediated cleavage. The large deletions contain mostly microhomology sequences at the break junctions and are generated via DNA resection, indicating an alternative end joining mechanism underlies their generation. This study reveals an underestimated yet common repair outcome that researchers should be aware of to avoid genotyping misinterpretation.

Collectively, the studies in this thesis have contributed to the development of CRISPR/Cas9 genome editing technology by providing valuable tools and new knowledge about DNA repair. In addition, I demonstrate that CRISPR/Cas9 technology can be applied to diverse biological questions including functional redundancy of developmental transcription factors and targeted chromosomal ablation.

Thesis Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree. I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968. I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works. I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Fatwa Adikusuma

Acknowledgment

I would like to firstly thank God for His guidance so I could end up into this wonderful journey.

I am so grateful to Prof. Paul Thomas for allowing me to do PhD in your lab. Thank you for nurturing me, for being very supportive, for permitting me to do silly experiments on your cost (e.g. RNA template HDR in mouse injection), for being very patient with all my problems in the lab, for giving your trust on me to extend my study. I am sorry I can't write down all your kindness to me, it is just too many to say.

Thanks to Thomas lab members (including the past), particularly James Hughes who has been acting as my mentor during my PhD. Most of my lab skills are taught by you, James. Many thanks to Sandie for the best microinjection in the world, three quarter of this thesis projects have your involvement in it, that means a lot. Thanks to Nick Rogers, Dale and Dan for the great advice. Thanks to Louise, Mel, Ella, Ruby, Chan and Gervais for being very nice colleagues. Louise, thanks for spending weekend with me to prove that NgAgo can't perform genome editing. Ruby, thanks for being God's messenger to show me the gRNA sequences that I can use for my SSA-trapping assay. Mel, I like betting with you, a cup of Nespresso isn't risky at all. Ella and Gervais, thanks for being partners in SBS, Fzd3 and Nestin projects. Chan, thanks for your equal contribution in dual-gRNA project. I would also thank collaborators from Michael Beard's, Frank Grutzner's and Shaun McColl's labs for the involvement in my studies. I really appreciate the staff members of Adelaide Uni, Animal house, the FACS centre and the microarray centre that have been really helpful to me.

Outside the lab world, I would like to thank my family. My wife Nia, thank you so much for your love, for being very supportive, the motivation for me to finish my PhD, for being very patient as sometimes I acted like a zombie due to this PhD burden. I would like to also thank my Mom and Dad, my brother Tata and Nissa, and my parents in law for their support and prayers.

I would also like to thank Prof. Sultana for her suggestion to pursue my PhD in Adelaide, Australia, and for assisting me to get the scholarship from Beasiswa Unggulan DIKTI. My gratitude also goes to the scholarship sponsors, mainly Beasiswa Unggulan DIKTI, and to

Publications

Adikusuma, F., Pederick, D., McAninch, D., Hughes, J. & Thomas, P. Functional Equivalence of the SOX2 and SOX3 Transcription Factors in the Developing Mouse Brain and Testes. *Genetics* 206, 1495-1503 (2017).

Adikusuma, F., Williams, N., Grutzner, F., Hughes, J. & Thomas, P. Targeted Deletion of an Entire Chromosome Using CRISPR/Cas9. *Molecular Therapy* (2017).

Presentations and Communications

Studies on SOXB1 redundancy (oral presentation)

The University of Adelaide School of Molecular and Biomedical Science Postgraduate Symposium. Australia (2013).

Efficient generation of mutant mice using CRISPR/Cas9 technology (oral presentation)

The University of Adelaide School of Molecular and Biomedical Science Postgraduate Symposium. Australia (2014).

Different strategies for generating mouse models using CRISPR (poster presentation)

The 4th Adelaide ANZSCDB Cell and Developmental Biology Meeting. Australia (2014).

High frequency of large deletions generated by targeted CRISPR/Cas9 DNA cleavage in mouse zygotes (poster presentation)

36th Lorne Genome Conference. Australia (2015).

Frequent large deletions generated by targeted CRISPR/Cas9 cleavage indicate a novel DNA repair mechanism (poster presentation)

Cold Spring Harbor meeting on Genome Engineering: The CRISPR-Cas Revolution. USA (2015).

CRISPR/CAS9-mediated replacement of Sox3 with Sox2 partially rescues Sox3 null phenotype: implications for functional redundancy (oral presentation)

The 5th Adelaide ANZSCDB Cell and Developmental Biology Meeting. Australia (2015).

Targeted deletion of an entire chromosome using CRISPR/Cas genome editing (poster presentation)

The 1st Conference & Symposium on Human Genetics. Indonesia (2016).

Awards

Best student talk (third prize). The University of Adelaide School of Molecular and Biomedical Science Postgraduate Symposium. Australia (2014).

Conference stipend. Cold Spring Harbor meeting on Genome Engineering: The CRISPR-Cas Revolution. USA (2015).

Best poster presentation. The 1st Conference & Symposium on Human Genetics. Indonesia (2016).

Chapter 1:
Introduction

1.1. Genome editing

Genome editing refers to the process of modifying (deleting, inserting or substituting) the genomic DNA sequence from its original form. This ability to modify genomic sequences is crucial for life science research. It can be used to mimic human diseases driven by genetic changes in cell lines and animal models, thereby increasing our understanding of the basic science of these diseases so that therapeutic options can be developed. Genomic modifications allow scientists to knock out genes or delete genetic elements to study their functions in biological systems. In an advanced form, genome editing can be used for therapy of diseases such as cancers, infectious and genetic diseases. Genome editing in the farming and crop industries can result in increased production by manipulating genes to generate plants or animals that are resistant to infection, can be harvested earlier or produce greater yields [1-8].

DNA modification in cells began decades ago with a method known as 'gene targeting', which takes advantage of the natural homologous recombination process. In gene targeting, exogenous DNA sequences are integrated into the desired genomic location by providing them with DNA donor containing sequences homologous to the DNA target. This technique has revolutionised biological research as it can be used in embryonic stem (ES) cells, and the genetically modified ES cells can be implanted into embryos to ultimately generate mutant animals. Many gene functions have been revealed by this technique. However, the gene targeting process is complicated. The incidence of targeted homologous recombination is very low (1 in 10^6–10^9 cells). Generation of genetically modified animals also requires multiple steps, which are prone to failure. Overall, this technique is time-consuming and costly, which hampers its use in research [4, 9].

Experiments with endonucleases have provided the insight that site-specific double strand breaks (DSB) can alter DNA sequences and lead to the deletion or insertion of sequences

at the break sites. More importantly, if exogenous DNA donor is provided while the DSB is present, the donor DNA can be integrated into the break sites more efficiently through the homology-directed repair (HDR) mechanism, allowing specific modification of DNA sequences (as is also the case for gene targeting). Given the importance of genome editing, scientists have been developing site-specific endonucleases to allow the efficient targeted modification of DNA sequences. To date, there are at least three major technologies that have been generated for this purpose: zinc finger nuclease (ZFN), transcription activator-like effector nuclease (TALEN) and clustered regularly interspaced short palindromic repeats (CRISPR) [1-7]

ZFN and TALEN recognise specific sequences through protein-DNA binding interactions via their DNA binding domains, which are customisable to enable the targeting of specific sequences (Figure 1.1). Their DNA binding domains consist of modules, with each module binding to specific nucleotides. Each module of ZFN or TALEN recognises three or one nucleotides respectively, and must be assembled in an order that corresponds to the target sequence. By fusing these sequence-specific binding domains to the dimer-dependent Fok1 nuclease domain, ZFN and TALEN can induce site-specific DNA cleavage. This involves dimerisation of Fok1 to induce DNA breaks; therefore, to target as specific site for cleavage, a pair of ZFN or TALEN molecules with the correct orientation and spacing is required (Figure 1.1). Both ZFN and TALEN have been shown to efficiently induce targeted DNA DSBs for genome editing purposes. However, designing sequence-specific binding modules for ZFN is complicated and expensive. TALEN design is easier, but synthesising its modules is still costly. Due to this complexity, both technologies are almost impossible to use for multiplex targeting of different sites. Furthermore, they are also thought to have problems with context-dependent specificity [1-7].

ZFN

TALEN

TALE repeat domains

Fokl

3'
5'

TACTAGGATCCGATTCGCNNNNNNNNNNNNTCATGCTAGACTGAGCA
CTGATCCTAGGCTAAGCGNNNNNNNNNNNNNAGTACGATCTGACTCGT

Fokl

Fokl

3'
5'

CCCCAACTGNNNNNNGCCGTAGAG
GGGGTTGACNNNNNNCGGCATCTC

Zinc finger domains

Fokl

CRISPR/Cas9

Cas9

GTCCATGATAGGTGCCATAT

3'
5'

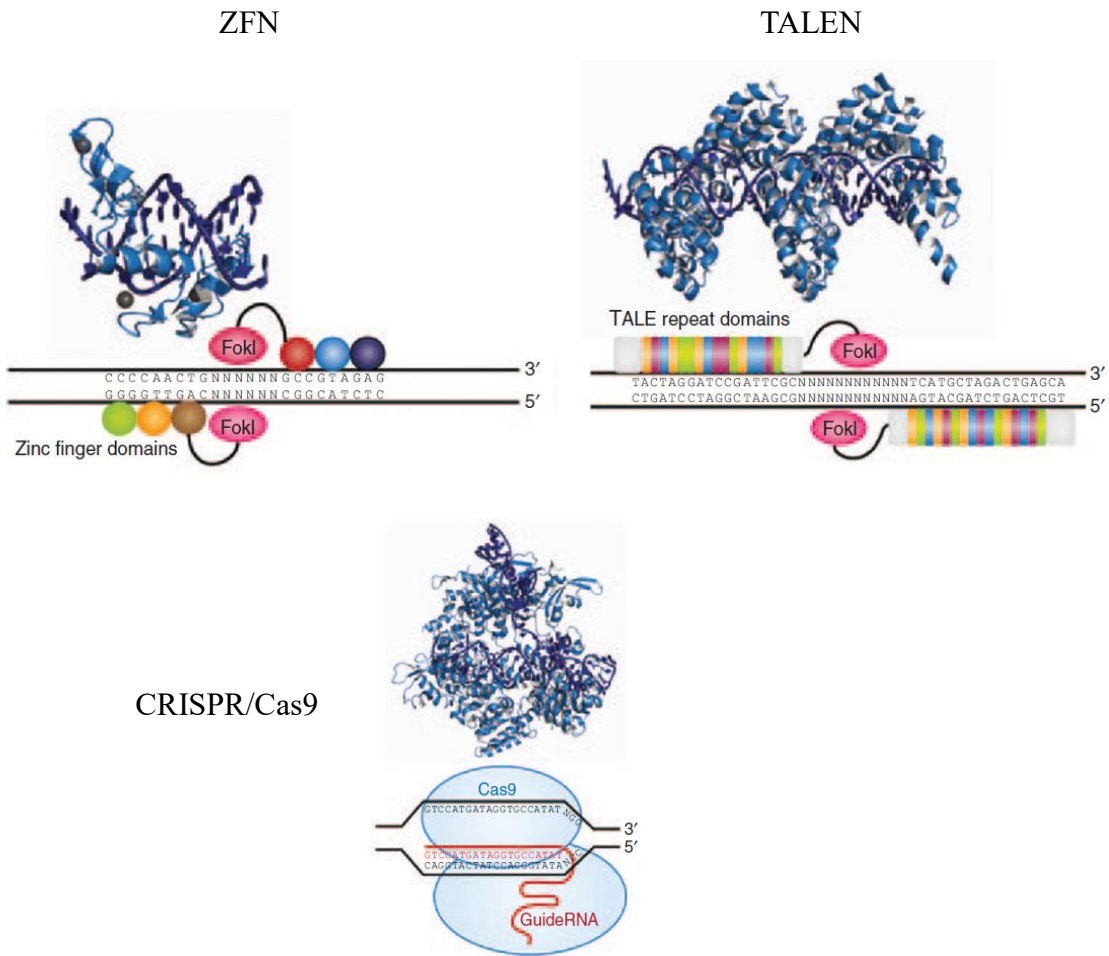GTCCATGATAGGTGCCATAT
CAGGTACTATCCACGGTATA

GuideRNA

Figure 1.1 | Major players in genome editing technologies: ZFN, TALEN and CRISPR/Cas9. ZFN and TALEN rely on proteins to bind to specific sequences. DSB are achieved through dimerisation of the Fok1 nuclease. CRISPR/Cas9 technology uses short RNA to recognise specific sequences with Cas9 as the nuclease. Adapted from Maeder and Gerbach [7].

On the other hand, the recently developed CRISPR/Cas system uses RNA:DNA Watson-Crick base pairing to provide specificity (Figure 1.1). The endonuclease, Cas9, induces highly efficient DSBs to DNA target sequences by forming a complex with a short RNA molecule termed the guide-RNA (gRNA), that acts to guide Cas9 to target sequences. This makes genome editing much simpler because different genomic sites can be targeted for cleavage simply by changing the gRNA sequence. Furthermore, multiple gRNAs can be expressed or produced easily, which allows multiplex targeting of genomic regions. This versatility has made CRISPR/Cas9 technology a favourite tool for genome editing, and has revolutionised the genome engineering field. The simplicity of this system has attracted many researchers to this technology to study gene function or for other purposes, which has led to many CRISPR-related publications being produced (Figure 1.2) [1-7].
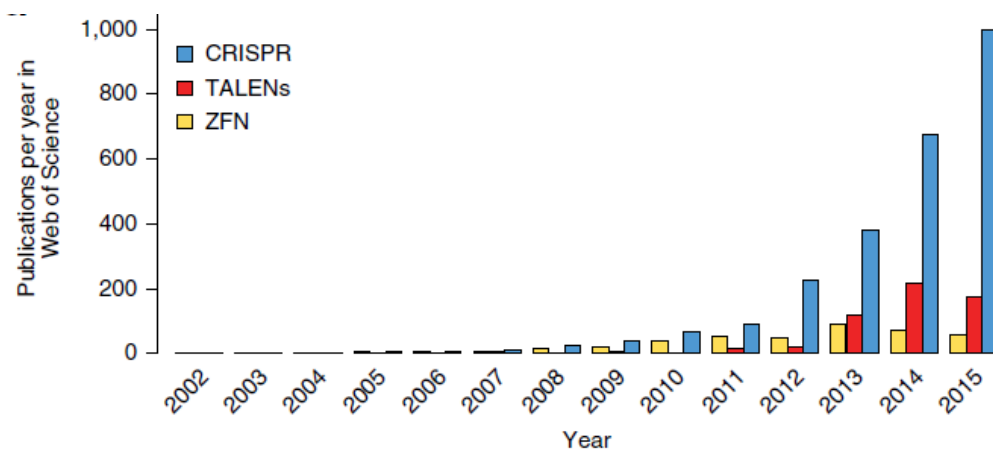

Figure 1.2 | Number of publications produced describing the use of ZFN, TALEN and CRISPR/Cas9 technologies. CRISPR/Cas9 technology has been rapidly adopted by researchers and has generated many publications over the past few years. Adapted from Barrangou and Doudna [2].

1.2. CRISPR/Cas9 technology

CRISPR is an adaptive immune system found in prokaryotes. It was noted many years ago when scientists observed that *Escherichia coli* bacteria contained repetitive sequences separated by non-repetitive sequences, although they did not understand the function of these sequences [10]. As more sequencing data became available, scientists noticed that these clustered repeat elements were surprisingly common in bacteria and archaea genomes. More than 40% of bacteria and >90% of archaea have these clustered repeat elements, which were later named CRISPR [11, 12].

CRISPR loci contain an array of repeat sequences separated by spacer sequences and genes called *Cas* (CRISPR-associated) genes, which are mostly located adjacent to the CRISPR elements [12]. The spacer sequences were found to be short DNA sequences found in phage, which were acquired and integrated into CRISPR loci during the adaptation/immunisation process (Figure 1.3) [13-15]. In Type II CRISPR systems, such as that of *Streptococcus pyogenes* (Sp), the CRISPR array is transcribed as a single long transcript called pre-crRNA and then converted to several small CRISPR-RNA (crRNAs) by RNase III, where each single crRNA targets a unique foreign sequence complementary to the spacer sequences. To cleave the invader's DNA, this crRNA requires two more components: Cas9 as the endonuclease and a trans-activating crRNA or tracrRNA, which forms a complex with the crRNAs to guide and activate the Cas9 endonuclease (Figure 1.3). Notably, Cas9 can only cleave DNA when the target sequences contain a certain PAM (protospacer adjacent motif) adjacent to the foreign DNA target sequence (the PAM for *Streptococcus pyogenes* Cas9 or SpCas9 is NGG). This PAM requirement is crucial for the prokaryote CRISPR/Cas9 system to distinguish between the foreign DNA target and their own CRISPR arrays [13, 16-22]. In the genome, the crRNA-tracrRNA-Cas9 complex will randomly bind to sites containing a PAM and dissociate from non-PAM sites. Once

the PAM is recognised by the complex, it then interrogates the DNA target sequences for complementarity with the gRNA sequences, starting from the sequences adjacent to the PAM followed by the remaining sequences in a sequential manner ('zip-up' mechanism). Binding to the PAM and the complementarity of crRNA-DNA target will activate the nuclease activity of Cas9 and induce DNA cleavage facilitated by two nuclease domains, HNH and RuvC, which generate a DSB at a position three nucleotides distal to the PAM sequence [23-25].

In 2012, Jinek et al. [26] showed that they could adapt the CRISPR/Cas9 system from *Streptococcus pyogenes* to cleave DNA in vitro. Six months after this publication, in 2013, Cong et al. and Mali et al. [27, 28] published their work showing this system could induce targeted cleavage of DNA in vivo in mammalian genomes and that they would be useful for genome editing applications. These studies all also indicated that the three components required for the CRISPR/Cas9 system (crRNA-tracrRNA-Cas9) could be simplified to two components by merging the crRNA and tracrRNA into a single guide RNA (gRNA) (Figure 1.4). This single gRNA of the SpCas9 system consists of ~100 nt with the first ~20 nt providing DNA target specificity. This means that targeting different target sequences requires alteration of only these ~20 nt, which can be easily performed using standard molecular biological techniques. Moreover, RNAs of this size can also be synthesised enzymatically or chemically. Currently, many vectors for performing genome editing using the CRISPR/Cas9 system are available and can be obtained easily through plasmid repositories such as Addgene (https://www.addgene.org/crispr/). Some of those vectors have been designed to be user-friendly; for example, by having 'golden gate' sites to simplify the generation of gRNAs of interest through the 'golden gate' cloning strategy [27, 29]. The choice of target sequences is also flexible as DNA target screening is only limited by the requirement for NGG PAM sequences, which are abundant in the genome.

The simplicity of the *Streptococcus pyogenes* CRISPR/Cas9 system has made this

technology popular for research requiring modification of genomic sequences in various
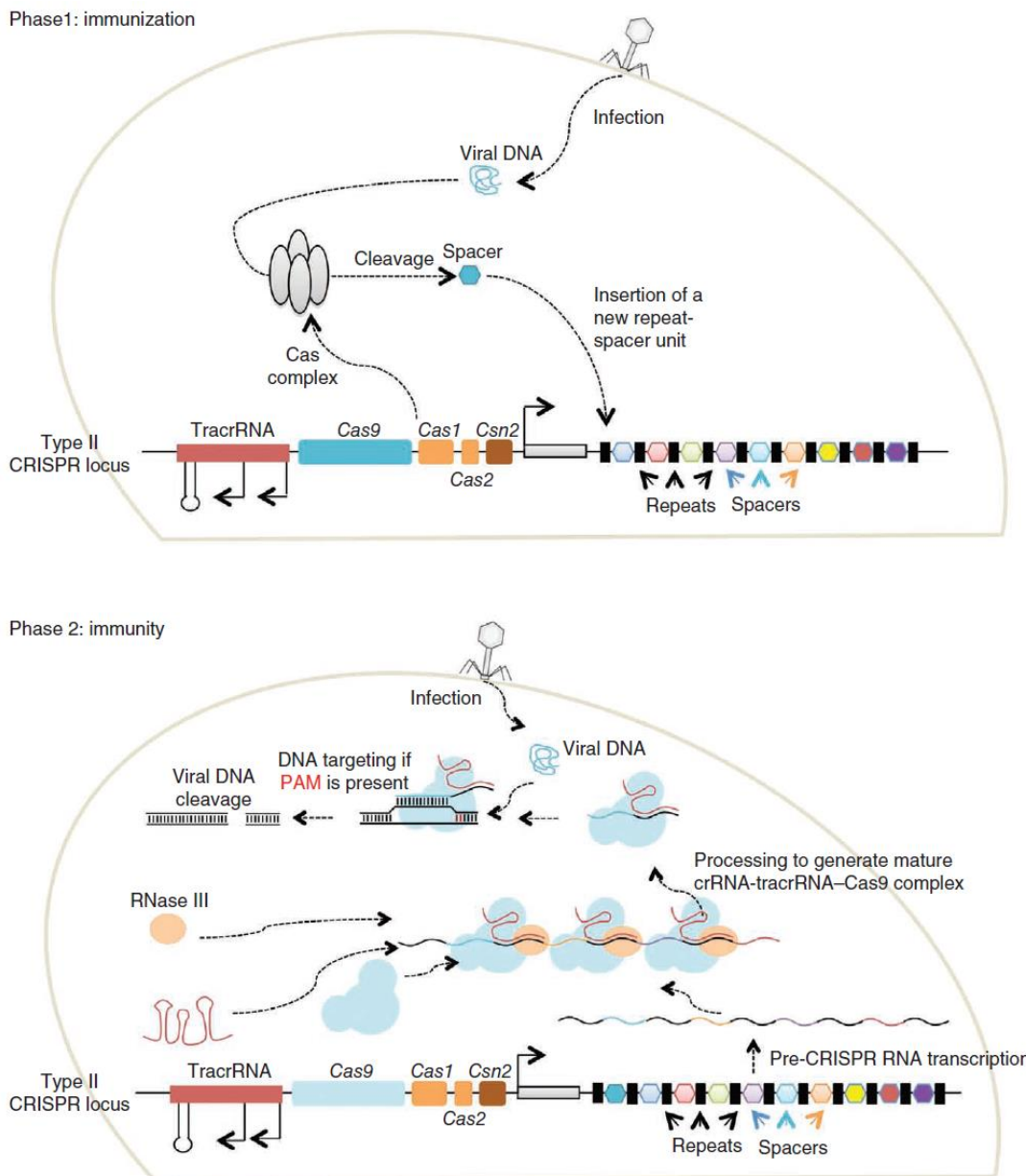
Figure 1.3 | Mechanism of CRISPR adaptation and immunity in bacteria with a Type II CRISPR locus. During adaptation, invading viral sequences are cleaved and the resulting sequences are inserted as spacers between CRISPR repetitive sequences. When the bacterium is infected by the same virus, this CRISPR array is transcribed, producing RNAs that guide Cas9 endonuclease to cleave the viral DNA sequences. Adapted from Mali et al. [5].

organisms for a range of applications. Organisms that have been modified via CRISPR/Cas9 include mice, rats, rabbits, non-human primates, fish, farm animals, plants, yeast, bacteria, and controversially, human embryos [1, 2, 30-32].

1.3. Editing strategies using the CRISPR/Cas9 system

Once CRISPR/Cas9 generates DSBs, the cellular DNA repair system immediately fixes the breaks. The primary repair pathway is non-homologous end joining (NHEJ), which results in precise joining of the breaks and therefore is non-mutagenic [33, 34]. However, in the case of endonuclease-mediated DSBs, these error-free ligations will be continuously cleaved by the endonucleases (when they are still present) as they are still recognised by the endonuclease as their target for cleavage. Cleavage and repair will endure until the endonucleases are no longer present or the target sequences are modified (through mis-repair) and not recognised by the endonuclease. The DNA modifications in NHEJ usually produce small deletions (<20 bp) and/or small insertions (indels) at the site of the breaks, which are thought to occur randomly (Figure 1.4) [5, 27, 29, 35, 36]. These indel mutations potentially generate frameshift alleles that are useful for knocking out genes or can be used to disrupt protein binding sites or other DNA elements. In brief, the NHEJ mechanism involves high-affinity binding of the Ku70/Ku80 heterodimer to DSB to protect the DNA ends from degradation and to recruit other factors, such as DNA-PKcs. The MRN complex is also recruited to the break site together with Artemis to allow modification of the DNA ends. Finally, DNA ligase IV, together with XRRC4 and XLF, ligate the breaks [34, 37].

Interestingly, if a single break from a gRNA/Cas9 induces a small indel mutation, a pair of DNA breaks generated using two gRNAs flanking a sequence region may result in the deletion of the intervening sequence when the NHEJ process joins the distal ends [38-41]. This phenomenon is commonly used to delete large sequences in a semi-predictable

manner. To date this approach has been used to delete sequences of up to 30 Mb [42]. Occasionally, the intervening sequences re-ligate back in the reverse orientation, creating inversions of the DNA region, which may be useful to model diseases caused by chromosomal inversions [40, 41, 43, 44]. In addition to generating large deletions and inversions, two simultaneous breaks can also induce chromosome translocations when the breaks are induced in two different chromosomes [43, 45, 46].

Recent studies have indicated that the distribution of small indels after CRISPR/Cas9 breaks are not random, as was previously thought, but that this is influenced by the sequences around the Cas9 target sites. These non-random repair outcomes presumably result from both NHEJ and micro-homology end joining (MMEJ) repair mechanisms. The distribution of indels can thus be perturbed by inhibition of factors involved in NHEJ or MMEJ [47]. The MMEJ mechanism (sometimes called alternative end joining/Alt-EJ) is distinct from NHEJ. In MMEJ, DNA breaks are followed by 5'–3' end resections generating single-stranded overhangs at both broken ends. The overhangs anneal to each other through their micro-homology sequences (1–16 nt), followed by DNA extension and ligation processes that result in the deletion of the sequences between the micro-homologies, along with the micro-homology sequence from one side. Since the micro-homology sequences engage during MMEJ, the product of this repair process is relatively predictable (Figure 1.4) [48, 49].

Specific changes in DNA sequences can be accomplished by providing a single-stranded oligodeoxynucleotide (ssODN) donor containing the intended sequence change flanked by homology sequences. Broken DNA will use this oligo donor as an HDR template, presumably via the same pathway as the single strand annealing (SSA) mechanism. The SSA mechanism is similar to MMEJ; however, DNA synapse/annealing after DNA resection occurs using longer homology sequences (Figure 1.4). The size of the homology

sequences of ssODN commonly used for genome editing is usually around 40–80 nt, although shorter homology sequences of 20–35 nt can still be used as efficient HDR templates (Figure 1.4) [35, 50-52]. This ssODN donor approach has been shown to be efficient for inducing intended changes. To date, the maximum size of commercially available synthetic oligonucleotides is ~200 nt, which is sufficient when attempting to perform substitutions of a few base pairs or inserting short sequences such as LoxP and HA/Flag tag sequences [29, 51]. However, this size limitation restricts the insertion of longer sequences. Insertion of long sequences is usually done by providing a double-stranded DNA (dsDNA) donor with long homology arms in the expectation that DNA repair employs the homologous recombination (HR) pathway and uses the donor as the HR template (Figure 1.4). The HR repair mechanism is an error-free repair pathway available in cellular systems. In natural HR, a DSB in one copy of a chromosome will be repaired using the sister chromatid as the repair template which is known as inter-homologue repair. The DSB will be followed by 5'–3' long resection of DNA, generating 3' overhangs that will search, invade and bind to the complementary sequences from the sister chromosome. After strand invasion, the broken string will be extended by polymerase, which copies the sequence of the sister chromatid before ligation, thus maintaining the integrity of the sequences [53]. The dsDNA donor for HR is co-transfected as a plasmid or a viral vector containing long homology sequences (usually >500 bp). However, the efficiency of this dsDNA donor approach is very low [29, 51, 54]. Inter-homologue repair itself can also be utilised to edit an allele in cells with heterozygous alleles by specifically targeting the to-be-modified allele for cleavage while leaving another allele intact, thus the broken allele can use the sister chromatid as the repair template [32, 55].

Given that ssODN donors are much more efficient for performing insertions, scientists have been interested in performing insertions of long sequences using long single-stranded DNA (ssDNA) donors to efficiently insert long DNA fragments. Long ssDNA donors can

be generated using techniques such as reverse transcriptase- or nicking endonuclease-based techniques. This long ssDNA approach has been shown to efficiently induce the insertion of long DNA fragments [56, 57].

Knocking-in large inserts using dsDNA donors can also take advantage of efficient MMEJ repair. This method, known as PITCh (precise integration into target chromosome), designs the dsDNA donor to contain short homology arms (~20 bp) (Figure 1.4). Delivery of dsDNA donors relies on a circular plasmid that is also cleaved by CRISPR/Cas9 at the end of the homology arms. Simultaneous DSBs to the insertion site and the PITCh donor result in efficient PITCh donor integration to the target region via the MMEJ mechanism [58, 59].

A recent study has shown that inserting long DNA sequences can be achieved efficiently by providing dsDNA donors without homology sequences. This homology-independent technique, called HITI (homology-independent targeted integration), harnesses the efficient NHEJ repair pathway to join a linear dsDNA donor to break sites (Figure 1.4). These researchers showed that this technique is more efficient than the HR-based dsDNA donor system (with long homology arms) and the PITCh technique. As the NHEJ repair pathway is available in non-dividing cells, this technique is useful to correct mutations requiring DNA insertions in vivo in adult tissues where most cells no longer divide [60].

All intended changes by co-delivery of ssDNA or dsDNA donors should always consider the ability of CRISPR/Cas9 to re-cleave the same or similar target sequences. Therefore, the final DNA sequences after the knock-in should not contain target sequences, for instance, by altering the PAM to sequences that are not recognisable by Cas9, or by creating more mismatches to the new sequences to abolish the binding ability of gRNAs to the DNA target.

Figure 1.4 | Genome editing strategies using the CRISPR/Cas9 system. DSBs, including those generated by Cas9, are repaired using pathways that either require DNA end resection (right box) or DNA end protection (left box). Repair by NHEJ will generate small indels (red highlights). dsDNA donor can be inserted to the breaks by the HITI technique. The pathway with resection can be repaired by MMEJ if the breaks occur between micro-homologies (black highlights), or by SSA if long repeat sequences are present (yellow highlights). The ssODN (oligo) donor can be efficiently inserted through the same mechanism as SSA. dsDNA can act as a donor template through the HR mechanism or the MMEJ-mediated mechanism (PITCh).

1.4. Delivery of CRISPR/Cas9 components

The CRISPR/Cas9 components, gRNA and Cas9, can be delivered to cells in the form of DNA (plasmid, viral vector), RNA or RNP (ribonucleoproteins). Delivery of CRISPR/Cas9 components can be accomplished in various ways depending on the type of experiment and its purpose. Transfections (including electroporation) or transductions of gRNA and Cas9 in DNA form are commonly performed for editing genomic DNA of cells in culture (Figure 1.5C). The vectors expressing the Cas9 and gRNA may also contain selection markers, such as drug resistance or fluorescent markers, to allow for enrichment of successfully transfected cells and thereby increase efficiency. DNA donor template can be co-transfected when performing HDR. Some researchers prefer transfection in RNA or RNP form for various reasons, even though this technique does not use selection markers [29, 61, 62].

CRISPR/Cas9 components (including the DNA donor) can also be microinjected to the cytoplasm or pronucleus of fertilised eggs, as is commonly performed when generating mice carrying mutations (Figure 1.5D) [35, 51, 54]. Microinjection requires expensive apparatus and highly skilled technicians. Recent advances allow the generation of CRISPR/Cas9-mediated mutant mice without microinjection. These techniques involve electroporation of CRISPR/Cas9 components into mouse zygotes in a culture dish before implantation into the uterus or direct electroporation of embryos within the intact mouse oviduct [63-65]. Electroporation can also be performed on developing mouse embryos for tissue/organ specific genomic modifications, particularly in the brain, by a technique called in utero electroporation [60, 66].

Delivery of CRISPR/Cas9 components to post-mitotic tissues in vivo is commonly performed by the injection of viral vectors into the target tissue or via nanoparticles (Figure 1.5E) [62, 67-69]. A recent study indicated that RNP injected locally to adult mouse brains

could also induce mutations in the injected tissue [70]. Delivery of CRISPR/Cas9 components can also be achieved by hydrodynamic injection. This technique involves the rapid injection of a large volume of CRISPR/Cas9 solution through a blood vessel, after which it will end up in the liver [68, 71]. However, this technique has not been proven to be safe for humans.

1.5. CRISPR/Cas9 efficiency and specificity

CRISPR/Cas9 can efficiently induce indel mutations in its target sequences. When combined with a good transfection and selection method, mutation induction by CRISPR/Cas9 can reach near 100% efficiency [36, 41]. Lower mutation efficiencies may result from lower transfection efficiencies. However, studies by Doench et al. [72, 73] suggest that different gRNAs may have different effectiveness depending on the features of their sequences. Based on these authors' recommendations, criteria for effective gRNAs include having a balanced GC content, having adenine in the central sequence of the gRNA, having guanine and avoiding cytosine at the gRNA sequence position 20 (adjacent to the PAM), and having a cytosine and avoiding guanine at gRNA sequence position 16. To facilitate the design of effective gRNAs, they also generated an algorithm for predicting the on-target efficiency of gRNAs, which is available as a gRNA design tool at http://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design, at Benchling CRISPR design tools, https://benchling.com/ and at CRISPOR http://crispor.org/.

Although generating indel mutations is typically very efficient using CRISPR/Cas9 technology, performing genomic modifications, such as insertions and substitutions, still requires enormous improvement to increase the efficiency of these processes. Improved HDR efficiency can be achieved by inhibiting the NHEJ pathway, which is expected to shift the repair mechanism to the HDR pathway (Figure 1.4). Researchers have observed
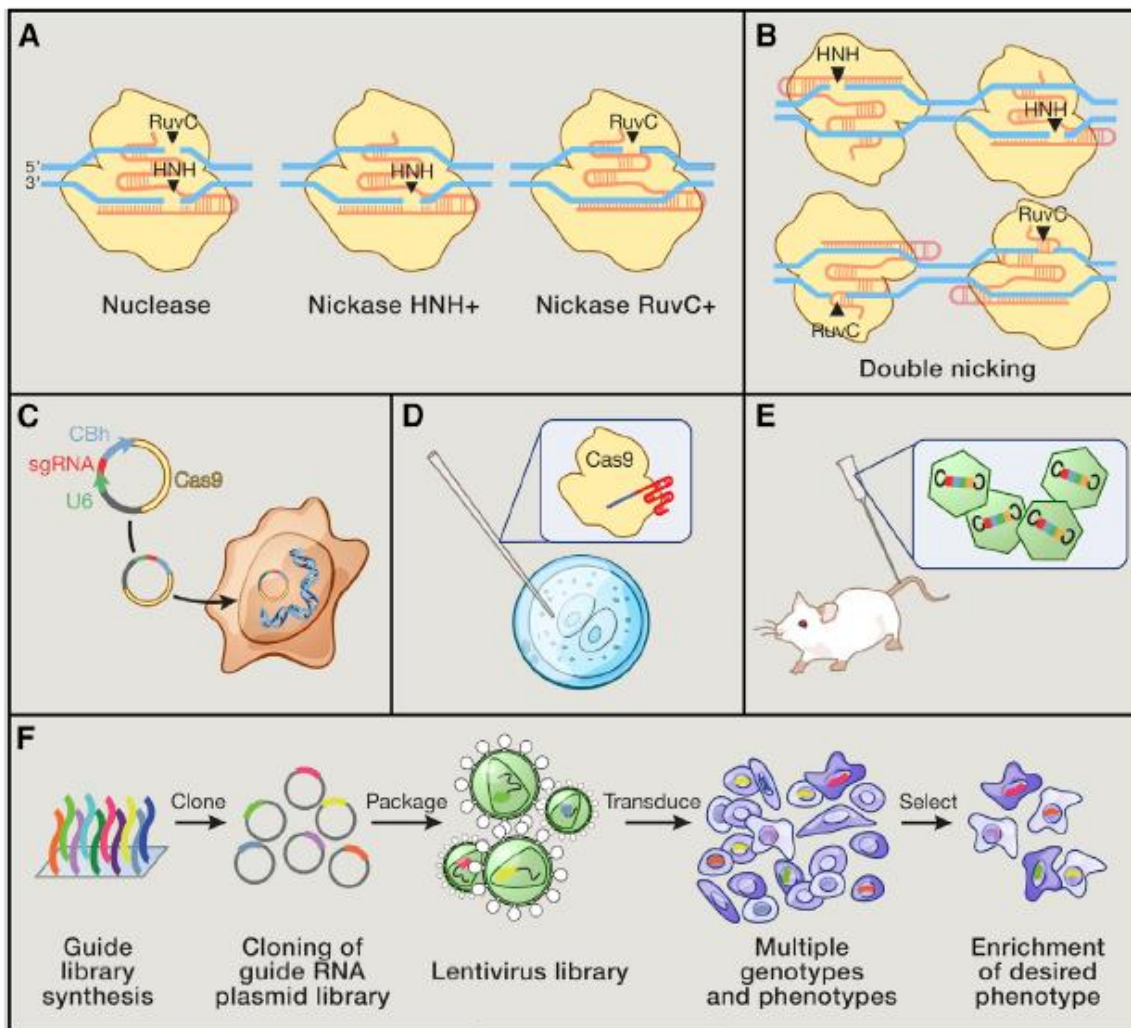
Figure 1.5 | Various applications of CRISPR/Cas9 genome editing. (A) Cas9 nuclease can be modified to become a nickase. (B) Paired nickases using two gRNAs located nearby with the correct orientations could be used to generate DSB with less off-targets. (C) Genome modification in cells can be achieved by delivery of CRISPR/Cas9 to cells via plasmid or viral vectors. (D) CRISPR/Cas9 can be microinjected into fertilised egg to generate a mutant animal. (E) CRISPR/Cas9 can be delivered in vivo through viral injection. (F) Schematic of a CRISPR/Cas9-based forward genetic screening experiment. Adapted from Hsu et al. [4].

increased HDR when they inhibited NHEJ factors such as Lig4, DNA-PKcs and Ku using chemical inhibitors, such as Scr7, NU7441 and KU-0060648, or by using a siRNA knock-down approach [74-76]. One study screened 4,000 small molecules and found two molecules, L755507 and Brefeldin-A, which could enhance HDR activity [77]. The other strategies to increase insertion efficiency are utilising the more efficient NHEJ and MMEJ pathways by providing the donor suitable for these repair mechanisms, such as those in the HITI and PITCh methods discussed earlier (Figure 1.4) [58-60]. Increased insertion efficiency can also be achieved by using modified oligo donors, such as phosphorothioate-modified oligonucleotides to slow down the process of ssODN degradation inside the cells [78], or by using donors containing selection markers to allow for enrichment when they are transfected into cultured cells [79].

Unfortunately, efficient cleavage mediated by simple RNA-DNA binding of CRISPR/Cas9 system comes at a cost: unwanted cleavage at non-target sites (off-targets). Fu et al. [80, 81] observed that CRISPR/Cas9 could target sites with few mismatches to the on-target sequences. The mismatch positions also determine the likelihood that a site will be a potential off-target, with mismatches at positions distal from the PAM likely to be better tolerated by Cas9. They also found that SpCas9 appears to tolerate NAG PAM sequences, which should be considered when choosing gRNA sequences with reduced off-target activities. To assess the off-target effects of Cas9 cleavage accurately and globally, some studies have performed unbiased genome-wide off-target analysis using techniques such as IDLV-capture, GUIDE-seq, BLESS, digenome-seq, HTGTS and CIRCLE-seq. Genome-wide unbiased off-target analysis revealed that off-targets in the CRISPR/Cas9 system were more significant than was previously thought. Off-targets could be detected at sites with up to six mismatches, at sequences with a 1 bp bulge mismatch, and even at sites with non-canonical PAM sequences, such as NAG, NAA, NGT, NGC and NCG sequences [69, 82-86].

Since off-targets can be undesirable in experiments, particularly in experiments for therapeutic uses, scientists have developed ways to reduce off-targets and increase specificity. One way is to choose unique targets with minimal potential off-targets through computational off-target predictions. While choosing 20 bp target sequences that contain NGG PAM sequences is an easy task, choosing targets that are not only unique in the genome, but also contain many mismatches to avoid mis-binding by CRISPR/Cas9, is challenging. Therefore, scientists have developed gRNA design tools that can predict potential off-targets by computationally gathering sites with few mismatches to the on-target sequences. Many software applications have been generated for designing gRNAs, including the commonly used tool generated by Feng Zhang's lab at MIT http://crispr.mit.edu/. A complete list of the available CRISPR design tools can be found from this link http://goo.gl/R0gANl. One of the newest tools that claims to better predict off-targets and incorporate comprehensive scoring of other aspects, such as on-target and micro-homology, is CRISPOR at http://crispor.org/ [87]. Although these CRISPR design tools help to reduce potential off-targets, they cannot guarantee that gRNAs with good off-target scores will prevent off-target effects. Genome-wide off-target studies have indicated that many of the detectable off-targets were missed by CRISPR design tool predictions, and thus off-targets should be determined experimentally in an unbiased way [82].

It has been shown that high concentrations of Cas9 and gRNAs create more off-targets. Therefore, some studies suggest delivering CRISPR/Cas9 components in RNP or RNA form instead of in DNA form because of their short half-lives. When expressing Cas9 and gRNAs from the DNA form, off-target effects can be minimised by reducing the amount of Cas9 and gRNA expression; for example, by transient transfection of low amounts of plasmids (which might also reduce the on-target efficiency), or by controlled expression, such as by using a drug inducible strategy or restricting Cas9 expression by using spatially- and temporally-specific promoters [61, 88-91].

Other studies have suggested gRNA modifications as an alternative way to reduce off-targets. Keith Joung's lab showed that truncated gRNAs (17–18 nt instead of 20 nt) resulted in reduced off-targets when tested by unbiased genome-wide off-target analysis [82, 92]. Another study showed reduced off-targets when adding two G nucleotides to the 5' end of gRNA sequences [93].

In addition, Cas9 itself has been modified to reduce off-target effects. One of these modifications involved altering Cas9 to create a nickase rather than a nuclease. To induce a DSB, Cas9 relies on two nuclease domains, HNH and RuvC, which cut the gRNA target and non-target strands, respectively (Figure 1.5A). By slight base substitutions to either the RuvC (D10A) or HNH domain (H840A), the nuclease activity of these domains can be inactivated. Inactivation of one nuclease domain generates Cas9 with nickase function, meaning that it cuts only one strand of DNA, which can efficiently be repaired without inducing mutations in cellular systems (Figure 1.5A and B). This nickase activity can be exploited to induce HDR without a break. However, the efficiency is very low compared to nuclease-mediated HDR, although this can be slightly improved by using asymmetric ssODN donors [28, 94, 95]. When Cas9 nickase is delivered with two closely spaced gRNAs targeting different strands (pairing) (Figure 1.5B), the two nickases will generate DSBs, which leads to mutations. Since pairing is compulsory for the dual-nickase to induce DSBs while the single-nickase lacks mutagenic outcomes, the paired-nickase strategy can dramatically reduce off-targets [28, 83, 93, 95].

This pairing strategy can also utilise another modification of Cas9, known as dCa9-Fok1. In dCas9-Fok1 technology, both nuclease domains of Cas9 have been inactivated by mutations D10A/H840A, thus resulting in catalytically dead Cas9 (dCas9), which cannot cleave DNA but still retains RNA-guided DNA binding activity. This dCas9 is then fused to the Fok1 nuclease (the nuclease used in ZFN and TALEN technologies), which is able

to induce DSBs when dimerised. Studies have shown that paired-dCas9-Fok1 guided by two gRNAs can induce mutations efficiently, while reducing mutagenic effects at off-target sites [96, 97].

Guided by the structure of the Cas9 protein, two groups have engineered SpCas9 to improve its specificity, thereby reducing off-target effects. The first group generated Cas9 with enhanced specificity, eSpCas9, by mutating Cas9 at K848A, K1003A and R1060A in the expectation that this would reduce helicase activity by weakening the interaction between Cas9 and the non-target DNA strand. The second group attempted to reduce the interaction between Cas9 and its target strand by creating Cas9 with high fidelity, SpCas9-HF1, which harbours mutations N497A, R661A, Q695A and Q926A. Both groups showed that their engineered Cas9 could decrease or even eliminate off-targets as measured by unbiased genome-wide off-target analyses while maintaining good cleavage activities at on-target sites [98, 99].

1.6. Variations of the CRISPR/Cas system

The most commonly used CRISPR/Cas system is SpCas9, which natively recognises NGG PAM sequences. NGG sequences are abundant in the human genome, occurring on average every 8–12 bp, which makes it relatively simple to identify Cas9 target sites [27, 80]. However, some applications require cleavage at certain positions, such as those involving HDR, which is most efficient when the cleavage site located in close proximity to the sequence to be modified. This can be a problem if an NGG PAM sequence is not available. Therefore, it is necessary to expand the number of editing tools. This can be achieved by identifying additional CRISPR endonucleases with different PAM specificities. Fortunately, the CRISPR/Cas9 system is found not only in *Streptococcus pyogenes* (Sp), but is also present in many other bacteria/archaea. Some Cas9 orthologues

that have been characterised include those from *Francisella novicida (Fn)*, *Staphylococcus aureus (Sa), Streptococcus thermophilus (St)* and *Neisseria meningitidis (Nm)*, which all recognise different type of PAM sequences (apart from *Fn*) and thus can broaden the number of targetable regions in the genome (Figure 1.6) [3, 4, 69]. Furthermore, some have smaller amino acid size, meaning that delivery using versatile AAV vectors with limited packaging capacity becomes possible [69].

A recent development in the genome engineering field is the discovery of a novel RNA-guided endonuclease, Cpf1, which is distinct from Cas9. The Cpf1 from *Acidaminococcus sp.* (AsCpf1) and from *Lachnospiraceae bacterium* (LbCpf1) have been shown to be capable of inducing mutations in mammalian genomes, with comparable efficiency to SpCas9 [100-102]. There are several interesting differences between the Cpf1 and Cas9 systems. The PAM requirement for Cpf1 is TTTN and is located at the 5' end of target sequences, in contrast to Cas9 system, where the PAM is located at the 3' end (Figure 1.6). Unlike the native CRISPR/Cas9 system, which is guided by two short RNA components (crRNA and tracrRNA), native Cpf1 only uses one short crRNA (~42 nt) without the requirement for a tracrRNA. While Cas9 cleaves DNA at a site near the PAM, generating blunt-ended products, Cpf1 cleaves sequences 18–23 bases away from the PAM sequence in a staggered conformation, generating 5' overhang products (Figure 1.6). These characteristics have numerous advantages for genome editing. The unique PAM requirement may be able to target thymidine-rich genomic regions, thus expanding the targetable sites. The short gRNA required by Cpf1 makes it less expensive when commercially purchased. Cpf1 can also cleave its own pre-crRNA to generate mature crRNAs, therefore allowing multiplex expression of the gRNAs from a single promoter [103, 104]. Since the cutting sites of Cpf1 are further away from the PAM sequences and are located outside the target sequence (if a 20 nt gRNA is used), it is thought that mutations would also be generated around the cutting sites that do not modify the target

sequence and make it prone to Cpf1 re-targeting. This means that the process of re-cleavage and repair takes more rounds compared to Cas9 cleavage. Multiple rounds of re-cleavage by Cpf1 due to NHEJ failure to modify the target sequence theoretically increases the chance of HDR repair when a DNA donor is provided, thus improving knock-in (KI) efficiency [100, 105]. Furthermore, Cpf1 has been shown to generate less off-targets compared to Cas9 [106, 107].

Besides screening for other natural RNA-guided endonuclease systems that cut different PAM sequences to broaden the range of targets, scientists have also engineered Cas9 so it can recognise different or relaxed PAM sequences. It was discovered that SpCas9 harbouring the mutations D1335V/R1335Q/T1337R (SpCas9-VQR), D1135E/R1335Q/T1337R (SpCas9-EQR) and D1135V/G1218R/R1335E /T1337R (SpCas9-VRER) could recognise NGA, NGAG and NGCG PAMs, respectively [98]. Engineering the FnCas9 by mutations E1369R/E1449H/R1556A (FnCas9-RAH) relaxes the PAM recognition from NGG to YG despite a low targeting efficiency when tested by mouse zygote injections [108]. Relaxed PAM recognition was also achieved by engineering SaCas9 with mutations E782K/N968K/R1015H (SaCas9-KKH), which led to the recognition of the more flexible NNNRRT PAM compared to the original PAM NNGRRT (Figure 1.6) [109].

Recently a novel system, NgAgo from *Natronobacterium gregoryi*, has sparked interest in the genome editing field since it was reported to induce DSBs by an endonuclease guided by short oligos without a requirement for PAM sequences [110]. However, this new system cannot be replicated by other scientists [111, 112], which led to the original paper being retracted.

Figure 1.6 | CRISPR systems. The commonly used CRISPR system is from SpCas9, which recognises NGG PAM sequences and generates blunt-ended DSB. Alternative CRISPR systems from different species can be utilised for genome editing purposes and recognise different PAM sequences. Note that CRISPR/Cpf1 generates staggered breaks away from the PAM. Cas9 can also be engineered to recognise different or relaxed PAM sequences. Adapted from Komor et al. [3].

1.7. Catalytically dead Cas9 for non-cleaving applications

As mentioned above, the nuclease domains of SpCas9 (RuvC and HNH) can be inactivated without abrogating its binding ability, producing catalytically dead Cas9 (dCas9) [26]. By fusing dCas9 with other proteins, scientists are able to direct proteins of interest to specific loci through gRNA-dCas9 binding activity. dCas9 has been fused with transactivator domains, such as VP64 and P65, for targeted gene activation, called CRISPR activation (CRISPRa) [28, 113-116], or with repressor domains, such as KRAB, to repress transcription of target genes, known as CRISPR interference (CRISPRi) [117, 118]. Targeted epigenetic changes of chromatin can be achieved by fusion of dCas9 with methyltransferase (DNMT3A), DNA demethylase (Tet1) or acetyltransferase (p300) [119-124]. Recent studies fused dCas9 with the rat cytidine deaminase rat APOBEC1 to performing genomic nucleotide substitution without DSB and HDR (cytidine deaminase catalyses C $\rightarrow$ T or G $\rightarrow$ A exchange. dCas9-APOBEC1 was shown to be able to produce these point mutations in a targeted manner. The base editing activity was improved by using a fusion of APOBEC1 with Cas9-nickase (H840A) and the uracil glycosylase inhibitor (Base Editor 3, or BE3). While the uracil glycosylase inhibitor inhibits the process of base excision repair (BER) that removes uracil from the DNA, the use of Cas9-nickase can stimulate mismatch repair, which will assist nucleotide replacement [125, 126]. Furthermore, dCas9 can also be fused to fluorophores to enable the visualisation of DNA in a sequence-specific manner. More interestingly, this sequence-specific labelling can be performed in live cells. Ma et al. have developed the CRISPRainbow technique, which can visualise six different loci simultaneously, greatly improving chromosomal imaging techniques [127, 128].

1.8. Biomedical applications of CRISPR/Cas9 genome editing

Since CRISPR/Cas9 was shown to be useful for genetically manipulating mammalian cells [27, 129], researchers have used this technology to alter the genomic sequences of their species of interest in many ways. Researchers mainly use CRISPR/Cas9 to knock out genes in cells to study gene function. Certain modifications designed to mimic mutations found in human diseases can also be created using the CRISPR/Cas9 system to model and study human diseases. The insertion of fluorescent markers or a HA/Flag tag to label genes of interest is useful to track their expression or for immunoprecipitation purposes. In addition, many researchers aim to develop molecular therapies for a range of diseases. Therefore, many experiments have been conducted to show that CRISPR/Cas9 is able to repair mutations in primary or stem cells, including induced pluripotent stem cells (iPScs), which possess the ability to differentiate into many kinds of tissues. The expectation of these kinds of experiments is that the edited cells can be used to replace faulty cells [1, 2, 4, 6-8, 130].

One important application of CRISPR/Cas9 genome editing is for the generation of genetically modified animal models, particularly mice, to study genetic function or model diseases in vivo. Unlike the generation of mutant mice using conventional gene targeting in ES cells, which may take years, the generation of genetically modified mice using CRISPR/Cas9 can be accomplished quickly and relatively easily (Figure 1.7). This involves the microinjection of CRISPR/Cas9 components into fertilised mouse zygotes to edit the genome before transplanting the zygotes to pseudo-pregnant females to generate founder animals. Generating small indels that may constitute frameshift alleles can reach nearly 100% efficiency. Targeting multiple genes for simultaneous knockout is also possible using the CRISPR system with simultaneous injection of multiple gRNAs.
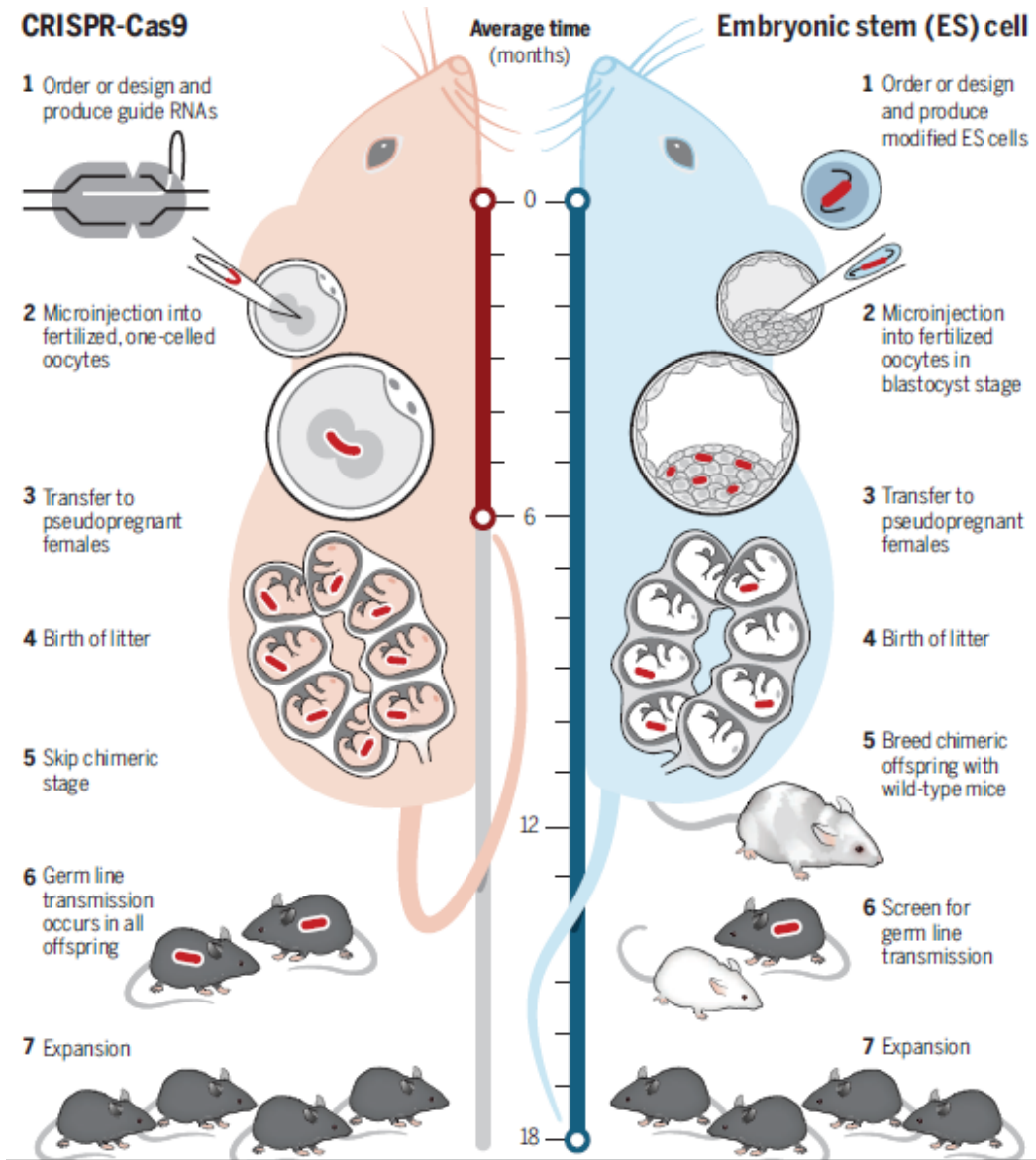
Figure 1.7 | Comparison of genetically modified mouse generation using CRISPR/Cas9 technology (left) and an approach using gene targeting in mouse ES cells (right). CRISPR/Cas9 injection to mouse zygotes can efficiently induce mutations and thus rapidly generate mutant mice. Adapted from Cohen [9].

Making small changes, such as point mutations or inserting an epitope tag using a ssODN donor is less efficient but readily achievable. However, generating large insertions with a plasmid donor via HR can be problematic and is difficult to achieve in some cases. Generating conditional KO mice with Cre-Lox system, an approach suitable for targeting genes important in embryonic development, is also still inefficient using CRISPR/Cas9 technique. This limitation forces some researchers to go back to the ES cell gene targeting approach when generating conditional KO mice or mice containing large insertions. Despite these current limitations, the simplicity offered by the CRISPR/Cas9 system is undoubtly a breakthrough in the mouse transgenesis field [9, 35, 51, 54]. Further optimisation will likely lead to increased efficiency of knock-in approaches.

CRISPR/Cas9 genome editing is envisioned to be used for ex vivo or in vivo disease therapies, particularly for diseases caused by mutations (Figure 1.8). Studies of the potential of the CRISPR/Cas9 system for therapeutic use have shown promising results. Correction of mutations by CRISPR/Cas9 can be performed in zygotes to generate a healthy individual. This has been tested in mice by microinjection of zygotes to treat diseases. For example, the CRISPR/Cas9 components were injected into a zygote heterozygous for a dominant-negative point mutation in the *CRYGC* gene, which causes cataracts. By targeting Cas9-mediated DSB to the mutant allele only, the defective point mutation was repaired by copying the correct allele (inter-homologue repair), resulting in cataract-free mice [55]. This inter-homologue repair mechanism is also utilised to correct heterozygous mutation in human preimplantation embryos using CRISPR/Cas9 platform [32]. In the study, scientists targeted *MYBPC3* gene in which the heterozygous mutation in human can lead to hypertrophic cardiomyopathy. By injecting CRISPR/Cas9 RNP to the human fertilised eggs resulting from in vitro fertilisation between $MYBPC3^{WT/\Delta GAGT}$ male and $MYBPC3^{WT/WT}$ female, they were able to fix the mutant alleles efficiently. Despite robust embryo screening to select the healthy embryos using preimplantation genetic

diagnosis (PGD), their approach could potentially be useful to improve the number of healthy embryos for more successful pregnancy rate [32].

Zygote injection of CRISPR/Cas9 and donor template has been shown to correct the point mutation that causes Duchenne muscular dystrophy (DMD) in a mouse model of DMD [131]. Post-natal treatment for this mouse model was achieved by AAV-mediated delivery of SaCas9 or SpCas9 and gRNAs to skeletal and cardiac muscle cells to delete the mutated exon in an exon skipping strategy. This strategy successfully led to expression of dystrophin and improved muscle function [132-134].

Defective point mutations in the ornithine transcarbamylase (*OTC*) gene, which cause metabolic disease, were shown to be corrected in the livers of newborn mice by intravenously infusing AAVs expressing Cas9, gRNA and a donor DNA. Mutation corrections occurred in 10% of hepatocytes, which resulted in increased survival when the mice were challenged with high-protein diet [135]. CRISPR/Cas9-mediated mutation corrections have also been performed in mice with the liver disease hereditary tyrosinaemia type I (HTI), which results from a point mutation in the *Fah* gene [71, 136]. By systemic delivery of Cas9 mRNA coated in lipid nanoparticles and sgRNA/HDR template by AAV, researchers were able to correct 6% of hepatocytes and cure HTI mice [136].
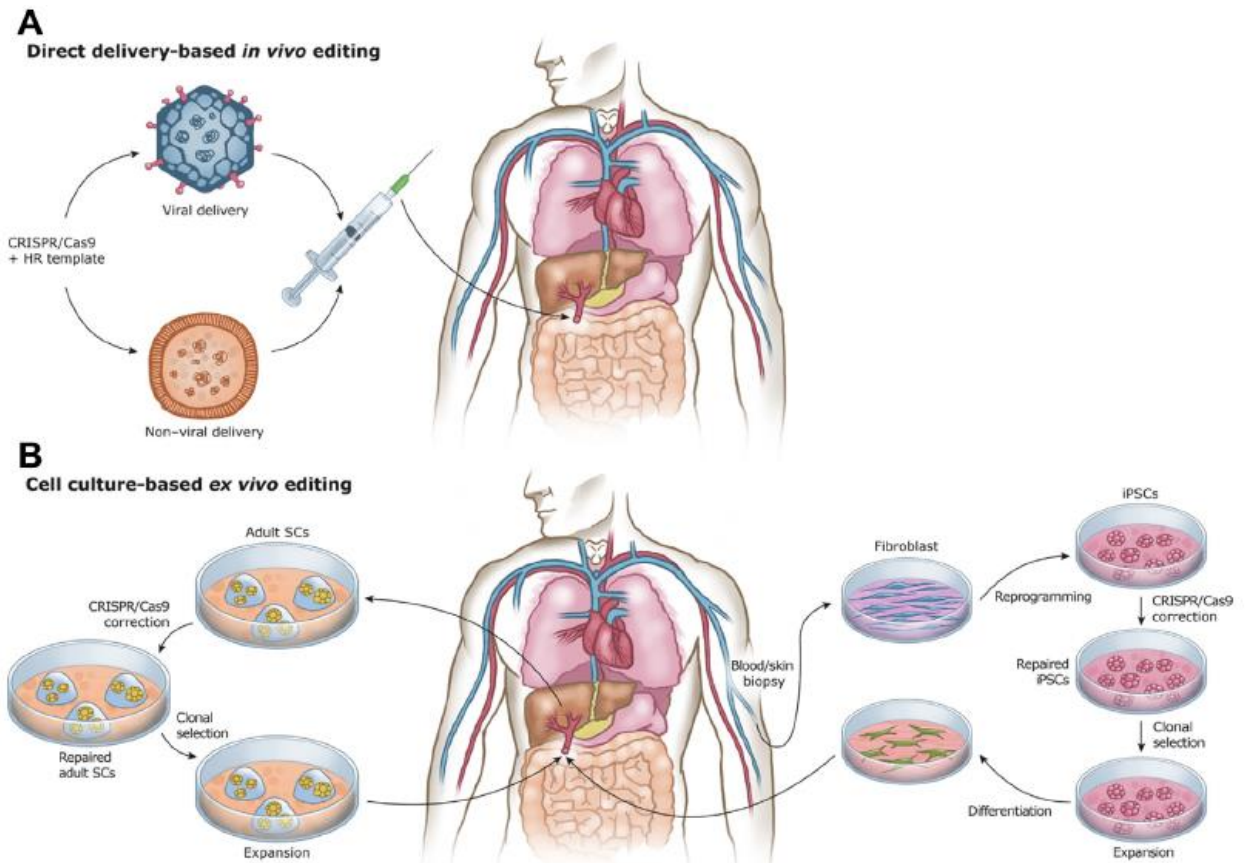
Figure 1.8 | Schematic of in vivo and ex vivo CRISPR/Cas9 therapy. CRISPR/Cas9 components can be injected as a viral vector or via nanoparticles to reach targeting tissues for in vivo therapeutic applications. Cells from patients can also be harvested and genetically modified/repaired in vitro before transplanted back into the patient. Cells, such as fibroblasts, from patients can also be induced to become iPSCs, then genetically modified in vitro. Modified iPSCs can then be differentiated to become the desired type of cells for transplantation. Adapted from Savic and Schwank [137].

CRISPR/Cas9 has been tested for the correction of a rat model of retinitis pigmentosa (a visual degeneration disease), where mice a carry a homozygous 1.9-kb deletion from intron 1 to exon 2 in the *Mertk* gene. Retinal injection of AAVs carrying CRISPR/Cas9 and donor DNA resulted in correction of the mutation in the eyes, expression of the functional *Mertk* gene and improved visual responses [60]. AAV delivery of SpCas9 or SaCas9 to mouse liver, targeting the PCSK9 gene for loss of function, successfully lowered the cholesterol levels in mutant mice, which could be useful to protect against cardiovascular disease [69, 138]. In another model, electroporation of RNP and a DNA donor into haematopoietic stem/progenitor cells (HSPCs) could correct the point mutation in the *HBB* gene that causes sickle cell disease [139, 140]. The edited HSPCs could be maintained for 16 weeks when engrafted into immunocompromised mice [139].

CRISPR/Cas9 genome editing could be used to knock out the *CCR5* gene in T cells and thus confer resistance to HIV infection [141-143]. This strategy has been used with ZFN in a clinical trial, with promising results. CD4+ T cells from patients with HIV were edited ex vivo by ZFN to knock out the *CCR5* gene, and the cells were transplanted back into the patients [144]. Eliminating the HIV genome via CRISPR/Cas9 cuts has also been proposed as a strategy for HIV therapy [145-147]. Cancer treatment with CRISPR/Cas9 genome editing could be achieved by cancer immunotherapy; for example, by knocking out the *PD1* gene in the patient's T cells ex vivo, followed by engraftment of the edited cells in the patient. These PD1-KO T cells are expected to robustly kill the cancer cells [148]. Killing cancer cells could also be realised by CRISPR/Cas9-mediated insertion of genes that cause apoptosis. A recent study used the paired-nickase strategy to insert the *HSV1-tk* gene into a cancer-specific genomic locus by adenovirus delivery of the CRISPR/Cas9-nickase components and the HSV1-tk construct. Treatment with ganciclovir forces the cancer cells containing the HSV1-tk cassettes to suicide, thus decreasing tumour size, as shown in mouse xenograft experiments [149]. Furthermore, CRISPR/Cas9 can be harnessed to kill

harmful bacteria in vivo by delivering bacteriophage carrying CRISPR/Cas9 that uniquely cleave the bacterial genome or plasmids in certain bacteria to cause lethality or antibiotic re-sensitisation [150-152].

The ability of CRISPR/Cas9 technology to easily knock out genes by frameshifting the open reading frame (ORF) via indel mutations has been used for genome-wide forward genetic screening to identify genes that play a role in producing a phenotype of interest. Lentiviral vectors carrying Cas9 and gRNA libraries targeting all genes, or clusters of genes, for loss of function are pooled and cells are infected in the expectation that only one gRNA is present in each cell to knock out a certain gene. The pooled cells are then subjected to treatment with a certain phenotypic consequence; for example, resistance to certain drugs. Cells that display the phenotype of interest are checked by sequencing to determine which gRNAs are enriched (or which genes have been knocked out) to give rise to the phenotype (Figure 1.5F) [153-157]. Genome-wide screening using dual gRNA libraries could be conducted to study the impact of deletion of regulatory elements or lncRNA, which requires larger deletions rather than small indels, or to uncover the interaction between two genes that give rise to a certain phenotype via its ability to simultaneously knock out two different genes [158-160].

CRISPR/Cas9 can be harnessed to create a gene drive system. Using gene drive (sometimes called the 'mutagenic chain reaction' or MCR), desired genetic modifications can be rapidly spread through a population by sexual reproduction. Normally, in biallelic organisms, a parent will transmit one copy of its genes to the offspring, while another copy comes from the other parent. Crossing a WT parent with a parent carrying one copy of a modified gene will result in 50% of the offspring being heterozygous for the modified gene. Further crossing of the heterozygotes with WT will again produce only heterozygotes in 50% of offspring (Figure 1.9). Interestingly, using CRISPR/Cas9 gene

drive system, crossing of a WT parent with a parent with one copy of a modified gene will produce homozygotes in all offspring, instead of heterozygotes, breaking the Mendelian law of inheritance. Further crossing of the homozygotes with the WT will always produce 100% homozygous offspring, thus rapidly spreading the modified genes through the population (Figure 1.9). This is possible because the CRISPR/Cas9 gene drive system will generate DSB in the WT allele during conception or specifically in the germ cell lineage and these breaks will be mostly repaired by the HR mechanism, which uses the sister chromatid containing the modified allele and the gene drive cassette as the HR template, thus copying the modified genes and the gene drive cassette to the WT allele (Figure 1.9). This gene drive system has been tested in flies, yeast and mosquitoes, and was able to spread the gene of interest in the population with an efficiency of almost 100% [161]. This is opens the possibility of controlling populations that are harmful, such as mosquitoes that cause malaria and dengue, by spreading genes/cassettes that cause gender imbalance or sterility thus reducing their populations [162]. Another use for the gene drive system may be to eradicate diseases, such as malaria, by spreading anti-malaria genes within mosquito populations [163].
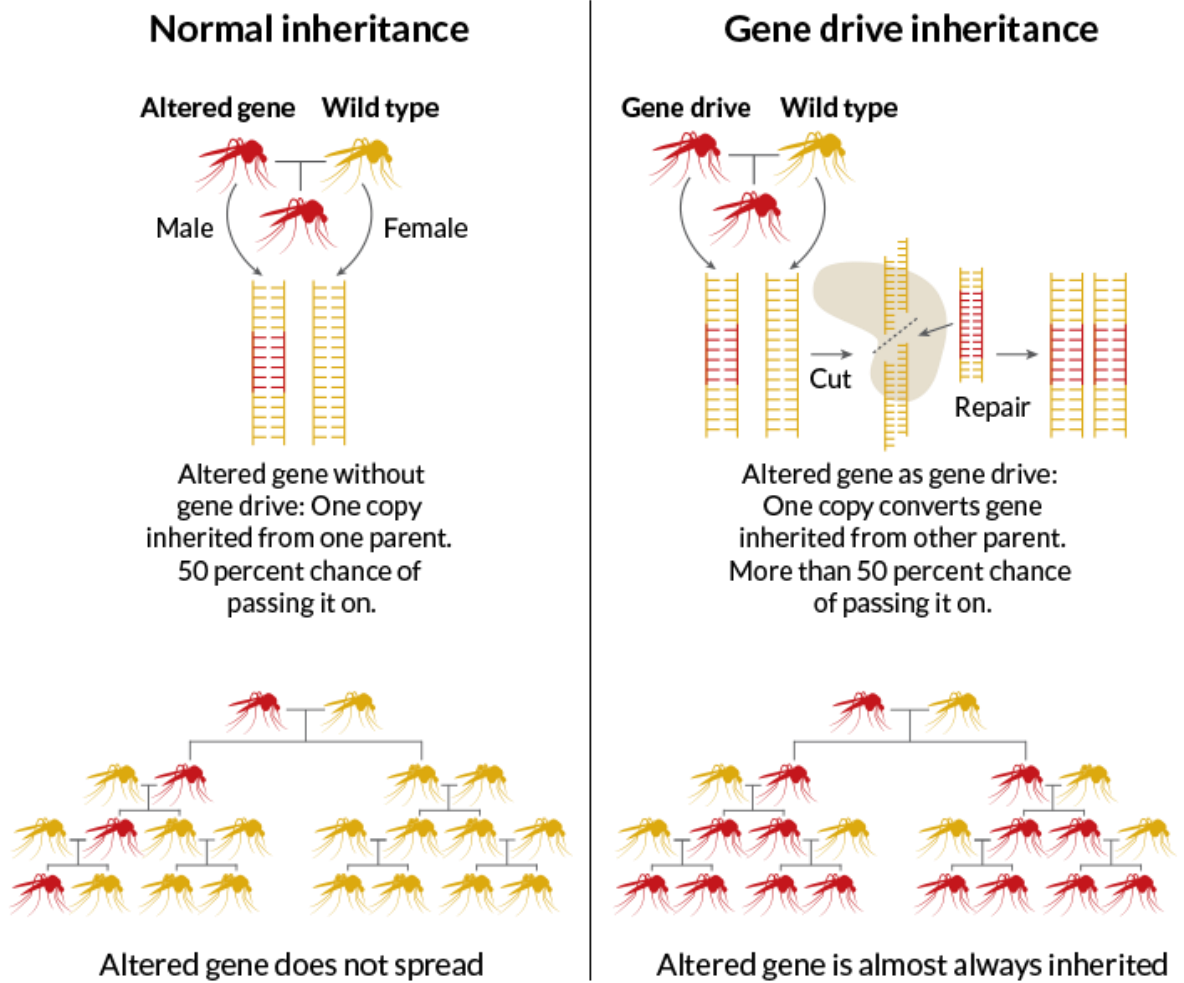
Figure 1.9 | Schematic of gene drive (mutagenic chain reaction). Normal Mendelian inheritance (left figure) cannot spread genetic sequences of interest (red mosquitoes) rapidly. While using CRISPR/Cas9 gene drive system (right figure), the WT alleles will always be forced to convert to the intended genetic modification and thus modified genes can rapidly spread through a population. Adapted from https://www.sciencenews.org/article/gene-drives-spread-their-wings

1.9. Project rationale

CRISPR/Cas9 genome editing technology has produced significant advances in life science research and holds great potential for disease modelling and therapeutics. The development of CRISPR/Cas9 technology has been very rapid. However, there is still room for improvements. The study reported in this thesis aims to improve the usefulness of CRISPR/Cas9 technology. The specific aims include:

1. Applying CRISPR/Cas9 technology for gene swapping in mice to study functional redundancy between two genes.
2. Developing strategies for deleting an entire chromosome using the CRISPR/Cas9 system.
3. Developing versatile plasmid vectors that help researchers to perform CRISPR/Cas9 genome editing.
4. A detailed characterisation of mutation outcomes after CRISPR/Cas9-mediated breaks.

Chapter 2:

Functional equivalence of the SOX2 and SOX3 transcription factors in the developing mouse brain and testes

2.1. Summary

One of the major advantages of CRISPR/Cas9 technology lies in its ability to generate genetically modified mice quickly and easily. In this chapter, this property is exploited to create a unique mouse model to study the functional redundancy between two genes is described. The study presented in this chapter has been published in *Genetics* as a paper entitled 'Functional equivalence of the SOX2 and SOX3 transcription factors in the developing mouse brain and testes'.

This paper addresses a long-standing question in the developmental biology field: whether the SOXB1 transcription factor family members (SOX1, SOX2 and SOX3) function redundantly and can compensate for the loss of another member, leading to phenotypic rescue in individual KO mice. Although they are structurally and functionally similar, until the current study there was no strong and direct evidence to claim that they were functionally equivalent. The most robust and physiological approach to study functional redundancy between two genes is by performing gene replacement in vivo. Therefore, we attempted to generate a mouse model where the *Sox3* ORF was replaced with that of *Sox2*. These mice, called $Sox3^{Sox2KI}$, lack Sox3, but express extra Sox2 in the same spatio-temporal pattern as that of Sox3. If the ectopic Sox2 can rescue the Sox3-null phenotype, this means that Sox2 and Sox3 function redundantly.

Previous efforts to generate this mouse model using gene targeting in ES cells failed due to unknown reasons. This is not unusual, as the generation of mouse models by this conventional technique requires many steps that are prone to failure. CRISPR/Cas9 technology was therefore chosen for generating the $Sox3^{Sox2KI}$ mice. Despite a low KI frequency, $Sox3^{Sox2KI}$ mice were successfully generated and used to answer the research question on the functional equivalence of SOX2 and SOX3.

We showed that Sox3-null phenotypes, such as testis defects and pituitary dysmorphology, are largely rescued in $Sox3^{Sox2KI}$ mice. Microarray analyses comparing genome-wide expression in $Sox3^{KO}$ vs WT testes revealed the presence of widespread genetic dysregulation in Sox3-null testes. These $Sox3^{KO}$ genetic alterations are normalised in $Sox3^{Sox2KI}$ testes, establishing that the Sox3-null phenotypic rescue by Sox2 results from its ability to rescue at the molecular level. Overall, CRISPR/Cas9 technology has facilitated this study, which provides robust evidence of functional equivalence between the SOX2 and SOX3 transcription factors.

Statement of Authorship

Title of Paper	Functional Equivalence of the SOX2 and SOX3 Transcription Factors in the Developing Mouse Brain and Testes
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Adikusuma, F., Pederick, D., McAninch, D., Hughes, J. & Thomas, P. Functional Equivalence of the SOX2 and SOX3 Transcription Factors in the Developing Mouse Brain and Testes. *Genetics* (2017)

Principal Author

Name of Principal Author (Candidate)	Fatwa Adikusuma
Contribution to the Paper	Designed and conceived the study, conducted most of the experiments, analysed data, generated the Figures and wrote the manuscript.
Overall percentage (%)	90%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.

Signature		Date	8/8/2017

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

    i.   the candidate's stated contribution to the publication is accurate (as detailed above);

    ii.   permission is granted for the candidate in include the publication in the thesis; and

    iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Daniel Pederick		
Contribution to the Paper	Analysed data for Figure 1G and Figure 2D, reviewed and edited the manuscript.		
Signature		Date	15.7.2017

Name of Co-Author	Dale McAninch		
Contribution to the Paper	Contributed to Supplementary Figure 3.		
Signature		Date	

Name of Co-Author	James Hughes		
Contribution to the Paper	Designed and conceived the study, wrote the manuscript.		
Signature		Date	25/07/2017

Name of Co-Author	Paul Thomas		
Contribution to the Paper	Designed and conceived the study, supervised the study and wrote the manuscript.		
Signature		Date	8/8/2017

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i.   the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Daniel Pederick		
Contribution to the Paper	Analysed data for Figure 1G and Figure 2D, reviewed and edited the manuscript.		
Signature		Date	

Name of Co-Author	Dale McAninch		
Contribution to the Paper	Contributed to Supplementary Figure 3.		
Signature		Date	18 7 17

Name of Co-Author	James Hughes		
Contribution to the Paper	Designed and conceived the study, wrote the manuscript.		
Signature		Date	

Name of Co-Author	Paul Thomas		
Contribution to the Paper	Designed and conceived the study, supervised the study and wrote the manuscript.		
Signature		Date	

Functional Equivalence of the SOX2 and SOX3 Transcription Factors in the Developing Mouse Brain and Testes

Fatwa Adikusuma,[*,†,‡] **Daniel Pederick,**[*,†] **Dale McAninch,**[*,†] **James Hughes,**[*,†] **and Paul Thomas**[*,†,§,1]

*School of Biological Sciences and †The Robinson Research Institute, University of Adelaide, South Australia, Australia 5005, ‡Center for Biomedical Research (CEBIOR), Faculty of Medicine, Diponegoro University, Semarang, Indonesia 50271, and §South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia 5000

ABSTRACT Gene duplication provides spare genetic material that evolution can craft into new functions. *Sox2* and *Sox3* are evolutionarily related genes with overlapping and unique sites of expression during embryogenesis. It is currently unclear whether SOX2 and SOX3 have identical or different functions. Here, we use CRISPR/Cas9-assisted mutagenesis to perform a gene-swap, replacing the *Sox3* ORF with the *Sox2* ORF to investigate their functional equivalence in the brain and testes. We show that increased expression of SOX2 can functionally replace SOX3 in the development of the infundibular recess/ventral diencephalon, and largely rescues pituitary gland defects that occur in *Sox3* null mice. We also show that ectopic expression of SOX2 in the testes functionally rescues the spermatogenic defect of *Sox3* null mice, and restores gene expression to near normal levels. Together, these *in vivo* data provide strong evidence that SOX2 and SOX3 proteins are functionally equivalent.

KEYWORDS SOXB1 genes; CRISPR/CAS9 mutagenesis; gene swap

ONE of the driving forces for the evolution of complex life is the duplication of genes, chromosomes, or entire genomes, providing the genetic material upon which natural selection can operate. Evolutionary theory predicts that having duplicated, a gene pair will be relieved from selective constraints, thereby enabling the accumulation of genetic alterations that can alter protein function (Force *et al.* 1999; Lynch and Conery 2000). The consequences of this are thought to favor loss of one copy (nonfunctionalization). Alternatively, gene functions can be divided between the paralogues (subfunctionalization), or one copy can acquire a novel advantageous function (neofunctionalization) while the other copy retains its original function (Force *et al.* 1999; Lynch and Conery 2000). Under this paradigm, it is expected that shared function within a given tissue will not be preserved by natural selection, and should therefore be lost over time. This stands in contrast to many observations of genetic redundancy that have emerged in the age of molecular genetics as gene deletions in seemingly important genes routinely yield no, or mild, phenotypes, and appear to be compensated for by paralogous partner genes (Wagner 2005). Estimates suggest that as many as 10–15% of mouse gene knockouts may have no or mild phenotypes (Barbaric *et al.* 2007). What forces allow the persistence of genetic redundancy are unclear, but genetic robustness that acts to maintain and bolster important processes may play a role (Force *et al.* 1999; Wagner 2005; Barbaric *et al.* 2007).

Persistent genetic redundancy is particularly striking in the *SoxB1* subfamily, which consists of *Sox1*, *Sox2*, and *Sox3*. These genes share highly similar sequences, both within and, to a lesser extent, outside of the DNA-binding HMG box. Several studies suggest that SOXB1 proteins have similar, if not identical, functional capabilities. For example, over-expression of chick or mouse *SoxB1* genes in chick neural tube results in inhibition of neural differentiation with cells retaining a progenitor identity (Bylund *et al.* 2003; Graham *et al.* 2003). Similarly, mouse *Sox1* and *Sox3* are able to replace *Sox2* for reprogramming of iPS cells (Nakagawa *et al.* 2008). Loss of function studies also generally support functional equivalence, particularly in the developing CNS where the

41

SoxB1 genes exhibit extensive overlapping expression. For example, Sox3 deletion in mice results in relatively mild neural defects, indicating that SOX2 and/or SOX1 can compensate for the absence of SOX3 in most neuroprogenitor contexts. However, one notable exception is the infundibulum, a ventral evagination of the ventral diencephalon that is responsible for induction of the anterior pituitary primordium (Rathke's Pouch). Despite coexpression of Sox2 and Sox3, pituitary induction and development is severely compromised in Sox2 and Sox3 single mutants (Rizzoti et al. 2004; Kelberman et al. 2006). It is thought that this is due to reduced dosage of SOX2 or SOX3, as opposed to unique roles of these proteins (Zhao et al. 2012). However, to date, experimental approaches that distinguish between these possibilities have not been published.

Restricted zones of SoxB1 expression outside of the nervous system have also been described, many of which are in stem/progenitor cells of developing organs. For example, Sox3 is uniquely expressed in the spermatogonial stem/progenitor cells of the postnatal testes (Rizzoti et al. 2004; Raverot et al. 2005). Consistent with a model of limited subfunctionalization, more severe phenotypes occur in knockout mice at sites of unique expression. For example, Sox3 null mice have spermatogenic defects likely due to the absence of Sox1 and Sox2 (Raverot et al. 2005). However, it is not known whether SoxB1 genes are functionally interchangeable at these unique zones of expression.

Herein, we describe an in vivo gene swap experiment in which Sox3 open reading frame (ORF) was deleted and replaced with Sox2 ORF to investigate their functional similarities. We show that SOX2 can functionally replace SOX3 in both the developing pituitary and testes, thereby rescuing phenotypes associated with SOX3-null mice.

Materials and Methods

Generation of CRISPR/Cas9 modified mice

CRISPR gRNAs were designed either side of the Sox3 ORF (5′-CCTGATGCGTTCTCTCGAGC-3′ and 5′-GACAGTTACGGC CAAACTTT-3′) using CRISPR Design tool (http://crispr. mit.edu/) and generated according to the protocol described in Wang et al. (2013). gRNA IVT was performed using HiScribe T7 Quick High Yield RNA Synthesis Kit. Cas9 mRNA was generated by IVT using the mMESSAGE mMACHINE T7 ULTRA Transcription Kit (Ambion) from pCMV/T7-hCas9 (Toolgen) digested with Xho1. gRNAs and Cas9 mRNA were purified using a MEGAclear Transcription Clean-Up Kit (Ambion). Our previously published Sox3 targeting vector (Hughes et al. 2013) was modified to replace Sox3 ORF with Sox2 ORF. Cas9 mRNA (100 ng/µl), gRNAs (50 ng/µl each) and donor plasmid (200 ng/µl) were injected into the cytoplasm of C57BL/6N zygotes using a FemtoJet microinjector, transferred to pseudo pregnant recipients, and allowed to develop to term. Homology directed repair (HDR) from the vector resulted in the Sox3^Sox2KI mice carrying a neomycin resistance cassette 1 kb downstream from the Sox2-KI stop codon. The Sox3^Sox2KI mice also contain a 2 bp deletion in the 5′UTR at the upstream gRNA site, and a 1 bp in the 3′UTR at the downstream gRNA site, presumably as a result of CRISPR/Cas9 recutting after HDR.

Microarray analysis

Microarray expression profiling was performed using Affymetrix GeneChip Mouse Gene 1.0 ST Arrays on three Sox3 null and three Sox3^Sox2KI 2-week testes. A total of six wild type age matched samples was included, comprising two groups of three matched to the same genetic background as the Sox3 null and Sox3^Sox2KI samples. Two way-ANOVA, using batch as a factor, was used to identify the significantly regulated genes. ANOVA was performed comparing to matched WT samples and comparing to pooled WT samples with similar results. We have presented data comparing two pooled WT samples.

Sperm counting

Cauda epididymis were isolated and minced in 1 ml of 37° DMEM media. Sperm were allowed to disperse for 10–15 min at 37°; 10 µl of the resuspension was diluted with 10 µl of 1 M Tris pH 9.5 solution to immobilize sperm before counting with a hemocytometer.

Data availability

Microarray data has been submitted to Gene Expression Omnibus (GEO) with accession number GSE96805. All other reagents can be made available upon request. Supporting data can be found in Supplemental Material, File S1.

Results

Generation of Sox3^Sox2KI mice using CRISPR/ Cas9 mutagenesis

To investigate the functional redundancy within the SoxB1 subgroup, we replaced the Sox3 ORF with that of Sox2 while leaving the remaining native Sox3 flanking sequences including the promoter and untranslated regions (UTR) intact. This mouse model, which we refer to as Sox3^Sox2KI, therefore lacks SOX3, and expresses SOX2 from the Sox3 locus on the X-chromosome. To generate Sox3^Sox2KI mice, we initially modified an existing Sox3 KO targeting construct (Rizzoti et al. 2004; Hughes et al. 2013) by replacing the Sox3 ORF with Sox2. Attempts to generate Sox3^Sox2KI mice by conventional gene targeting in mouse ES cells failed to produce any chimeras despite multiple rounds of injections using germline competent cells (data not shown). To circumvent this issue, we employed CRISPR/Cas9 technology to generate Sox3^Sox2KI mice by zygotic injection of Cas9 mRNA, as well as gRNA pairs targeting either side of the Sox3 ORF, and the donor (targeting) plasmid (Figure 1A). PCR screening of 17 founders identified a Sox3^Sox2KI female harboring the intended replacement event (Figure 1B). The other allele of this Sox3^Sox2KI female, as well as seven other founders, lacked the entire Sox3 ORF due to deletion of the interval between the two gRNA cuts. These were used to generate Sox3-null mice (Sox3^KO). No gross
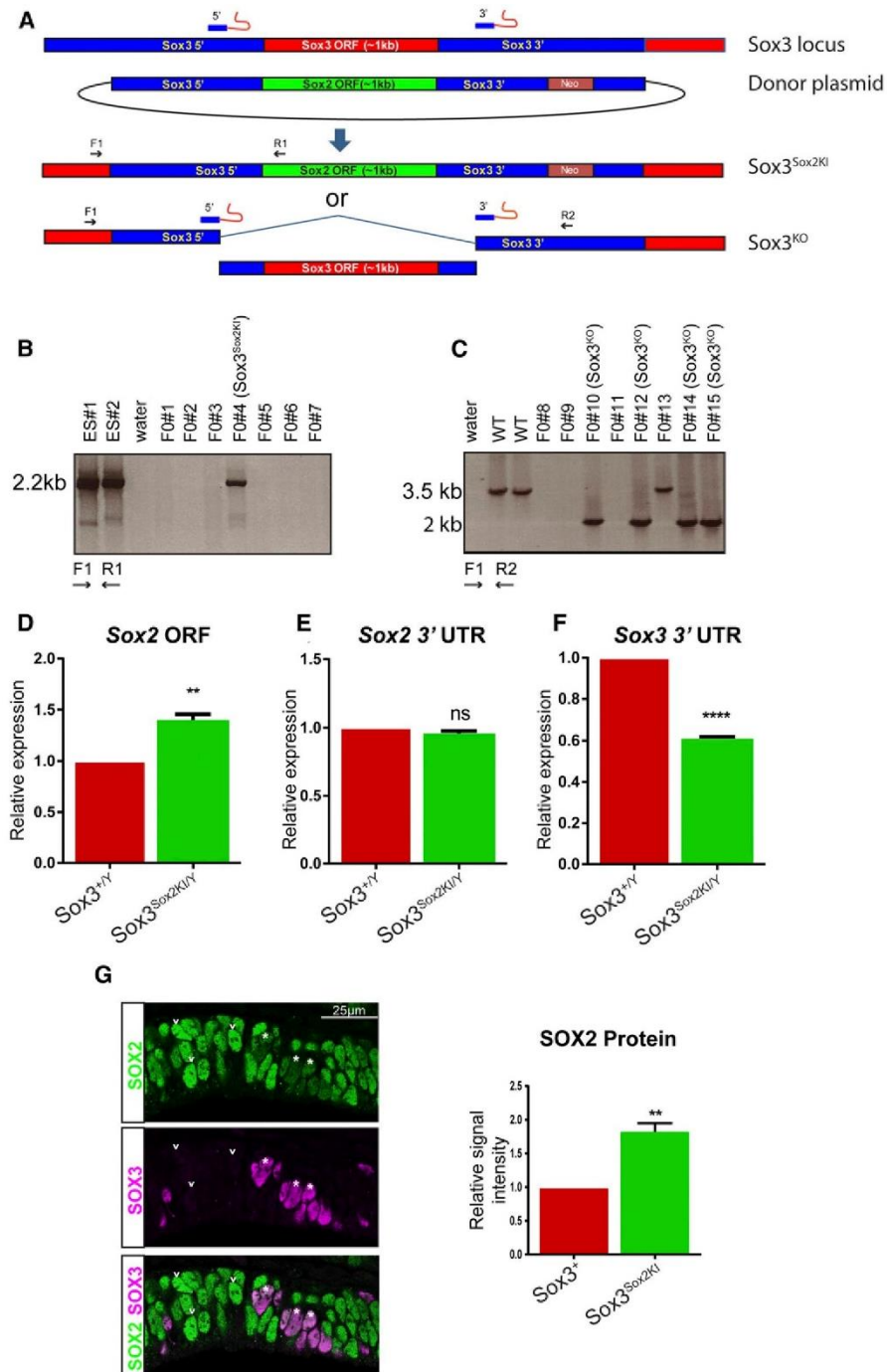
Figure 1 Generation of *Sox3^Sox2KI* and *Sox3^KO* mice with CRISPR mutagenesis. (A) Schematic showing the strategy for generation of the *Sox3^Sox2KI* mice. (B, C) PCR screening of founder mice. (D) *Sox2* transcript levels were measured in 11.5 dpc brains by qPCR with primers located in the *Sox2* ORF. (E) qPCR showing expression from the endogenous *Sox2* locus using primers located in the 3′UTR. (F) Transcription from the *Sox3* locus was reduced in *Sox3^Sox2KI/Y* as determined by qPCR with primers located in the *Sox3* 3′UTR. (G) SOX3 and SOX2 coimmunostaining in the ventral diencephalon of 12.5 dpc *Sox3^Sox2KI/+* embryos. Asterisks indicate SOX3 positive cells, and arrowheads indicate SOX3 negative cells. Quantification of SOX2 staining intensity in SOX3 positive and SOX3 negative cells (*n* = 3 embryos).

abnormalities were observed in *Sox3^Sox2KI* or *Sox3^KO* adult mice (data not shown), and their body weights were comparable with wild type (WT) littermates (Figure S1 in File S1).

To determine whether *Sox2* was expressed from the *Sox3* locus in the developing brain, we performed qRT-PCR using primers that amplify the *Sox2* ORF, the *Sox2* 3′UTR, and the *Sox3* 3′UTR. The level of *Sox2*-ORF-containing transcripts was significantly elevated in *Sox3^Sox2KI/Y* embryos *vs.* *Sox3^+/Y* (Figure 1D). *Sox2* 3′UTR-containing transcript levels were equivalent (Figure 1E), indicating that the lack of *Sox3* in *Sox3^Sox2KI* embryos does not result in a compensatory increase in *Sox2* transcript levels. Notably, the level of
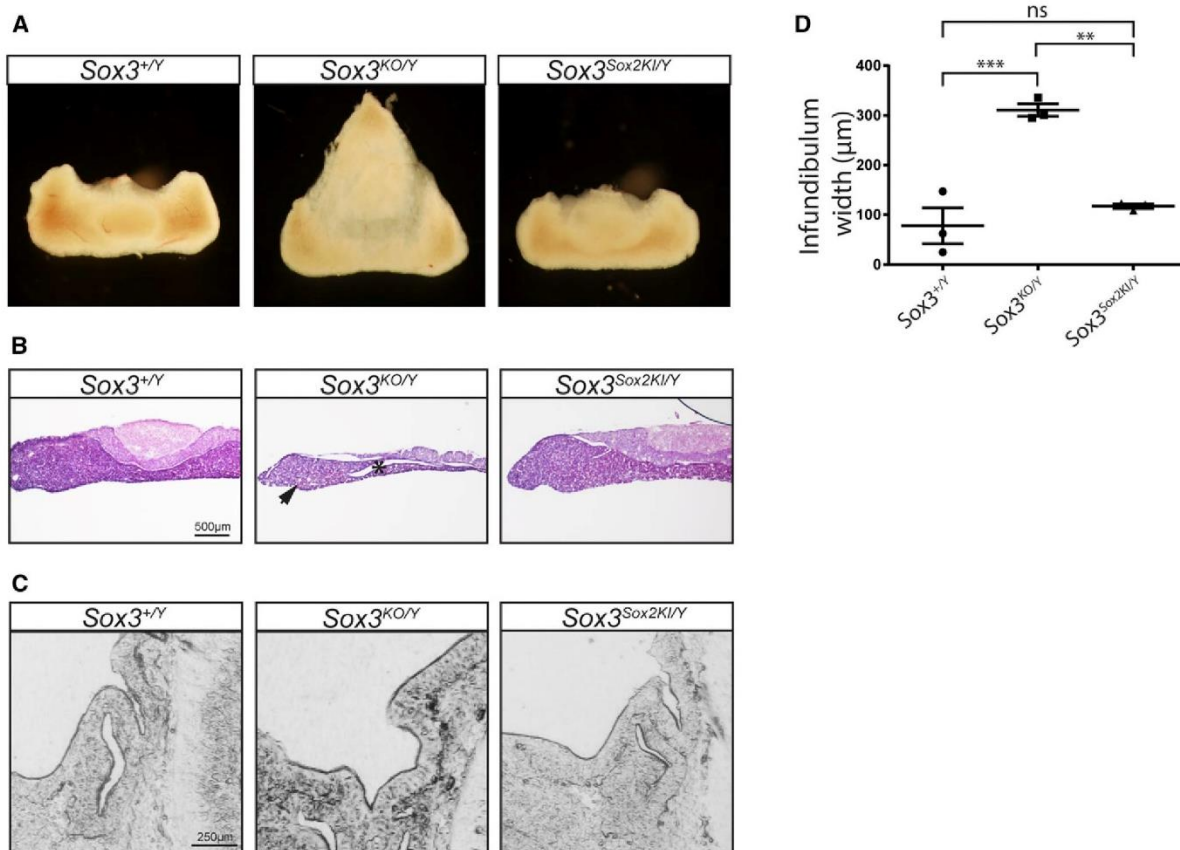
Figure 2 Increased expression of *Sox2* can rescue *Sox3* null pituitary defects. (A) Adult pituitaries showing dorsal displacement of the neural lobe in *Sox3^KO/Y* mice. (B) Haemotoxylin and eosin staining of 8-week-old pituitary coronal sections. Asterisk highlights abnormal clefting. Arrowhead indicates hypoplastic anterior lobe. (C) Phase contrast images of 12.5 dpc developing pituitaries (saggital sections). (D) Quantification of infundibulum width of 12.5 dpc embryos (*n* = 3). One-way ANOVA using Tukey's Multiple comparison test. Mean ± SEM.

transcripts from the *Sox3* locus is lower in *Sox3^Sox2KI* embryonic heads compared to *Sox3^+/Y* heads (Figure 1F). This reduction may reflect small differences in transcription rate caused by transgene elements, or reduction in the stability of the *Sox3^Sox2KI/Y* chimeric transcript.

To confirm that the observed increase in *Sox2* transcript levels in *Sox3^Sox2KI/Y* embryos resulted in increased protein levels, we performed immunostaining on the ventral diencephalon, where SOX3 and SOX2 are coexpressed in the infundibulum (Wood and Episkopou 1999; Rogers *et al.* 2013). We examined *Sox3^Sox2KI/+* embryos in which random X-inactivation results in a mixture of cells expressing either the *Sox2KI* or *Sox3* allele. SOX2 immunostaining intensity was significantly elevated in *Sox2KI* cells (SOX3 negative) in comparison to neighboring SOX3 positive cells, consistent with additional expression of SOX2 from *Sox3^Sox2KI* allele (Figure 1G). Collectively, these results demonstrate successful gene replacement in our *Sox3^Sox2KI* mice, and show that *Sox3* has been removed and replaced with a copy of *Sox2* that is regulated in a *Sox3*-specific fashion.

Rescue of pituitary induction defect in Sox3^Sox2KI mice

We next sought to assess whether SOX2 protein could functionally rescue *Sox3*-null phenotypes. Given that pituitary development is extremely sensitive to *SoxB1* gene dosage (Zhao *et al.* 2012), we examined adult pituitaries to determine whether the replacement of *Sox3* with *Sox2* was able to rescue *Sox3* null pituitary defects (Rizzoti *et al.* 2004). To assess rescue by *Sox2KI*, we collected pituitaries from 8-week-old mice. As expected, we found malformations in *Sox3^KO/Y* pituitaries; indeed, these were even more severe than previously reported on a mixed genetic background (Figure 2A) (Rizzoti *et al.* 2004; Hughes *et al.* 2013). The dorsal aspect of *Sox3* null pituitaries was triangular, such that the neural lobe was rostrally displaced forming a tenuous connection with the anterior lobe. Sectioning revealed hypoplasia of the anterior lobe, and deep clefting of the residual lumen of Rathke's Pouch in *Sox3^KO/Y* pituitaries, similar to previous reports (Figure 2B) (Rizzoti *et al.* 2004; Hughes *et al.* 2013). In contrast, the gross morphology of
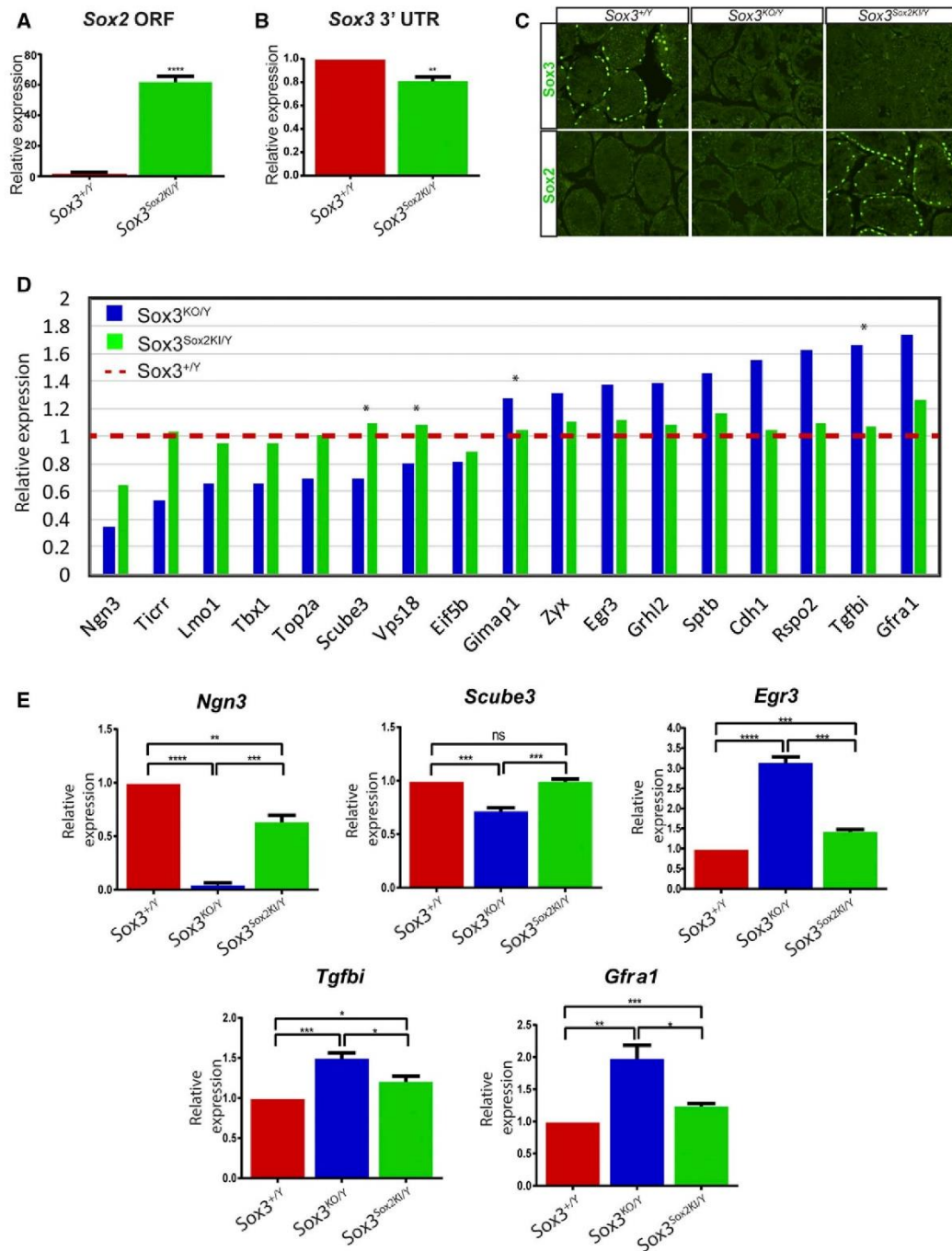
Figure 3 SOX2 regulates SOX3 target genes in the mouse testes. (A) *Sox2* qPCR analysis in 2 week *Sox3*$^{+/Y}$ and *Sox3*$^{Sox2KI/Y}$ testes [Student's two tailed unpaired *t*-tests (****<0.0001)]. (B) qPCR analysis of *Sox3* 3' UTR in 2 week *Sox3*$^{+/Y}$ and *Sox3*$^{Sox2KI/Y}$ testes [Student's two-tailed unpaired *t*-tests (**<0.01)]. (C) SOX2 and SOX3 immunostaining on *Sox3*$^{+/Y}$, *Sox3*$^{KO/Y}$ and *Sox3*$^{Sox2KI/Y}$ 4-week-old testes. (D) Microarray analysis was performed on 2 week testes from *Sox3*$^{+/Y}$, *Sox3*$^{KO/Y}$ and *Sox3*$^{Sox2KI/Y}$, and two-way comparisons were performed between *Sox3*$^{+/Y}$ and either *Sox3*$^{Sox2KI/Y}$ or *Sox3*$^{KO/Y}$. Genes presented were significantly different between *Sox3*$^{+/Y}$ and *Sox3*$^{KO/Y}$ testes (Step-up *P* value <0.05). * <0.05 step-up *P* value between *Sox3*$^{KO/Y}$ and *Sox3*$^{Sox2KI/Y}$. (E) qPCR validation of SOX2 regulation of SOX3 target genes in 2 week testes. ANOVA multiple comparisons with Bonferroni's correction were performed using Graphpad Prism (****<0.0001, ***<0.001, **<0.01, *<0.05, ns, not significant).
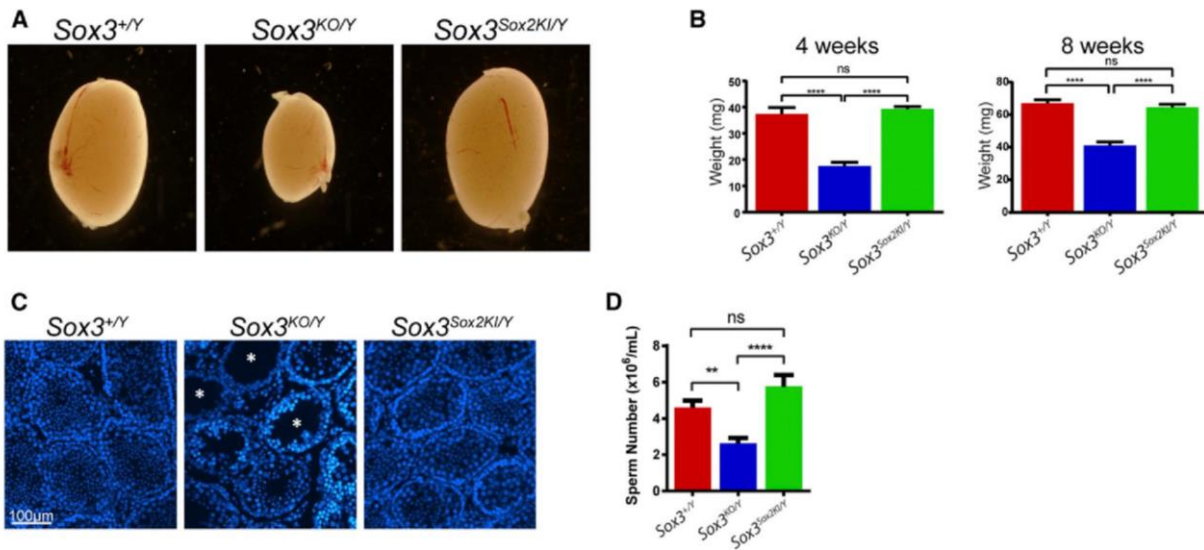
Figure 4 SOX2 can functionally replace SOX3 in the testes. (A) Gross morphology of $Sox3^{+/Y}$, $Sox3^{KO/Y}$ and $Sox3^{Sox2KI/Y}$ 4-week testes. (B) Quantification of testes weight at 4 and 8 weeks old [at least 18 testes were weighed for each genotype, and results were compared using one way ANOVA multiple comparisons with Bonferroni's correction (****<0.0001, ns, not significant)]. (C) DAPI staining of $Sox3^{+/Y}$, $Sox3^{KO/Y}$ and $Sox3^{Sox2KI}$ 4 week old testes. Asterisks indicate empty tubules. (D) Sperm counts from 8-week $Sox3^{+/Y}$, $Sox3^{KO/Y}$ and $Sox3^{Sox2KI/Y}$ epididymis [at least 14 samples were counted for each genotype, and results were compared using one way ANOVA multiple comparisons with Bonferroni's correction (****<0.0001, **<0.01 ns, not significant)].

$Sox3^{Sox2KI/Y}$ pituitaries was indistinguishable from $Sox3^{+/Y}$ controls (Figure 2A), apart from very subtle clefting between the intermediate and anterior lobe (Figure 2, A and B). We next compared anterior pituitary induction in $Sox3^{KO/Y}$ and $Sox3^{KI/Y}$ embryos. $Sox3^{KO/Y}$ embryos had an infundibular recess that was significantly wider than $Sox3^{+/Y}$ embryos, and a dysmorphic Rathke's Pouch with extensive branching, some of which (three of four) had failed to detach from the oral cavity (Figure 2, C and D and Figure S2 in File S1). In contrast, the width of the infundibulum in $Sox3^{Sox2KI}$ embryos was completely rescued. Furthermore, Rathke's Pouch was always separated from the oral cavity, although a mild branching phenotype was observed (Figure S2 in File S1). Together, this data indicates that SOX2 is able to functionally replace SOX3 in the developing brain.

Ectopic SOX2 in the testes can regulate SOX3 target genes

We next assessed whether SOX2 is able to replace SOX3 in a tissue where it is not normally expressed. For these experiments, we turned to the testes, where $Sox3$ is the only $SoxB1$ member expressed in spermatogonial stem/progenitor cells, and is functionally required for normal spermatogenesis (Raverot et al. 2005). After first confirming the absence of $Sox1$ and $Sox2$ in the testes by qRT-PCR (Figure S3 in File S1), we then examined testes from $Sox3^{Sox2KI/Y}$ mice to determine whether $Sox2$ was expressed in a $Sox3$-specific manner. Robust expression of $Sox2$ was detected by qRT-PCR in $Sox3^{Sox2KI/Y}$ testes, but not in $Sox3^{+/Y}$ testes (Figure 3A). To determine whether the level of $Sox2KI$ expression in

$Sox3^{Sox2KI/Y}$ mice was similar to endogenous $Sox3$ expression in $Sox3^{+/Y}$ mice, we compared $Sox3$ 3' UTR transcript levels in testes (as $Sox3^{Sox2KI/Y}$ retains an intact $Sox3$ UTR; Figure 1F). $Sox3$ locus expression in $Sox3^{Sox2KI/Y}$ was slightly lower than levels seen in $Sox3^{+/Y}$, but was a closer match than that observed in the developing brain (Figure 1F and Figure 3B, respectively). In addition, SOX2 immunostaining could not be detected in $Sox3^{+/Y}$ and $Sox3^{KO/Y}$ testes, but was present in $Sox3^{Sox2KI/Y}$ testes in undifferentiated spermatogonia (Figure 3C). SOX3 immunostaining was present in $Sox3^{+/Y}$ testes in undifferentiated spermatogonia, but, as expected, not in $Sox3^{KO/Y}$ and $Sox3^{Sox2KI/Y}$ testes (Figure 3C).

Since SOX2 and SOX3 bind and regulate target genes via highly similar HMG-box DNA binding domain (Bergsland et al. 2011), we reasoned that SOX2 could be capable of regulating SOX3 target genes in the testes of $Sox3^{Sox2KI/Y}$ mice. To test this possibility, we performed microarray analysis comparing $Sox3^{+/Y}$, $Sox3^{KO/Y}$ and $Sox3^{Sox2KI/Y}$ 2-week testes. A total of 17 genes (excluding $Sox3$) were significantly altered when comparing $Sox3^{KO/Y}$ to $Sox3^{+/Y}$ (step-up P value ≤0.05), of which nine were upregulated and eight were downregulated. These genes included $Ngn3$, which has previously been shown to be downregulated in $Sox3$ null testes (Raverot et al. 2005). Expression levels of all 17 genes were statistically returned to $Sox3^{+/Y}$ levels in $Sox3^{Sox2KI/Y}$ testes (Figure 3D), although only four were significantly different in $Sox3^{KO/Y}$ vs. $Sox3^{Sox2KI/Y}$ testes. As step-up P values are known to be a very strict measure of microarray changes, and often underestimate changes in order to avoid high false discovery rates (Benjamini and Hochberg 1995), we sought

46

to validate the degree of expression rescue using qPCR on independent biological samples. Five of the 17 genes were selected for validation by qPCR, and fold changes were all in the same direction as the microarray data (Figure 3E). Notably, all five genes assessed by qPCR were statistically closer to $Sox3^{+/Y}$ levels in $Sox3^{Sox2KI/Y}$ in comparison to $Sox3^{KO/Y}$, but, in four of five cases fell short of complete restoration, with only *Scube3* showing no significant difference between $Sox3^{+/Y}$ and $Sox3^{Sox2KI/Y}$. Collectively, this analysis indicates that ectopically expressed SOX2 is able to regulate SOX3 genes in the testes, and can largely restore normal gene expression patterns. However, a small subset of those genes most heavily deregulated by the loss of *Sox3* fail to achieve complete rescue in $Sox3^{Sox2K/Y}$ mice, likely due to the slightly decreased levels of SOX2 expression in the testes.

Sox2 can functionally replace Sox3 in the testes

Given the high degree of expression restoration in $Sox3^{Sox2KI/Y}$ testes, we next sought to assess whether this correlated with functional rescue of testes size and sperm production, which has previously been reported for $Sox3^{KO/Y}$ animals (Rizzoti *et al.* 2004; Raverot *et al.* 2005). $Sox3^{KO/Y}$ mice exhibited significantly reduced testes size and weight at 4 and 8 weeks (Figure 4, A and B). Empty seminiferous tubules were observed in sections of 4-week-old $Sox3^{KO/Y}$ testes (Figure 4C), and $Sox3^{KO/Y}$ mice had a significantly reduced sperm count at 8 weeks of age (Figure 4D). In contrast, $Sox3^{Sox2KI/Y}$ males had normal testes size and sperm count with no evidence of empty tubules (Figure 4, A–D). Thus, testes morphology and function were completely rescued in $Sox3^{Sox2KI/Y}$ mice. These data provide unequivocal evidence of *SoxB1* functional redundancy *in vivo*.

Discussion

Gene swap experiments provide an ideal approach to investigate functional overlap of related genes under physiological conditions. To our knowledge, this is the first report describing the use of CRISPR/Cas9-assisted mutagenesis to perform an *in vivo* gene-swap of two closely related genes. By cutting either side of *Sox3* in the presence of a *Sox2*-containing repair template, we were able to simultaneously remove *Sox3* and insert the *Sox2* ORF in its place. From a practical standpoint, this provides a feasible and rapid pathway to generate gene-swap mice. It should be noted that the efficiency of generation was lower than anticipated based on previous reports of CRISPR/Cas9-assisted insertional mutagenesis (Yang *et al.* 2013), with only a single $Sox3^{Sox2KI}$ founder generated from a total of 17 live born mice. Generation of null alleles was more efficient, with almost half of the founders having a large deletion extending between the two CRISPR cut sites. The reason for the low efficiency of $Sox3^{Sox2KI}$ alleles most likely reflects low efficiency of HDR in comparison to nonhomologous end joining, as noted by others (Yang *et al.* 2013). In the future, it will be interesting to test whether strategies to promote HDR, such as knockdown of the Ku protein, can be used to increase knock-in efficiency (Basu *et al.* 2015). Any future optimization should be sure to allow the coincidental production of KO alleles, as this enables simultaneous assessment of KI and KO phenotypes, and, as such, circumvents the significant confounding influence of genetic background. Indeed, we have illustrated the importance of this consideration with our description of a more severe pituitary phenotype in *Sox3* null mice on a C57Bl/6 background in comparison with the previously published phenotype on a mixed genetic background (Rizzoti *et al.* 2004).

The testes provide a particularly tractable setting for this investigation, as SOX3 is the only SOXB1 protein to be expressed in this tissue, and its absence results in a relatively severe phenotype. Importantly, we have shown that the morphogenetic and spermatogenic defects that result from *Sox3* deletion were completely rescued when *Sox2* was ectopically expressed in its place. At the molecular level, these data indicate that ectopic SOX2 is able to bind and regulate SOX3 target genes in spermatogonial stem/progenitor cells. This finding is consistent with published ChIP-seq studies showing that the binding sites of these proteins overlap extensively in cultured neural progenitor cells (Bergsland *et al.* 2011). While a modest difference in spermatogonial marker expression remained in $Sox3^{Sox2KI/Y}$ testes, it seems likely this results from the slightly lower level of $Sox3^{Sox2KI}$ mRNA compared with wild type *Sox3*. However, potential differences in the affinity of SOX2 and SOX3 antibodies make it near impossible to directly compare levels of SOX3 and SOX2 protein in $Sox3^{+/Y}$ and $Sox3^{Sox2KI/Y}$ testes, respectively. Therefore, we cannot completely exclude the possibility that the residual differences in $Sox3^{Sox2KI/Y}$ testes are due to slightly different functions of SOX3 and SOX2 proteins. It is also conceivable that functional differences may be exposed under nonlaboratory conditions as observed in *HoxA1/B1* gene swap mice (Ruff *et al.* 2015). Nevertheless, these gene swap data provide the most compelling evidence to date that SOX2 and SOX3 are functionally interchangeable proteins. Given the complete rescue of SOX3 morphological defects by SOX2, we predict that this functional redundancy should operate in both directions, such that SOX3 should be equally well suited to replacing SOX2 if expressed at the correct time and place. Thus, it would be interesting to perform the complementary gene swap to determine whether, for example, *Sox3* expression from the *Sox2* locus is capable of rescuing the early lethality phenotype of *Sox2* null mice.

Unlike the testis, where a single SOXB1 protein (SOX3) is expressed in a small population of cells, all three SOXB1 proteins are expressed in neural progenitors across the developing CNS. Direct comparison of SOX3 and SOX2 expression in the embryonic brain has shown that virtually all SOX3+ cells also express SOX2 (Rogers *et al.* 2013; Cheah and Thomas 2015). Given our gene swap data showing SOX2/3 functional redundancy in the testes, as well as overexpression data indicating their functional equivalence (Bylund *et al.* 2003; Graham *et al.* 2003), it is not surprising that most of the CNS develops normally in *Sox3* null mice.

However, a notable exception is the developing ventral diencephalon where *Sox3* loss-of-function results in CNS/pituitary defects in both mice and humans despite expression of *Sox2*. Here, we show that these *Sox3* null defects are almost completely rescued by expression of *Sox2* from the *Sox3* locus, indicating that SOX2 and SOX3 proteins are functionally interchangeable in pituitary induction. It is likely that the residual minor clefting in the $Sox3^{Sox2KI}$ adult pituitary is due to a decreased level of $Sox3^{Sox2KI}$ mRNA, reinforcing that pituitary induction is extremely sensitive to *Sox2/3* levels, which, in $Sox3^{Sox2KI/Y}$ brains, are reduced by only ~16% compared to wild type (see File S1). Alternatively, these minor abnormalities may be due to a slight difference in the functionality of SOX3 and SOX2 in pituitary development.

In the mouse, only a handful of examples exist in which paralogous gene function has been examined using a KI gene swap approach. In these examples, the degree of functional equivalence has been complete (*Otx1/2*, *En1/2*, *Osr1/2*, and *HoxA3/D3*) (Hanks *et al.* 1995; Greer *et al.* 2000; Acampora *et al.* 2003; Gao *et al.* 2009), partial (*Sox8/10* and *HoxA1/B1*) (Kellerer *et al.* 2006; Ruff *et al.* 2015), or limited (*Phox2a/2b*) (Coppola *et al.* 2005). Given the profound rescue of testes defects in $Sox3^{Sox2KI}$ mice, we suggest that SOX3 and SOX2 proteins are biochemically equivalent. The observed differences in their null mutant phenotypes therefore likely reflects unique zones of expression, presumably generated through evolutionary subfunctionalization. However, why two biochemically equivalent proteins have retained widespread overlapping expression within the developing brain across an extensive evolutionary period remains an intriguing question. This does not appear to be an accident of evolution, as ancestral SOX B proteins in *Drosophila* (SoxNeuro and Diachete) exhibit similar partial redundancy, and active conservation of binding sites that allow both paralogues to bind (Carl and Russell 2015). One explanation is that it may be difficult to fully disentangle regulatory elements that direct *Sox3* expression in the ventral diencephalon from other neuroprogenitor zones. Further analysis of *SoxB1* gene regulation, for example, though CRISPR/Cas9-mediated enhancer deletion, may reveal insights into this interesting biological phenomenon.

Literature Cited

Acampora, D., A. Annino, E. Puelles, I. Alfano, F. Tuorto *et al.*, 2003 OTX1 compensates for OTX2 requirement in regionalisation of anterior neuroectoderm. Gene Expr. Patterns GEP 3: 497–501.

Barbaric, I., G. Miller, and T. N. Dear, 2007 Appearances can be deceiving: phenotypes of knockout mice. Brief. Funct. Genomics Proteomics 6: 91–103.

Basu, S., A. Aryan, J. M. Overcash, G. H. Samuel, M. A. E. Anderson *et al.*, 2015 Silencing of end-joining repair for efficient site-specific gene insertion after TALEN/CRISPR mutagenesis in Aedes aegypti. Proc. Natl. Acad. Sci. USA 112: 4038–4043.

Benjamini, Y., and Y. Hochberg, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. 57: 289–300.

Bergsland, M., D. Ramsköld, C. Zaouter, S. Klum, R. Sandberg *et al.*, 2011 Sequentially acting Sox transcription factors in neural lineage development. Genes Dev. 25: 2453–2464.

Bylund, M., E. Andersson, B. G. Novitch, and J. Muhr, 2003 Vertebrate neurogenesis is counteracted by Sox1–3 activity. Nat. Neurosci. 6: 1162–1168.

Carl, S. H., and S. Russell, 2015 Common binding by redundant group B Sox proteins is evolutionarily conserved in Drosophila. BMC Genomics 16: 292.

Cheah, P.-S., and P. Q. Thomas, 2015 SOX3 expression in the glial system of the developing and adult mouse cerebellum. Springerplus 4: 400.

Coppola, E., A. Pattyn, S. C. Guthrie, C. Goridis, and M. Studer, 2005 Reciprocal gene replacements reveal unique functions for Phox2 genes during neural differentiation. EMBO J. 24: 4392–4403.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan *et al.*, 1999 Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151: 1531–1545.

Gao, Y., Y. Lan, C. E. Ovitt, and R. Jiang, 2009 Functional equivalence of the zinc finger transcription factors Osr1 and Osr2 in mouse development. Dev. Biol. 328: 200–209.

Graham, V., J. Khudyakov, P. Ellis, and L. Pevny, 2003 SOX2 functions to maintain neural progenitor identity. Neuron 39: 749–765.

Greer, J. M., J. Puetz, K. R. Thomas, and M. R. Capecchi, 2000 Maintenance of functional equivalence during paralogous Hox gene evolution. Nature 403: 661–665.

Hanks, M., W. Wurst, L. Anson-Cartwright, A. B. Auerbach, and A. L. Joyner, 1995 Rescue of the En-1 mutant phenotype by replacement of En-1 with En-2. Science 269: 679–682.

Hughes, J., S. Piltz, N. Rogers, D. McAninch, L. Rowley *et al.*, 2013 Mechanistic insight into the pathology of polyalanine expansion disorders revealed by a mouse model for X linked hypopituitarism. PLoS Genet. 9: e1003290.

Kelberman, D., K. Rizzoti, A. Avilion, M. Bitner-Glindzicz, S. Cianfarani *et al.*, 2006 Mutations within Sox2/SOX2 are associated with

abnormalities in the hypothalamo-pituitary-gonadal axis in mice and humans. J. Clin. Invest. 116: 2442–2455.

Kellerer, S., S. Schreiner, C. C. Stolt, S. Scholz, M. R. Bösl *et al.*, 2006 Replacement of the Sox10 transcription factor by Sox8 reveals incomplete functional equivalence. Development 133: 2875–2886.

Lynch, M., and J. S. Conery, 2000 The evolutionary fate and consequences of duplicate genes. Science 290: 1151–1155.

Nakagawa, M., M. Koyanagi, K. Tanabe, K. Takahashi, T. Ichisaka *et al.*, 2008 Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat. Biotechnol. 26 (1): 101–106.

Raverot, G., J. Weiss, S. Y. Park, L. Hurley, and J. L. Jameson, 2005 Sox3 expression in undifferentiated spermatogonia is required for the progression of spermatogenesis. Dev. Biol. 283: 215–225.

Rizzoti, K., S. Brunelli, D. Carmignac, P. Q. Thomas, I. C. Robinson *et al.*, 2004 SOX3 is required during the formation of the hypothalamo-pituitary axis. Nat. Genet. 36: 247–255.

Rogers, N., P.-S. Cheah, E. Szarek, K. Banerjee, J. Schwartz *et al.*, 2013 Expression of the murine transcription factor SOX3 during embryonic and adult neurogenesis. Gene Expr. Patterns GEP 13: 240–248.

Ruff, J. S., R. B. Saffarini, L. L. Ramoz, L. C. Morrison, S. Baker *et al.*, 2015 Fitness assays reveal incomplete functional redundancy of the HoxA1 and HoxB1 paralogs of mice. Genetics 201: 727–736.

Wagner, A., 2005 Distributed robustness *vs.* redundancy as causes of mutational robustness. Bioessays 27: 176–188.

Wang, H., H. Yang, C. S. Shivalila, M. M. Dawlaty, A. W. Cheng *et al.*, 2013 One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell 153: 910–918.

Wood, H. B., and V. Episkopou, 1999 Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages. Mech. Dev. 86: 197–201.

Yang, H., H. Wang, C. S. Shivalila, A. W. Cheng, L. Shi *et al.*, 2013 One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell 154: 1370–1379.

Zhao, L., S. E. Zevallos, K. Rizzoti, Y. Jeong, R. Lovell-Badge *et al.*, 2012 Disruption of SoxB1-dependent Sonic hedgehog expression in the hypothalamus causes septo-optic dysplasia. Dev. Cell 22: 585–596.

Communicating editor: T. R. Magnuson

1 # SUPPLEMENTAL INFORMATION

2 **Functional equivalence of the SOX2 and SOX3 transcription factors in**

3 **the developing mouse brain and testes**

4 Fatwa Adikusuma[1,2], Daniel Pederick[1], Dale McAninch[1], James Hughes[1] and Paul

5 Thomas[1*]

6

7 [1]School of Biological Sciences and The Robinson Research Institute, University of

8 Adelaide, Adelaide, SA, AUS 5005

9 [2]CEBIOR, Faculty of Medicine, Diponegoro University, Semarang, Indonesia, 50271

10 [3]South Australian Health and Medical Research Institute, Adelaide, SA, AUS, 5000

11

12 *Corresponding author:*

13 Email: paul.thomas@adelaide.edu.au

14

15

16

17

18

19   **Supplemental Experimental Procedures**

20   **gDNA preparation and PCR genotyping**

21   gDNA was collected from tail tissues and extracted using High Pure PCR Template

22   Preparation Kit (Roche). To screen for $Sox3^{Sox2KI}$ founders, PCR was performed using

23   F: 5'-CCTGCTGAAACATTCCCTGT-3' and R: 5'-TTCAGCTCCGTCTCCATCAT-3'. PCR

24   products from desired founders were Sanger sequenced to identify specific

25   mutations. Large deletions resulting from two gRNA cuts were detected using

26   primers F: 5'-CCTGCTGAAACATTCCCTGT-3' and R: 5'-ACAAAACCCCGACAGTTACG-3'.

27   For routine genotyping of $Sox3^{Sox2KI}$ mice, multiplex PCR was performed using

28   primers F1: 5'-CACAACTCCGAGATCAGCAA-3', F2: 5'-GAACGCATCAGGTGAGAGAAG-

29   3', R1: 5'-CGGCGTTCATGTAGCTCTG-3' and R2: 5'-TTCAGCTCCGTCTCCATCAT-3'. For

30   routine genotyping of $Sox3^{KO}$ mice, primers F1: 5'-CAGCATGTACCTGCCACCT-3', F2:

31   5'-CCCGGATCTGAGCAGGTAT-3' and R: 5'-ACAAAACCCCGACAGTTACG-3' were used.

32

33   **RNA preparation and qPCR**

34   RNA was extracted using Trizol (Invitrogen), purified using Rneasy mini kit (Qiagen)

35   after DNAse treatment on column using RNase-Free DNase Set (Qiagen). cDNA was

36   generated using a High Capacity RNA-to-cDNA kit (Applied Biosystems) for qPCR

37   analyses using Fast SYBR Green Master Mix (Applied Biosystems) on an Applied

38   Biosystems 7500 StepOnePlus machine. Primers used for qPCR are as follows:

39    $Act\beta$ F: 5'-CTGCCTGACGGCCAGG-3', R: 5'-GATTCCATACCCAAGAAGGAAGG-3'

51

40 *Sox2* ORF F: 5'-ACCAGCTCGCAGACCTACAT-3' R: 5'-TCGGACTTGACCACAGAGC-3'

41 *Sox3* 3'UTR F: 5'-AACCTAGGAATCCGGGAAGA-3' R: 5'-CGTAACTGTCGGGGTTTTGT-3'

42 *Sox2* 3'UTR F: 5'-TTCGAGGAAAGGGTTCTTGCTG-3' R: 5'-

43 CCTTCCTTGTTTGTAACGGTCCT-3' *Sox1* F: 5'- CCCTCGGATCTCTGGTCA -3' R: 5'-

44 GCAGGTACATGCTGATCATCTC-3'

45 *Ngn3*: F: 5'-CCCCAGAGACACAACAACCT-3' R: 5'-AGTCACCCACTTCTGCTTCG-3'

46 *Gfra1* F: 5'-ATCGGGCAGTACACATCTCTG-3' R: 5'-TGTGGTTATGTGGCTGGAGG-3'

47 *Tgfbi* F: 5'-TGAAGCGTTCCAAGCCATGC-3' R: 5'-GATGCCTCCGCTAACCAGGATT-3'

48 *Egr3* F: 5'-TCAACCTCTTCTCCGGCAGC-3' R: 5'-GATTGGGCTTCTCGTTGGTCA-3'

49 *Scube3* F: 5'-CTGGCACATGACGGACACAAC-3' R: 5'-CGTAGCTGCCCATCATGTTGAC-3'

50

51 **Tissue collection and Immunofluorescence**

52 Tissues were fixed overnight in 4% PFA and equilibrated in 30% sucrose. Tissues

53 were then embedded in OCT compound (Sakura Finetek) for cryosectioning using

54 Leica CM1900 at 10-14 μm. For immunohistochemistry, tissue slides were blocked

55 using PBS/0.3% Triton X-100/10% horse serum before primary antibodies were

56 added for overnight incubation at 4 ^{0}C. Slides were then washed in PBS and

57 incubated with secondary antibodies for 2 hours at room temperature. Slides were

58 mounted using Prolong Gold Antifade plus DAPI (Molecular Probes, Invitrogen) and

59 imaged using a Nikon Eclipse Ti microscope. Primary antibodies were SOX3 (R&D

60 AF2569, 1:200), SOX2 (Santa Cruz SOX2 Y-17 sc-17320, 1:100, used for testes), SOX2

61 (Millipore Ab5603, 1:1000, used for embryonic brain). Secondary antibodies were

62 donkey anti-goat 488 (Life Technologies) and donkey anti-rabbit TxRed (Life

63 Technologies).

64

Quantification of SOX2 protein using immunohistochemistry

66 Immunostaining was described as above and images acquired using an Leica SP5

67 spectral scanning confocal microscope. The mean SOX3 and SOX2 staining intensity

68 was calculated in at least 100 cells/embryo using NIS Elements Advance Research Software

69 (Nikon). Raw data was transferred to GraphPad Prism 7 where it was transformed such that

70 cells with mean SOX3 intensity 5x above background were deemed SOX3 positive cells. The

71 remaining cells were deemed SOX3 negative. Students unpaired T-test was performed tp

72 test significance.

73

Calculation of relative *Sox2*/*Sox3* levels

75 Quantitation of transcripts from the *Sox2* and *Sox3* locus in WT and *Sox3^{Sox2KI}* reveals

76 that the *Sox3^{Sox2KI}* locus is producing *Sox2* ORF transcripts at 60% of the normal level

77 of *Sox3* transcript based on a 5'UTR PCR. This additional *Sox2* ORF increases the total

78 *Sox2* ORF transcript pool by approximately 40% in *Sox3^{Sox2KI}* embryonic heads. Since

79 the rise in *Sox2* transcripts is less than expected based on the contribution from the

80 *Sox3^{Sox2KI}* locus this implies that the *Sox2* locus contributes approximately 60% of the

81 total *Sox2*/*3* transcript pool in the developing brain. Combined Sox2/3 levels in

53

82   *Sox3$^{Sox2KI/Y}$* mice can be calculated as follows: *Sox2* locus (60%) + *Sox3* locus (60% x

83   40%) = 60% + 24% = 84%.   This amounts to a 16% reduction compared with WT

84   levels.

85

86   **Purification of spermatogonia**

87   Testes from post-natal day 6 mice were isolated, de-capsulated, and incubated for

88   15 min each in 0.5 mg/ml collagenase/DMEM with agitation and then in 0.25%

89   trypsin/EDTA in DMEM. Tubules were dissociated manually by pipetting and washed

90   in 0.5% BSA in DMEM by centrifugation. Cell pellets were resuspended in DMEM and

91   filtered twice through a 70 μm membrane, then separated over a 2-4% BSA gradient.

92   Purified spermatogonia were identified by GCNA1 staining.
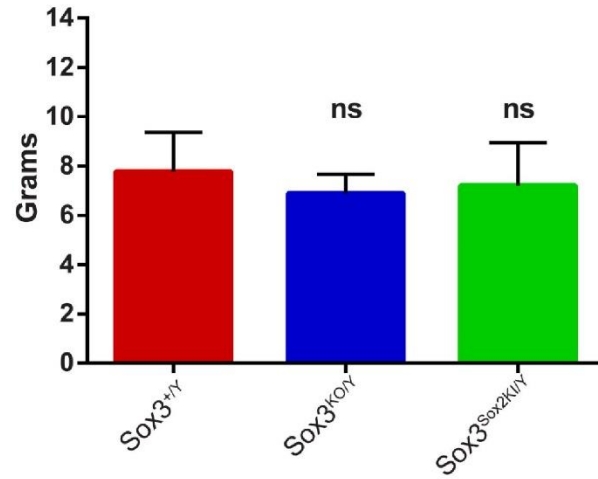
93

94

95

96

97

98

99

100

101 **Supplementary Figures**
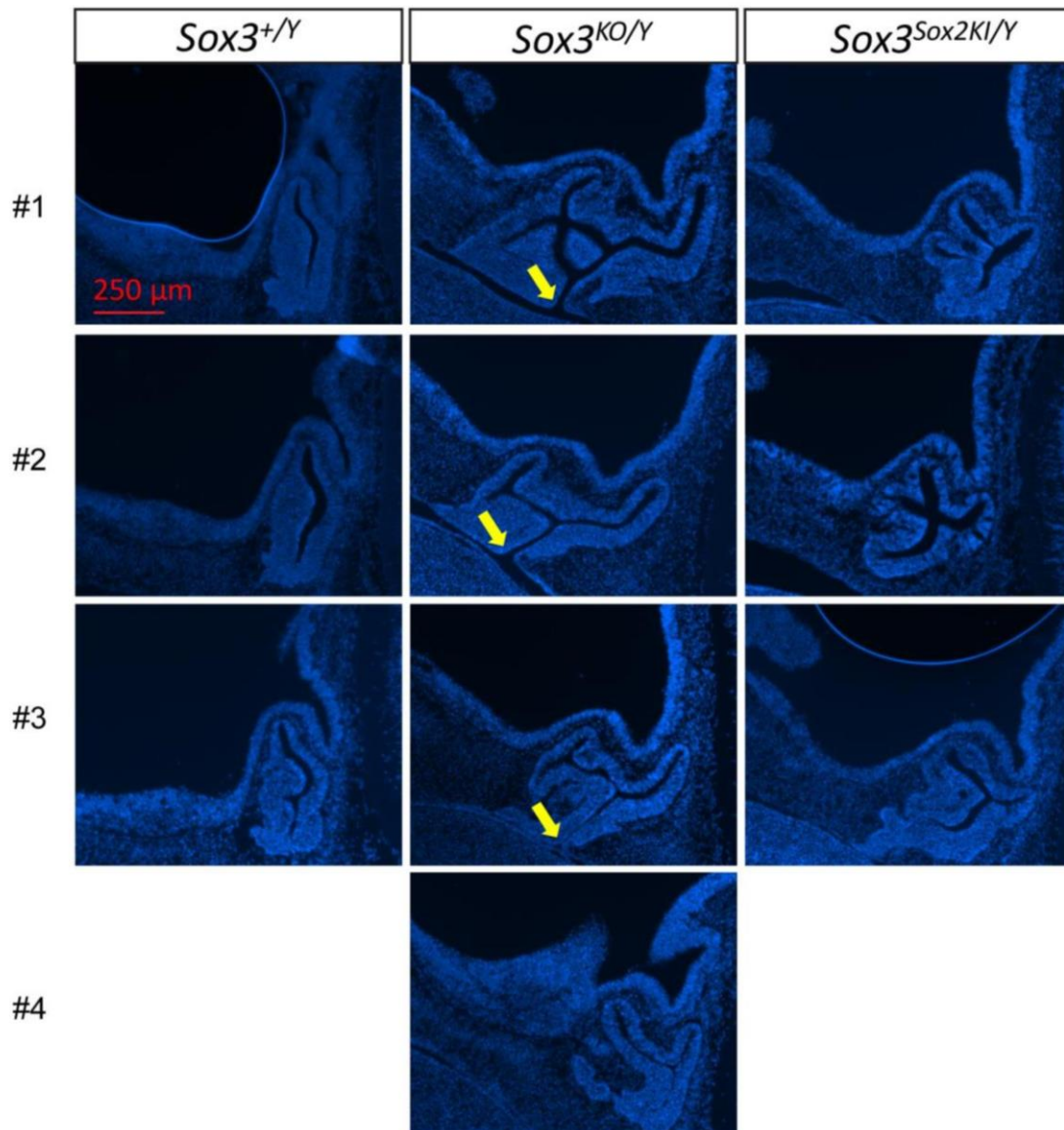

102

103 **Supplementary Figure 1**

104 **Male wean weights were unchanged in mutant mice.**

105 Male mice were weighed at weaning (3 weeks). At least 14 mice of each genotype

106 were weighed and unpaired t-tests were performed to assess pairwise differences

107 with respect to $Sox3^{+/}$

108

	Sox3^(+/Y)	*Sox3*^(KO/Y)	*Sox3*^(Sox2KI/Y)
#1			
#2			
#3			
#4			

250 μm

109

110

111 **Supplementary Figure 2**

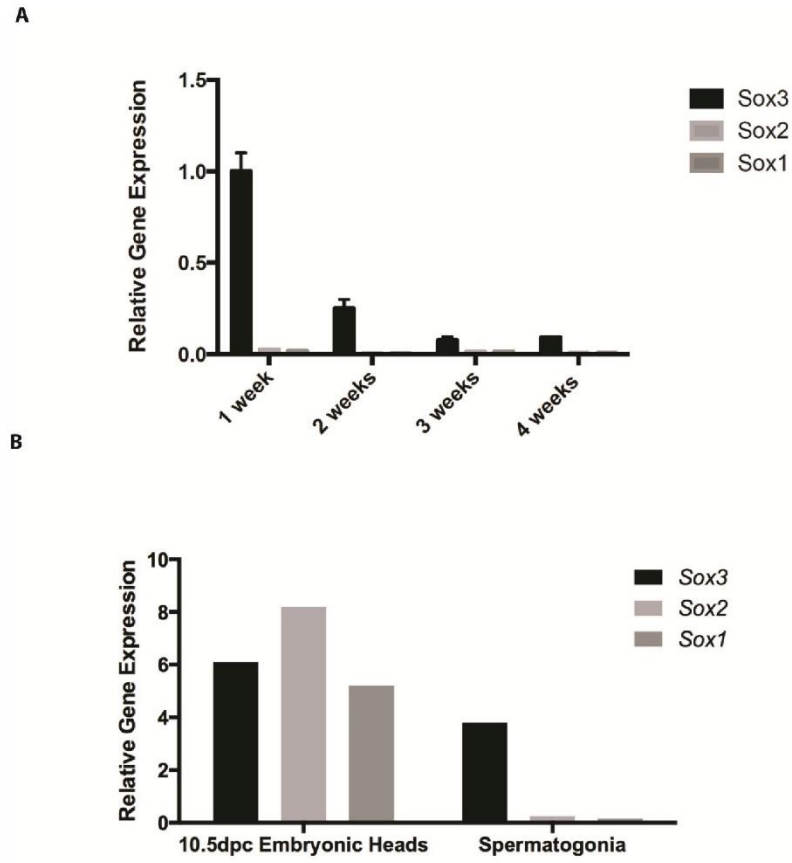112 **Sox2KI largely rescues embryonic *Sox3*-null Rathke's pouch**

113 DAPI staining of midline sagittal sections of Sox3^(+/Y), Sox3^(KO/Y) and Sox3^(Sox2KI/Y) 12.5 dpc

114 embryos (*n*=3 per genotype) showing rescue of the Sox3^(KO/Y) pituitary induction

115 phenotype. Arrows indicate failure of the Rathke's pouch to separate from the oral

116 cavity. See Fig. 2C for further details.

A

B

119 **Supplementary Figure 3**

120 **Sox1 and Sox2 are not detectable in the testes.**

121 (A) Expression of *Sox1*, *Sox2* and *Sox3* was measured by qPCR in the testes at various

122 ages (*n*=3). Expression levels were normalised to *Sox3* levels in 1 week old testes. (B)

123 Expression levels were also compared between 10.5 dpc embryonic heads and

124 purified postnatal day 6 spermatogonia.

This page intentionally left blank

Chapter 3:

Targeted deletion of an entire chromosome using CRISPR/Cas9

3.1. Summary

Genome editing has been used to model genetic diseases and holds great promise for their treatment. This may be applicable for monogenic diseases, but remains a challenge for diseases caused by supernumerary chromosomes (aneuploidy), such as Down syndrome, as this requires modification of an entire chromosome, which has not been shown to be possible with existing genome editing technologies, including CRISPR/Cas9. Therefore, this project aimed to develop strategies to facilitate genome editing at the chromosomal level.

It was hypothesised that deletion of an entire chromosome could be achieved by deleting its centromere, or by bombardment of the chromosome arm with multiple cuts (shredding). These hypotheses were tested in vitro in mouse ES cells and in vivo by mouse zygote injection, by attempting to delete the Y chromosome (~90 MB in size) as loss of this chromosome does not affect cell/mouse growth and viability. To efficiently delete the centromere, gRNAs that target repetitive sequences that cut the centromere 140X and 41X were employed. Surprisingly, efficient Y chromosome depletion of up to 85%, as measured by qPCR analyses, was observed. Two cuts flanking the Y centromere for intervening deletion could still efficiently delete the Y chromosome with ~40% efficiency. To test whether shredding the Y arm could result in Y loss, gRNAs that target repetitive sequences in the Y long arm were employed. These gRNAs enabled the Y long arm to be cut 298X, 116X, 45X, 8X and 2X. All gRNAs, apart from gRNA 2X, could obviously induce Y deletion with an efficiency correlating with the number of cuts, ranging from 27–82%, whereas gRNA 298X was the most efficient. This highly efficient Y loss was confirmed by Y painting FISH analysis.

This strategy was then tested in vivo by microinjection of gRNA centromere 41X and Cas9 mRNA into mouse zygotes. Successful deletion of Y would produce female XO mice. The

gonadal phenotypes and the genotypes of the mice were assessed, and females with an XO genotype were detected, indicating the successful application of this strategy in vivo. Collectively, the study described in this chapter has established strategies to delete an entire chromosome using CRISPR/Cas9 genome editing that should be applicable to other chromosomes for modelling and therapeutic intervention in aneuploidy diseases. This study has been published in *Molecular Therapy* as a paper entitled 'Targeted deletion of an entire chromosome using CRISPR/Cas9'.

Statement of Authorship

Title of Paper	Targeted deletion of an entire chromosome using CRISPR/Cas9
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Adikusuma, F., Williams, N., Grutzner, F., Hughes, J. & Thomas, P. Targeted Deletion of an Entire Chromosome Using CRISPR/Cas9. *Mol Ther*, doi:10.1016/j.ymthe.2017.05.021 (2017)

Principal Author

Name of Principal Author (Candidate)	Fatwa Adikusuma
Contribution to the Paper	Designed and conceived the study, conducted most of the experiments, analysed data, generated the Figures and wrote the manuscript.
Overall percentage (%)	90%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date	8/8/2017

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

    i.   the candidate's stated contribution to the publication is accurate (as detailed above);

    ii.   permission is granted for the candidate in include the publication in the thesis; and

    iii.   the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Nicole Williams		
Contribution to the Paper	Performed and analysed FISH analysis		
Signature		Date	

Name of Co-Author	Frank Grutzner		
Contribution to the Paper	Performed and analysed FISH analysis		
Signature		Date	13 July 2017

Name of Co-Author	James Hughes		
Contribution to the Paper	Designed and conceived the study, wrote the manuscript.		
Signature		Date	25/07/2017

Name of Co-Author	Paul Thomas		
Contribution to the Paper	Designed and conceived the study, supervised the study and wrote the manuscript.		
Signature		Date	8/8/2017

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

   i.   the candidate's stated contribution to the publication is accurate (as detailed above);

   ii.   permission is granted for the candidate in include the publication in the thesis; and

   iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Nicole Williams	
Contribution to the Paper	Performed and analysed FISH analysis	
Signature		Date 17/07/17

Name of Co-Author	Frank Grutzner	
Contribution to the Paper	Performed and analysed FISH analysis	
Signature		Date

Name of Co-Author	James Hughes	
Contribution to the Paper	Designed and conceived the study, wrote the manuscript.	
Signature		Date

Name of Co-Author	Paul Thomas	
Contribution to the Paper	Designed and conceived the study, supervised the study and wrote the manuscript.	
Signature		Date

Molecular Therapy

Letter to the Editor

Targeted Deletion of an Entire Chromosome Using CRISPR/Cas9

The recent emergence of gene editing technologies, in particular CRISPR/Cas, has enabled rapid generation of disease models and provides a novel approach for the treatment of monogenic disorders through correction of disease-causing mutations.[1,2] In contrast, the therapeutic potential of CRIPSR/Cas technology for aneuploidies, such as Down syndrome (Trisomy 21), remains unexplored. Indeed, disorders that are caused by supernumerary chromosomes represent a significant challenge, because genetic correction requires targeted ablation of an entire chromosome, which, to our knowledge, has not been demonstrated using genome editing technology.[1]

To assess the potential of CRISPR/Cas technology to effect chromosomal loss, we investigated the hypothesis that simultaneous generation of multiple DNA double-strand breaks (DSBs) at targeted chromosomal locations can induce directed chromosomal deletion.[3] We selected the 90 Mb acrocentric mouse Y chromosome for deletion because loss of this chromosome does not overtly impact cell/mouse viability and it is only present in one copy in male cells, thus facilitating screening.[4]

Our first strategy used CRISPR/Cas to fragment the centromere, which is indispensable for chromosome segregation during mitosis.[5] We screened the 90 kb Y centromere for guide RNA (gRNA) sequences in repetitive elements that would enable targeted cleavage at multiple sites. We identified two gRNA candidates that target the centromere 140 or 41 times (centro 140X and centro 41X, respectively; Figure 1A). For comparison, we also tested a gRNA pair targeting two unique sequences immediately flanking the centromere (centro 2X; Figure 1A). Cas9 and single-guide RNA (sgRNA) were expressed in R1 XY mouse embryonic stem cells (ESCs) using plasmid PX459 V.2, followed by transient puromycin selection, to ensure only transfectants were

harvested.[6] Quantification of Y chromosome dosage was performed by genomic qPCR amplification of Uba1y and Erdr1, genes located at the end of the Y chromosome short and long arm, respectively (Figure 1A). Strikingly, Uba1y and Erdr1 qPCR signal was reduced by 80%–85% for both centro 140X and centro 41X compared with the sgRNA-expressing negative control (Neo-gRNA; Figure 1B). Further, a reduction of ~40% was achieved using the centro 2X gRNA (Figure 1B). To confirm that the reduction of qPCR signal was caused by Y chromosome loss, we performed fluorescence in situ hybridization (FISH) using Y chromosome paint on centro 41X-treated samples. Consistent with the qPCR data, the Y chromosome was not detected in 90% of centro 41X cells compared to 13% of control cells (Figure 1C). We also noted that 6% of control cells had two Y chromosomes, and this was reduced to less than 1% in centro 41X-treated cells. These findings confirm that CRISPR/Cas-mediated centromere cleavage leads to Y chromosome loss at high efficiency.

Next, we tested an alternative strategy for chromosome deletion in which the long arm is targeted for fragmentation by cleavage at multiple sites. As this approach does not target the centromere, it has potential for application in both dividing and non-dividing cells. We again identified gRNAs that targeted repetitive sequences in the Y chromosome (Figure 1A). However, the selected gRNAs sequences were specific to the long arm to ensure the centromere was left intact. Expression of sgRNAs that targeted the long arm 298X, 116X, 45X, 8X, and 2X resulted in Uba1y qPCR signal loss of 69%, 40%, 26%, 27%, and 3%, respectively, and Erdr1 qPCR signal loss of 82%, 68%, 68%, 52%, and 27%, respectively (Figure 1B). These data indicate that targeted fragmentation of a chromosomal arm can induce chromosome deletion and the frequency of deletion is proportional to the number of cuts. Notably, apart from long arm 298X Edr1, all long gRNAs resulted in significantly higher Uba1y and Edr1 signals than the centro 41X and 140X gRNAs (Table S3). Given that Uba1y qPCR signal was significantly higher than Erdr1 for all long gRNAs (Table

S3), we speculate that fragmentation of the long arm occasionally results in chromosome truncation or translocation, with retention of the Y short arm sequence containing Uba1y. FISH Y painting analysis in 298X-treated samples revealed 95% of cells contained no Y chromosome signal, confirming that the long arm fragmentation strategy was indeed effective (Figure 1C).

Notably, the degree of Y chromosome depletion induced by 8X and 45X are similar. This is significant, because targeted deletion of potentially any chromosome could be achieved relatively easily by transfection of a single vector expressing eight unique gRNAs.[7] We were also impressed with the activity of the long arm 2X gRNA. Although this gRNA induced negligible loss of short arm signal (3%), it appears to truncate the Y long arm relatively efficiently based on an Erdr1 qPCR signal loss of 27%.

Having successfully deleted an entire chromosome in vitro, we next tested our centromere deletion strategy in vivo in mouse zygotes with the expectation that successful Y chromosome deletion in male zygotes would result in an XO female phenotype.[4] We selected gRNA centro 41X due to its high efficiency in vitro and low off-target prediction (Table S2).[8]

After zygote injection of centro 41X gRNA and Cas9 mRNA, we collected 27 E15.5 embryos, of which 11 were phenotypically male and 16 were female based on gonadal assessment (Figure S1). We then screened the female embryos for X chromosome dosage and identified five embryos with only one X chromosome (Figures 1D and 1E). No evidence of XO karyotype was detected in control females injected with autosomal targeted gRNAs (Figure S2).

To directly assess Y chromosome loss in the five single X females, we performed Y chromosome genomic qPCR. Y short and long arm signals were undetectable in two of these embryos, indicating an XO karyotype. The remaining three embryos contained approximately 50% Y short arm signal and no long arm signal, suggesting that these mice were mosaic, with half of the cells containing
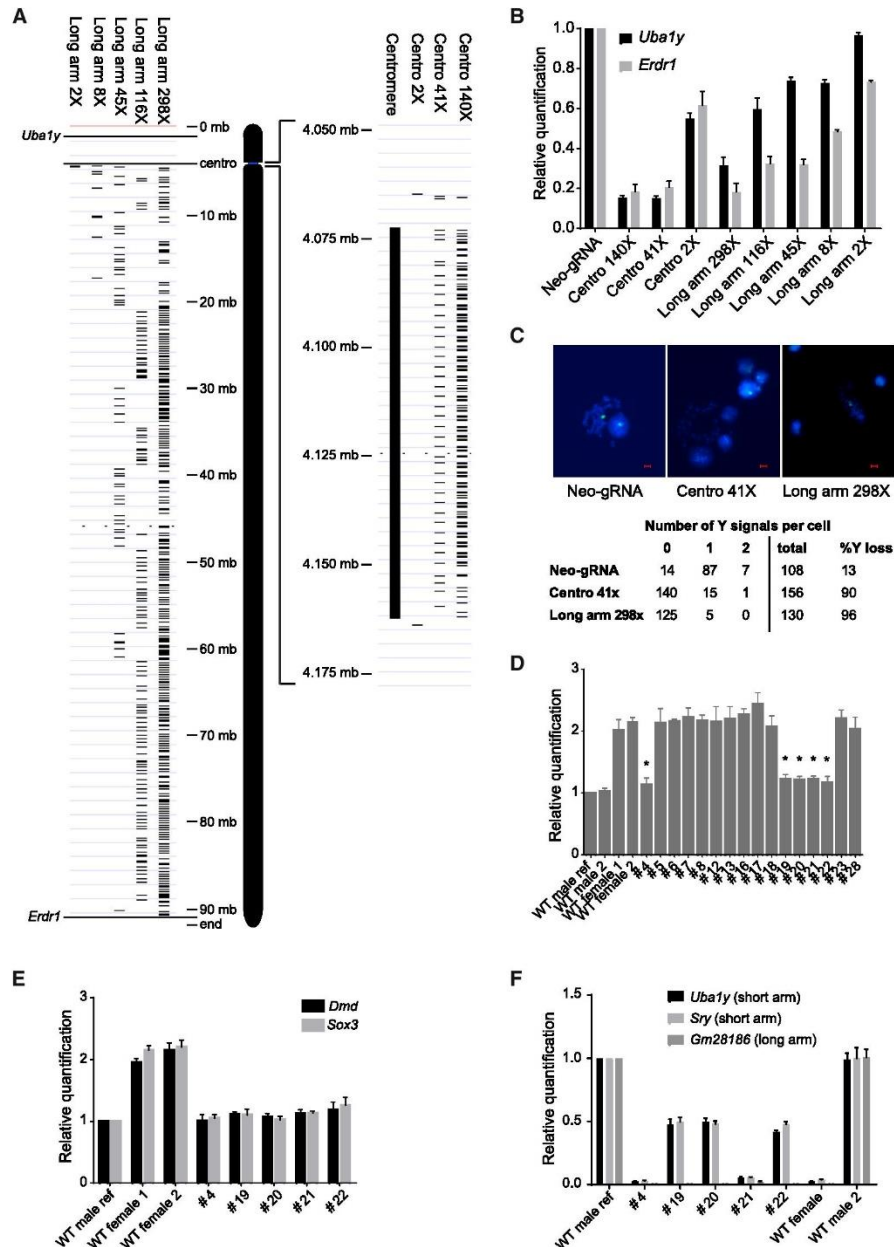
Figure 1. Deletion of Y Chromosome Using CRISPR/Cas9 in Mouse ESCs and In Vivo Mouse Zygote Injection

(A) Schematic showing the position of gRNA target sites in the long arm and centromere of the Y chromosome. (B) qPCR of genomic DNA to quantify Y chromosome dosage. *Sox1* qPCR was used as the internal reference control. Data are presented as mean ± SEM from n ≥ 3 biological replicates. Statistical analysis using two-way ANOVA is presented in Table S3. (C) FISH analysis detection of Y chromosome loss. Y chromosome and DAPI staining was indicated by green and blue signals, respectively. Scale bar, 5 μm. (D) *Xist* genomic qPCR of phenotypically female mice generated through zygote injection of centro 41X gRNA. Asterisks indicate female candidates with single X. (E) *Dmd* and *Sox3* genomic qPCR confirming single X chromosome in female XO candidates. (F) Genomic qPCR quantifying dosage of Y short and long arms. *Sox1* qPCR was used as the internal reference control. Results are presented as mean ± SD from n ≥ 3 replicates.

Letter to the Editor

translocated/truncated Y short arm and the other half containing no detectable Y (Figure 1F).

Given mosaic outcomes are common following CRISPR/Cas zygote injection,[9] we extended our screening to look for phenotypic males that were mosaic for Y chromosome loss. We identified 3 of 11 males with 10%–20% reduction of Y dosage (Figures S3A and S3B). Testis development in these embryos is unsurprising given this level of XY cells.[10] In summary, from 27 embryos, we identified 11 XX females, 8 XY males, 2 XO females, 3 mosaic XO females, and 3 mosaic XO males (Table S1). These results provide proof of concept for efficient chromosome deletion in vivo.

This study shows that targeted chromosome deletion is achievable and relatively efficient both in vitro and in vivo using CRISPR/Cas genome editing. This approach should be applicable for other chromosomes and could be utilized in a variety of cellular contexts and species. Accordingly, we envisage that this strategy will be applied to modeling of aneuploidy syndromes and therapeutic intervention by targeting parental-specific polymorphisms.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Materials and Methods, three figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.ymthe.2017.05.021.

AUTHOR CONTRIBUTIONS

F.A., J.H., and P.T. conceived the study. F.A. performed all of the experiments apart from the FISH analysis, which was performed by N.W. and F.G. F.A., J.H., and P.T. drafted the manuscript, which was reviewed and edited by all authors.

CONFLICTS OF INTEREST

The authors declare that they have no competing financial interests.

ACKNOWLEDGMENTS

Fatwa Adikusuma,[1,4]
Nicole Williams,[1] Frank Grutzner,[1,2]
James Hughes,[1]
and Paul Thomas[1,2,3]

[1]School of Biological Science, The University of Adelaide, Adelaide, SA 5005, Australia; [2]Robinson Research Institute, The University of Adelaide, Adelaide, SA 5005, Australia; [3]South Australian Health and Medical Research Institute, Adelaide, SA 5000, Australia; [4]Centre for Biomedical Research, Faculty of Medicine, Diponegoro University, Semarang 50271, Indonesia

http://dx.doi.org/10.1016/j.ymthe.2017.05.021

Correspondence: Paul Thomas, School of Biological Science, The University of Adelaide, Adelaide, SA 5005.
E-mail: paul.thomas@adelaide.edu.au

REFERENCES

1. Barrangou, R., and Doudna, J.A. (2016). Applications of CRISPR technologies in research and beyond. Nat. Biotechnol. 34, 933–941.

2. Mali, P., Esvelt, K.M., and Church, G.M. (2013). Cas9 as a versatile tool for engineering biology. Nat. Methods 10, 957–963.

3. Dumanski, J.P., Rasi, C., Lönn, M., Davies, H., Ingelsson, M., Giedraitis, V., Lannfelt, L., Magnusson, P.K., Lindgren, C.M., Morris, A.P., et al. (2015). Mutagenesis. Smoking is associated with mosaic loss of chromosome Y. Science 347, 81–83.

4. Probst, F.J., Cooper, M.L., Cheung, S.W., and Justice, M.J. (2008). Genotype, phenotype, and karyotype correlation in the XO mouse model of Turner Syndrome. J. Hered. 99, 512–517.

5. Westhorpe, F.G., and Straight, A.F. (2013). Functions of the centromere and kinetochore in chromosome segregation. Curr. Opin. Cell Biol. 25, 334–340.

6. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. 8, 2281–2308.

7. Sakuma, T., Nishikawa, A., Kume, S., Chayama, K., and Yamamoto, T. (2014). Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. Sci. Rep. 4, 5400.

8. Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J., and Mateo, J.L. (2015). CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. PLoS ONE 10, e0124633.

9. Yang, H., Wang, H., and Jaenisch, R. (2014). Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. Nat. Protoc. 9, 1956–1968.

10. Chang, H.J., Clark, R.D., and Bachman, H. (1990). The phenotype of 45,X/46,XY mosaicism: an analysis of 92 prenatally diagnosed cases. Am. J. Hum. Genet. 46, 156–167.

Supplemental Information

Targeted Deletion of

an Entire Chromosome

Using CRISPR/Cas9

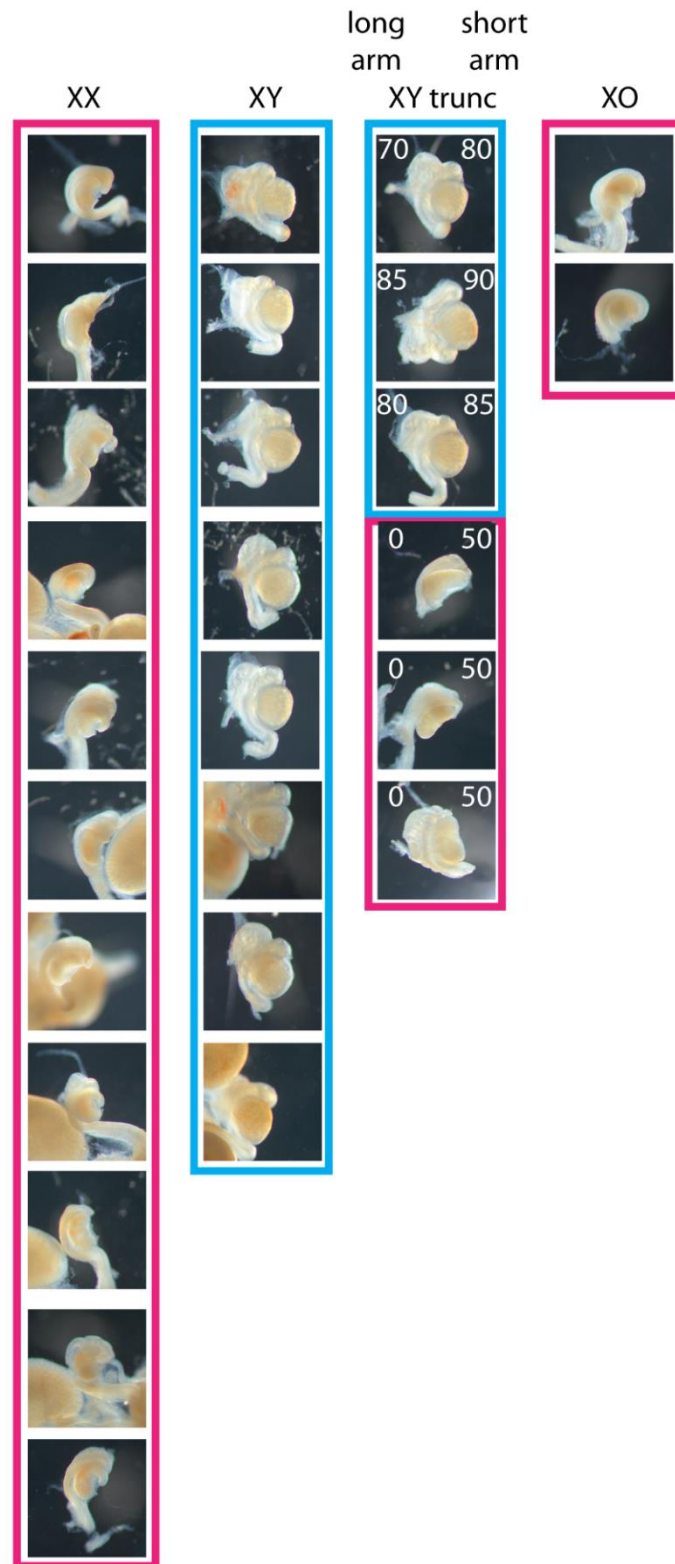Fatwa Adikusuma, Nicole Williams, Frank Grutzner, James Hughes, and Paul Thomas

Figure S1 Gonadal phenotypes of E15.5 mouse embryos from CRISPR/Cas9 zygote injection.
Phenotypically male (blue box) and female (pink box) gonads are shown with their sex chromosome dosage.
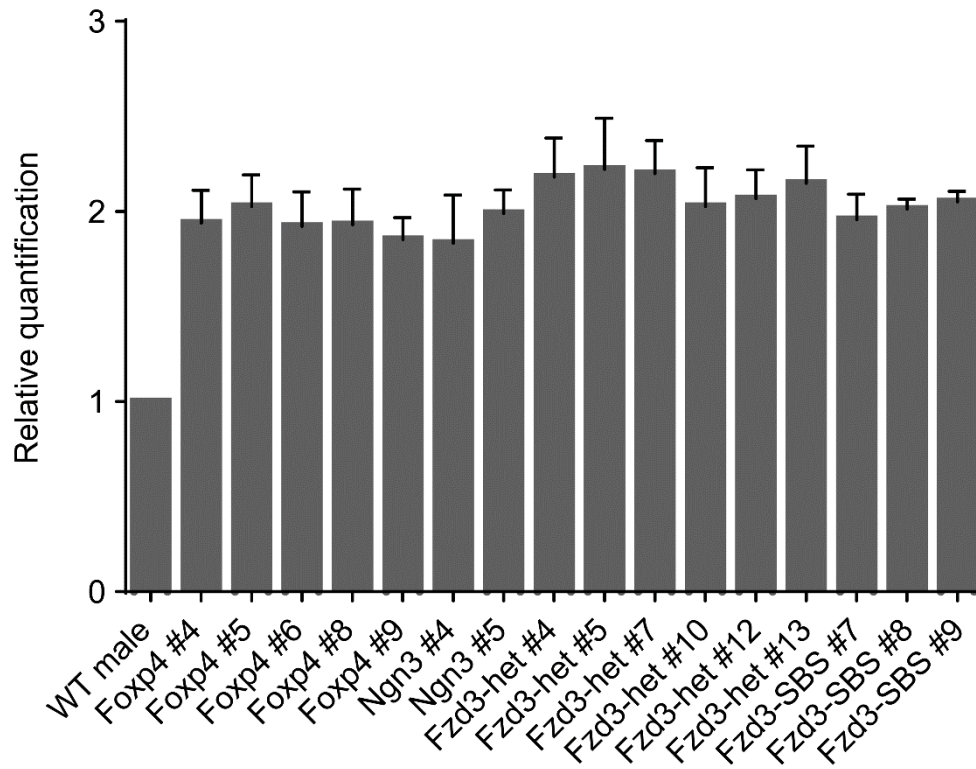
Figure S2 Assessment of X chromosome copy number by *Sox3* qPCR in female mice injected with autosomal gRNAs**.** Sixteen phenotypically female founder mice generated from injection of autosomal gRNAs targeting *Ngn3, Foxp4* and *Fzd3* genes had two copies of the X chromosome. These data indicate that spontaneous loss of the X chromosome does not occur in zygotes injected with CRISPR/Cas9 reagents. Data were presented as mean ± SD from n ≥ 3 replicates.

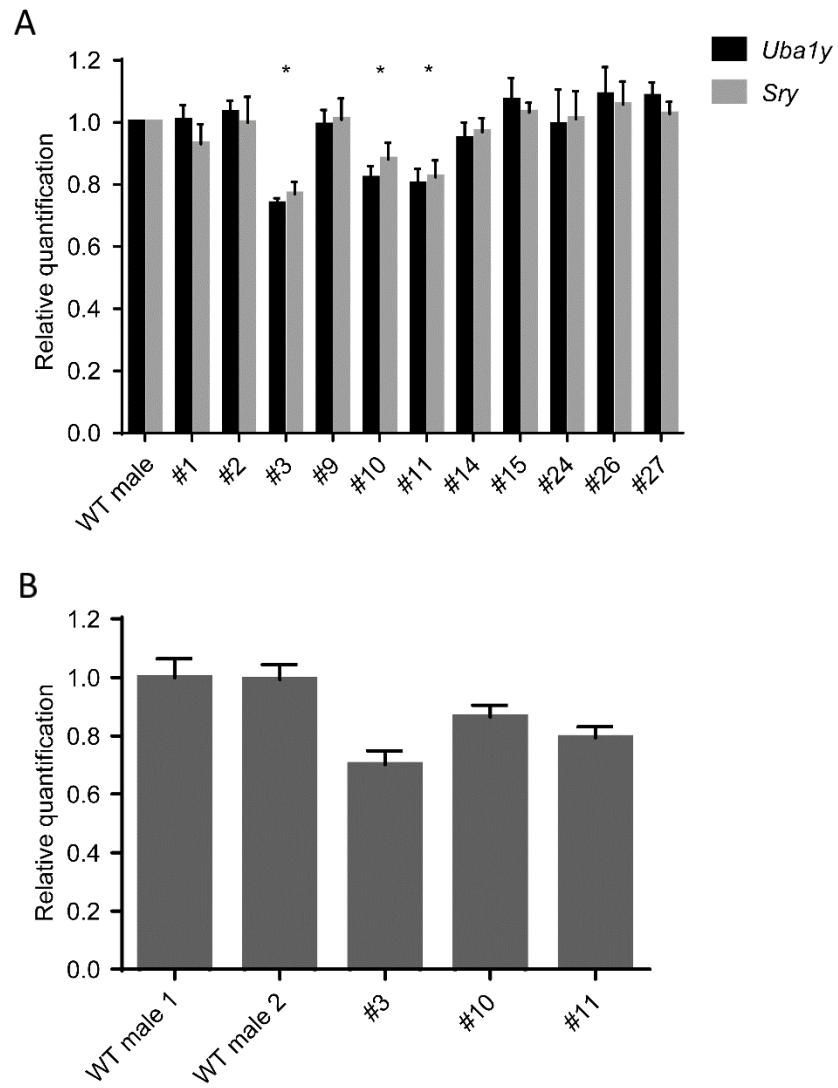**Figure S3** Assessment of Y chromosome dosage in males generated from gRNA centro 41X zygote injection. (**A**) qPCR analysis of *Uba1y* and *Sry* (which are located on Y short arm) revealed males with reduced Y dosage (asterisks) suggesting mosaic XY-XO. (**B**) Confirmation of the mosaicism by qPCR of *Gm28186* (which is located on Y long arm). Data were presented as mean ± SD from n ≥ 3 replicates.

Table S1 | List of gonadal phenotypes and genotypes of all mice generated from gRNA centro 41X injection

Identifier	Gonadal phenotype	Genotype	Additional info
#1	Male	XY	
#2	Male	XY	
#3	Male	XY, XO	Mosaic, with more than 70% XY cells
#4	Female	XO	
#5	Female	XX	
#6	Female	XX	
#7	Female	XX	
#8	Female	XX	
#9	Male	XY	
#10	Male	XY, XO	Mosaic, with more than 70% XY cells
#11	Male	XY, XO	Mosaic, with more than 70% XY cells
#12	Female	XX	
#13	Female	XX	
#14	Male	XY	
#15	Male	XY	
#16	Female	XX	
#17	Female	XX	
#18	Female	XX	
#19	Female	XO, XY_{short_arm}	Mosaic, half XO, half contain truncated Y short arm
#20	Female	XO, XY_{short_arm}	Mosaic, half XO, half contain truncated Y short arm
#21	Female	XO	
#22	Female	XO, XY_{short_arm}	Mosaic, half XO, half contain truncated Y short arm
#23	Female	XX	
#24	Male	XY	
#26	Male	XY	
#27	Male	XY	
#28	Female	XX	

Table S2 | On-targets and potential off-targets all the gRNAs used in this study (provided in excel file that can be downloaded from the publication source)

Table S3 | The two-way ANOVA statistical analysis of *Uba1Y* and *Erdr1* qPCR related to Figure 1B (provided in excel file that can be downloaded from the publication source)

Materials and Methods

gRNA screening and plasmid construction. sgRNAs were identified by manual screening of Y chromosome sequences using the CCTop gRNA design tool http://crispr.cos.uni-heidelberg.de/ provided by Stemmer *et al*. (2015). This tool was also used to predict the off-target potentials containing PAM sequences NGG and NAG (Supplementary information, Table S2). PX459.V2.0 (pSpCas9(BB)-2A-Puro, Addgene #62988) plasmid was used for Cas9 and sgRNA expression. PX459.V2.0 containing sgRNA was prepared as previously described by Ran *et al*. (2013). For dual gRNA centro 2X and long arm 2X, an additional U6-sgRNA cassette was added to the *NotI* site to allow simultaneous expression of two different gRNAs from single plasmid. Plasmid preparations were performed using PureLink® HiPure Plasmid Midiprep Kit (Life Technologies).

Cell culture and transfection. R1 mouse embryonic stem cells were cultured in 15% FCS/DMEM supplemented with 2 mM Glutamax (Gibco), 100 µM non-essential amino acid (Gibco), 100 µM 2-mercaptoethanol (Sigma), 3 µM CHIR99021 (Sigma) 1 µM PD0325901 (Sigma) and LIF (generated in-house). One million of ES cells were nucleofected with 3 µg of plasmid DNA using the Neon™ Transfection System 100 µL Kit (Life technologies) at 1400 V, 10 ms and 3 pulses according to the manufacturer's protocol. 24 hr after transfection, selection was conducted by adding puromycin (2 µg/ml) to the media for the next 48 hr. Surviving cells were cultured for 4-7 days without selection before harvesting.

DNA extraction and qPCR. Genomic DNA was extracted from 1-2 million ES cells or tail tissue using High Pure PCR Template Preparation Kit (Roche) according to the manufacturer's instructions. qPCRs were performed using Fast SYBR Green Master Mix (Applied Biosystems) on an Applied Biosystems 7500 StepOnePlus machine. *Sox1* qPCR was used as internal reference control to normalize qPCR value in all qPCR analyses.

FISH analysis. Cells were cultured in media containing 0.1 µg colcemid (Roche) for 1-2 hours, harvested and incubated in 0.075 M KCl hypotonic solution for 20 minutes. The cells were then fixed using methanol-acetic acid (3:1) solution, dropped onto slides and dried. FISH staining was performed using Mouse IDetect™ Chromosome Y Paint Probe (Empire Genomics) according to the manufacturer's instructions. Y signals were counted from both metaphase and interphase spreads.

Mouse zygote injection. All the experiments involving animal use have been approved by the University of Adelaide Animal Ethics Committee. Cas9 mRNA was produced by *in vitro* transcription of *XhoI*-linearized pCMV/T7-hCas9 (Toolgen) using mMESSAGE mMACHINE® T7 ULTRA Transcription Kit (Ambion). sgRNA centro 41X was generated according to a previously described protocol.[35] In brief, PCR was performed using a T7 containing forward primer 5'-TTAATACGACTCACTATAGAGGAGTTAATATAAAAAACA-3' and a reverse primer 5'-AAAAGCACCGACTCGGTGCC-3' and the PX459.V2 centro 41X plasmid template. The product was purified by QIAquick PCR Purification Kit (Qiagen) and used as a template for in vitro transcription using the HiScribe™ T7 Quick High Yield RNA Synthesis Kit (NEB). RNA purification was conducted using RNeasy Mini Kit (Qiagen). Cas9 mRNA (200 ng/µl) and sgRNA centro 41X (100 ng/µl) were injected to the cytoplasm of C57BL/6N zygotes using a Femtojet microinjector. The survival rate of injected zygotes was 89.4% (93/104). 57 zygotes were transferred into 3 pseudo pregnant females (19/recipient). 27 embryos with normal appearance were harvested at E15.5 for gonadal assessment and tissue collection.

List of gRNA sequences used

Name	gRNA sequences 5'-3'	Position in Y
Centro 2X	Left: GGATAAATGTTACATGCAA	4.064.613
	Right: GATAATAGTTTACTATTCTAA	4.163.810
Centro 41X	GGAGTTAATATAAAAAACA	4.065.169 to 4.159.436
Centro 140X	GAAGAATTACAATGAAAAATA	4.065.349 to 4.161.710
Long arm 2X	Left: GTCCTCTACGTCTATCAGGA	4.312.132
	Right: GTTTCCAGCCGGGTTTCTTAC	4.412.892
Long arm 8X	GTTCTATGTCAATTTAGGTGG	4.313.453 to 17.275.105
Long arm 45X	GACTGGGTTCTCCTAATCCTT	4.417.594 to 90.167.758
Long arm 116X	GTGGAATTGTGATCTAGATA	5.726.265 to 88.887.822
Long arm 298X	GGCAAAGCACTTCTGCACC	4.596.490 to 90.662.856
Neo	GGCAGCGCGGCTATCGTGGC	None in mouse

Red highlight indicates additional G was added to the gRNAs

List of qPCR primers

qPCR	F (5'-3')	R (5'-3')	Position
Ubaly (Y short arm)	GGCCACAGAGACTTGGGCCGAC	TGCCTTGTGGTGCCTGTGGC	chrY: 831.667 - 831.891 & 681.963 - 682.187
Erdr1 (Y long arm)	CTGACTGCGTACAGAGAAATGTCC	GGAAGACACACACACATCTGCA	chrY: 90.822.353 - 90.822.426 & 90.816.416 - 90.816.489
Sry (Y short arm)	CATTTATGGTGTGGTCCCGTG	ATCTTCAATCTCTGTGCCTCCT	chrY: 2.663.492 - 2.663.607
Gm28186 (Y long arm)	CATGCCTCAGACCCCTCAAG	TCTGACCATTGTGACTCAGAC	chrY: 4.372.957 - 4.373.114
Xist	GCCATCCTCCCTACCTCAGAA	CCTGACATTGTTTTCCCCCTAA	chrX
Dmd	ACAGCAAGCAAGCAAGGCTTC	CTCACCAGTGACCCTATGATTGC	chrX
Sox3	CGTTGCCTTGTACCGAAGAT	CGGGACTTCTCGCTTTTGTA	chrX
Sox1	GACTTGCAGGCTATGTACAACATC	CCTCTCAGACGGTGGAGTTATATT	Chr8

Supplemental Reference:

1. Wang, H. et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910-918 (2013).

Chapter 4:

Versatile single-step-assembly CRISPR/Cas9 vectors for dual gRNA expression

4.1. Summary

The availability of vast array of CRISPR/Cas9 plasmid vectors via repositories such as Addgene has been invaluable for researchers who are interested in genome editing. Popular CRISPR/Cas9 constructs include vectors from Zhang laboratory at the Broad Institute, such as the Cas9 nuclease plasmids pX330, pX458 and pX459.V2 and the Cas9 nickase plasmids pX335, pX461 and pX462. These vectors are very user-friendly as they allow expression of all CRISPR/Cas9 components (Cas9 and a single gRNA), including selection markers (pX458 and pX461 contain a GFP cassette; pX459.V2.0 and pX462.V2.0 contain a puromycin resistance cassette) from a single plasmid (all-in-one). Moreover, these vectors contain a 'golden gate' cloning site at the gRNA site, rendering these vectors easily customisable for facile generation of gRNAs of interest by a one-step digestion ligation protocol (see http://www.genome-engineering.org/crispr/ for more detail).

However, some applications require simultaneous expression of two gRNAs, such as inducing DSBs using a paired-nickase strategy, generating large deletions or generating double KO lines, and there are no vectors available for the simple generation of all-in-one plasmids expressing dual gRNAs. Existing multiplex CRISPR/Cas9 vectors require multiple cloning or PCR steps to generate all-in-one vectors expressing dual gRNAs.

This chapter describes the design and construction of versatile CRISPR/Cas9 plasmid vectors that allow a one-step cloning reaction for the generation of all-in-one vectors expressing dual gRNAs. The previously mentioned CRISPR vectors from Zhang laboratory were modified by adding an extra gRNA expression cassette with slight sequence modifications at the golden gate cloning site. These modified vectors were named pDG330, pDG458, pDG459, pDG335, pDG461 and pDG462, based on the name of the original vectors (DG stands for dual gRNAs). Generation of all-in-one vectors expressing dual gRNAs of interest can be achieved easily using a one-step cloning

protocol. The ability of these vectors to perform tasks requiring dual gRNA expression, such generating mutations at two different sites, generating intervening large deletions and inducing DSBs using the paired-nickase strategy was tested, and all could be accomplished efficiently using these dual gRNA vectors. This study is presented as a manuscript that will be submitted to a scientific journal. Finally, these vectors will be made available to the scientific community through Addgene to help researchers performing experiments requiring simultaneous expression of dual gRNAs.

Statement of Authorship

Title of Paper	Versatile single-step-assembly CRISPR/Cas9 vectors for dual gRNA expression
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Fatwa Adikusuma*, Chandran Pfitzner* and Paul Thomas. *Versatile single-step-assembly CRISPR/Cas9 vectors for dual gRNA expression.* 2017

Principal Author

Name of Principal Author (Candidate)	Fatwa Adikusuma		
Contribution to the Paper	Designed and conceived the study, conducted the experiments and wrote the manuscript.		
Overall percentage (%)	65%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	8/8/2017

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i.   the candidate's stated contribution to the publication is accurate (as detailed above);

ii.   permission is granted for the candidate in include the publication in the thesis; and

iii.   the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Chandran Pfizner		
Contribution to the Paper	Performed experiments and edited the manuscript		
Signature		Date	

Name of Co-Author	Paul Thomas		
Contribution to the Paper	Designed, conceived, supervised the study, and wrote the manuscript		
Signature		Date	8/8/2017

Statement of Authorship

Title of Paper	Versatile CRISPR/Cas9 vectors for facile generation of paired gRNAs
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Fatwa Adikusuma*, Chandran Pfitzner* and Paul Thomas. *Versatile CRISPR/Cas9 vectors for facile generation of paired gRNAs*. 2017

Principal Author

Name of Principal Author (Candidate)	Fatwa Adikusuma
Contribution to the Paper	Designed and conceived the study, conducted the experiments and wrote the manuscript.
Overall percentage (%)	65%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature		Date	

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

     i.   the candidate's stated contribution to the publication is accurate (as detailed above);

     ii.   permission is granted for the candidate in include the publication in the thesis; and

     iii.   the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Chandran Pfitzner
Contribution to the Paper	Performed experiments and edited the manuscript
Signature	Date 22/7/2017

Name of Co-Author	Paul Thomas
Contribution to the Paper	Designed, conceived, supervised the study, and wrote the manuscript
Signature	Date

Versatile single-step-assembly CRISPR/Cas9 vectors for dual gRNA expression

Fatwa Adikusuma[1,4,5], Chandran Pfitzner[1,5] and Paul Thomas[1,2,3,*]

[1] School of Biological Sciences, The University of Adelaide, South Australia, Australia, 5005

[2] Robinson Research Institute, The University of Adelaide, South Australia, Australia, 5005

[3] South Australian Health and Medical Research Institute, South Australia, Australia, 5000

[4] Centre for Biomedical Research, Faculty of Medicine, Diponegoro University, Semarang, Indonesia, 50271

[5] Contributed equally

* Corresponding author paul.thomas@adelaide.edu.au

Abstract

CRISPR/Cas9 technology enables efficient, rapid and cost-effective targeted genomic modification in a wide variety of cellular contexts including cultured cells. Some applications such as generation of double knock-outs, large deletions and paired-nickase cleavage require simultaneous expression of two gRNAs. Although single plasmids that enable multiplex expression of gRNAs have been developed, these require multiple rounds of cloning and/or PCR for generation of the desired construct. Here, we describe a series of vectors that enable generation of customized dual-gRNA expression constructs via an easy one-step golden gate cloning reaction using two annealed oligonucleotide inserts with different overhangs. Through nucleofection of mouse embryonic stem cells, we demonstrate highly efficient cleavage of the target loci using the dual-guide plasmids, which are available as Cas9-nuclease or Cas9-nickase expression constructs, with or without selection markers. These vectors are a valuable addition to the CRISPR/Cas9 toolbox and will be made available to all interested researchers via the Addgene plasmid repository.

Introduction

CRISPR/Cas9 technology is a powerful genome editing tool that has become widely used by researchers to generate targeted genetic modifications in many contexts including cultured cell lines and zygotes. CRISPR/Cas9 offers several advantages over preexisting genome editing technologies including ease of use, relatively low cost and high activity (Barrangou and Doudna, 2016; Hsu et al., 2014; Komor et al., 2017; Mali et al., 2013b; Sander and Joung, 2014). The CRIPSR/Cas9 platform comprises two components; Cas9, which functions as a programmable endonuclease that generates a blunt-ended double-stranded break (DSB) and a ~100 nt guide RNA (gRNA), in which the ~20 nt at the 5' end directs Cas9 to the target site via RNA:DNA complementary base pairing (Cong et al., 2013; Jinek et al., 2012; Mali et al., 2013c). Generation of a targeted DSB can be achieved by delivery of Cas9 and gRNA components in plasmid, RNA or ribonucleoprotein (RNP) forms. For some applications, such as cultured cells, plasmids are generally preferred due to their ease of generation and stability. Commonly used plasmids for expression of Cas9 or Cas9-nickase (D10A) and single gRNA are available from the Zhang laboratory and can be obtained through the Addgene plasmid repository. These plasmids contain both gRNA and Cas9 expression cassettes in a single plasmid with optional selection markers such as puromycin or GFP to facilitate screening. Importantly, generation of a unique customized gRNA of interest can be performed easily as the gRNA cloning site contains BbsI restriction sites, allowing a one-step golden gate cloning approach for insertion of a pair of annealed oligonucleotides containing the specific ~20 bp guide sequence (Cong et al., 2013; Ran et al., 2013b).

To simultaneously target a pair of genomic regions, expression of two gRNAs is required. While this can be achieved by co-transfection of two plasmids, this process can be inefficient. To achieve efficient dual cuts, all CRISPR/Cas9 components with dual-gRNAs

should be expressed from a single plasmid. Single plasmids expressing multiple gRNAs have been developed, however generation of the desired constructs using those existing plasmids require multiple cloning and/or PCR steps. Here we modify commonly-used vectors from the Zhang laboratory so that each plasmid can express two gRNAs and can be generated via a simple one-step cloning method. We show that these plasmids, termed dual-gRNA plasmids, provide an efficient tool for experiments requiring simultaneous expression of two gRNAs such as multiplexed knock-out of two genes, generation of large deletions and generation of indels using Cas9-nickase. These vectors are a valuable addition to the CRISPR/Cas toolbox and will be made available through the Addgene plasmid repository.

Results

Generation of vectors

To generate plasmids that permit simultaneous expression of two gRNAs, we inserted an additional hU6-gRNA expression cassette into the available CRISPR plasmids from the Zhang laboratory. The second cassette was positioned in the opposite orientation to the original hU6-gRNA expression cassette to reduce the possibility of recombination (Fig. 1A). The additional cassette also contains a BbsI golden gate site at the guide insertion site as per the original cassette. However, unlike the original BbsI site which generates GTTT and GGTG overhangs, the new site generates CGGT and TTTA overhangs (Fig. 1B) allowing simultaneous targeted insertion of two annealed oligonucleotides with different complementary overhangs in a one-step digestion-ligation reaction (Fig. 1C; see below). We added the extra gRNA cassette to the following Cas9 nuclease vectors: pX330 (no selection marker), pX458 (GFP selection marker) and pX459.V2.0 (puromycin selection

marker), and to the following Cas9-nickase vectors: pX335 (no selection marker), pX461 (GFP selection marker) and pX462.V2.0 (puromycin selection marker). Those vectors were named pDG330, pDG458, pDG459, pDG335, pDG461 and pDG462, respectively.

Efficient generation of custom dual-gRNA vector using a one-step cloning protocol

Having generated the dual-gRNA vectors, we next tested whether we could simultaneously insert two annealed oligonucleotide duplexes in a one-step cloning process. We designed two gRNA oligonucleotide inserts targeting the mouse *Sox1* and *Sox3* genes. These inserts carried BspMI and SacI restriction sites at the original and second hU6-gRNA sites, respectively. Annealed oligonucleotide duplex pairs and pDG459 vector were subjected to a one-step digestion-ligation cycling protocol followed by bacterial transformation (Fig. 1C). All 12 colonies analyzed contained vectors with correct assembly based on their RFLP pattern (Fig. 1D). This demonstrates that our dual-gRNA vector design combined with the one-step cloning protocol can allow easy and efficient generation of CRISPR/Cas9 vectors with dual-gRNA expression cassettes.

Efficient generation of DSB at two sites using vectors expressing Cas9 nuclease and dual-gRNAs

We next tested whether the dual-gRNA Cas9-nuclease vectors could efficiently induce indels or deletions through simultaneous digestion at two target sites. Four different pDG459 derivatives were initially generated; the first targeted *Sox1* site A and *Sox3* site A (pDG459 Sox1A/Sox3A), the second targeted *Sox1* site B and *Sox3* site B (pDG459 Sox1B/Sox3B), the third targeted *Sox1* site A and *Sox1* site B (pDG459 Sox1A/Sox1B) which are separated by 51 bp and the last targeted *Sox3* site A and *Sox3* site B (pDG459

Sox3A/Sox3B) which are separated by 47 bp (Fig. 2A). All target sequences contained restriction sites and hence indel generation at each site could be assayed by RFLP analyses. In addition, efficient digestion by pDG459 Sox1A/Sox1B or pDG459 Sox3A/Sox3B gRNAs should cause a deletion of ~50 bp which can be readily detected by PCR. Each of the four constructs were separately transfected to the mouse ES cells followed by puromycin selection to ensure only transfectants were harvested. *Sox1* and *Sox3* PCRs were performed on Sox1A/Sox3A-treated samples followed by a BfuAI (isoschizomer of BspMI) and SacI RFLP assay to assess indel generation at Sox1A and Sox3A sites, respectively. Both RFLP analyses indicated that pDG459 Sox1A/Sox3A plasmid induced mutations with ~100% efficiency at both Sox1A and Sox3A sites (Fig. 2B and Fig. S1A). Highly efficient mutagenesis of the Sox1B and Sox3B sites was also detected by ApaI and SfoI RFLP assays in pDG459 Sox1B/Sox3B-trasfected cells (Fig. 2B and S1B). We next examined whether deletion of the sequences between the cut sites could be induced by pDG459 Sox1A/Sox1B or Sox3A/Sox3B transfection. PCR products corresponding to deletion alleles were readily generated in pDG459 Sox1A/Sox1B- or Sox3A/Sox3B-treated samples but not in the WT and the unpaired controls upon *Sox1* or *Sox3* PCR (Fig. 2B, S1C-D). Efficient dual nuclease activity was also demonstrated using pDG330- and pDG458-derived constructs (Fig. S2A-B). Together, these data indicate that all-in-one dual-gRNA Cas9 nuclease vectors can facilitate efficient simultaneous cutting at two gRNA target sites.

Efficient DSBs induced by plasmids expressing Cas9-nickase and dual paired-gRNAs

Expression of Cas9-nickase with a single gRNA results in a ssDNA break that is typically repaired without causing a mutation. In contrast, expression of Cas9-nickase and two gRNAs targeting closely spaced sites on opposite DNA strands will generate a staggered

DSB, repair of which results in indel mutations (Mali et al., 2013a; Ran et al., 2013a). We next tested the dual-gRNA Cas9-nickase vectors to assess whether they could efficiently induce DSBs via expression of gRNA pairs. We generated pDG462 derivatives targeting Sox1A/Sox1B and Sox3A/Sox3B which have the requisite orientation and spacing to permit mutagenesis by paired-nickase activity (Fig. 2A). As negative controls, we also generated pDG462 targeting Sox1A/Sox3A and Sox1B/Sox3B which are not paired therefore should not generate indel mutations. Vectors were transfected to mouse ES cells followed by puromycin selection. T7E1 heteroduplex assays revealed that pDG462 Sox1A/Sox1B and Sox3A/Sox3B efficiently generated mutations at *Sox1* and *Sox3*, respectively (Fig. 3 and S3). In contrast, there was no evidence of mutations after transfection of the non-paired control plasmids (Fig. 3 and S3). Efficient mutation of *Sox3* was also achieved using dual-gRNA nickase vectors pDG335 and pDG461 expressing Sox3A/Sox3B (Fig. S4). Together, these data demonstrate efficient targeted mutagenesis using dual-gRNA paired-nickase vectors.

Discussion

Plasmids from the Zhang laboratory have greatly simplified generation of customized gRNA-Cas9/Cas9-nickase expression constructs through utilization of the golden gate cloning strategy. Users only need to anneal a pair of oligonucleotides and ligate them into the vectors via a one-step cloning process, circumventing multiple rounds of PCR and cloning (Cong et al., 2013; Ran et al., 2013b). We modified available plasmids to allow simultaneous insertion of two oligonucleotide duplex inserts using the simple one-step cloning method. These modified vectors provide a user-friendly and cost effective system to perform experiments that require simultaneous expression of two gRNAs. Additionally, we have shown that both gRNA cassettes are active and induce mutations with high

efficiency at both target sites when combined with reliable transfection and selection methods.

Other recent studies have also generated CRISPR/Cas9 vectors that are able to express dual-gRNAs simultaneously, most of which also take advantage of golden gate cloning. However, unlike the dual-gRNA vectors described herein, these require multiple rounds of cloning and/or PCR (Kabadi et al., 2014; Maddalo et al., 2014; Sakuma et al., 2014; Vidigal and Ventura, 2015). Additionally, the strategy to express dual-gRNA as a polycistronic transcript that is split by Csy4 RNA polymerase (Tsai et al., 2014) has been shown to have low efficiency (Han et al., 2017). Furthermore, our dual-gRNA vectors are available with Cas9 nuclease or nickase, and with or without selection markers, and can therefore be utilized in a broad range of experimental contexts. Vectors from other studies, although more complicated, are useful when conducting experiments requiring more than 2 gRNAs since those vectors can bear up to 7 gRNAs in a single vector (Kabadi et al., 2014; Sakuma et al., 2014).

Our one-step cloning strategy could be applied to generate multiple gRNAs by adding more hU6-gRNA cassettes. To do so, the BbsI sites of the new cassettes would need to be modified to produce different unique overhangs upon digestion. This cloning approach could also be combined with other commonly used CRISPR platform variants such as Cpf1, dCas9-Fok1, Cas9-HF, eSpCas9, and other Cas9 orthologs or mutants that recognize different PAM sequences.

Off-target mutagenesis is one of the most significant issues of CRISPR/Cas9 genome editing (Fu et al., 2013; Hsu et al., 2013), particularly for therapeutic applications. The paired-nickase strategy has previously been shown to minimize the off-target effects that are a feature of Cas9 nuclease (Cho et al., 2014; Frock et al., 2015). We therefore

anticipate that the dual-gRNA nickase vectors will be an attractive option for users who require efficient mutagenesis and with maximum specificity.

Efficient dual nuclease cuts are useful for generating targeted large deletions for many purposes such as studying the function of enhancers or long non-coding RNA. In some situations, targeted large deletions are required to delete an exon such as for DMD therapeutics via exon skipping (Long et al., 2016; Nelson et al., 2016; Tabebordbar et al., 2016) or to delete a centromere for chromosome removal (Adikusuma et al., 2017b). Dual-gRNA Cas9 vectors could also be used for simultaneous KO of two different genes. We also offer our dual-gRNA nuclease vectors for efficient generation of chromosome translocations to model diseases such as Burkitt's lymphoma or acute myeloid leukemia (Maddalo et al., 2014). Dual DSBs may also aid insertion of flanking loxP sequences for conditional deletion and for insertion of gene swap constructs (Adikusuma et al., 2017a; Quadros et al., 2017). Furthermore, these vectors can also be used for injection into mouse zygotes for the generation of mutant mice (Mashiko et al., 2013). Taken together our vectors are a valuable addition to the CRISPR/Cas9 toolbox and should be useful for many CRISPR/Cas9-based applications.

Materials and Methods

Plasmid and gRNA design

Plasmids pX330, pX335, pX458, pX459.V2.0, pX461 and pX462.V2.0 were gifts from Feng Zhang (Addgene plasmid 42230, 42335, 48138, 62988, 48140 and 62987, respectively). The Cas9 or Cas9-nickase of those plasmids are derived from *Streptococcus pyogenes* Cas9 which recognizes NGG PAM sequences. The BbsI sequences from pX330 were replaced with the second version of BbsI sequences (see Fig 1B). The hU6-gRNA

region was then amplified using primers containing NotI sites. PCR products were then ligated to original plasmids at the NotI site. Guide sequences targeting Sox1A, Sox1B, Sox3A and Sox3B were 5'-GCCGCCGGGCGAGTGCAGGT-3', 5'-GCCCACGAACCTCTCGGGCC-3', 5'-GCTGACCCACATCTGAGCTC-3' and 5'-GACCGCAGTCCCGGCGCCC-3', respectively, which were designed using online CRISPR design tool http://crispr.mit.edu/.

One step cloning for the generation of customized dual-gRNA plasmid

Forward and reverse oligonucleotides containing the guide sequences for Sox1A, Sox1B, Sox3A and Sox3B with appropriate overhangs (see Fig 1B) were phosphorylated and annealed by mixing 100 pmol of each pair and 0.5 µL T4 PNK (NEB) then incubated at 37 °C for 30 minutes, 95 °C for 5 minutes and slowly ramped to RT. Annealed oligonucleotides were diluted 1 in 125. Pairs of oligonucleotide duplexes were ligated into the empty vectors in a one-step digestion ligation reaction by mixing the diluted duplex oligonucleotide pairs (1 µL each) with 100 ng empty vector, 100 µmol of DTT, 10 µmol of ATP, 1 µL of BbsI (NEB), 0.5 µL of T4 ligase (NEB) and NEB-2 buffer in 20 µL of reaction. The mixture was placed in a thermocycler and cycled 6 times at 37 °C for 5 minutes and 16 °C for 5 minutes before bacterial transformation. Plasmids were prepared using miniprep kit (Qiagen) or PureLink® HiPure Plasmid Midiprep Kit (Life Technologies). Correct insertion of oligonucleotide duplexes into the vectors was confirmed by Sanger sequencing using the following primers: GGTTTCGCCACCTCTGACTTG (first insert) and TGCATCGCATTGTCTGAGTAGG (second insert). It is recommended to digest the vectors using BbsI before sequencing as correct insertion should remove the BbsI sites. List of oligo sequences that were used to generate dual-gRNA plasmids can be seen in Supplementary Table S1.

Cell culture and transfection

R1 mouse embryonic stem cells were used for all experiments. Cells were cultured in 15% FCS/DMEM supplemented with LIF, 3 μM CHIR99021 (Sigma), 1 μM PD0325901 (Sigma), 2 mM Glutamax (Gibco), 100 μM non-essential amino acids (Gibco) and 100 μM 2-mercaptoethanol (Sigma). One million ES cells were nucleofected with 3 μg of plasmid DNA using the Neon™ Transfection System 100 μL Kit (Life technologies) at 1400 V, 10 ms and 3 pulses according to the manufacturer's protocol. For transfection of pDG459 and pDG462, puromycin selection (2 μg/mL was initiated 24 hours post transfection for 48 hours. GFP FACS was performed on cells transfected with pDG458 and pDG461 48 hours post transfection. Surviving cells were cultured for 4-7 days without selection before harvesting. Cells transfected with plasmid pDG330 and pDG335 did not undergo any selection.

DNA extraction, PCR, RFLP and T7E1 assay

Genomic DNA was extracted from 1-2 million cells using High Pure PCR Template Preparation Kit (Roche) according to the manufacturer's instructions. *Sox1* PCR was performed using primers F: 5'-CCCTTCTCTCCGCTAGGC-3' and R: 5'-GTTGTGCATCTTGGGGTTTT-3'. *Sox3* PCR used primers F: 5'-CAGCATGTACCTGCCACCT-3' and R: 5'-ACAAAACCCCGACAGTTACG-3'. RFLP or T7E1 assay was performed by mixing 5 μL of PCR products (without purification) with the restriction enzymes or T7E1 enzyme (NEB) in a total volume of 20 μL and incubated for 1 hour at the suggested optimal temperatures. Prior to T7E1 assay, PCR products were slowly re-annealed to form heteroduplex products by heating the PCR products at 95° C for 5 minutes and slowly ramped down to room temperature.

Acknowledgments

Author contributions

F.A & P.T conceived and designed the study. F.A & C.P performed the experiments. F.A and P.T wrote the manuscript with the input from all authors.
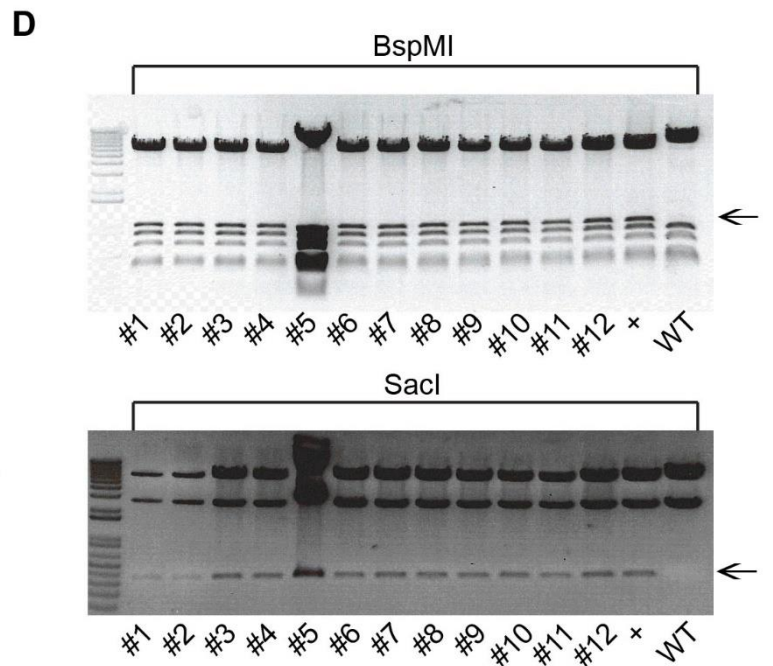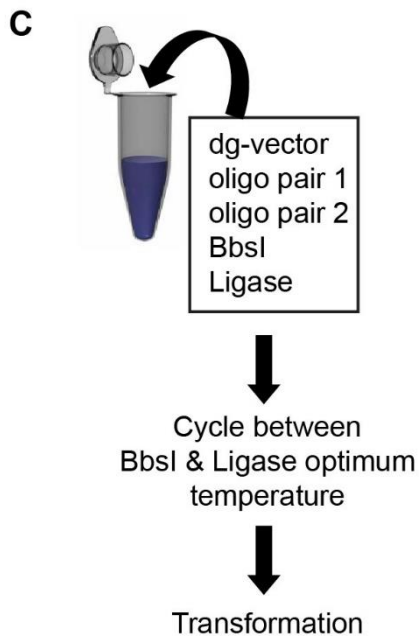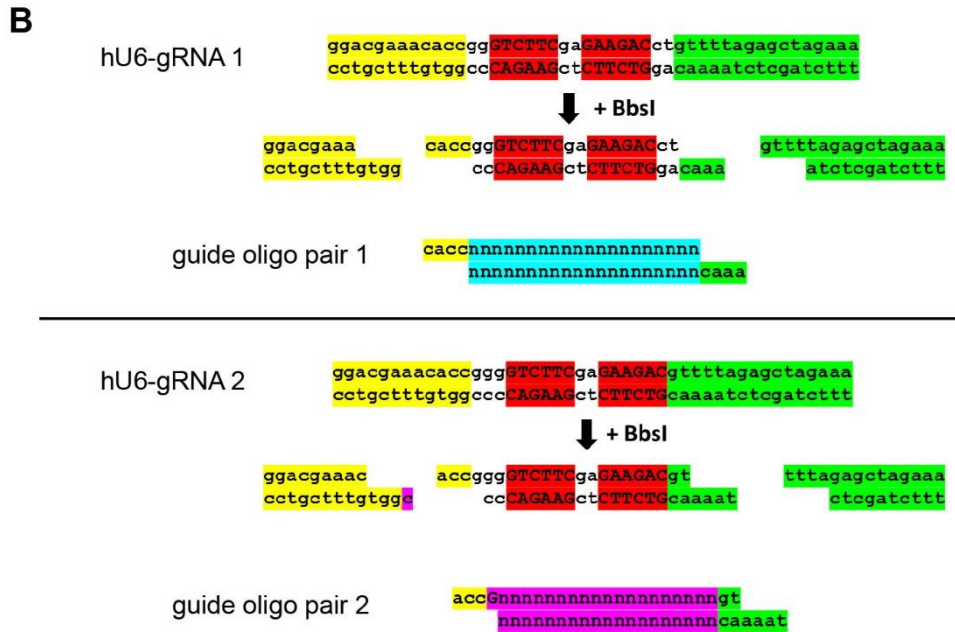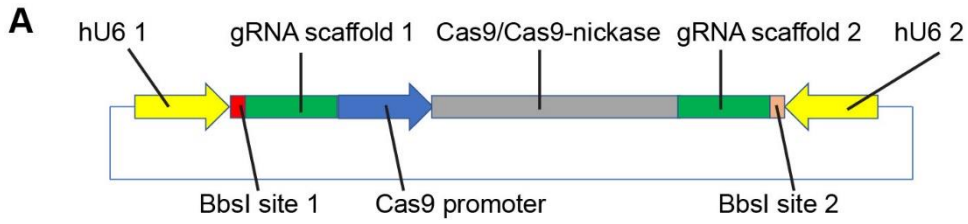
Figure 1 | Generation of dual-gRNA expressing vectors. (A) Schematic of dual-gRNA vectors. (B) Golden gate cloning strategy for insertion of specific guide sequences into each cassette. Note that the BbsI sites generate different overhangs after restriction digest. Red highlights indicate the BbsI sites, yellow and green highlights are part of hU6 promoter and gRNA, respectively, that are necessarily present in the plasmid. Blue and purple highlights indicate the unique customized guide sequences (C) One-step cloning protocol for the generation of customized dual-gRNA vectors. (D) Insertion of Sox1A and Sox3A oligonucleotide duplexes into pDG459 resulted in correct insertions in all 12 colonies as indicated by BspMI and SacI restriction digest. The black arrow indicates the diagnostic band for correct insertion.
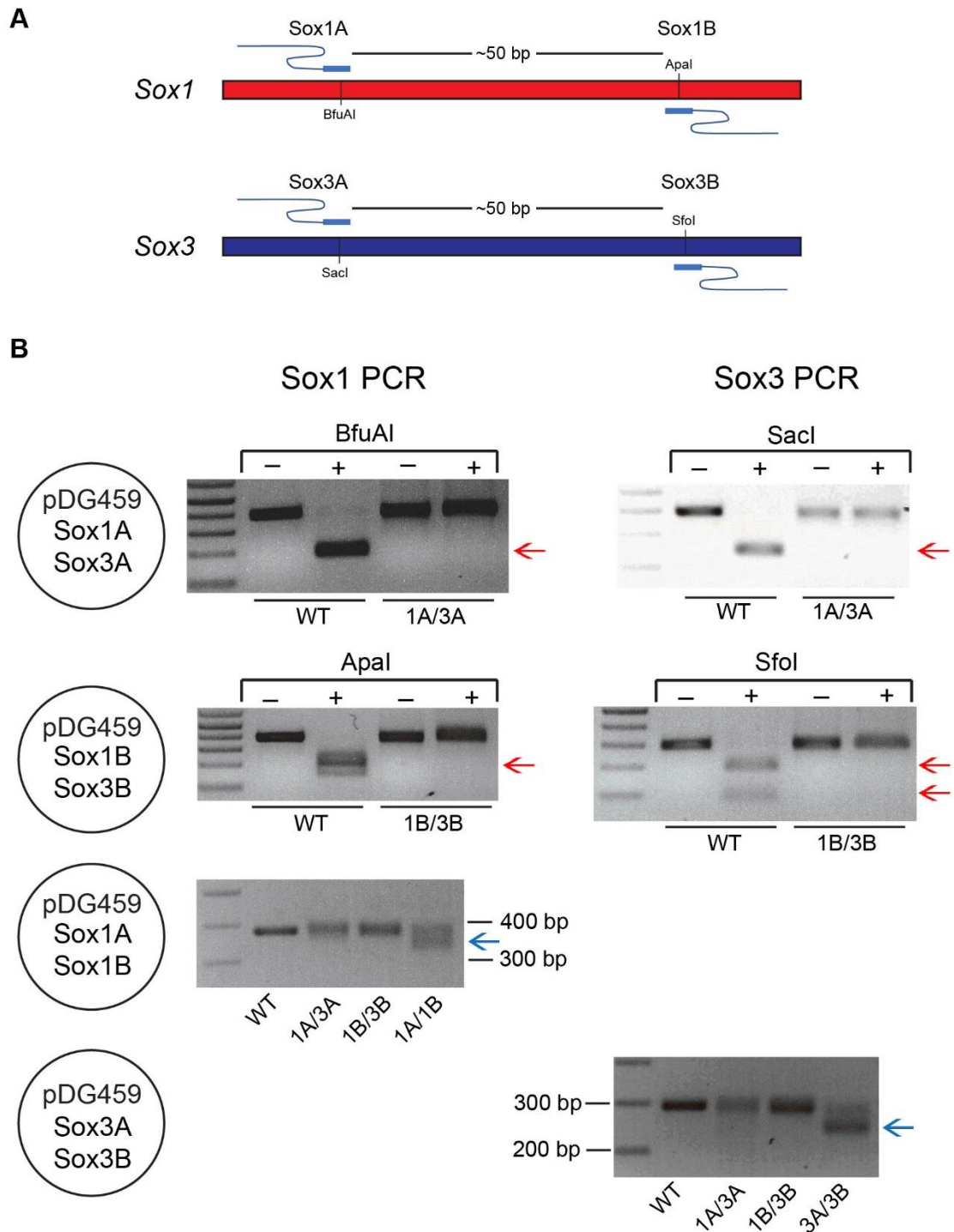
Figure 2 | Efficient dual cutting mediated by pDG459 vector. (A) Schematic of gRNA target sites in the *Sox1* and *Sox3* genes. (B) Highly efficient dual cuts induced by vectors derived from pDG459 as indicated by PCR and RFLP analyses. WT products were cut by restriction enzymes resulting in bands indicated by the red arrows. Absence of these bands in dual-gRNA vector-treated samples indicated that the Cas9 nuclease and the gRNAs efficiently induced mutations thus destroying the restrictions sites. Efficient cuts from pDG459 Sox1A/Sox1B and pDG459 Sox3A/Sox3B were indicated by deletion of ~50 bp regions between cuts (blue arrows). Complete figures with more independent samples can be found in Supplementary Figure S1.
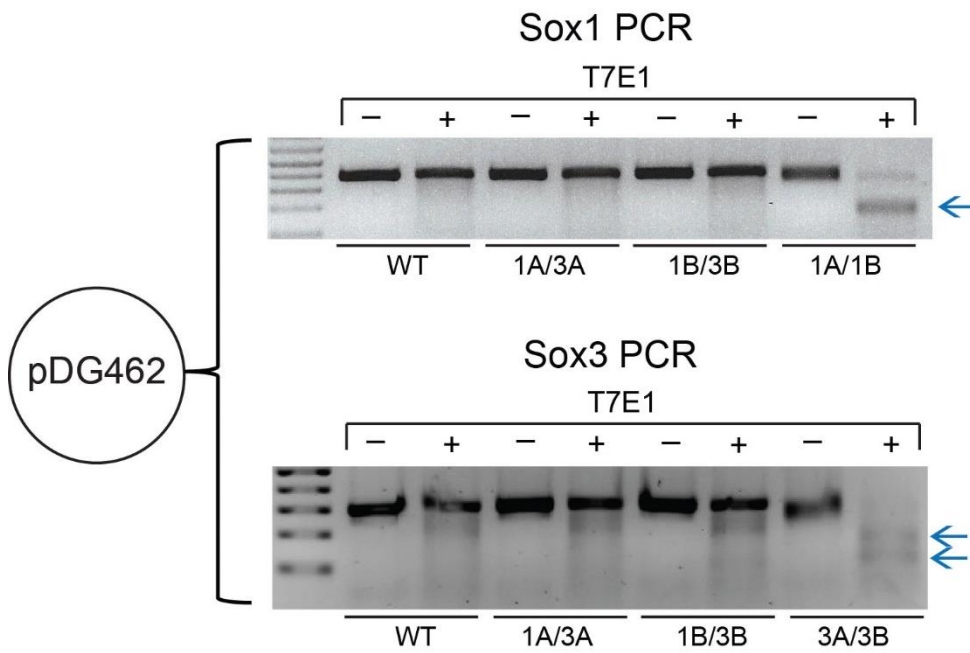
Figure 3 | Paired-nickase DSB induction by pDG462. *Sox1* or *Sox3* PCR followed by T7E1 assay was performed on pDG462-transfected samples. Mutations in *Sox1* and *Sox3* were induced by pDG462 Sox1A/Sox1B or pDG462 Sox3A/Sox3B, respectively, as indicated by the digested products after T7E1 treatment (blue arrows). Mutations were not induced by non-paired-nickase control plasmids (pDG462 Sox1A/Sox3A or pDG462 Sox1B/Sox3B). Complete figures with more independent samples can be found in Supplementary Figure S3.
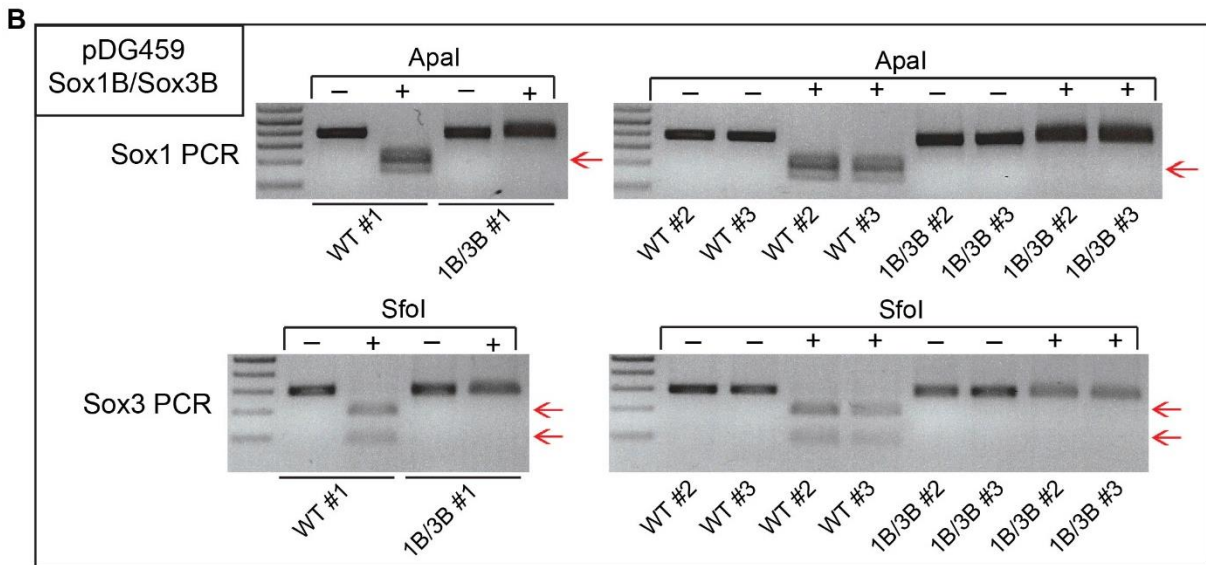
Supplemental information

Figure S1-S4 and Table S1

Versatile single-step-assembly CRISPR/Cas9 vectors for dual gRNA expression

Fatwa Adikusuma, Chandran Pfitzner and Paul Thomas
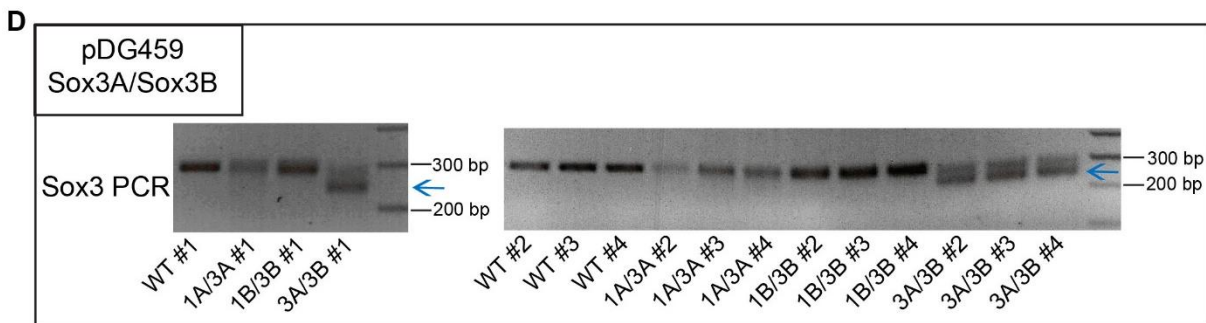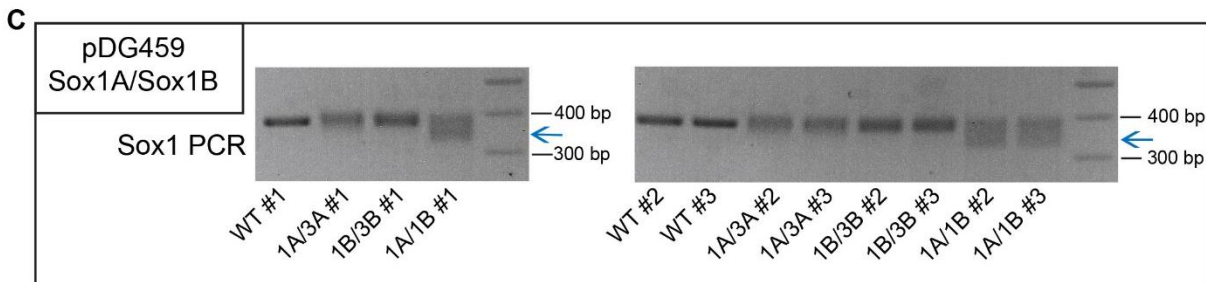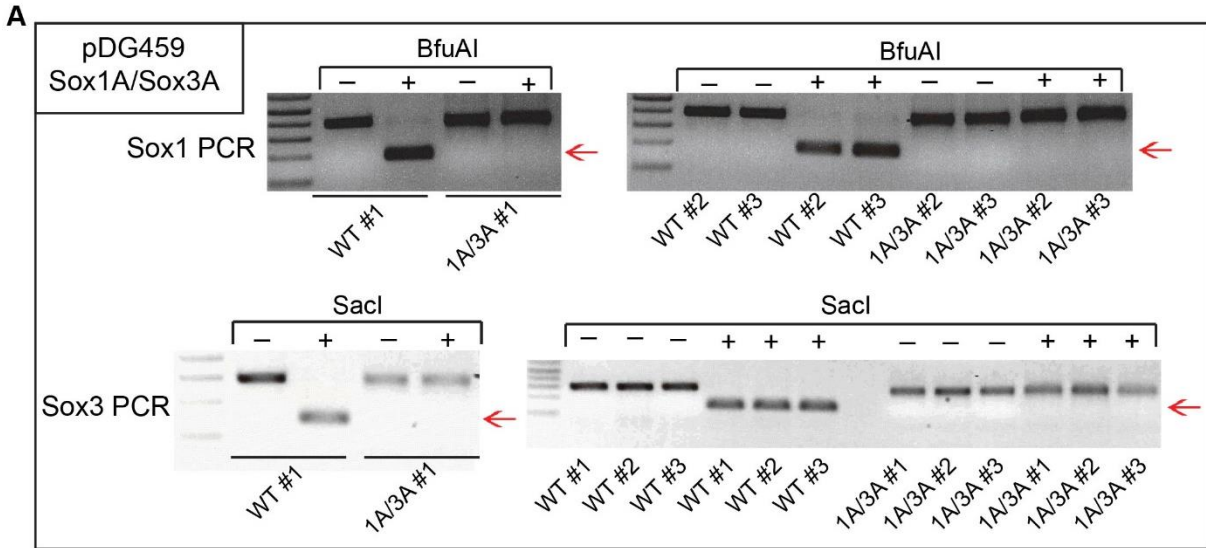
Supplementary Figure S1 | Efficient dual cutting mediated by pDG459 vector, extended figures of Fig. 2B with more independent samples. (A) BfuAI and SacI RFLP analyses indicated efficient dual cuts from pDG459 Sox1A/Sox3A. (B) ApaI and SfoI RFLP analyses indicated efficient dual cuts from pDG459 Sox1B/Sox3B. WT products after digestions (red arrows) were absent in pDG459-treated samples. (C) Large deletions were induced in the *Sox1* region in pDG459 Sox1A/Sox1B-treated samples. (D) Large deletions were induced in the *Sox3* region in pDG459 Sox3A/Sox3B-treated samples. Large deletion fragments are indicated with blue arrows. Each sample came from independent transfection (n ≥ 3).
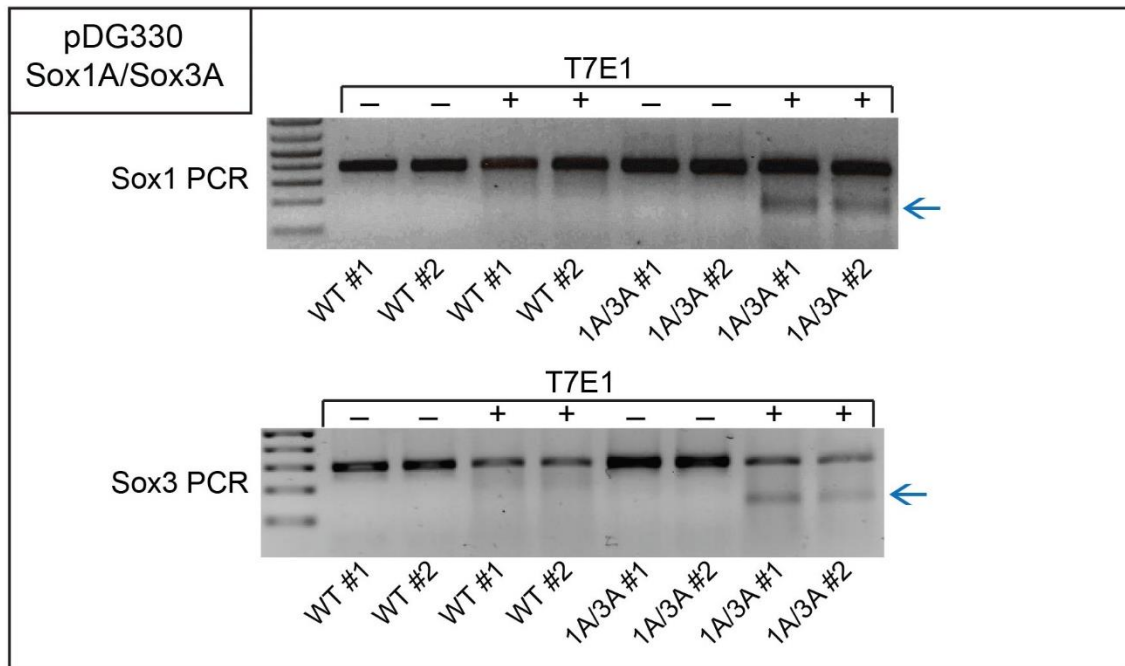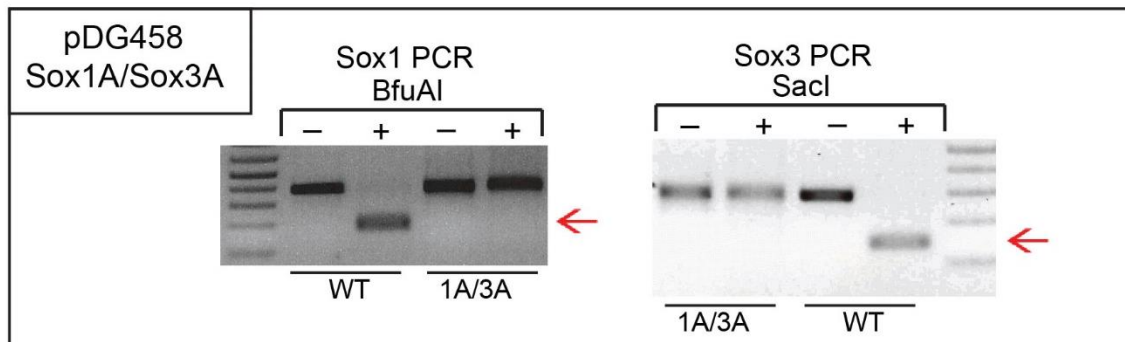
Supplementary Figure S2 | Mutation inductions mediated by vectors pDG330 and pDG458. (A) Transfection of pDG330 Sox1A/Sox3A into mouse ES cells induced mutations at both targets which were indicated by smaller fragments after T7E1 assay (blue arrows). (B) BfuAI and SacI RFLP were used to assess the mutation induction in Sox1A and Sox3A sites, respectively, after treatment of pDG458 Sox1A/Sox3A followed by GFP FACS enrichment. Presence of WT products produced smaller bands after restriction digestions (red arrows) which were absent in pDG458 Sox1A/Sox3A-treated samples. Each sample came from independent transfection.

A

B
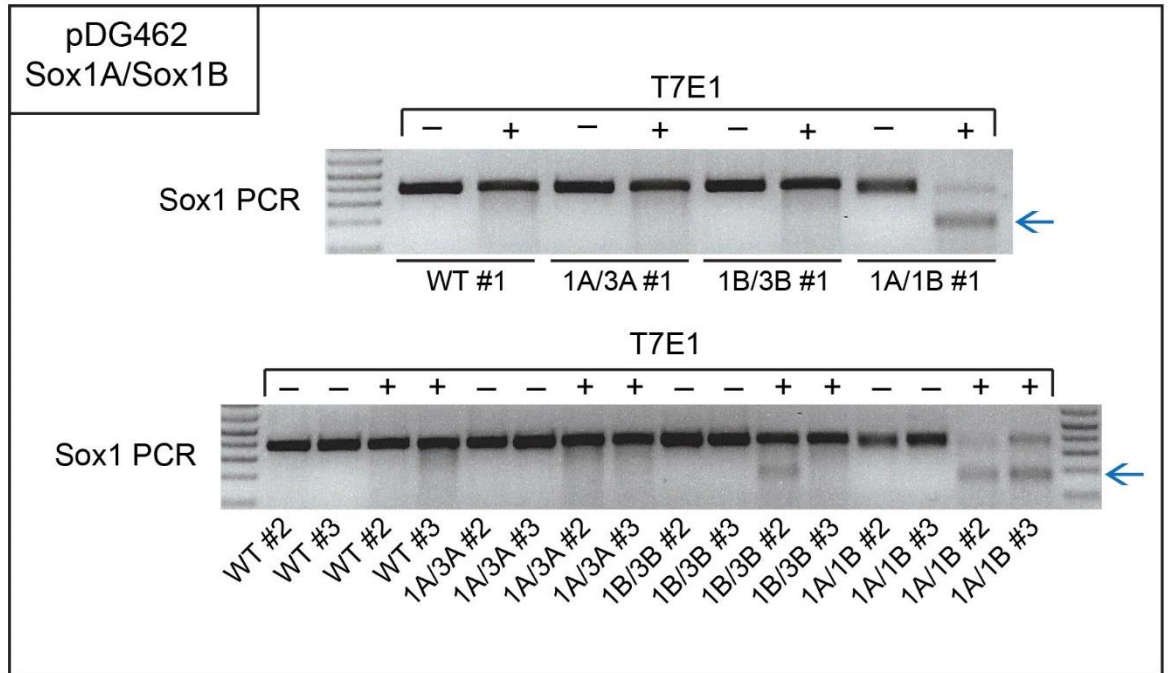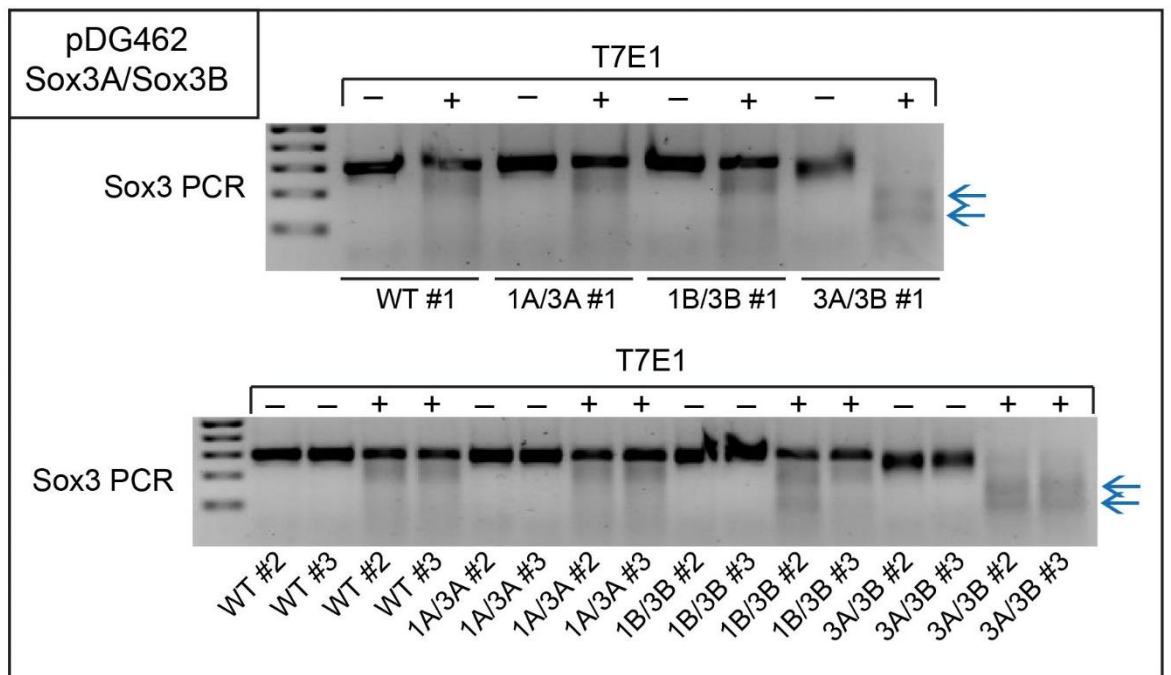
**Supplementary Figure S3 | Paired-nickase DSB induction by pDG462, extended
figures of Fig. 3 with more independent samples**. Smaller bands produced after T7E1
digestion (blue arrows) indicated presence of mutation in samples treated with paired-
nickase pDG462 Sox1A/Sox1B (A) or Sox3A/Sox3B (B). Each sample came from
independent transfections.

A

B
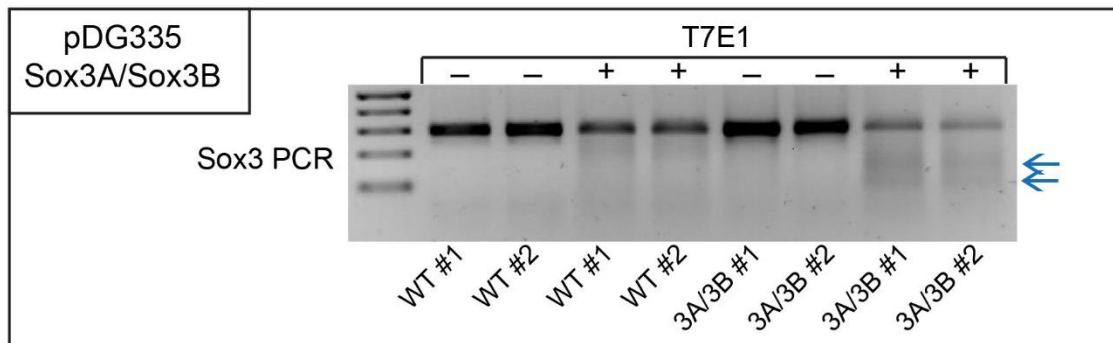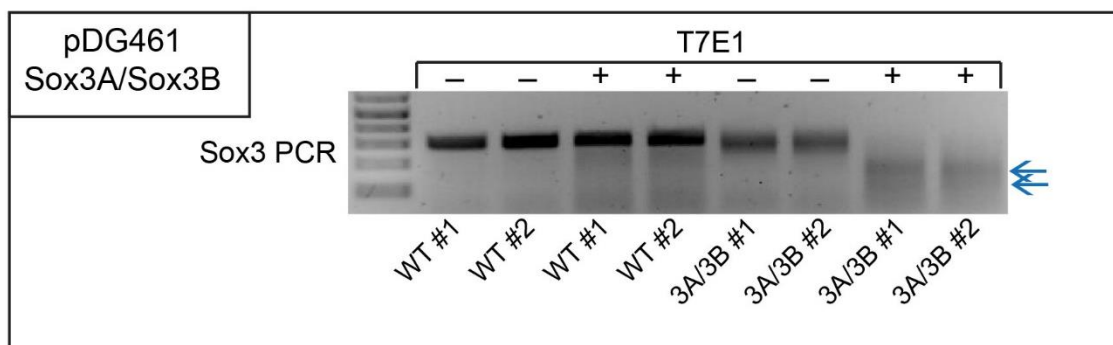
Supplementary Figure S4 | Paired-nickase-mediated mutation inductions by pDG335 and pDG461 vectors. T7E1 assay showed that expression of paired-nickase gRNAs Sox3A/Sox3B from pDG335 (A) or pDG461 (B) induced mutations in the *Sox3* locus as indicated by the presence of cut products (blue arrows). Each sample came from independent transfections.

Supplementary Table S1 | List of oligos used to generate the dual-gRNA targeting plasmids

Target	Oligo pair 1 (5'-3')	Oligo pair 2 (5'-3')
Sox1A/Sox3A	CACCGCCGCCGGGCGAGTGCAGGT	ACCGCTGACCCACATCTGAGCTCGT
	AAACACCTGCACTCGCCCGGCGGC	TAAAACGAGCTCAGATGTGGGTCAG
Sox1B/Sox3B	CACCGCCCACGAACCTCTCGGGCC	ACCGACCGCAGTCCCGGCGCCCGT
	AAACGGCCCGAGAGGTTCGTGGGC	TAAAACGGGCGCCGGGACTGCGGT
Sox1A/Sox1B	CACCGCCGCCGGGCGAGTGCAGGT	ACCGCCCACGAACCTCTCGGGCCGT
	AAACACCTGCACTCGCCCGGCGGC	TAAAACGGCCCGAGAGGTTCGTGGG
Sox3A/Sox3B	CACCGCTGACCCACATCTGAGCTC	ACCGACCGCAGTCCCGGCGCCCGT
	AAACGAGCTCAGATGTGGGTCAGC	TAAAACGGGCGCCGGGACTGCGGT

Grey highlights indicate the sequence of the guides.

References

Adikusuma, F., Pederick, D., McAninch, D., Hughes, J., and Thomas, P. (2017a). Functional Equivalence of the SOX2 and SOX3 Transcription Factors in the Developing Mouse Brain and Testes. Genetics *206*, 1495-1503.

Adikusuma, F., Williams, N., Grutzner, F., Hughes, J., and Thomas, P. (2017b). Targeted Deletion of an Entire Chromosome Using CRISPR/Cas9. Mol Ther.

Barrangou, R., and Doudna, J.A. (2016). Applications of CRISPR technologies in research and beyond. Nat Biotechnol *34*, 933-941.

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res *24*, 132-141.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A.*, et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.J., Kii, E., and Alt, F.W. (2015). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nat Biotechnol *33*, 179-186.

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. Nat Biotechnol *31*, 822-826.

Han, K., Jeng, E.E., Hess, G.T., Morgens, D.W., Li, A., and Bassik, M.C. (2017). Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. Nat Biotechnol *35*, 463-474.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262-1278.

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. Nat Biotechnol *31*, 827-832.

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.

Kabadi, A.M., Ousterout, D.G., Hilton, I.B., and Gersbach, C.A. (2014). Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. Nucleic Acids Res *42*, e147.

Komor, A.C., Badran, A.H., and Liu, D.R. (2017). CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell *168*, 20-36.

Long, C., Amoasii, L., Mireault, A.A., McAnally, J.R., Li, H., Sanchez-Ortiz, E., Bhattacharyya, S., Shelton, J.M., Bassel-Duby, R., and Olson, E.N. (2016). Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. Science *351*, 400-403.

Maddalo, D., Manchado, E., Concepcion, C.P., Bonetti, C., Vidigal, J.A., Han, Y.C., Ogrodowski, P., Crippa, A., Rekhtman, N., de Stanchina, E.*, et al.* (2014). In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. Nature *516*, 423-427.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013a). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol *31*, 833-838.

Mali, P., Esvelt, K.M., and Church, G.M. (2013b). Cas9 as a versatile tool for engineering biology. Nat Methods *10*, 957-963.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013c). RNA-guided human genome engineering via Cas9. Science *339*, 823-826.

Mashiko, D., Fujihara, Y., Satouh, Y., Miyata, H., Isotani, A., and Ikawa, M. (2013). Generation of mutant mice by pronuclear injection of circular plasmid expressing Cas9 and single guided RNA. Sci Rep *3*, 3355.

Nelson, C.E., Hakim, C.H., Ousterout, D.G., Thakore, P.I., Moreb, E.A., Castellanos Rivera, R.M., Madhavan, S., Pan, X., Ran, F.A., Yan, W.X.*, et al.* (2016). In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. Science *351*, 403-407.

Quadros, R.M., Miura, H., Harms, D.W., Akatsuka, H., Sato, T., Aida, T., Redder, R., Richardson, G.P., Inagaki, Y., Sakai, D.*, et al.* (2017). Easi-CRISPR: a robust method for one-step generation of mice carrying conditional and insertion alleles using long ssDNA donors and CRISPR ribonucleoproteins. Genome Biol *18*, 92.

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.* (2013a). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell *154*, 1380-1389.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013b). Genome engineering using the CRISPR-Cas9 system. Nat Protoc *8*, 2281-2308.

Sakuma, T., Nishikawa, A., Kume, S., Chayama, K., and Yamamoto, T. (2014). Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. Sci Rep *4*, 5400.

Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol *32*, 347-355.

Tabebordbar, M., Zhu, K., Cheng, J.K., Chew, W.L., Widrick, J.J., Yan, W.X., Maesner, C., Wu, E.Y., Xiao, R., Ran, F.A.*, et al.* (2016). In vivo gene editing in dystrophic mouse muscle and muscle stem cells. Science *351*, 407-411.

Tsai, S.Q., Wyvekens, N., Khayter, C., Foden, J.A., Thapar, V., Reyon, D., Goodwin, M.J., Aryee, M.J., and Joung, J.K. (2014). Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nat Biotechnol *32*, 569-576.

Vidigal, J.A., and Ventura, A. (2015). Rapid and efficient one-step generation of paired gRNA CRISPR-Cas9 libraries. Nat Commun *6*, 8083.

Chapter 5:

CRISPR/Cas9-induced breaks frequently generate random large deletions via DNA resection

5.1. Summary

Given that Thomas laboratory routinely uses CRISPR/Cas9 to generate mutant mouse model by zygote injection, comprehensive assessment of the mutation efficiency needs to be established. In this study, six different biallelic loci were assessed and screened for mutations using heteroduplex assays and Sanger sequencing, both of which involved 300–600 bp PCR amplification. Heteroduplex screening of 137 samples indicated that mutations were present in only 61% of samples. Surprisingly, more comprehensive screening, including Sanger sequencing, revealed that 97% of samples contained mutations. The high false negative rate in the heteroduplex assay was caused, unexpectedly, by large deletion mutations in one allele, which prevented PCR amplification and therefore heteroduplex formation. By performing larger PCR (~1.6-3.2 kb) amplifications of all samples, it was observed that 41% samples had large deletions of >100 bp. This is unusual, as repair of a single DSB generated by Cas9 is thought to use non-homologous end joining mechanism resulting in small indel mutations. The formation of these large deletions induced by CRISPR/Cas9 DSB was further studied in mouse ES cells. With the five gRNAs tested, frequent large deletion mutations were consistently observed, as measured by qPCR analyses. Whole genome sequencing (WGS), which examines mutation outcomes in an unbiased manner, was also conducted to confirm these frequent large deletions. Surprisingly, WGS also detected frequent insertions of the CRISPR/Cas9 expression plasmid into the break sites.

The large deletions found in mouse zygote injection samples were sequenced to characterise these deletions. Interestingly, the deletions were asymmetric relative to the break site and the break junctions contained predominantly 1–3 bp microhomology sequences, and in some cases small insertions or no microhomologies. Using a novel assay called SSA trapping, these large deletions were found to require the occurrence of DNA

end resections. These characteristics suggested that alternative end joining (Alt-EJ) was the repair mechanism underlying the formation of these large deletions.

A manuscript describing this study will be submitted for a publication to inform the scientific community that a single DSB can result in large deletions or plasmid insertion. This is crucial as failure to notice these frequent repair outcomes could result in misjudging the true genotype, leading to misinterpretation of the data.

Statement of Authorship

Title of Paper	CRISPR/Cas9-induced breaks frequently generate random large deletions via DNA resection
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Fatwa Adikusuma, Sandra Piltz, Mark Corbett, Michelle Turvey, Shaun McColl, Karla Helbig, Michael Beard, James Hughes & Paul Thomas. *CRISPR/Cas9-induced breaks frequently generate random large deletions via DNA resection.* 2017

Principal Author

Name of Principal Author (Candidate)	Fatwa Adikusuma
Contribution to the Paper	Designed and conceived the study, conducted most of the experiments, analysed data, generated the Figures and wrote the manuscript.
Overall percentage (%)	90%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature		Date	8/8/2017

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

  i. the candidate's stated contribution to the publication is accurate (as detailed above);

  ii. permission is granted for the candidate in include the publication in the thesis; and

  iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Sandra (Sandie) Piltz		
Contribution to the Paper	Performed all microinjections		
Signature		Date	

Name of Co-Author	Mark Corbett		
Contribution to the Paper	Analysed whole genome sequencing data		
Signature		Date	20/07/2017

Name of Co-Author	Michelle Turvey		
Contribution to the Paper	Performed PA gel heteroduplex assay and sequencing samples from Pik3r6 microinjection		
Signature		Date	20/07/2017

Name of Co-Author	Shaun McColl		
Contribution to the Paper	Contributed to PA gel heteroduplex assay and sequencing samples from Pik3r6 microinjection		
Signature		Date	20/07/2017

Name of Co-Author	Karla Helbig		
Contribution to the Paper	Performed PA gel heteroduplex assay and sequencing samples from Viperin microinjection		
Signature		Date	21/07/2017

Name of Co-Author	Michael Beard		
Contribution to the Paper	Contributed to PA gel heteroduplex assay and sequencing samples from Viperin microinjection		
Signature		Date	25/0717

Name of Co-Author	James Hughes		
Contribution to the Paper	Performed PA gel heteroduplex assay and sequencing samples from Hmgcs2 microinjection		
Signature		Date	25/07/2017

Name of Co-Author	Paul Thomas		
Contribution to the Paper	Designed, conceived, supervised the study, and wrote the manuscript.		
Signature		Date	8/8/2017

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

   i.   the candidate's stated contribution to the publication is accurate (as detailed above);
   ii.   permission is granted for the candidate in include the publication in the thesis; and
   iii.   the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Sandra (Sandie) Piltz		
Contribution to the Paper	Performed all microinjections		
Signature		Date	20/07/17

Name of Co-Author	Mark Corbett		
Contribution to the Paper	Analysed whole genome sequencing data		
Signature		Date	

Name of Co-Author	Michelle Turvey		
Contribution to the Paper	Performed PA gel heteroduplex assay and sequencing samples from Pik3r6 microinjection		
Signature		Date	

Name of Co-Author	Shaun McColl		
Contribution to the Paper	Contributed to PA gel heteroduplex assay and sequencing samples from Pik3r6 microinjection		
Signature		Date	

Name of Co-Author	Karla Helbig		
Contribution to the Paper	Performed PA gel heteroduplex assay and sequencing samples from Viperin microinjection		
Signature		Date	

CRISPR/Cas9-induced breaks frequently generate random large deletions via DNA resection

Fatwa Adikusuma [1,2,5]**, Sandra Piltz** [1,2,3]**, Mark A. Corbett** [2]**, Michelle Turvey** [1]**, Shaun McColl** [1]**, Karla Helbig** [4]**, Michael Beard** [1]**, James Hughes** [1] **& Paul Thomas** [1,2,3,*]

[1] School of Biological Sciences, University of Adelaide, South Australia, Australia, 5005

[2] Robinson Research Institute, The University of Adelaide, South Australia, Australia, 5005

[3] South Australian Health and Medical Research Institute, South Australia, Australia, 5000

[4] School of Life Sciences, La Trobe University, Melbourne, Victoria, Australia, 3086

[5] Centre for Biomedical Research (CEBIOR), Faculty of Medicine, Diponegoro University, Semarang, Indonesia, 50271

* Corresponding author: paul.thomas@adelaide.edu.au

Abstract

CRISPR/Cas9 genome editing has emerged as a powerful technology for generating targeted genomic modifications in a vast array of cellular contexts including cultured cell lines, zygotes and somatic cells. DNA double-stranded breaks (DSB) generated by CRISPR/Cas9 are commonly repaired by non-homologous end-joining (NHEJ) mechanism resulting in small indels. Unexpectedly, through analysis of 137 mice generated from CRISPR/Cas9 zygote injections, we identified large deletions (>100 bp) of up to 2.3 kb in 56 (41%) of founders. Frequent large deletions were also generated in ES cells in response to CRISPR/Cas9-induced single-site DSBs. Unbiased whole genome sequencing (WGS) in pooled ES cells indicated that large deletions occurred in 37.5% of alleles. Using a novel single-strand annealing (SSA) capture assay, we also show that large deletions are generated via DNA resection and that their breakpoint junctions contain microhomologies or insertions, indicating an alternative end-joining (Alt-EJ) mechanism. This study unveils underestimated yet common repair outcomes of endonuclease-mediated cleavage. Researchers should consider the possibility of large deletions to avoid misinterpretation of PCR-based genotyping assays.

Introduction

The CRIPSR/Cas9 genome editing platform has been developed from an adaptive immunity system in bacteria and archaea and comprises two components; Cas9, which functions as a programmable endonuclease that generates a blunt-ended double-stranded break (DSB) and a ~100 nt guide RNA (gRNA). The latter contains a ~20 nt guide sequence at the 5' end that directs Cas9 to the target site via RNA:DNA complementary base pairing and an invariant ~80 nt sequence required for stabilization of the gRNA/Cas9 complex. While modification of the guide sequence provides considerable flexibility in target site selection, there is also a strict requirement for a protospacer adjacent motif (PAM) sequence (NGG for SpCas9) adjacent to the DNA target sequence to which Cas9 binds. CRISPR/Cas9 technology has emerged as the system of choice for many researchers to perform genome editing in many organisms due to its high efficiency, flexibility, ease of use and low cost (Cong et al., 2013; Hsu et al., 2014; Jinek et al., 2012; Komor et al., 2017; Mali et al., 2013; Sander and Joung, 2014).

Genomic modification by CRISPR/Cas9 is achieved by relying on the cellular DNA repair mechanism to fix the breaks induced by the Cas9. The mechanism that is thought to predominate the repair of these DSBs is non-homologous end-joining (NHEJ) which typically generates small deletions (~1-20 bp) and in some cases short insertions ($\leq$5 nt) at the cut site. Alternatively, specific targeted modifications can be introduced at the breaks through an accurate DNA repair mechanism called homologous recombination (HR) or homology directed repair (HDR) which requires addition of a ssDNA or dsDNA donor repair template containing homology arms flanking the mutation sequence to be copied (Cong et al., 2013; Hsu et al., 2014; Komor et al., 2017; Mali et al., 2013; Sander and Joung, 2014).

In this study, we show that DNA repair after Cas9-induced DSBs does not exclusively result in small insertions and deletions (indels) as commonly perceived. Instead, our analyses reveal that large deletions (>100 bp) also appear to be a common outcome of CRISPR/Cas9 single-site cleavage in zygotes and pluripotent cells. These large deletions are generated via DNA end resection and often contain microhomology sequence tracts flanking their breakpoints suggesting an alternative end-joining (Alt-EJ) mechanism, also referred to as microhomology-mediated end-joining (MMEJ), which is promoted by DNA polymerase θ (Polθ) (Kent et al., 2015). Together, these findings expand our knowledge of the repair outcomes that occur in response to CRISPR/Cas9 cleavage and provide a useful caution to researchers to consider large deletions when genotyping mutant cells or founder animals.

Results

High frequency of large deletions generated by CRISPR-Cas9 cleavage in zygotes

To gain a more comprehensive understanding of the efficiency of CRISPR mutagenesis and modalities of DNA repair, we screened founder embryos and mice derived from CRISPR/Cas9-injected zygotes for mutant alleles. Six loci were independently targeted for indel mutations using a single gRNA. Three of the gRNAs targeted coding regions of autosomal genes which do not cause nullizygous lethality (*Viperin, Pik3r6 and Hmgcs2*) and three targeted intronic/flanking regions in *Ngn3, Fzd3* and *Foxp4* (Figure 1A). 137 samples (embryos and mice) were initially screened using a polyacrylamide (PA) gel heteroduplex assay of a 0.3-0.6 kb PCR product spanning the gRNA target site (Chen et al., 2012). In our experience, the PA heteroduplex assay provides similar reliability to the commonly used T7E1 heteroduplex assay but is easier to perform and less expensive. PA gel screening detected mutations in 83/137 (61%) of samples (Figure 1B, S1 and Table

S1). These included samples with smaller PCR products than expected, suggesting that they contain deletions in the order of 100-300 bp. Notably, four samples failed to amplify (Figure S1, sample *Foxp4* #1, #2, #9 and *Fzd3* #29), indicating that they may have contained biallelic deletions that encompassed the PCR primer binding sites. To further characterize the mutations in each sample, we sequenced the PCR products including the "mutation-negative" samples. As expected, small indels were detected in all PA gel heteroduplex-positive samples. 87% of the sequenced samples contained at least two mutant alleles, while heterozygous (i.e. "single hit)" mutations were only observed in the remaining 13% (Table S1). Surprisingly, 88% of heteroduplex-negative samples that we sequenced contained a small indel as opposed to the WT sequence (Table S1). Only one type of mutation was observed in these samples which explained why they were not detected by heteroduplex assay (false negatives). While it is possible that both alleles carry the same mutation, this is unlikely given the range of indels that were detected in each group of founders.

An alternative possibility is that one allele has a small indel while another has a large deletion that eliminates one or both primer binding sites thereby preventing PCR. To investigate this possibility, we performed ~1.6 kb PCR using primers equidistant from gRNA-PAM sequences. Surprisingly, we found that 34% of samples generated amplicons that were smaller than expected indicating alleles with large (>100 bp) deletions (Figure 1C, Figure S2A and Table S2). Additional large deletion products were generated when we performed a 3.2 kb PCR on *Ngn3* and *Foxp4* samples (Figure S2B and Table S2). Large deletions were detected in heteroduplex-negative samples as well as heteroduplex-positive samples, indicating mosaicism alleles in the latter samples. We could also generate products from samples that failed to amplify in the initial 0.3-0.6 kb PCR, confirming that these contained large deletion alleles (Figure S1 and S2, sample *Foxp4* #1, #2, #9 and *Fzd3* #29). Deleted sequences were confirmed by direct sequencing of gel-purified PCR

products. Up to this point, we could identify large deletion alleles in 41% of samples with some founders harboring more than one large deletion event (Table S2). The size of large deletions (after ~1.6 kb PCR) ranged from 100-800 bp (Figure 3 and Figure S2). The largest deletion found in this experiment was 2.3 kb that we detected in a *Foxp4* sample after 3.2 kb PCR (Figure S2B, sample *Foxp4* #19). Combining these large deletion data with the heteroduplex assay and Sanger sequencing analyses, we found that 133 of 137 (97%) of founders contained at least one mutation, a very high efficiency compared to the other published studies (Shen et al., 2013; Wang et al., 2013; Yang et al., 2013; Zhang et al., 2016).

High frequency of large deletions generated by CRISPR-Cas9 cleavage in mouse ES cells

To further study the frequency and large deletions, we performed additional experiments using mouse embryonic stem cells (mESC). PX459.V2.0 expression constructs encoding a single gRNA and Cas9 were transiently transfected and puromycin-resistant transfectants were harvested for genomic DNA (gDNA) extraction. Single gRNAs targeting autosomal (*Ngn3*, *Sox1* and *Viperin*) and single copy X-linked (*Sox3*) loci were selected for analysis. RFLP-based genotyping (for *Sox1, Viperin* and *Sox3*) and TIDE analysis (for *Ngn3*) indicated close to 100% cutting efficiency in harvested cell pools (Figure S3A-D). This high cutting efficiency permitted accurate quantification of large deletion events by qPCR of pooled genomic DNA using primers that bind >100 bp from the cutting sites. For each gRNA, two qPCRs were performed to cover large deletions occurred at both upstream and downstream regions of the cutting sites (Figure 2A). We reasoned that deletion of primer binding sites due to a large deletion would result in reduction of qPCR signal. Interestingly, qPCR analyses for all four gRNAs showed significant reduction compared to

a negative control gRNA targeting the bacterial neomycin resistance gene (*Neo* gRNA). The reduction in qPCR signal ranged from 14-30% with an average of 24% and 23% reduction in the upstream and downstream qPCR, respectively (Figure 2A). Given that unidirectional large deletions would reduce qPCR signal for only one of the flanking qPCR reactions, whereas bidirectional deletions would reduce signal for both qPCR reactions, we conclude that the frequency of large deletion events is between ~23% (all bidirectional) and ~47% (all unidirectional). Notably, these large deletions were also generated when DSBs were induced with 5' overhang sticky-ended cuts via double Cas9-nickase (Cas9n) as observed by qPCR analyses (Figure S4).

As detection of large deletions using PCR is prone to amplification bias and is confounded by deletion of primer sequences, we next sought to use an unbiased approach to determine the frequency and extent of CRISPR/Cas9-induced large deletions in ESC. We performed PCR-free paired-end whole genome sequencing (WGS) analysis on gDNA from *Viperin* gRNA-transfected ESCs. From a total of 88 sequence reads, only two reads (2.3%) corresponded to WT alleles, consistent with the high mutagenesis efficiency as shown by RFLP analysis (Fig. S3D). Small indels and substitutions were found in 33 reads (37.5%), while large deletions inferred from discordant mapping of paired-end reads and split read mapping over the breakpoint were also detected in 33 reads (37.5%) (Figure 2B). The remaining 20 (23%) unexpectedly contained insertions of the PX459.V2.0 expression plasmid. Taken together, these data confirm that large deletions are frequently generated after CRISPR/Cas9-mediated DNA cleavage in ESC and mouse zygotes.

Random large deletions contain microhomology and are initiated by DNA resection

To investigate the mechanism of large deletion generation, we sequenced 59 of the large deletion alleles detected by ~1.6 kb PCR from F0 mice/embryos generated via the *Ngn3*,

FoxP4, *Fzd3*, *Viperin*, *Pik3r6* and *Hmgcs2* gRNAs described (Figure S2A), including the large deletions generated by two other gRNAs targeting *Kcnt1* and the *Sox3*. Each gRNA generated a range of unique deletions indicating that the process underpinning generation of large deletions is stochastic (Figure 3A). Indeed, we only found two examples of independent founders carrying the same deletion. Assessment of the breakpoint sequences revealed that the vast majority of junctions contained microhomology (MH) of 1-3 bp (63%) or longer (22%). Junction sequences with short insertions (predominantly 1-2 bp insertions) or no homology were also present in low abundance (Figure 3A and 3B). We observed, interestingly, that the orientation of the large deletions was noticeably asymmetric/unidirectional with respect to the cutting site (Figure 3A). Indeed, only 4 of 59 random large deletions were obviously bidirectional (>100 bp to both directions) (Figure 3A).

Together, these characteristics suggested Alt-EJ as the repair process underlying the formation of large deletions. Alt-EJ is characterized by the presence of MH or insertion at the repair junction. It usually results in relatively large deletions and is thought to be promoted by DNA end resection (Black et al., 2016; Chang et al., 2017; Deriano and Roth, 2013). To determine whether the generation of large deletion requires end resection, we developed a single-strand annealing (SSA) trapping assay at an endogenous locus in mouse ES cells. In this assay, DNA end resection following a single DSB traps the cell into using SSA repair mechanism due to the presence of flanking repeat sequences (Bhargava et al., 2016). We used the *Nprl2*-SSA gRNA to generate a DSB within a 9 bp spacer sequence located between two 34 bp perfect repeats in mouse *Nprl2* gene. We reasoned that DNA resections of >43 nt would be trapped into a SSA repair event generating a 43 (34 + 9) bp deletion (Figure 3C). Indeed, PCR analysis of *Nprl2*-SSA gRNA expressing cells indicated a high frequency of the 43 bp SSA-deletion repair event which was confirmed by Sanger sequencing and TIDE analysis (Figure 3D and S5A-B). As the SSA negative control,

another highly efficient DSB was generated outside of the repeat sequences using gRNA *Nprl2*-out (Figure 3C and S5C), which did not generate the 43 bp SSA-deletion repair event (Figure 3D). As expected, DNA breaks from gRNA *Nprl2*-out produced frequent large deletions as determined by qPCR (Figure 3E). In contrast, DSB induced by gRNA *Nprl2*-SSA failed to generate large deletions, indicating that trapping the DNA resection intermediate into SSA repair prevents the generation of the large deletions (Figure 3E). Taken together, these data suggest that the Alt-EJ mechanism underpins the generation of large deletions.

Discussion

From this study, we again witness the remarkable efficiency of the CRISPR/Cas9 system in generating DSB-induced mutations in mammalian cells. WT samples/alleles were rarely detected after mouse zygote injections or ESC transfections, as has also been observed by others (Bell et al., 2014; Wang et al., 2013). The high mutagenic efficiency of all gRNAs used in this study contrasts with their average on-target score (Table S3 and S4), arguing against the necessity for selecting gRNAs based on their on-target efficiency score (Doench et al., 2016). Rather, this study underlines the importance of microinjectionist skill in delivering high quality CRISPR reagents into zygotes and the development of effective transfection strategies for efficient mutagenesis in mice and cells, respectively.

Here, we also learnt about the significant disadvantage of relying exclusively upon PCR heteroduplex formation assays for mutation screening of founder animals. The high frequency of biallelic mutations that included a large deletion event resulted in a very high false negative rate. Mixing and reannealing of founder and WT PCR products would circumvent this problem and should be considered as the primary screening assay for founder animals. RFLP analysis, whereby the WT allele is cleaved at or near the gRNA

cut site, is also a useful screening tool, provided that a suitable restriction site is available. Direct sequencing of the PCR product is probably the most definitive method, although this is significantly more expensive than other options.

A key finding of this study is that DSB generated by CRISPR/Cas9 endonuclease activity are frequently repaired via a mechanism that generates large deletions flanking the cleavage site. This finding contrasts with the common perception that CRISPR/Cas9-mediated breaks are almost exclusively repaired by NHEJ resulting in small indels. Deletion of PCR primer binding sites is probably the most significant factor contributing to the underappreciation of large deletion frequency. We find that large deletions are unexpectedly common, occurring in 41% of founder mice across 6 loci. This frequency is likely an underestimate as further separation of the genotyping PCR primers (beyond 3.2 kb) would likely reveal additional large deletions. Mouse ES cells also exhibit a high frequency of large deletions, as determined by qPCR (23 - 47%) and unbiased WGS analysis (37.5%.), demonstrating that the mechanism that leads to large deletions is not restricted to totipotent cells. Indeed, recent reports have also identified frequent large deletions in mouse zygotes (Shin et al., 2017; Zhang et al., 2016), as well as mouse embryonic brains (Zuckermann et al., 2015) and mouse NIH-3T3 (transformed fibroblast) cells (Mou et al., 2017). Moreover, large deletion repair outcomes are also found in DSB induced by other nucleases such as TALENs and IsceI, suggesting they are general feature of DSB repair (Cheng et al., 2016; Honma et al., 2003). Together these studies suggest that large deletions will be generated in many cell contexts and researchers should be alert to the possibility that their occurrence may lead to inaccurate genotyping, especially in cell lines (as opposed to mice) where it is not possible to segregate mutant alleles via breeding.

How are large deletions generated in zygotes and ESC? One well-known DNA repair mechanism that generates large deletions is single-strand annealing (SSA) (Bhargava et al.,

2016). This mechanism can be used to repair DSBs that occur between long repeat sequences. 5'-3' resection of the DSB allows the complementary single-stranded repeat sequences to bind each other creating a deletion of the intervening region. Thus, SSA repair products contain a long homology sequences in the breakpoint junctions and the deletion size depends on the distance between the flanking repeats (Bhargava et al., 2016). In contrast, the large deletions observed in this study were random in size and position, and the junctions contain only short tracts of microhomology (mostly 2-3 bp) or small insertions, which are typical signatures of Alt-EJ. Through development of an endogenous SSA trapping assay, we demonstrated that DNA repair initiated by end resection is frequent, and that these large deletions also require DNA end resection to form. Together, these data suggest that the repair mechanism underlying the formation of large deletions is a poorly defined Alt-EJ repair mechanism (Bhargava et al., 2016; Black et al., 2016; Chang et al., 2017; Deriano and Roth, 2013). The high frequency of large deletions supports the notion that Alt-EJ is not exclusively a backup pathway, but it is also active when NHEJ repair process is available (Sfeir and Symington, 2015). The high frequency of apparent Alt-EJ may be influenced by the absence of HR pathway in our experiments. It might be interesting to test whether large deletion frequency is altered when HR is available since Alt-EJ and HR repair both share the initial end resection process and are active at the same phases of the cell cycle, S and G2 (Bhargava et al., 2016; Deriano and Roth, 2013). In Alt-EJ, repair of DSBs is initiated with 5'-3' DNA end resection by the MRN complex together with CtIP resulting in 3' ssDNA overhangs (Chang et al., 2017; Sfeir and Symington, 2015). The key factor in Alt-EJ, Polθ, promotes transient annealing between 3' ssDNA overhangs via the use of microhomology sequences (Kent et al., 2015). Any excessive remaining 3' overhangs at the DNA synapse are likely cleaved by an unknown nuclease before the next steps take place. The minimally paired overhangs are extended by Polθ activity, allowing for stabilization of the DNA synapse and gap filling. Once the gaps

are filled, the DNA end-joining intermediate is ligated by Lig1 or Lig3 (Black et al., 2016; Deriano and Roth, 2013; Simsek et al., 2011). One study estimated that Alt-EJ was 270 times more common than HR repair of DNA DSB in human cells (Honma et al., 2003). It is therefore possible that the low frequency of intended HDR that is commonly observed in mice and cells (Adikusuma et al., 2017a) is due to successful competition by Alt-EJ (Zelensky et al., 2017). Therefore, attempts to increase HDR frequency should consider repressing this repair pathway (Zelensky et al., 2017) in addition to repressing NHEJ (Chu et al., 2015; Maruyama et al., 2015; Robert et al., 2015; Yu et al., 2015).

It is intriguing to consider how the 3' overhang sequences generated by DNA resection select target sequences for microhomology base pairing given the abundant options that are available. We found that 2-3 nt microhomologies predominated at the breakpoint junctions of large deletions, although 1 nt microhomology was also observed. While longer microhomologies ≥2 bp generate more stable transient DNA synapses that are rapidly accommodated by Polθ to intiate Alt-EJ (Kent et al., 2015), the infrequent availability of these microhomologies at the right time and position might account for their relative paucity in large deletion sequence junctions. Besides promoting microhomology-mediated end-joining between DNA overhangs, Polθ also possesses terminal transferase activity that can extend the overhangs during the end-joining process (Kent et al., 2016). Terminal extension before microhomology annealing and DNA replication will create novel sequences that result in insertions at DNA repair junctions. This terminal transferase activity is likely responsible for the occasional small insertions we detected at the large deletion junctions (Kent et al., 2016). In the case of 0 microhomology, it is possible that the Polθ -mediated extension generates chance microhomology, such annealing of the microextended sequence does not leave a trace of microhomology or insertion (Simsek et al., 2011).

One interesting characteristic of the large deletions is that they typically extend in only one direction from the cut site (i.e. they are unidirectional/asymmetric). Relatively asymmetric large deletions were also recently observed in *C. elegans* (at G4 break site) (Koole et al., 2014), in human cells (Isce1-induced DSB) (Honma et al., 2003) and in mice (Cas9 breaks) (Zuckermann et al., 2015). Studies in *C. elegans* indicated that Polθ was responsible for this unidirectional large deletion, again supporting Alt-EJ as the repair mechanism underlying the large deletions identified in our study (Koole et al., 2014; van Schendel et al., 2015). These unidirectional deletions suggest that annealing occurs between the terminal sequences of 3' overhang with the microhomology complementary sequence located a significant distance away from the break site. This is supported by in vitro analysis which showed efficient Polθ-mediated microhomology annealing between terminal and internal sequences (Kent et al., 2015; Wyatt et al., 2016). This activity requires trimming of DNA overhang flap on only one side during the process. The directionality of the deletion (upstream or downstream from the break) depends on which 3' overhang uses the terminal sequences and which 3' overhang use the internal sequences during the microhomology annealing step. When annealing occurs between terminal sequences of 3' overhang sense strand with microhomology internal sequences of 3' overhang antisense, a large deletion of the downstream sequence is generated, and vice versa. We hypothesize that there is competition between sense and antisense 3' overhang strands for annealing of their terminal sequences which we term the Strand Competition for Annealing after Resection (SCAR) effect (Figure 4A). Stochasticity of the proposed SCAR effect provides an explanation for the variation in orientation and extent of the large deletions that we identified for each gRNA.

We propose two possible mechanisms to explain the (relatively rare) bidirectional large deletions (Figure 4B-C). Microhomology annealing after resection could occur between both internal sequences of 3' overhangs (Figure 4B). In this case, excessive DNA flaps in

122

both sides need to be trimmed to allow subsequent steps of DNA extension and ligation. In vitro study showed this scenario was inefficient (Kent et al., 2015; Wyatt et al., 2016). The second possibility is that relatively longer resection can occur in 3'-5' direction (Dorsett et al., 2014). Microhomology-mediated annealing of terminal sequences of the 3'-5' resected 3' overhang with internal sequences of opposite 3' overhang would result in bidirectional large deletions. In this scenario, excessive flap trimming only occurs on one side (Figure 4C).

In addition to the observation of frequent large deletion, we also surprisingly observed frequent plasmid integration to the break site from WGS analysis. It is still unclear how circular DNA integrates to the DSB, but it likely involves HDR. Targeted integration of homology-independent linear DNA has been known to be quite efficient in combination with CRISPR/Cas9 breaks (Hisano et al., 2015; Suzuki et al., 2016). Interestingly, Polθ was recently shown to play a major role in facilitating off-target or random DNA integration in mice and human cells (Zelensky et al., 2017). We suspect the plasmids underwent linearization inside the cells before integrating to the break sites which likely explains the phenomenon of random stable integration of circular plasmids into the genome. This integration of homology independent plasmid to the break sites should be considered as a DNA repair outcome when conducting experiments using plasmid-based expression vectors. It will be interesting to know how the pattern of DNA repair outcomes in the absence of plasmids, whether repairs resulting in more large deletions, more small indels or equally distributed to both repair outcomes. Further study by delivery of Cas9 and gRNAs in RNA or protein form is needed to address this question.

In summary, our data revealed that in addition to small indels, DNA DSBs in particular Cas9-mediated breaks are frequently repaired resulting in large deletions and plasmid

insertions. We caution researchers to consider these repair outcomes when performing CRISPR/Cas genome editing experiments to avoid misinterpretation of the true genotype.

Experimental procedures

Generation of gRNA, Cas9 mRNA and expression plasmids

gRNAs (Table S3 and S4) were designed using CRISPR designing tool from Zhang lab MIT http://crispr.mit.edu. On-target scores were calculated using CRISPOR gRNA design tool (Haeussler et al., 2016) based on algorithm developed by Doench et al. (Doench et al., 2016). Most of the gRNAs for zygote injections were generated in house using the protocol described by Wang et al. (Wang et al., 2013). In brief, a pair of oligo containing target sequences was ligated to BbsI-linearized PX330 plasmid and was used as PCR template to produce T7-gRNA PCR products. T7-gRNA PCR products were PCR purified (Qiagen) and were then used for IVT template to produce the gRNAs using HiScribe T7 (NEB). gRNAs for *Pik3r6* and *Hmgcs2* were purchased from ToolGen. Cas9 mRNA was produced by IVT of XhoI-linearized Cas9 vector (ToolGen) using mMessage T7 ultra kit (Ambion). All RNAs were purified using Rneasy mini kit (NEB). For experiments in mouse ES cells, PX459.V.2.0 was used as gRNA and Cas9 expression plasmid as described by Ran et al. (Ran et al., 2013). For expression of Cas9-nickase (Cas9n) and dual gRNAs, plasmid PX462.V.2.0 was modified by cloning an extra gRNA expression cassette to NotI site of the plasmid as described by Adikusuma et al. (Adikusuma et al., 2017b). Thus, all components (Cas9n, gRNA 1, gRNA 2 and puromycin selection marker) were expressed from a single plasmid. Plasmid preparations were performed using PureLink® HiPure Plasmid Midiprep Kit (Life Technologies).

Mouse zygote injections

All animal procedures have been approved by The University of Adelaide Animal Ethic Committee. Cas9 mRNA (100 ng/µL) and gRNAs (50 ng/µL each) were injected into the cytoplasm of C57BL/6N zygotes using a FemtoJet microinjector, transferred to pseudo pregnant recipients and allowed to develop to term or harvested as embryos. Injections of *Sox3* gRNA was accompanied with single-stranded oligo donor.

DNA extraction, PCR, genomic qPCR, heteroduplex assay and RFLP

Genomic DNA was extracted from tail tissue or ES cells (1-2 million) using High Pure PCR Template Preparation Kit (Roche) according to the manufacturer's instructions. PCRs were mostly conducted using FailSafe™ PCR System (Epicentre). qPCRs were performed using Fast SYBR Green Master Mix (Applied Biosystems) on an Applied Biosystems 7500 StepOnePlus machine. Statistical analysis was performed using Student's two tailed unpaired t-tests in GraphPad Prism software. Heteroduplex assay or RFLP was conducted following 0.3-0.6 kb PCR. For heteroduplex assay, PCR products were heated 95° C for 3 minutes and cooled down slowly to room temperature (0.1° C/s) prior to gel electrophoresis on 12% non-denaturing polyacrylamide gel for 2 hours at 150 V. RFLP was performed by incubating 5 uL of PCR products (without purification) in a total volume of 20 µL digestion reaction containing the relevant buffers and restriction enzymes (NEB) at the suggested optimal temperatures for 1 hour. PCR and qPCR primers are listed in Table S5.

Sequencing and TIDE analysis

For sequencing, PCR products were purified using QIAquick PCR Purification Kit
(Qiagen) and directly sequence using the forward or reverse primer of each PCR. For
TIDE analysis, sequencing data was entered to TIDE online tool at https://tide.nki.nl/
(Brinkman et al., 2014). For large deletion sequencing, PCR products were separated on
1% agarose gel. Visually smaller bands from the gel were cut and gel purified using
QIAquick Gel Extraction Kit (Qiagen) before subjected for big dye sequencing reaction
using the forward or reverse primers of each PCR.

CRISPR/Cas9 plasmid transfection in mESc

Cell culture and transfection followed the procedure described by Adikusuma et al.
(Adikusuma et al., 2017b). R1 mouse embryonic stem cells were cultured in 15%
FCS/DMEM supplemented with 2 mM Glutamax (Gibco), 100 µM non-essential amino
acid (Gibco), 100 µM 2-mercaptoethanol (Sigma), 3 µM CHIR99021 (Sigma) 1 µM
PD0325901 (Sigma) and LIF (generated in-house). One million of ES cells were
nucleofected with 3 µg of plasmid DNA using the Neon™ Transfection System 100 µL
Kit (Life technologies) at 1400 V, 10 ms and 3 pulses according to the manufacturer's
protocol. 24 hr after transfection, selection was conducted by adding puromycin (2 µg/mL)
to the media for the next 48 hr. Surviving cells were cultured for 4-7 days without selection
before harvesting.

Whole genome sequencing

Whole genome sequencing was performed using Illumina HiSeq X Ten system which was
conducted by Garvan Institute via Australian Genome Research Facility (AGRF) according

to their instructions. Sequence data were mapped to the GRCm38_68 build of the mouse genome with BWA-MEM 0.7.12-r1039. Indels realignment and base quality score recalibration was performed using the genome analysis toolkit v3.6 using the mouse genomes project 129/SvlmJ v5 merged indel and dbSNP142 vcf files (Keane et al., 2011). SNP and indels were called using the genome analysis toolkit haplotype caller. Larger insertions and deletions were called with DELLY (Rausch et al., 2012) and Manta (Chen et al., 2016). Reads with discordant mapping indicative of large deletions, insertions, inversions and translocations specifically covering the PAM site within *Viperin* were extracted from the BAM file based on bitwise mapping flags using samtools-v1.2 (*samtools view*) to quantify putative mutation events. The identity of novel sequence insertions was by BLAST to the NCBI non-redundant sequence database. Sequence alignments were visualized with the integrative genome viewer (Robinson et al., 2011).

Author contributions

F.A. & P.T. conceived and designed the study. F.A. performed all experiments apart from zygote microinjections (performed by S.P.), heteroduplex assay & sequencing of *Viperin*, *Pik3r6* and *Hmgcs2* mice (performed by K.H., M.B., M.T., S.M. & J.H.) and WGS data analysis (performed by M.C.). F.A. & P.T. wrote the manuscript with the input from all authors.

Conflict of interests

Authors declare that they have no competing of financial interests.
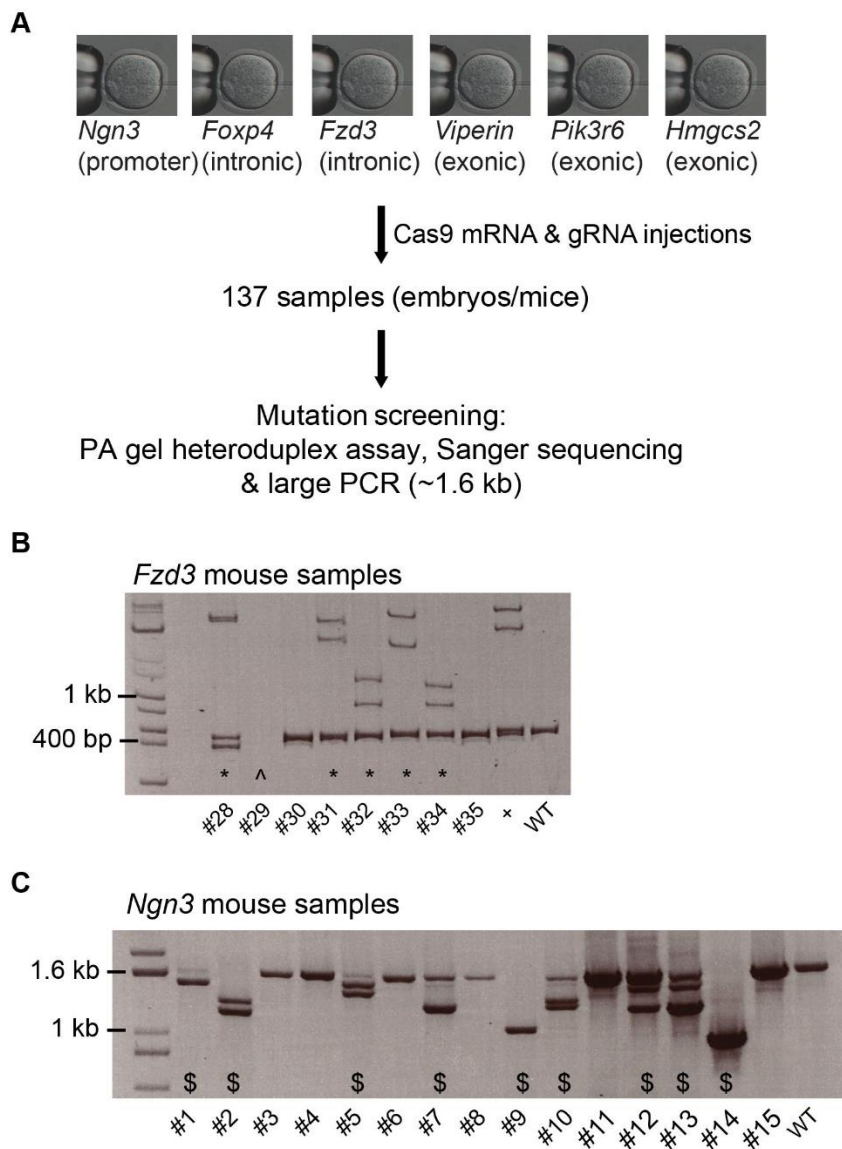
Acknowledgment

Figure 1. CRISPR/Cas9-mediated Breaks in Mouse Zygote Injection

(A) Experimental flow of mouse zygote injection experiment. Six independent zygote injections targeting biallelic sequences produced 137 samples that were analyzed by PA gel heteroduplex assay, Sanger sequencing and large PCR (~1.6 kb). (B) Example of PA gel heteroduplex assay with ~400 bp PCR in *Fzd3* mouse samples. Complete figures from all injections can be found in Figure S1. * = Samples containing heteroduplex generate extra bands besides the WT-sized bands. ^ = Sample that cannot be amplified which was found to have large deletion when larger PCR was performed. + = heteroduplex positive control. (C) Example of ~1.6 kb PCR in *Ngn3* samples. Complete figures can be found in Figure S2A. $ = Samples that were found to have smaller bands suggesting large deletions.
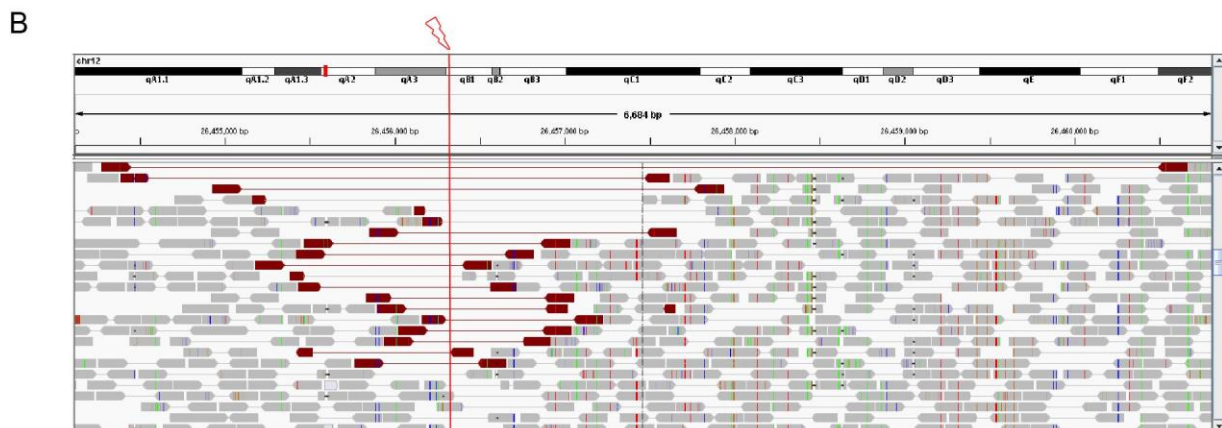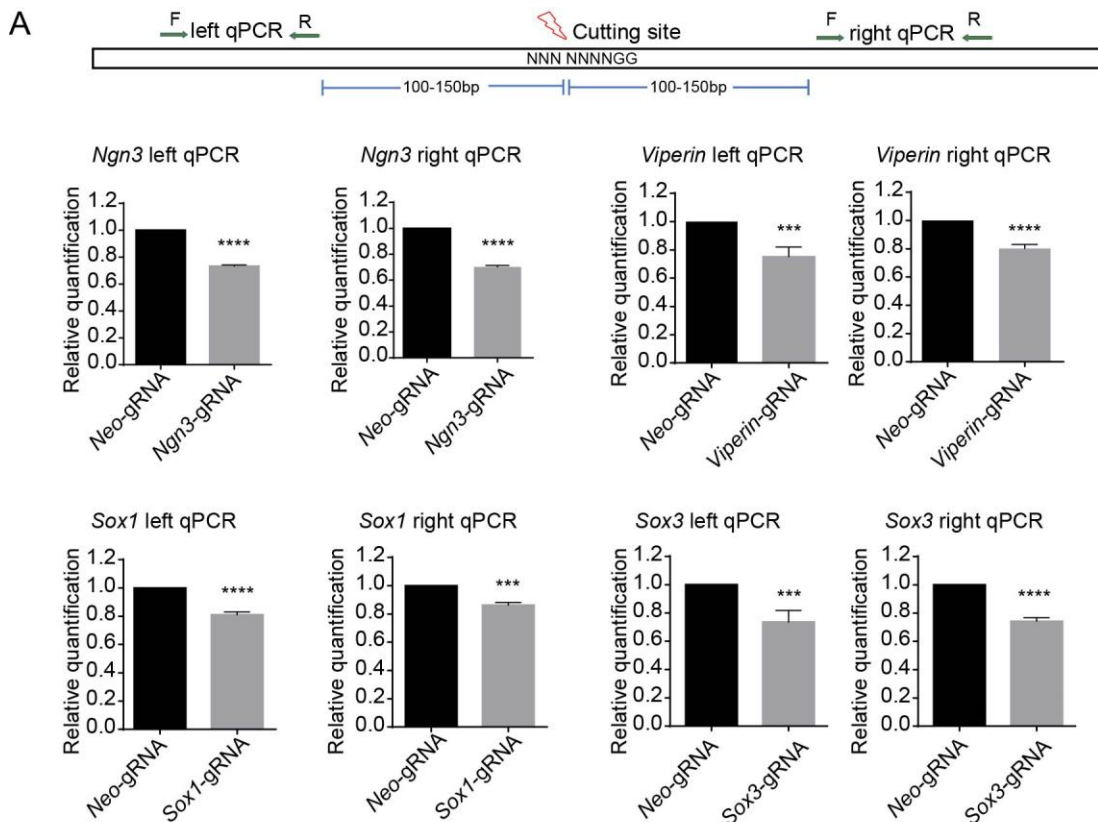
Figure 2. Detection of large deletions after CRISPR/Cas9-mediated breaks in pooled mouse ES cells

(A) Large deletions after CRISPR/Cas9 breaks were detected by qPCR using primers located >100 bp away from the cutting sites. Two qPCRs were conducted for each DSB, on the upstream (left) and downstream (right) relative to the NGG PAM sequences. *Neo* gRNA as control. Internal reference qPCR used *Sox1*-3'UTR qPCR, except for *Sox1*-left or *Sox1*-right qPCR which used *Sox2* qPCR. Data are presented as mean ± SEM (n ≥ 3). ***P<0.001; ****P<0.0001. (B) Whole genome sequencing analysis to *Viperin* gRNA-treated mouse ES cell pools. Integrative genome viewer snapshot shows reads paired in sequencing and sorted by insert size. Read pairs in red indicate discordantly mapping pairs indicative of large deletions
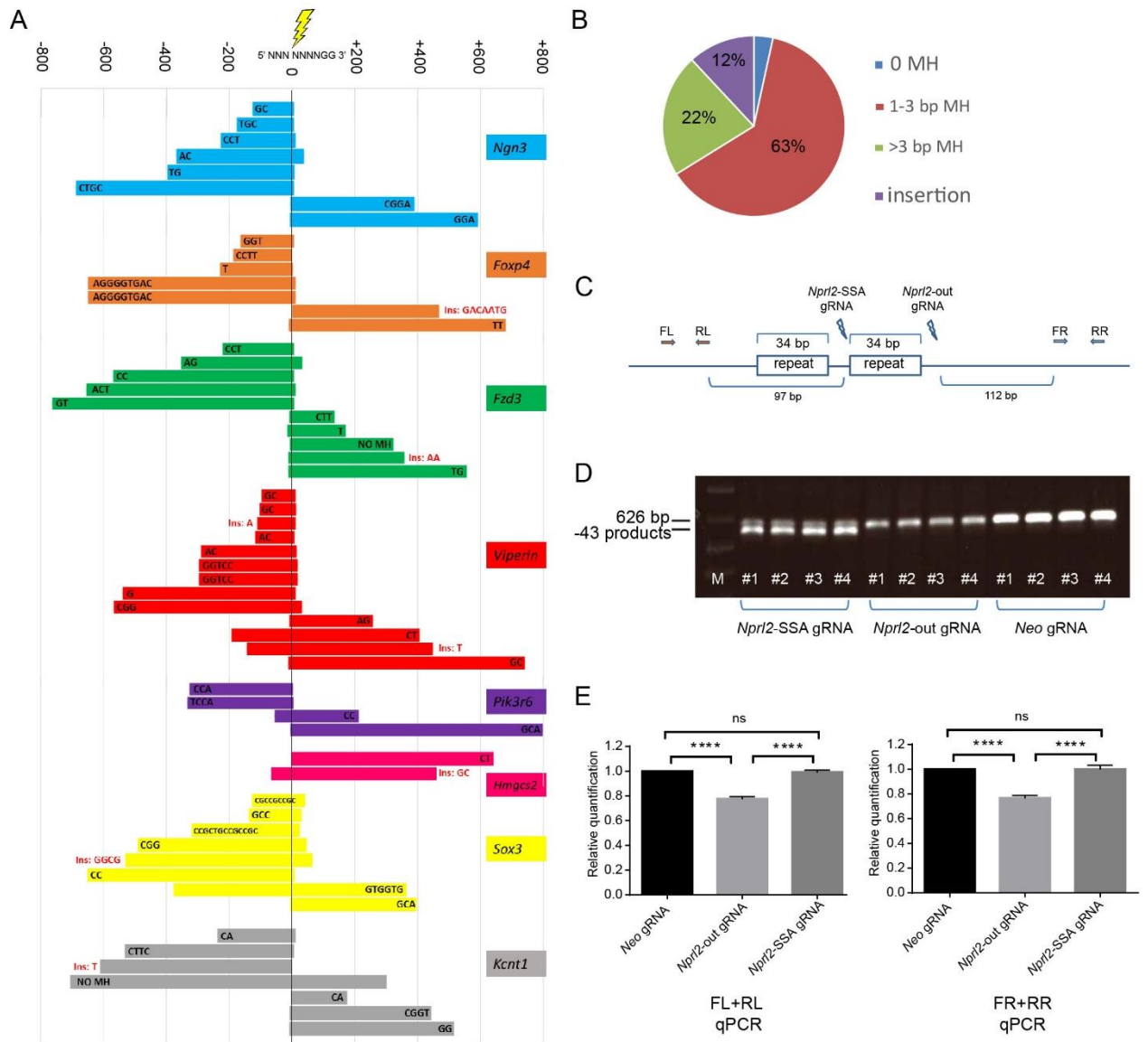
Figure 3. Characteristics of large deletions

(A) Sequencing of large deletion bands after ~1.6 kb PCR from mouse zygote injection samples. 0 represents the cutting site of Cas9. Each bar represents the deletion position relative to the NGG PAM sequences. Microhomologies or insertions in the break junction are indicated by sequences with black and red font, respectively. (B) Percentage of large deletion with MH or insertions. (C) Schematic of SSA trapping assay. (D) PCR across the cutting sites showed an extra band of 43 bp deletion after *Nprl2*-SSA gRNA cuts but not in other controls. Four biological replicates for each gRNA. M = marker. (E) Genomic DNA qPCR quantification using primers FL+RL (left figure) and FR+RR (right figure) as indicated in (C). Data were presented as mean ± SEM (n = 4). *Sox1* qPCR was used as internal reference.
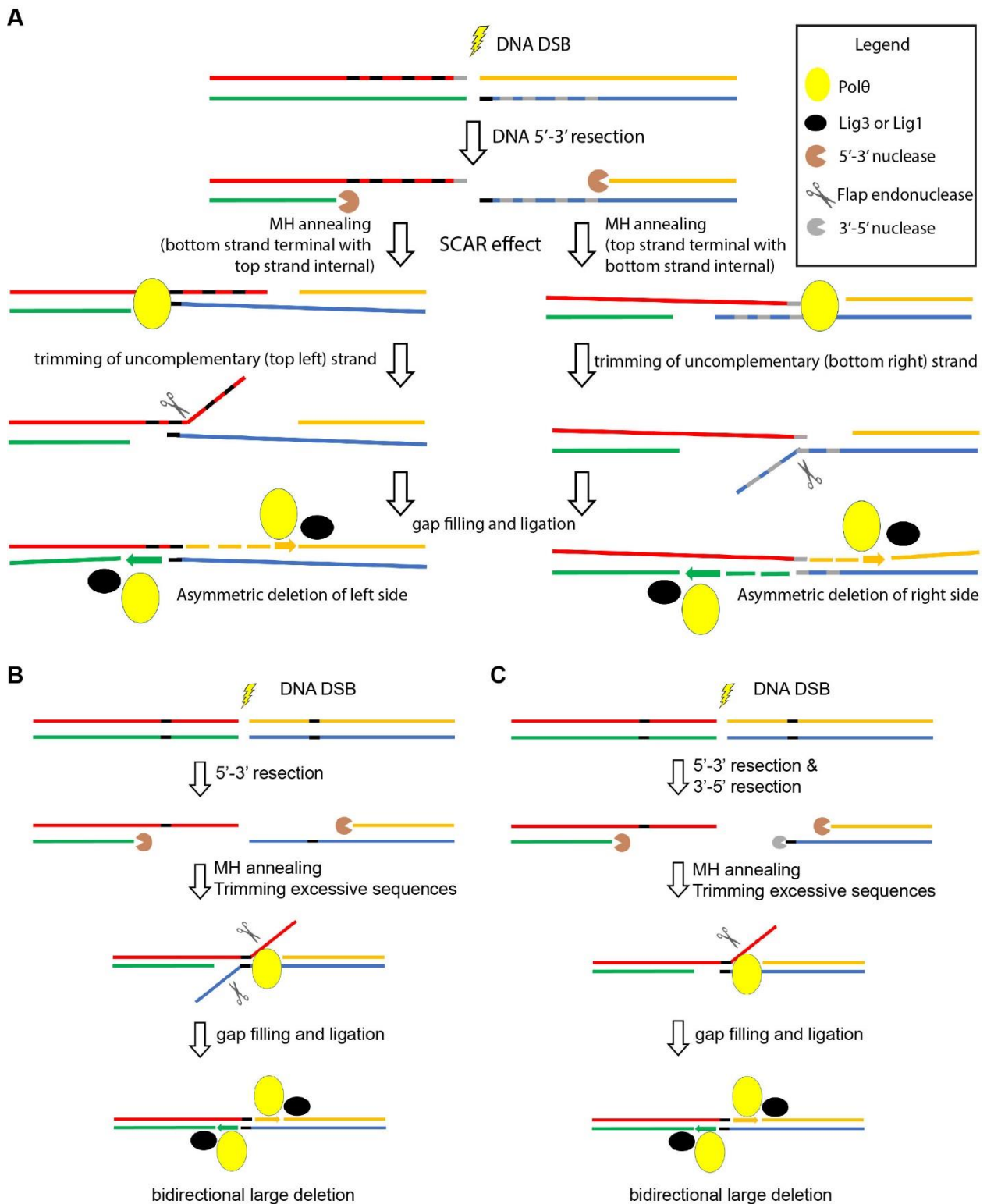
Figure 4. Mechanism of alternative end-joining generating large deletions

(A) Mechanism of unidirectional/asymmetric large deletions. (B) Mechanism of bidirectional large deletions via annealing between both internal sequences. (C) Mechanism of bidirectional large deletions that involves 3'-5' resection. (see discussion for details).

Supplementary Information

(Figure S1, S2, S3, S4, S5, Table S1, S2, S3, S4 and S5)

CRISPR/Cas9-induced breaks frequently generate random large deletions via DNA resection

Fatwa Adikusuma, Sandra Piltz, Mark A. Corbett, Michelle Turvey, Shaun McColl [1], Karla Helbig, Michael Beard, James Hughes & Paul Thomas

Ngn3 mouse samples

Foxp4 mouse samples

Fzd3 mouse samples
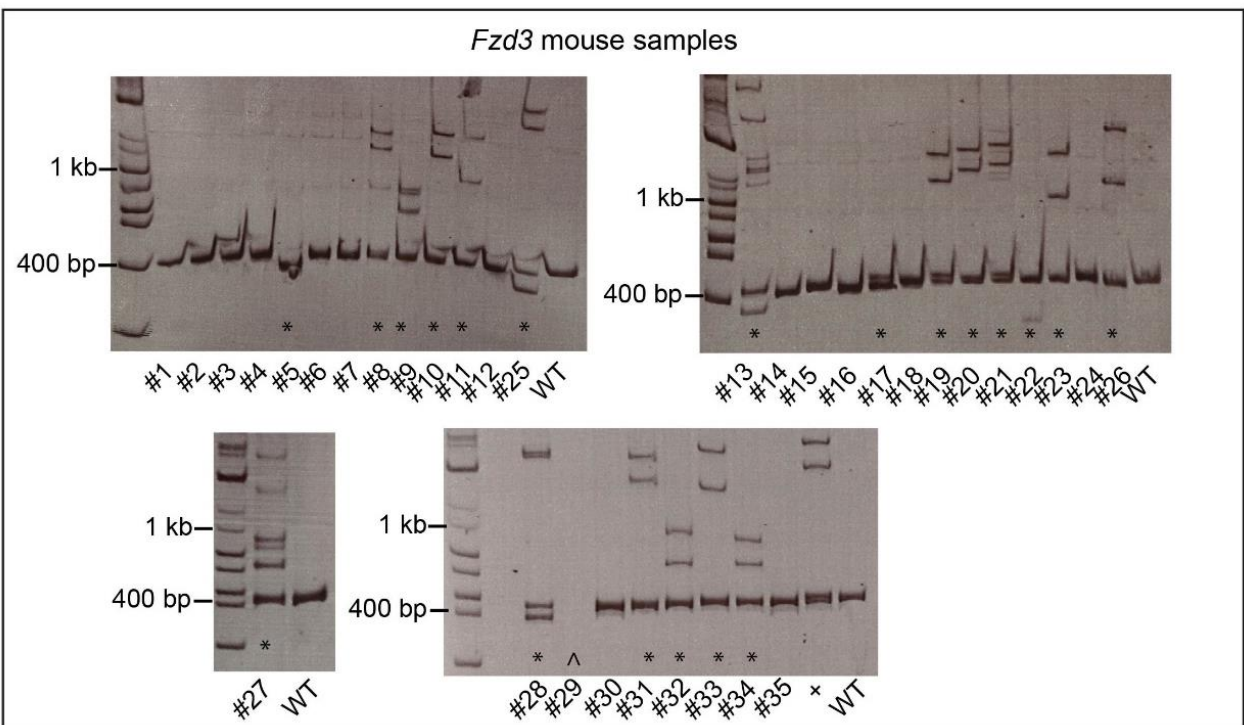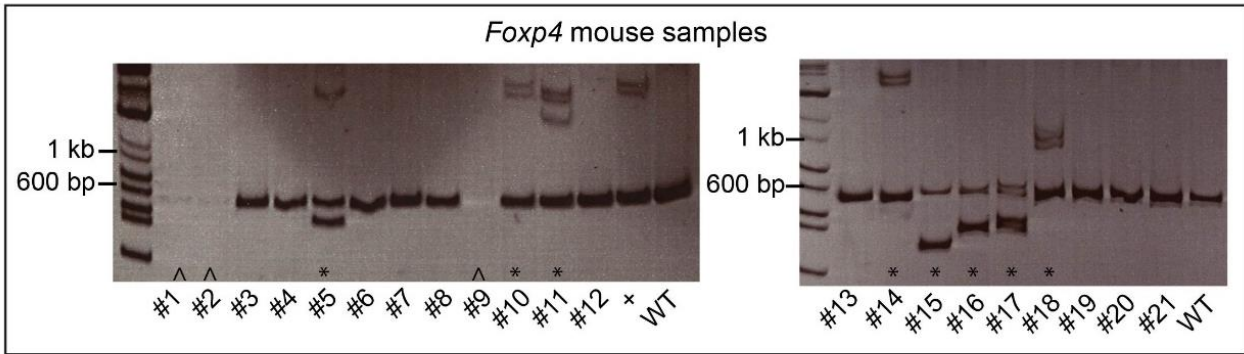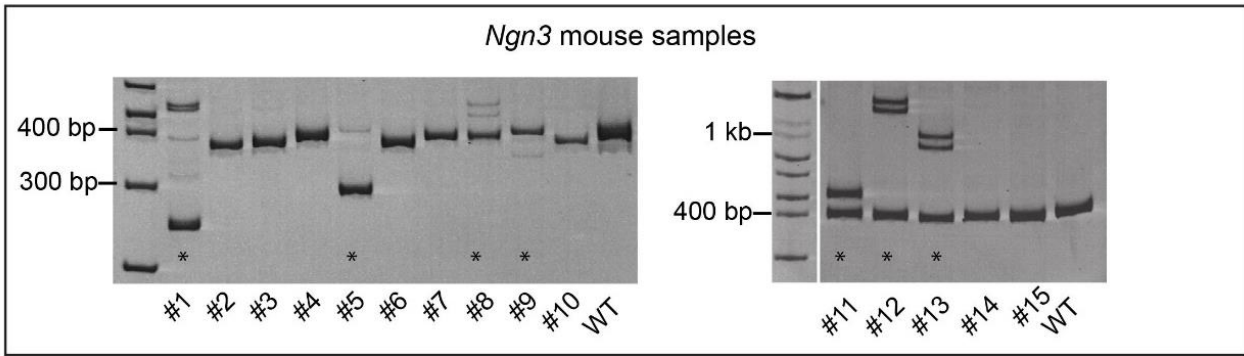
Figure continues to next page
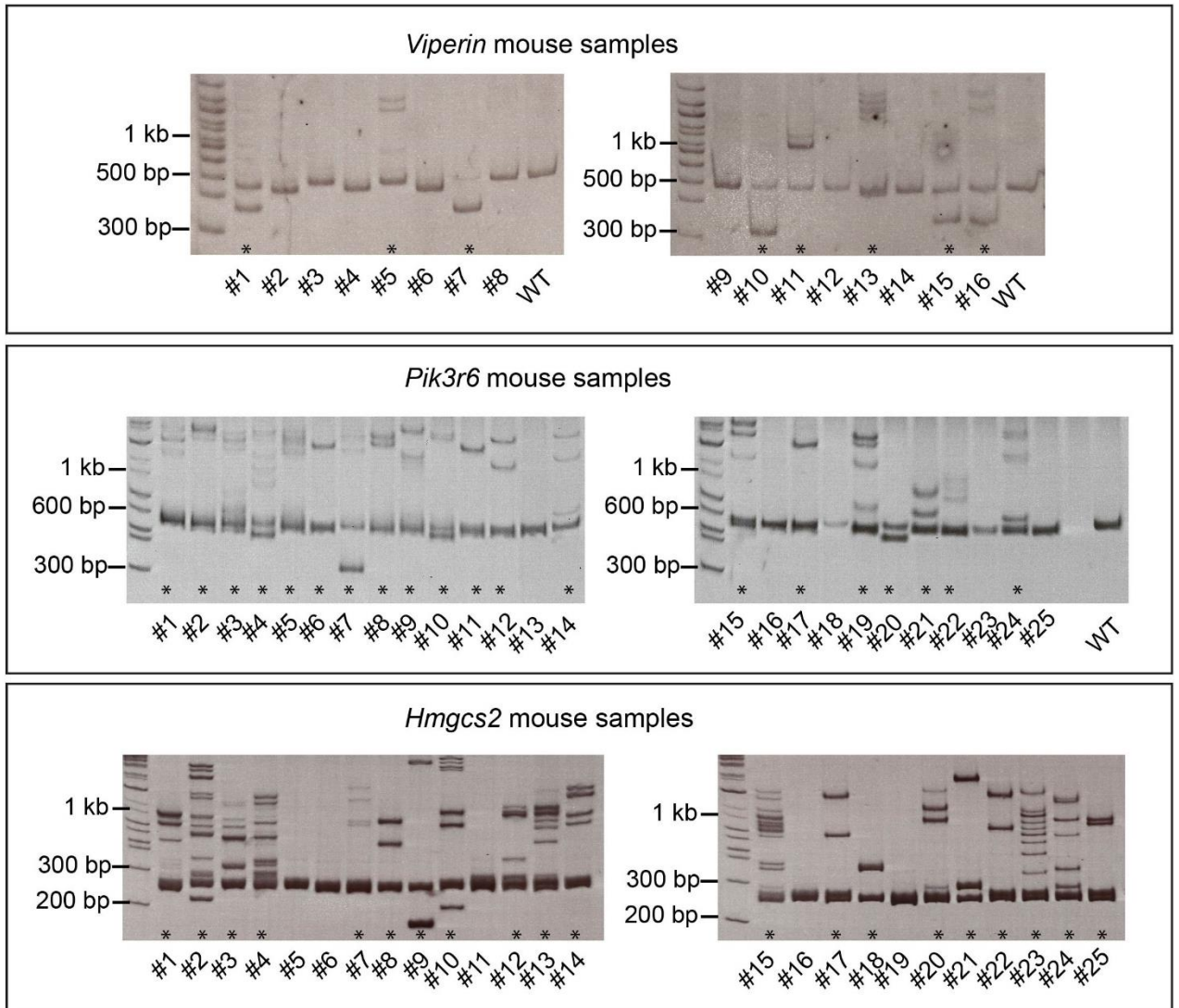
Figure S1. Complete figures of PA gel heteroduplex assay screening in mouse zygote injection samples (related to Figure 1B). * = Samples containing heteroduplex generate extra bands besides the WT-sized bands. ^ = Samples that cannot be amplified which was found to have large deletions when larger PCR was performed. + = heteroduplex positive control.

A

Figure continues to next page

136

Viperin mouse samples

Pik3r6 mouse samples

Hmgcs2 mouse samples
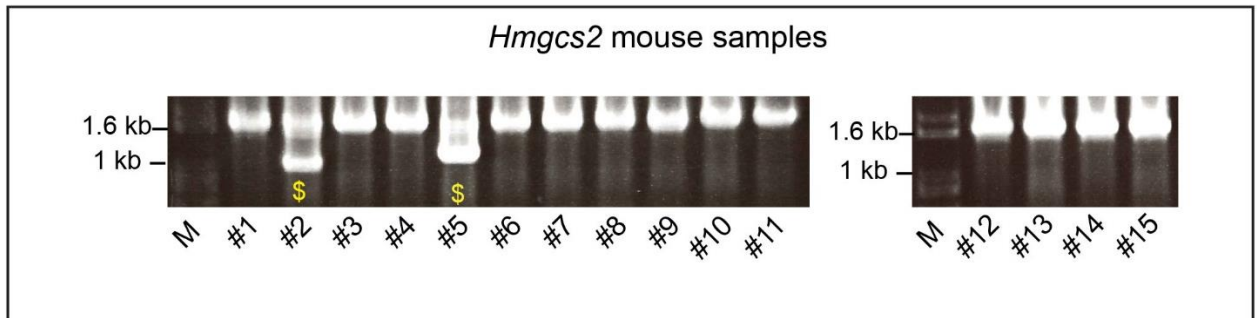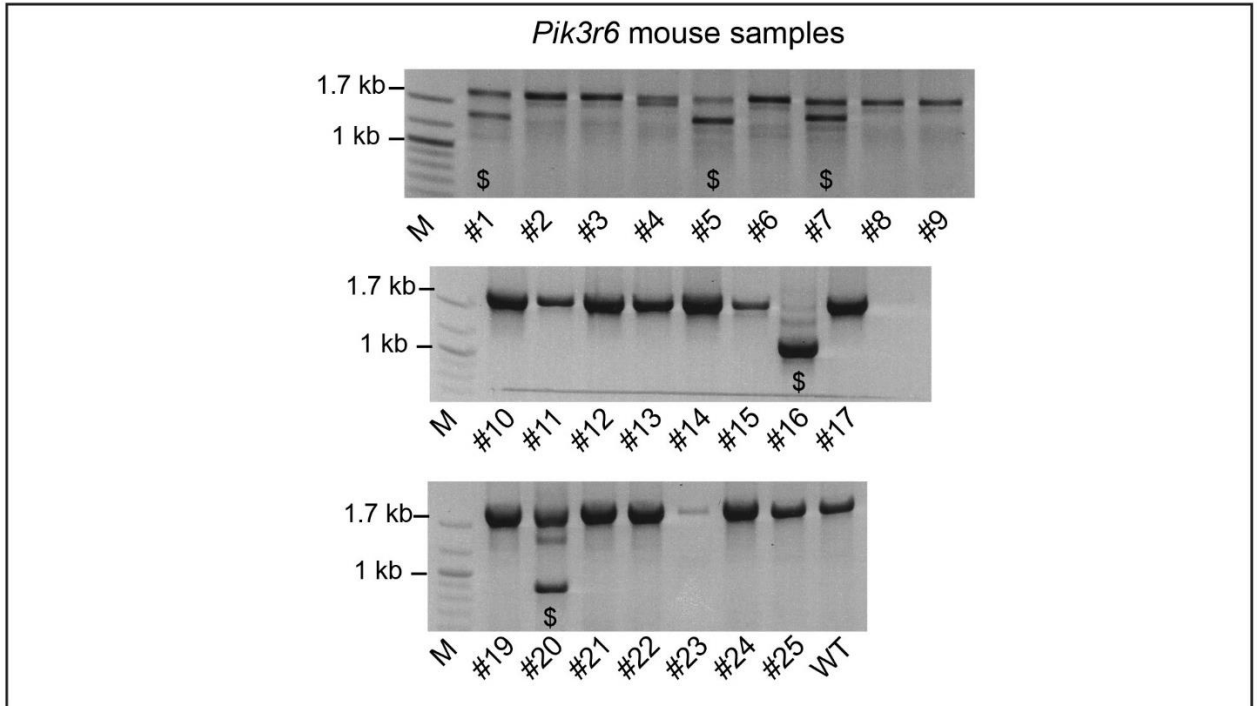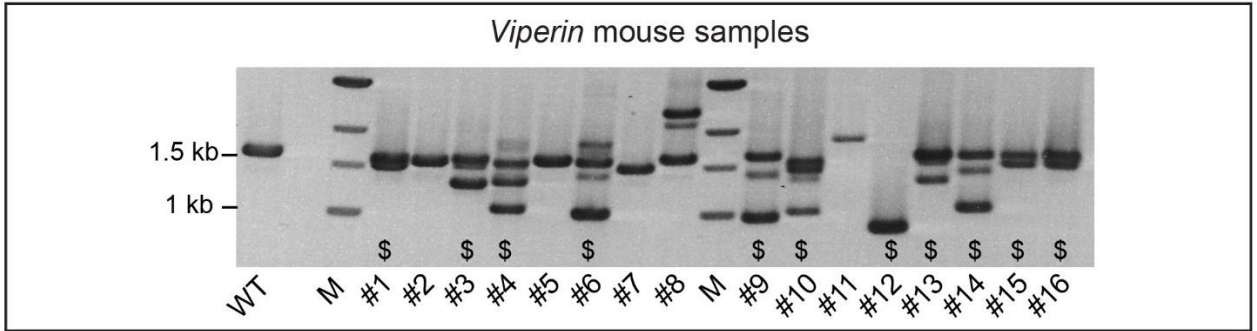
Figure continues to next page

B

Figure S2. Large deletion outcomes produced after Cas9-mediated breaks. (A) Complete figures of large PCR (~1.6 kb) (related to Figure 1C). $ = Samples that were found to have smaller bands suggesting large deletions. (B) ~3.2 kb PCR in *Ngn3* and *Foxp4* founders detected more large deletion bands. Newly found large deletions are indicated by the arrows.
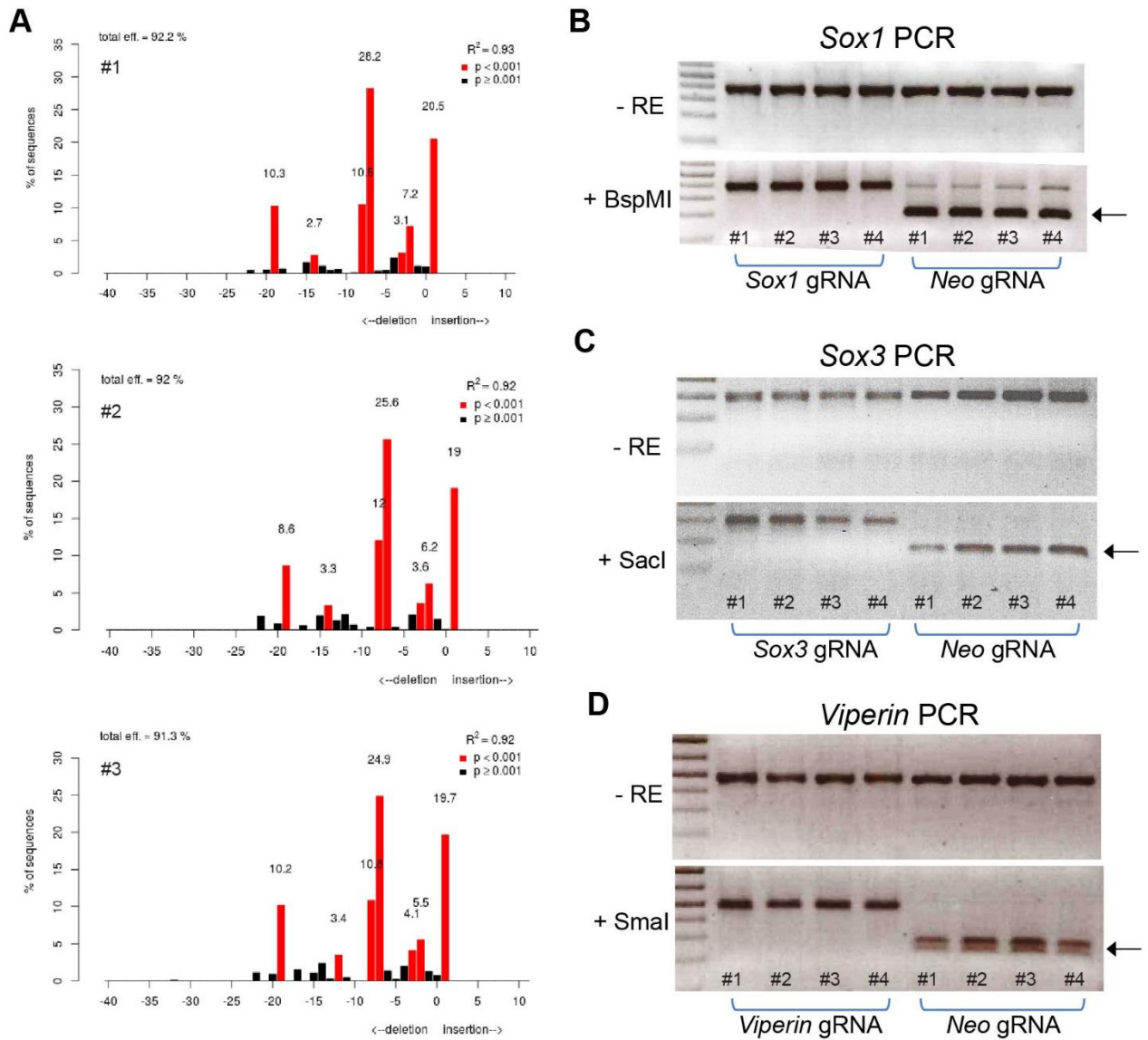
Figure S3. High cutting efficiency of CRISPR/Cas9-mediated breaks in mouse ES cells. (A) TIDE analysis on *Ngn3* gRNA-treated samples hardly contained WT alleles (n = 3). (B) RFLP analysis using BspMI on *Sox1* gRNA-treated samples (n = 4). (C) RFLP analysis using SacI on *Sox3* gRNA-treated samples (n = 4). (D) RFLP analysis using SmaI on *Viperin* gRNA-treated samples (n = 4). Each gRNA was designed to induce DSB at the corresponding restriction sites. Successful cuts should destroy the restriction sites that can be analyzed by restriction endonuclease (RE) digestion. Bands indicating the WT alleles (arrows) were clearly present in the negative controls (*Neo* gRNA), but were hardly detected in samples transfected with the gRNA treatments.
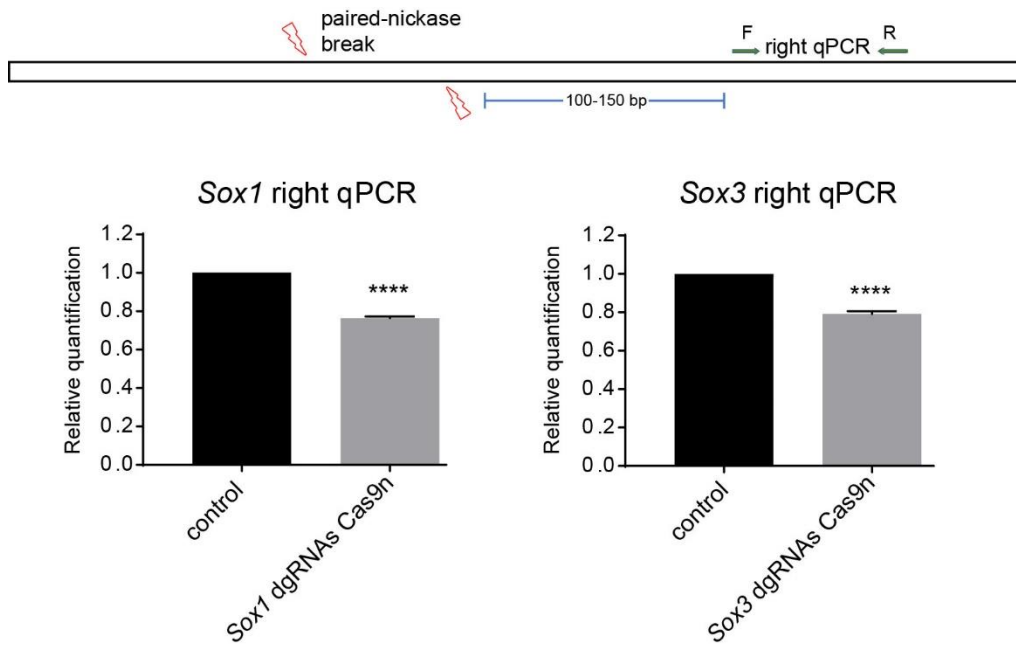
Figure S4. Sticky-ended breaks induced by paired Cas9 nickase also generate large deletions. Samples from paired-nickase targeting *Sox1* (left) or *Sox3* (right) were analyzed by qPCR using primers located >100 bp away from the breaks. The controls for *Sox1* dgRNA Cas9n were samples treated with *Sox3* dgRNA Cas9n, and vice versa.
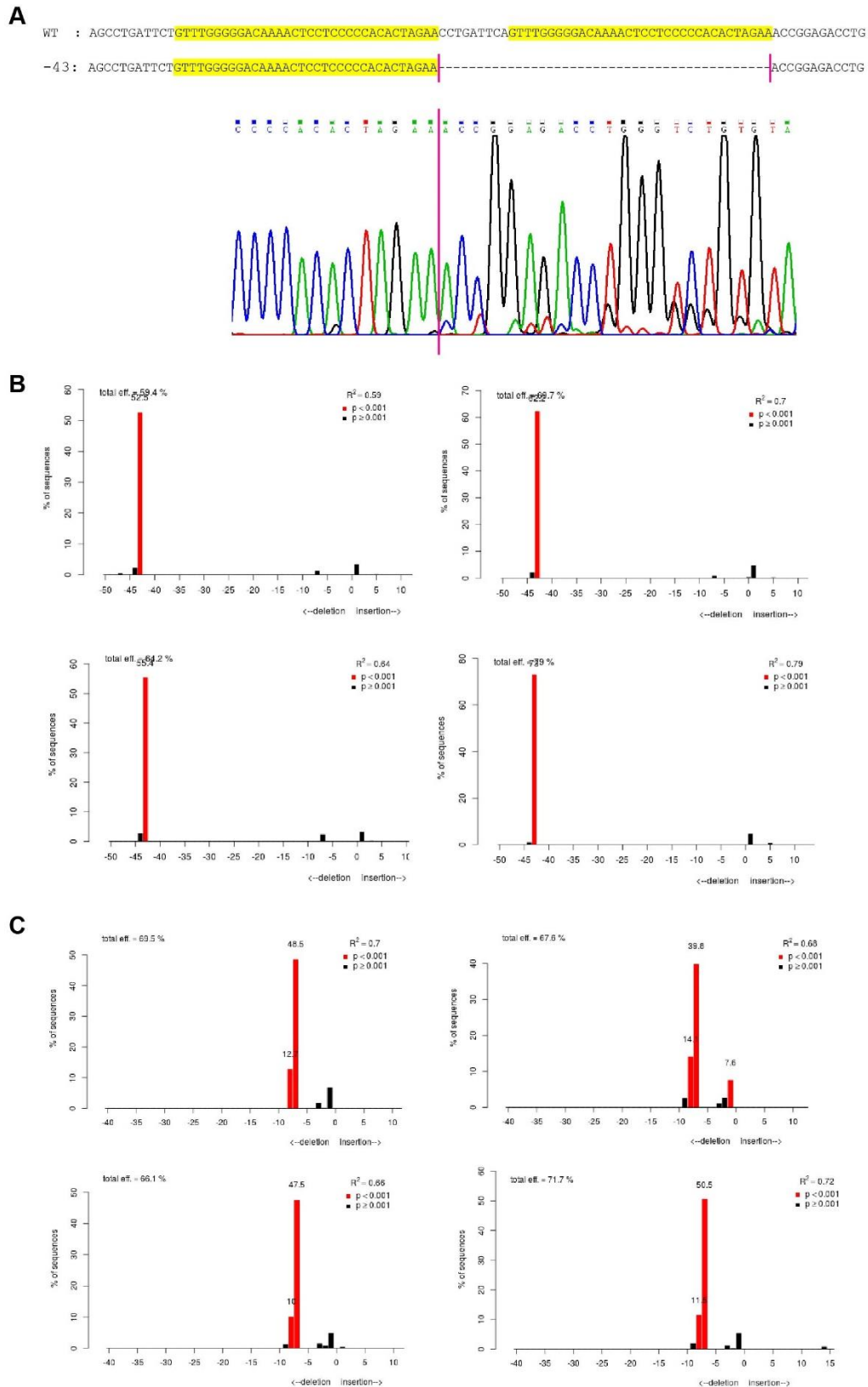
Figure S5. Sequencing and TIDE analyses of samples from SSA trapping experiment. (A) Chromatogram of *Nprl2*-SSA gRNA-treated samples clearly showed the abundance of 43 bp deletion as a result of SSA repair. (B) 43 bp deletion in *Nprl2*-SSA gRNA-treated samples shown by TIDE analyses (n = 4). (C) TIDE analyses of *Nprl2*-out gRNA-treated samples hardly detected WT alleles suggesting highly efficient mutation outcomes induced by this gRNA (n = 4).

Table S1. The summary of PA heteroduplex assay and direct Sanger sequencing mutation screening of zygote injected mouse founders

gRNA targets	Total samples	PA gel assay			Sequencing of PA+		Sequencing of PA -ve	
		+	-ve	fail	PA+ samples sequenced	mutants	PA -ve samples sequenced	mutants
Ngn3	15	7	8	0	7	7	8	7
Foxp4	21	8	10	3	8	8	10	10
Fzd3	35	20	14	1	15	15	8	8
Viperin	16	8	8	0	8	8	8	7
Pik3r6	25	20	5	0	20	20	5	4
Hmgcs2	25	20	5	0	19	19	4	2
total	137	83	50	4	77	77	43	38

Table S2. Summary of large deletions found after large PCR.

gRNA targets	Total samples	~1.6 kb PCR		~1.6 kb + ~3.2 kb PCR	
		Samples with large del	Total large del bands	Samples with large del	Total large del bands
NgN3	15	9	12	12	17
Foxp4	21	9	9	15	16
Fzd3	35	10	10	NA	NA
Viperin	16	11	13	NA	NA
Pik3r6	25	5	5	NA	NA
Hmgcs2	25	3	3	NA	NA
Total	**137**	**47**	**52**	**56**	**64**

Table S3. List of gRNAs used in zygote injection experiment

Targets	gRNA sequences	On-target score
Ngn3	GCACAGCTGGATTCCGGACAAA	56
Foxp4	CCAGCGTTCCCATTGTCCTT	37
Fzd3	CTTAGCAAGGGTGTGAAAAG	64
Viperin	GGGTGGCTAGATCCCGGGA	70
Pik3r6	CTTACCCTGATTGCTCTGGA	56
Hmgcs2	TACAATCCCTCCTGCTCCCC	39
Kcnt1	TGCATGAACCGCATGTTGGA	56
Sox3	GCCCACCGGGCTGCTGGCGG	31

Table S4. List of gRNAs used in ES cell experiment

Targets	gRNA sequences	On-target score
Ngn3	GCACAGCTGGATTCCGGACAAA	56
Sox1	GCCGCCGGGCGAGTGCAGGT	56
Viperin	GGGTGGCTAGATCCCGGGA	70
Sox3	GCTGACCCACATCTGAGCTC	58
Nprl2-SSA	GCACTAGAACCTGATTCAGTT	32
Nprl2-out	GACACTAGAAACCGGAGACCT	61
Neo	GGCAGCGCGGCTATCGTGGC	NA
Sox1 double nickase	GCCGCCGGGCGAGTGCAGGT & GCCCACGAACCTCTCGGGCC	NA
Sox3 double nickase	GCTGACCCACATCTGAGCTC & GACCGCAGTCCCGGCGCCC	NA

Table S5. The primers used in this study.

Target genes	F primer 5'-3'	R primer 5'-3'	purpose
Ngn3	CCCAAACCTCCTTCATGCTA	CTGCAGTGATGAGACCCAGA	PA gel assay PCR and TIDE analysis
Ngn3	TGCTAGTCCTCTCTGGTCTGTG	GCTCCCGATCATTGGCCTTC	Large 1.6 kb PCR
Ngn3	TGCTAGTCCTCTCTGGTCTGTG	CTCTCTGCCAACAGTCTGCC	Large 3.2 kb PCR
Ngn3	GGGCAGAGCAGATAAAGCGTG	CTGCAGTGATGAGACCCAGA	Left qPCR
Ngn3	CCCAAACCTCCTTCATGCTA	GTTCCTCAAAGAGCCTCGCC	Right qPCR
Foxp4	AGTTCAAGGCCATCTGCCAC	CACCCAGCCCTTCTAAGTAGC	PA gel assay PCR
Foxp4	TGGCAGGACAGAGCAAACAG	CCTCAATCCTCCTCAGTGGG	Large 1.6 kb PCR
Foxp4	CTGAGGGTTGTTCTCCCACTTC	CCTCAATCCTCCTCAGTGGG	Large 3.2 kb PCR
Fzd3	AGGCTGTTCCACATTGGTTC	GTGTTTCTCTAAGCAGGGATGT	PA gel assay PCR
Fzd3	CACACAGAGTATTGTTCGCAG	CATCTGCATAAACCCACACTC	Large 1.6 kb PCR
Viperin	ACTGAGTCAAGGGAGGTGTTTC	GTTTGAGCAGAAGCAGTCCTCG	PA gel assay PCR and RFLP analysis
Viperin	GTGTTTGCCTGGAATATACCAGTCTTGAGTCCT	GACAATCTGCAAGGATTGAATGCTA	Large 1.6 kb PCR
Viperin	ACTGAGTCAAGGGAGGTGTTTC	GATAGGCACACACCTGCTGCT	Left qPCR
Viperin	CTACCACTTCACTCGTCAGTGC	GTTTGAGCAGAAGCAGTCCTCG	Right qPCR
Pik3r6	ATCTTCCAGCTCAGAGCAATG	GGAAGGTAACCAGGAAAGAAGG	PA gel assay PCR
Pik3r6	AGTGGGAATGATACAGCAGGAC	AGGCTTGAGCACTCACCTTC	Large 1.6 kb PCR
Hmgcs2	AATTGCCACATTATTGGTTGG	ACTTCCCTGCTTCCACATTG	PA gel assay PCR
Hmgcs2	AATCAAGCCACCACTCTTGC	CCCTCCCTTTCCTAAGTTGC	Large PCR
Kcnt1	CTTCTGTCCTCACAGGCCTGA	ACAGAGCCTAGGGAGAGTTTGG	Large 1.6 kb PCR
Sox3	GAACGCATCAGGTGAGAGAA	AACCTAGGAATCCGGGAAGA	Large 1.6 kb PCR (zygotes)
Sox3	CAGCATGTACCTGCCACCT	ACAAAACCCCGACAGTTACG	RFLP analysis (ES cells)

Sox3	CATCGCTTCGCACTCGCA	GCAGGTACATGCTGATCATGTC	Left qPCR
Sox3	CGTTGCCTTGTACCGAAGAT	CGGGACTTCTCGCTTTTGTAC	Right qPCR
Sox1	CCCTTCTCTCCGCTAGGC	GTTGTGCATCTTGGGGTTTT	RFLP analysis
Sox1	GATCCTGGTTGGCCTTGGTG	GTTGTGCATCTTGGGGTTTT	Left qPCR
Sox1	CCCTTCTCTCCGCTAGGC	GGTGGGTGGAGAGAGGATCA	Right qPCR
Nprl2	TTCTTCAGCGAGTTCCACCC	TACACCTGGACCGTGTCAAA	TIDE analysis
Nprl2	CCCAAGAGAAACACTGGACCAAG	CAGTGGTCTACTTTGTGCTTCCA	Left qPCR
Nprl2	ATTCCAAGGCGTAGAGGCGATC	GGAGTGTGGAAGGCACCTGAT	Right qPCR
Sox1	GACTTGCAGGCTATGTACAACATC	CCTCTCAGACGGTGGAGTTATATT	Internal reference qPCR
Sox2	ACCAGCTCGCAGACCTACAT	TCGGACTTGACCACAGAGC	Internal reference qPCR

References

Adikusuma, F., Pederick, D., McAninch, D., Hughes, J., and Thomas, P. (2017a). Functional Equivalence of the SOX2 and SOX3 Transcription Factors in the Developing Mouse Brain and Testes. Genetics *206*, 1495-1503.

Adikusuma, F., Williams, N., Grutzner, F., Hughes, J., and Thomas, P. (2017b). Targeted Deletion of an Entire Chromosome Using CRISPR/Cas9. Mol Ther.

Bell, C.C., Magor, G.W., Gillinder, K.R., and Perkins, A.C. (2014). A high-throughput screening strategy for detecting CRISPR-Cas9 induced mutations using next-generation sequencing. BMC Genomics *15*, 1002.

Bhargava, R., Onyango, D.O., and Stark, J.M. (2016). Regulation of Single-Strand Annealing and its Role in Genome Maintenance. Trends Genet *32*, 566-575.

Black, S.J., Kashkina, E., Kent, T., and Pomerantz, R.T. (2016). DNA Polymerase theta: A Unique Multifunctional End-Joining Machine. Genes (Basel) *7*.

Brinkman, E.K., Chen, T., Amendola, M., and van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res *42*, e168.

Chang, H.H.Y., Pannunzio, N.R., Adachi, N., and Lieber, M.R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nat Rev Mol Cell Biol.

Chen, J., Zhang, X., Wang, T., Li, Z., Guan, G., and Hong, Y. (2012). Efficient detection, quantification and enrichment of subtle allelic alterations. DNA Res *19*, 423-433.

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Kallberg, M., Cox, A.J., Kruglyak, S., and Saunders, C.T. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics *32*, 1220-1222.

Cheng, W., Zhao, H., Yu, H., Xin, J., Wang, J., Zeng, L., Yuan, Z., Qing, Y., Li, H., Jia, B.*, et al.* (2016). Efficient generation of GGTA1-null Diannan miniature pigs using TALENs combined with somatic cell nuclear transfer. Reprod Biol Endocrinol *14*, 77.

Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., and Kuhn, R. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol *33*, 543-548.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A.*, et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

Deriano, L., and Roth, D.B. (2013). Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. Annu Rev Genet *47*, 433-455.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R.*, et al.* (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol *34*, 184-191.

Dorsett, Y., Zhou, Y., Tubbs, A.T., Chen, B.R., Purman, C., Lee, B.S., George, R., Bredemeyer, A.L., Zhao, J.Y., Sodergen, E.*, et al.* (2014). HCoDES reveals chromosomal DNA end structures with single-nucleotide resolution. Mol Cell *56*, 808-818.

Haeussler, M., Schonig, K., Eckert, H., Eschstruth, A., Mianne, J., Renaud, J.B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J.*, et al.* (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol *17*, 148.

Hisano, Y., Sakuma, T., Nakade, S., Ohga, R., Ota, S., Okamoto, H., Yamamoto, T., and Kawahara, A. (2015). Precise in-frame integration of exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. Sci Rep *5*, 8841.

Honma, M., Izumi, M., Sakuraba, M., Tadokoro, S., Sakamoto, H., Wang, W., Yatagai, F., and Hayashi, M. (2003). Deletion, rearrangement, and gene conversion; genetic consequences of chromosomal double-strand breaks in human cells. Environ Mol Mutagen *42*, 288-298.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262-1278.

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.

Keane, T.M., Goodstadt, L., Danecek, P., White, M.A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M.*, et al.* (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. Nature *477*, 289-294.

Kent, T., Chandramouly, G., McDevitt, S.M., Ozdemir, A.Y., and Pomerantz, R.T. (2015). Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase theta. Nat Struct Mol Biol *22*, 230-237.

Kent, T., Mateos-Gomez, P.A., Sfeir, A., and Pomerantz, R.T. (2016). Polymerase theta is a robust terminal transferase that oscillates between three different mechanisms during end-joining. Elife *5*.

Komor, A.C., Badran, A.H., and Liu, D.R. (2017). CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell *168*, 20-36.

Koole, W., van Schendel, R., Karambelas, A.E., van Heteren, J.T., Okihara, K.L., and Tijsterman, M. (2014). A Polymerase Theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. Nat Commun *5*, 3216.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823-826.

Maruyama, T., Dougan, S.K., Truttmann, M.C., Bilate, A.M., Ingram, J.R., and Ploegh, H.L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nat Biotechnol *33*, 538-542.

Mou, H., Smith, J.L., Peng, L., Yin, H., Moore, J., Zhang, X.O., Song, C.Q., Sheel, A., Wu, Q., Ozata, D.M.*, et al.* (2017). CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. Genome Biol *18*, 108.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat Protoc *8*, 2281-2308.

Rausch, T., Zichner, T., Schlattl, A., Stutz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics *28*, i333-i339.

Robert, F., Barbeau, M., Ethier, S., Dostie, J., and Pelletier, J. (2015). Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. Genome Med *7*, 93.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol *29*, 24-26.

Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol *32*, 347-355.

Sfeir, A., and Symington, L.S. (2015). Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? Trends Biochem Sci *40*, 701-714.

Shen, B., Zhang, J., Wu, H., Wang, J., Ma, K., Li, Z., Zhang, X., Zhang, P., and Huang, X. (2013). Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. Cell Res *23*, 720-723.

Shin, H.Y., Wang, C., Lee, H.K., Yoo, K.H., Zeng, X., Kuhns, T., Yang, C.M., Mohr, T., Liu, C., and Hennighausen, L. (2017). CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. Nat Commun *8*, 15464.

Simsek, D., Brunet, E., Wong, S.Y., Katyal, S., Gao, Y., McKinnon, P.J., Lou, J., Zhang, L., Li, J., Rebar, E.J.*, et al.* (2011). DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. PLoS Genet *7*, e1002080.

Suzuki, K., Tsunekawa, Y., Hernandez-Benitez, R., Wu, J., Zhu, J., Kim, E.J., Hatanaka, F., Yamamoto, M., Araoka, T., Li, Z.*, et al.* (2016). In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. Nature *540*, 144-149.

van Schendel, R., Roerink, S.F., Portegijs, V., van den Heuvel, S., and Tijsterman, M. (2015). Polymerase Theta is a key driver of genome evolution and of CRISPR/Cas9-mediated mutagenesis. Nat Commun *6*, 7394.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell *153*, 910-918.

Wyatt, D.W., Feng, W., Conlin, M.P., Yousefzadeh, M.J., Roberts, S.A., Mieczkowski, P., Wood, R.D., Gupta, G.P., and Ramsden, D.A. (2016). Essential Roles for Polymerase theta-Mediated End Joining in the Repair of Chromosome Breaks. Mol Cell *63*, 662-673.

Yang, H., Wang, H., Shivalila, C.S., Cheng, A.W., Shi, L., and Jaenisch, R. (2013). One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell *154*, 1370-1379.

Yu, C., Liu, Y., Ma, T., Liu, K., Xu, S., Zhang, Y., Liu, H., La Russa, M., Xie, M., Ding, S.*, et al.* (2015). Small molecules enhance CRISPR genome editing in pluripotent stem cells. Cell Stem Cell *16*, 142-147.

Zelensky, A.N., Schimmel, J., Kool, H., Kanaar, R., and Tijsterman, M. (2017). Inactivation of Pol theta and C-NHEJ eliminates off-target integration of exogenous DNA. Nat Commun *8*, 66.

Zhang, T., Yin, Y., Liu, H., Du, W., Ren, C., Wang, L., Lu, H., and Zhang, Z. (2016). Generation of VDR Knock-Out Mice via Zygote Injection of CRISPR/Cas9 System. PLoS One *11*, e0163551.

Zuckermann, M., Hovestadt, V., Knobbe-Thomsen, C.B., Zapatka, M., Northcott, P.A., Schramm, K., Belic, J., Jones, D.T., Tschida, B., Moriarity, B.*, et al.* (2015). Somatic CRISPR/Cas9-mediated tumour suppressor disruption enables versatile brain tumour modelling. Nat Commun *6*, 7391.

Chapter 6:
General discussion and future directions

6.1. Increasing the efficiency of genome editing

CRISPR/Cas9 technology has developed rapidly since the first successful modification of the mammalian genome [27, 129]. The generation of mutations can now be conducted with high efficiency. Chapter 5 has shown that generating genetically modified mice containing mutations, such as small indels, is very efficient, with almost all mice containing mutations. In Chapter 2, it was shown that the deletion of a large region flanked by two gRNAs is less efficient. Deletion of large intervening sequences has been proposed for the treatment of diseases that can be cured by such deletions. To translate this strategy into clinical practice, it is crucial to improve the efficiency of generating large deletions. This low efficiency may be caused by a low delivery efficiency of both gRNAs. Therefore, the optimisation of vectors and delivery techniques that can carry all components, including both gRNAs, in one vector that enables efficient in vivo delivery should be the next investigation priority. Low efficiency of intervening sequence deletion might also be caused by prominent NHEJ repair, which could generate indels at both cutting sites without snipping out the intervening sequences [41]. Studying the mechanism of large intervening sequence deletion, including characterising the molecular factors involved, might uncover a strategy to suppress indel NHEJ and increase the efficiency of large deletions. The vectors generated in this study described in Chapter 4 may facilitate the investigation of the mechanisms involved in the deletion of long intervening sequences, as these vectors allow the efficient simultaneous delivery of dual gRNAs.

Many disease-causing mutations may be repaired using DNA replacement or insertions generated through the HDR mechanism. However, as shown in Chapter 2, intentional genomic modification by DNA KI is inefficient, which is problematic for therapeutic interventions or experiments requiring this approach [8]. Therefore, future studies need to address this problem to better utilise CRISPR/Cas9 technology for the seamless correction of disease-causing mutations.

One strategy to increase HDR is to hijack the cellular repair mechanisms to favour HDR over NHEJ. These strategies have been tested in studies using inhibitors of NHEJ, such as Scr7, NU7441 and KU-0060648, to force the repair mechanisms towards the HDR pathway, which involves a DNA end resection step [74-76]. However, some of these studies reported no increase in HDR when these inhibitors were used [139, 164]. Furthermore, it is possible that these inhibitors could cause adverse effects or toxicity, which could create unforeseen problems.

The study presented in Chapter 5 indicated that DNA end resection occurs frequently and can be repaired by an alternative end joining mechanism, resulting in large deletions. Despite frequent DNA end resection, the efficiency of donor-templated repair remains low, presumably due to a preference for alternative end joining repair rather than donor-templated repair. Therefore, we need to bear in mind that inhibition of the NHEJ pathway might lead to DNA resection, but could possibly result in alternative end joining repair outcomes, neglecting the DNA donor provided. Therefore, increasing the KI efficiency might require additional inhibition of the alternative end joining pathway. It would be interesting to investigate competition between alternative end joining and donor-mediated repair in the HDR pathway.

It is also possible that the abandonment of available donor repair template after DNA end resection occurs because the donor DNA is not a suitable template for the repair machinery. Therefore, it is also crucial to determine the type of DNA donors preferred by the DNA repair system. To date, the most efficient integration is achieved by providing ssODN donor with homology arms [35, 50-52]. The use of ssODNs was previously limited by the maximum size at which they can be generated (200 nt), which has meant that the insertion of longer sequences was not possible. Insertion of longer sequences can now be achieved through HR or other recently developed techniques, such as HITI and PITCh, as

mentioned in Chapter 1 [58-60]. For performing DNA insertion using HITI, it is not recommended to use a plasmid to express the gRNA and Cas9, as frequent plasmid integration into the break site has been observed, possibly via the same pathway as HITI. The plasmid may compete with the donor DNA for insertion into the breaks, thus reducing the insertion efficiency. Insertion of long sequences can now be achieved by using long ssDNA donor which is generated by the in vitro transcription reverse transcription (iVT-RT) technique, or is obtained commercially from IDT with its product named Megamer ssDNA fragments. This KI approach, known as Easi-CRISPR, has been used to flox an exon of mouse gene through zygote injection by inducing two cuts at both floxing sites and co-delivery of the long ssDNA donor [56, 57]. Our dual gRNA vectors described in this thesis provide tools to help performing floxing strategy using long ssDNA in cell lines due to its ability to increase the dual cutting efficiency.

In this chapter, a strategy to easily generate long DNA donors with single strand homology arms, mimicking the ssODN donor homology arms, is proposed. This technique, named HD-ADI (heteroduplex donor-assisted DNA integration), is very simple as it involves only two PCR amplification steps for the generation of DNA donors. Two sets of PCRs are required: the first to amplify the insert sequences, while the second amplifies the insert with the addition of homology arm sequences (20–40 nt) to the PCR primers (Figure 6.1). The first PCR produces purely insert sequences, while the second PCR produces insert sequences flanked by homology arms. The products from the first and second PCRs are mixed and slowly re-annealed. This re-annealing should produce four different kinds of products; two homoduplex products and two heteroduplex products, in which the latter are the dsDNA inserts with single-stranded homology arms (Figure 6.1). Interestingly, although it was expected that the donor products that facilitate efficient insertions would be the heteroduplexes, the remaining homoduplexes can also act as donor templates. The first homoduplex from the first PCR could integrate into the DSB via the NHEJ mechanism,

similar to the HITI method. The second homoduplex from the second PCR could be inserted into the DSB via the MMEJ mechanism, similar to the PITCh method (Figure 6.1). Therefore, the chances of DNA insertions are increased by providing various kinds of donor templates for both kinds of repair pathways (NHEJ or HDR). Another advantage of this proposed technique compared to Easi-CRISPR is its ability to perform longer insertions, since PCR can produce longer inserts than the Easi-CRISPR can generate (~2 kb).

The other aspect necessary to consider when trying to increase the efficiency of DNA replacement/insertion is to induce the DSB very close to the insertion site [98]. The targetable sites of the current commonly used CRISPR SpCas9 system are limited by the requirement for a NGG PAM, which limits the available cleavable sites. To broaden the targetable regions, other CRISPR systems that recognise various PAM sequences need to be expanded and optimised and extensive research to screen or engineer these CRISPR systems must be carried out. It is still possible that an RNA-guided endonuclease system with a very relaxed, yet highly specific PAM requirement will be found. One study has engineered FnCas9 to recognise a relaxed PAM YG, however this engineered Cas9 was shown to be inefficient for generating DSB-mediated mutations [108].

Another way to undertake efficient mutation correction is by utilising the base editor in which dCas9 or Cas9-nickase is attached to cytidine deaminase, as discussed in Chapter 1. CRISPR base editing shows great potential to efficiently achieve substitutions without the requirement for a repair template and the HDR process, both of which contribute to the low efficiency observed when performing substitutions. The base editor could substitute C $\rightarrow$ T with an efficiency of up to 75% in mammalian cells [125, 126]. This is exciting, given that the efficiency of HDR using donor template is very low. However, currently the base editor still requires considerable improvement, such as increasing the bases that can be

PCR 1: amplification of the pure insert
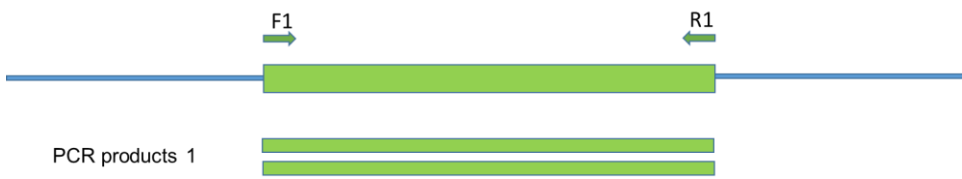
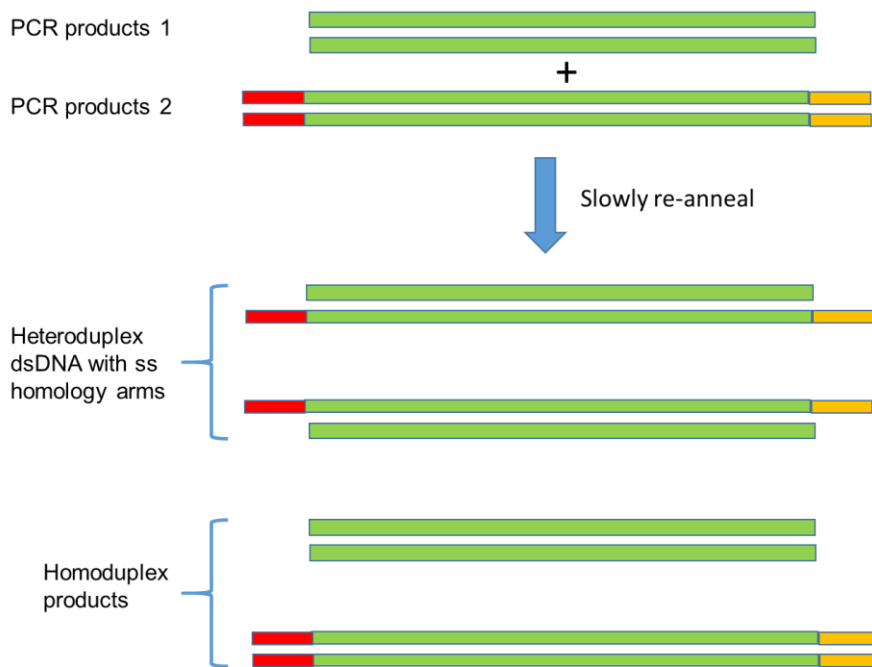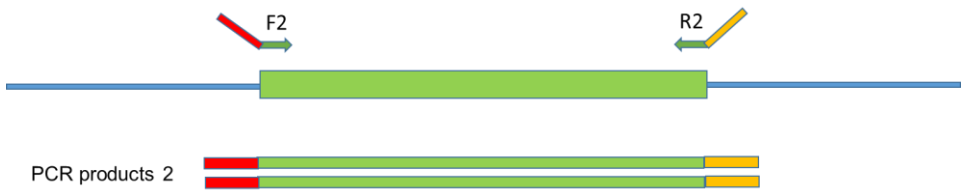PCR 2: amplification of insert using primers containing homology arms

Figure 6.1 | Generation of donor templates for the HD-ADI strategy.

edited as the current base editor can only edit C → T or G → A. The targetable sites that can be edited are also still limited, since the current base editor requires NGG PAM sequences, and only edits the targeted base within a small window at positions 4–8 of the targeting sequences. It would be an advantage if base editors could choose the bases they target flexibly. Therefore, future studies should engineer the current base editor to be able to target more a flexible window, and explore the utilisation of the other CRISPR systems that have various PAM requirements for their base editor agents. Currently, the most efficient base editor can still induce indels at the target sites, and can also induce substitutions at unintended sites (off-targets). Future studies must address this problem, for example by optimising Base Editor version 2 (BE2), which induces fewer indels, or by using other CRISPR systems with high specificity, such as Cpf1 or high fidelity engineered Cas9. Shortening the exposure of the base editor in the genome by using RNP delivery could also potentially reduce the off-target effects [61].

The editing efficiency is strongly affected by the efficiency of the delivery method. It has been demonstrated that when CRISPR/Cas9 components are supplied by microinjection or performing selections, they exhibit effective cutting activity resulting in mutations in the target sequences. Therefore, increased efficiency in therapeutic genome editing can be achieved by robust delivery of CRISPR/Cas9 components to the target cells. Delivery of CRISPR/Cas9 components for ex vivo therapeutic approaches should be relatively simple, as a result of the many effective in vitro delivery techniques available. However, challenges arise when performing delivery for in vivo therapeutics. The most promising in vivo delivery is by viral vectors, such as the AAV vectors, which have various serotypes. Although AAVs can deliver CRISPR/Cas9 components efficiently, studies showing successful delivery of CRISPR/Cas9 components using these viral vectors have been limited to certain tissues/organs such as the liver, muscle, brain and eye [8, 62]. Further investigations are required to optimise delivery using this approach with other

tissues/organs, such as the heart and pancreas. The AAV vectors that have been approved for clinical use have a restricted packaging capacity of ~4.5 kb, while SpCas9 is ~4.2 kb. Therefore, there is limited space remaining to carry the other elements required, such as a gRNA expression cassette. This limitation could be tackled by using smaller Cas9 orthologues, such as SaCas9 [69]. Exploring other RNA-guided nuclease systems with smaller sizes, particularly those that are highly efficient and specific with relaxed PAMs would be beneficial to improve the delivery and efficiency of this technology for in vivo therapeutic genome editing. Investigations could also focus on identifying other viral vectors with larger packaging capacity that are non-immunogenic and can infect target tissues effectively. Alternative delivery modes should also be considered, such as using nanoparticles to deliver the mRNA or protein of interest. This alternative has the advantage of transient presence of the nuclease, thus reducing off-target effects that might result from constitutive expression of Cas9 [8, 62, 68].

6.2. Therapeutic genome editing strategies

An important application of CRISPR genome editing is to cure disease. For genetic diseases, correction of the disease-causing mutation to the WT sequence is the ideal approach. Genetic correction of point mutations or deletions can be achieved by DNA substitutions or insertions. For loss of function diseases, therapeutic transgenes can be inserted to a 'safe harbour' site to compensate for the loss of protein in a controllable fashion. This approach may be more difficult, as it involves the less frequent HDR process. Point mutations or small deletions in the exonic region can be corrected more efficiently by an exon skipping approach using a large intervening sequence deletion strategy to remove the exon containing the mutations. Despite being a more efficient method, the resulting protein loses an exon, which might compromise its function. Intervening sequence large

deletions can also be used to delete repeat expansions for the treatment of the nucleotide expansion disorders such as fragile X syndrome, spinocerebellar ataxia, Huntington's disease and Friedrich's ataxia, or for diseases caused by duplications [6-8]. The therapeutic approach may be simpler for treatments that requires the generation of indel mutations, such as knocking out *CCR5* for the treatment of HIV infection or the *PCSK9* gene to treat hypercholesterolemia [69, 138, 141-143]. Therapeutic approaches for the diseases caused by dominant gain of function alleles, such as Huntington's disease, can also utilise efficient NHEJ-mediated indel mutations to knock out the mutant allele while retaining the WT allele [8].

Chapter 3 presented techniques for successful deletion of the Y chromosome using centromere removal and chromosome shredding strategies, which can potentially be used for the therapy of aneuploidy diseases. This study is the first to show that genome editing technologies are capable of deleting an entire chromosome, and to date this is the largest deletion (~90 MB) ever reported using genome editing [42]. Moreover, this study resulted in the generation of the first XO mouse model via genome editing that could be used as a model for Turner syndrome. From this study, we now know that DSBs can induce chromosome loss, and that these DSBs may also be the cause of chromosomal losses occurring in humans [165].

These strategies could be applied for disease modelling and the treatment of diseases caused by aneuploidy, which are relatively frequent, occurring in about 1 in 300 live births. However, this approach requires further optimisation to determine the best way to delete a chromosome. It is hypothesised that shredding the entire chromosome, including the short arm, centromere and long arm, will be the ideal way to achieve efficient deletion of the full chromosome. Future studies need to study the chromosome deletion in non-dividing cells, since this may be more relevant for studies directed towards in vivo therapeutic

applications. The centromere removal strategy may not be suitable for non-dividing cells as removal of the centromere would result in re-joining between the short and long arms and retention of the chromosome. In contrast, the chromosome shredding strategy may induce direct chromosome deletion in non-dividing cells. It will also be interesting to determine the ideal number of cuts and the cut positions that can induce chromosome deletion. It is possible that a small number of cuts in the ideal location could produce more efficient chromosome deletion than many cuts in non-ideal positions. This study showed that as few as eight cuts in the long arm of the Y chromosome were sufficient for inducing efficient chromosome deletion. The cutting sites of gRNA long arm 8X are located between positions ~4.3–17.3 MB of a ~90 MB Y chromosome. It may be interesting for future studies to examine whether the same number of cuts more widely dispersed such that they cover both the long and the short arms of the chromosome would result in more efficient deletion in non-dividing cells. Moreover, it is possible that efficient chromosome deletion could be achieved by targeting both telomeres for removal, which would destabilise the chromosome's integrity. Future studies could also address the question of whether other CRISPR systems, particularly those producing staggered cuts, such as Cpf1, could be more efficient for inducing chromosome deletions compared to the SpCas9 system, which produces blunt cuts. Additionally, it is possible that the chromosome fragments may translocate to other chromosome which can cause problems. Therefore, further studies need to characterise this possibility.

The mechanism of chromosome deletion via arm shredding is still unclear. It is puzzling how DSBs in the chromosome arms could trigger deletion of the remaining chromosome regions. Therefore, future studies also need to study the process and the factors involved in chromosome deletion. Understanding this mechanism could help revealing the best way to delete a chromosome efficiently. It is possible that inhibition of NHEJ factors could result

in increased chromosome deletion, while inhibition of DNA end resection factors may impede chromosome deletion.

One aneuploidy disease that could be prioritised for therapy using the chromosome deletion approach is Trisomy 21 Down syndrome, which occurs 1 in 700 livebirths. Down syndrome is the most common intellectual disability disorder, with many other devastating symptoms including congenital heart defects, early-onset Alzheimer's disease and haematopoietic disorders [166]. Some studies have addressed possible strategies to remove or silence a copy of Chromosome 21 in cells with trisomy 21. Trisomy correction in cells with trisomy 21 can happen naturally, but is rare in vitro. Performing selection by inserting a suicide gene, such as the *TK* gene, to the extra chromosome could simplify the identification of cells with euploid chromosomes. However, the selection process is complicated, and in vivo application is unlikely to be possible [167]. This in vitro natural aneuploidy correction could also be improved by overexpression of the ZSCAN4 protein [168]. Another treatment approach for Down syndrome is by methylation silencing of a copy of Chromosome 21 by inserting the *XIST* gene into that chromosome, although this is difficult for clinical application [169]. To date no genome editing studies have been able to correct trisomy 21. The chromosome deletion strategies presented in this thesis have opened an avenue to correct trisomy 21 by deleting one copy of the chromosome. However, these approaches will require extensive further study to bring them to real application, since there are a number of issues that need to be resolved. The first of these is how one copy of a chromosome can be specifically deleted without targeting the remaining two copies of Chromosome 21. One solution could be to scan for sequence variations in all copies of this chromosome to look for unique target sequences belong to only one copy of the chromosome. Furthermore, unlike the mouse or human Y chromosome, which contains abundant repetitive sequences, Chromosome 21 may not contain as many repetitive sequences, thus limiting the strategy of using a single gRNA for chromosome shredding.

Although a set of unique gRNAs can be used to facilitate chromosome deletion, targeting repetitive sequences might minimise the number of gRNAs required, which would simplify their delivery and increase the efficiency of this approach. Therefore, bioinformatic analysis is crucial to help identify repetitive target sequences. The other issues that need to be addressed are that assuming a copy of Chromosome 21 could successfully be eliminated, how could this be used in clinical applications? What cells/tissues would need to be targeted and how could the CRISPR/Cas9 components be delivered to the target cells/tissues? Ex vivo trisomy correction of haematopoietic cells and autologous transplantation of the corrected cell to the patients might rescue or prevent the haematopoietic difficulties faced by Down syndrome patients. Studies have shown that CRISPR/Cas9 component could be delivered to neuronal cells in the adult brain by local injection of CRISPR/Cas9 AAVs or RNP in vivo [67, 69, 70]. Combining this strategy with the chromosome deletion techniques explored here could potentially correct trisomy in neuronal cells in the brain, which might rescue aspects of the brain-related abnormal phenotype, such as intellectual disability. Overall, extensive consideration and studies are necessary to develop this genome editing technology for the therapy of Down syndrome and other diseases caused by aneuploidy.

6.3. Concluding remarks

Overall, the studies in this thesis have contributed to the development and application of the CRISPR/Cas9 genome editing platform. Specifically, this thesis describes the application of CRISPR/Cas9 genome editing for performing gene swaps in mice to study the functional equivalence between two closely related proteins. Strategies were developed that allow the deletion of an entire chromosome, which have the potential to model or treat diseases resulting from aneuploidy. Vectors were also generated to help other researchers

to perform experiments requiring simultaneous dual gRNA expression. Finally, in a result that is important for other researchers to be aware of, unexpected DNA repair outcomes after Cas9 breaks were demonstrated, which may lead false interpretations of edited genotypes.

The current development of CRISPR/Cas applications is still the tip of the iceberg. The future of CRISPR genome editing will be very bright, with the rapid development of CRISPR systems such as fusions of dCas9 or other catalytically dead RNA-guided endonucleases with various enzymatic domains to create tools for various functions. Engineered variants of CRISPR systems that recognise different PAM sequences for a broad range of targetable regions will also emerge, and many more useful vectors will be available to help researchers performing genome editing experiments. We will also see many more different applications conducted using CRISPR/Cas technology. This technology is predicted to uncover the unprecedented functions of gene(s), particularly with the use of CRISPR/Cas9 genome-wide screening. The endonuclease activity of Cas9 could be utilised for routine diagnostic applications, such as RFLP and Southern blots. CRISPR/Cas9-based gene drive will potentially be used to eradicate not only malaria- and dengue-carrying mosquitoes but also pest rodents. The ability of CRISPR/Cas9 to easily induce DSB will also help scientists to better understand the DNA repair mechanism that will eventually uncover the most efficient way to hijack the DNA repair machinery to generate predictable repair outcomes. Ultimately, further development of CRISPR/Cas technology, together with the advance of gene and stem cell therapy technology, will one day enable CRISPR technology to be widely used in clinical setting for disease therapeutics.

References

1. Sander, J.D. and J.K. Joung, *CRISPR-Cas systems for editing, regulating and targeting genomes.* Nat Biotechnol, 2014. **32**(4): p. 347-55.
2. Barrangou, R. and J.A. Doudna, *Applications of CRISPR technologies in research and beyond.* Nat Biotechnol, 2016. **34**(9): p. 933-941.
3. Komor, A.C., A.H. Badran, and D.R. Liu, *CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes.* Cell, 2017. **168**(1-2): p. 20-36.
4. Hsu, P.D., E.S. Lander, and F. Zhang, *Development and applications of CRISPR-Cas9 for genome engineering.* Cell, 2014. **157**(6): p. 1262-78.
5. Mali, P., K.M. Esvelt, and G.M. Church, *Cas9 as a versatile tool for engineering biology.* Nat Methods, 2013. **10**(10): p. 957-63.
6. Porteus, M.H., *Towards a new era in medicine: therapeutic genome editing.* Genome Biol, 2015. **16**: p. 286.
7. Maeder, M.L. and C.A. Gersbach, *Genome-editing Technologies for Gene and Cell Therapy.* Mol Ther, 2016. **24**(3): p. 430-46.
8. Cox, D.B., R.J. Platt, and F. Zhang, *Therapeutic genome editing: prospects and challenges.* Nat Med, 2015. **21**(2): p. 121-31.
9. Cohen, J., *Mice made easy.* Science, 2016. **354**(6312): p. 538-542.
10. Ishino, Y., et al., *Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product.* J Bacteriol, 1987. **169**(12): p. 5429-33.
11. Mojica, F.J., et al., *Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria.* Mol Microbiol, 2000. **36**(1): p. 244-6.
12. Jansen, R., et al., *Identification of genes that are associated with DNA repeats in prokaryotes.* Mol Microbiol, 2002. **43**(6): p. 1565-75.
13. Mojica, F.J., et al., *Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements.* J Mol Evol, 2005. **60**(2): p. 174-82.
14. Bolotin, A., et al., *Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin.* Microbiology, 2005. **151**(8): p. 2551-2561.
15. Horvath, P. and R. Barrangou, *CRISPR/Cas, the immune system of bacteria and archaea.* Science, 2010. **327**(5962): p. 167-70.
16. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-12.
17. Brouns, S.J., et al., *Small CRISPR RNAs guide antiviral defense in prokaryotes.* Science, 2008. **321**(5891): p. 960-4.
18. Marraffini, L.A. and E.J. Sontheimer, *CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA.* Science, 2008. **322**(5909): p. 1843-5.
19. Marraffini, L.A. and E.J. Sontheimer, *Self versus non-self discrimination during CRISPR RNA-directed immunity.* Nature, 2010. **463**(7280): p. 568-71.
20. Deveau, H., et al., *Phage response to CRISPR-encoded resistance in Streptococcus thermophilus.* J Bacteriol, 2008. **190**(4): p. 1390-400.
21. Garneau, J.E., et al., *The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA.* Nature, 2010. **468**(7320): p. 67-71.
22. Deltcheva, E., et al., *CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.* Nature, 2011. **471**(7340): p. 602-7.
23. Sternberg, S.H., et al., *DNA interrogation by the CRISPR RNA-guided endonuclease Cas9.* Nature, 2014. **507**(7490): p. 62-7.
24. Nishimasu, H., et al., *Crystal structure of Cas9 in complex with guide RNA and target DNA.* Cell, 2014. **156**(5): p. 935-49.
25. Jinek, M., et al., *Structures of Cas9 endonucleases reveal RNA-mediated conformational activation.* Science, 2014. **343**(6176): p. 1247997.
26. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-21.

27.   Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-23.

28.   Mali, P., et al., *CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering.* Nat Biotechnol, 2013. **31**(9): p. 833-8.

29.   Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system.* Nat Protoc, 2013. **8**(11): p. 2281-308.

30.   Liang, P., et al., *CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes.* Protein Cell, 2015. **6**(5): p. 363-72.

31.   Tang, L., et al., *CRISPR/Cas9-mediated gene editing in human zygotes using Cas9 protein.* Mol Genet Genomics, 2017.

32.   Ma, H., et al., *Correction of a pathogenic gene mutation in human embryos.* Nature, 2017. **548**(7668): p. 413-419.

33.   Guirouilh-Barbat, J., et al., *Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells.* Mol Cell, 2004. **14**(5): p. 611-23.

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

35.   Wang, H., et al., *One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering.* Cell, 2013. **153**(4): p. 910-8.

36.   Bell, C.C., et al., *A high-throughput screening strategy for detecting CRISPR-Cas9 induced mutations using next-generation sequencing.* BMC Genomics, 2014. **15**: p. 1002.

37.   Deriano, L. and D.B. Roth, *Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage.* Annu Rev Genet, 2013. **47**: p. 433-55.

38.   Zhang, L., et al., *Large genomic fragment deletions and insertions in mouse using CRISPR/Cas9.* PLoS One, 2015. **10**(3): p. e0120396.

39.   Fujii, W., et al., *Efficient generation of large-scale genome-modified mice using gRNA and CAS9 endonuclease.* Nucleic Acids Res, 2013. **41**(20): p. e187.

40.   Xiao, A., et al., *Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish.* Nucleic Acids Res, 2013. **41**(14): p. e141.

41.   Canver, M.C., et al., *Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells.* J Biol Chem, 2014. **289**(31): p. 21312-24.

42.   Essletzbichler, P., et al., *Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line.* Genome Res, 2014. **24**(12): p. 2059-65.

43.   Choi, P.S. and M. Meyerson, *Targeted genomic rearrangements using CRISPR/Cas technology.* Nat Commun, 2014. **5**: p. 3728.

44.   Maddalo, D., et al., *In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system.* Nature, 2014. **516**(7531): p. 423-7.

45.   Blasco, R.B., et al., *Simple and rapid in vivo generation of chromosomal rearrangements using CRISPR/Cas9 technology.* Cell Rep, 2014. **9**(4): p. 1219-27.

46.   Jiang, J., et al., *Induction of site-specific chromosomal translocations in embryonic stem cells by CRISPR/Cas9.* Sci Rep, 2016. **6**: p. 21918.

47.   van Overbeek, M., et al., *DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks.* Mol Cell, 2016. **63**(4): p. 633-46.

48.   Sfeir, A. and L.S. Symington, *Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway?* Trends Biochem Sci, 2015. **40**(11): p. 701-14.

49.   Truong, L.N., et al., *Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells.* Proc Natl Acad Sci U S A, 2013. **110**(19): p. 7720-5.

50.   Paix, A., et al., *Scalable and versatile genome editing using linear DNAs with microhomology to Cas9 Sites in Caenorhabditis elegans.* Genetics, 2014. **198**(4): p. 1347-56.

51.   Yang, H., et al., *One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering.* Cell, 2013. **154**(6): p. 1370-9.

52.   Storici, F., et al., *Conservative repair of a chromosomal double-strand break by single-strand DNA through two steps of annealing.* Mol Cell Biol, 2006. **26**(20): p. 7645-57.

53.   Jasin, M. and R. Rothstein, *Repair of strand breaks by homologous recombination.* Cold Spring Harb Perspect Biol, 2013. **5**(11): p. a012740.

54.   Yang, H., H. Wang, and R. Jaenisch, *Generating genetically modified mice using CRISPR/Cas-mediated genome engineering.* Nat Protoc, 2014. **9**(8): p. 1956-68.

55.   Wu, Y., et al., *Correction of a genetic disease in mouse via use of CRISPR-Cas9.* Cell Stem Cell, 2013. **13**(6): p. 659-62.

56.   Miura, H., et al., *CRISPR/Cas9-based generation of knockdown mice by intronic insertion of artificial microRNA using longer single-stranded DNA.* Sci Rep, 2015. **5**: p. 12799.

57.   Yoshimi, K., et al., *ssODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes.* Nat Commun, 2016. **7**: p. 10431.

58.   Sakuma, T., et al., *MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems.* Nat Protoc, 2016. **11**(1): p. 118-33.

59.   Nakade, S., et al., *Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9.* Nat Commun, 2014. **5**: p. 5560.

60.   Suzuki, K., et al., *In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration.* Nature, 2016. **540**(7631): p. 144-149.

61.   Kim, S., et al., *Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins.* Genome Res, 2014. **24**(6): p. 1012-9.

62.   Yin, H., K.J. Kauffman, and D.G. Anderson, *Delivery technologies for genome editing.* Nat Rev Drug Discov, 2017.

63.   Takahashi, G., et al., *GONAD: Genome-editing via Oviductal Nucleic Acids Delivery system: a novel microinjection independent genome engineering method in mice.* Sci Rep, 2015. **5**: p. 11406.

64.   Hashimoto, M. and T. Takemoto, *Electroporation enables the efficient mRNA delivery into the mouse zygotes and facilitates CRISPR/Cas9-based genome editing.* Sci Rep, 2015. **5**: p. 11315.

65.   Chen, S., et al., *Highly Efficient Mouse Genome Editing by CRISPR Ribonucleoprotein Electroporation of Zygotes.* J Biol Chem, 2016. **291**(28): p. 14457-67.

66.   Zuckermann, M., et al., *Somatic CRISPR/Cas9-mediated tumour suppressor disruption enables versatile brain tumour modelling.* Nat Commun, 2015. **6**: p. 7391.

67.   Swiech, L., et al., *In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9.* Nat Biotechnol, 2015. **33**(1): p. 102-6.

68.   Jiang, C., et al., *A non-viral CRISPR/Cas9 delivery system for therapeutically targeting HBV DNA and pcsk9 in vivo.* Cell Res, 2017. **27**(3): p. 440-443.

69.   Ran, F.A., et al., *In vivo genome editing using Staphylococcus aureus Cas9.* Nature, 2015. **520**(7546): p. 186-91.

70.   Staahl, B.T., et al., *Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes.* Nat Biotechnol, 2017. **35**(5): p. 431-434.

71.   Yin, H., et al., *Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype.* Nat Biotechnol, 2014. **32**(6): p. 551-3.

72.   Doench, J.G., et al., *Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9.* Nat Biotechnol, 2016. **34**(2): p. 184-91.

73.   Doench, J.G., et al., *Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation.* Nat Biotechnol, 2014. **32**(12): p. 1262-7.

74.   Maruyama, T., et al., *Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining.* Nat Biotechnol, 2015. **33**(5): p. 538-42.

75.   Chu, V.T., et al., *Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells.* Nat Biotechnol, 2015. **33**(5): p. 543-8.

76.   Robert, F., et al., *Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing.* Genome Med, 2015. **7**: p. 93.

77.   Yu, C., et al., *Small molecules enhance CRISPR genome editing in pluripotent stem cells.* Cell Stem Cell, 2015. **16**(2): p. 142-7.

78.   Renaud, J.B., et al., *Improved Genome Editing Efficiency and Flexibility Using Modified Oligonucleotides with TALEN and CRISPR-Cas9 Nucleases.* Cell Rep, 2016. **14**(9): p. 2263-72.

79. Lee, K., et al., *Synthetically modified guide RNA and donor DNA are a versatile platform for CRISPR-Cas9 engineering.* Elife, 2017. **6**.

80. Hsu, P.D., et al., *DNA targeting specificity of RNA-guided Cas9 nucleases.* Nat Biotechnol, 2013. **31**(9): p. 827-32.

81. Fu, Y., et al., *High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells.* Nat Biotechnol, 2013. **31**(9): p. 822-6.

82. Tsai, S.Q., et al., *GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases.* Nat Biotechnol, 2015. **33**(2): p. 187-97.

83. Frock, R.L., et al., *Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases.* Nat Biotechnol, 2015. **33**(2): p. 179-86.

84. Wang, X., et al., *Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors.* Nat Biotechnol, 2015. **33**(2): p. 175-8.

85. Kim, D., et al., *Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells.* Nat Methods, 2015. **12**(3): p. 237-43, 1 p following 243.

86. Tsai, S.Q., et al., *CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets.* Nat Methods, 2017.

87. Haeussler, M., et al., *Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR.* Genome Biol, 2016. **17**(1): p. 148.

88. Lin, S., et al., *Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery.* Elife, 2014. **3**: p. e04766.

89. Aubrey, B.J., et al., *An inducible lentiviral guide RNA platform enables the identification of tumor-essential genes and tumor-promoting mutations in vivo.* Cell Rep, 2015. **10**(8): p. 1422-32.

90. Chen, Y., et al., *Engineering Human Stem Cell Lines with Inducible Gene Knockout using CRISPR/Cas9.* Cell Stem Cell, 2015. **17**(2): p. 233-44.

91. Dow, L.E., et al., *Inducible in vivo genome editing with CRISPR-Cas9.* Nat Biotechnol, 2015. **33**(4): p. 390-4.

92. Fu, Y., et al., *Improving CRISPR-Cas nuclease specificity using truncated guide RNAs.* Nat Biotechnol, 2014. **32**(3): p. 279-84.

93. Cho, S.W., et al., *Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases.* Genome Res, 2014. **24**(1): p. 132-41.

94. Richardson, C.D., et al., *Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA.* Nat Biotechnol, 2016. **34**(3): p. 339-44.

95. Ran, F.A., et al., *Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity.* Cell, 2013. **154**(6): p. 1380-9.

96. Tsai, S.Q., et al., *Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing.* Nat Biotechnol, 2014. **32**(6): p. 569-76.

97. Guilinger, J.P., D.B. Thompson, and D.R. Liu, *Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification.* Nat Biotechnol, 2014. **32**(6): p. 577-82.

98. Kleinstiver, B.P., et al., *Engineered CRISPR-Cas9 nucleases with altered PAM specificities.* Nature, 2015. **523**(7561): p. 481-5.

99. Slaymaker, I.M., et al., *Rationally engineered Cas9 nucleases with improved specificity.* Science, 2016. **351**(6268): p. 84-8.

100. Zetsche, B., et al., *Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system.* Cell, 2015. **163**(3): p. 759-71.

101. Kim, Y., et al., *Generation of knockout mice by Cpf1-mediated gene targeting.* Nat Biotechnol, 2016. **34**(8): p. 808-10.

102. Watkins-Chow, D.E., et al., *Highly Efficient Cpf1-Mediated Gene Targeting in Mice Following High Concentration Pronuclear Injection.* G3 (Bethesda), 2017. **7**(2): p. 719-722.

103. Fonfara, I., et al., *The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA.* Nature, 2016. **532**(7600): p. 517-21.

104. Zetsche, B., et al., *Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array.* Nat Biotechnol, 2017. **35**(1): p. 31-34.

105.   Fagerlund, R.D., R.H. Staals, and P.C. Fineran, *The Cpf1 CRISPR-Cas protein expands genome-editing tools.* Genome Biol, 2015. **16**: p. 251.

106.   Kleinstiver, B.P., et al., *Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells.* Nat Biotechnol, 2016. **34**(8): p. 869-74.

107.   Kim, D., et al., *Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells.* Nat Biotechnol, 2016. **34**(8): p. 863-8.

108.   Hirano, H., et al., *Structure and Engineering of Francisella novicida Cas9.* Cell, 2016. **164**(5): p. 950-61.

109.   Kleinstiver, B.P., et al., *Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition.* Nat Biotechnol, 2015. **33**(12): p. 1293-1298.

110.   Gao, F., et al., *DNA-guided genome editing using the Natronobacterium gregoryi Argonaute.* Nat Biotechnol, 2016. **34**(7): p. 768-73.

111.   Lee, S.H., et al., *Failure to detect DNA-guided genome editing using Natronobacterium gregoryi Argonaute.* Nat Biotechnol, 2016. **35**(1): p. 17-18.

112.   Javidi-Parsijani, P., et al., *No evidence of genome editing activity from Natronobacterium gregoryi Argonaute (NgAgo) in human cells.* PLoS One, 2017. **12**(5): p. e0177444.

113.   Konermann, S., et al., *Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex.* Nature, 2015. **517**(7536): p. 583-8.

114.   Maeder, M.L., et al., *CRISPR RNA-guided activation of endogenous human genes.* Nat Methods, 2013. **10**(10): p. 977-9.

115.   Chavez, A., et al., *Highly efficient Cas9-mediated transcriptional programming.* Nat Methods, 2015. **12**(4): p. 326-8.

116.   Perez-Pinera, P., et al., *RNA-guided gene activation by CRISPR-Cas9-based transcription factors.* Nat Methods, 2013. **10**(10): p. 973-6.

117.   Gilbert, L.A., et al., *CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes.* Cell, 2013. **154**(2): p. 442-51.

118.   Mandegar, M.A., et al., *CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs.* Cell Stem Cell, 2016. **18**(4): p. 541-53.

119.   McDonald, J.I., et al., *Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation.* Biol Open, 2016. **5**(6): p. 866-74.

120.   Liu, X.S., et al., *Editing DNA Methylation in the Mammalian Genome.* Cell, 2016. **167**(1): p. 233-247 e17.

121.   Vojta, A., et al., *Repurposing the CRISPR-Cas9 system for targeted DNA methylation.* Nucleic Acids Res, 2016. **44**(12): p. 5615-28.

122.   Xu, X., et al., *A CRISPR-based approach for targeted DNA demethylation.* Cell Discov, 2016. **2**: p. 16009.

123.   Morita, S., et al., *Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions.* Nat Biotechnol, 2016. **34**(10): p. 1060-1065.

124.   Hilton, I.B., et al., *Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers.* Nat Biotechnol, 2015. **33**(5): p. 510-7.

125.   Komor, A.C., et al., *Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage.* Nature, 2016. **533**(7603): p. 420-4.

126.   Kim, K., et al., *Highly efficient RNA-guided base editing in mouse embryos.* Nat Biotechnol, 2017. **35**(5): p. 435-437.

127.   Chen, B., et al., *Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system.* Cell, 2013. **155**(7): p. 1479-91.

128.   Ma, H., et al., *Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow.* Nat Biotechnol, 2016. **34**(5): p. 528-30.

129.   Mali, P., et al., *RNA-guided human genome engineering via Cas9.* Science, 2013. **339**(6121): p. 823-6.

130.   Moreno, A.M. and P. Mali, *Therapeutic genome engineering via CRISPR-Cas systems.* Wiley Interdiscip Rev Syst Biol Med, 2017.

131.   Long, C., et al., *Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA.* Science, 2014. **345**(6201): p. 1184-8.

132. Nelson, C.E., et al., *In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy.* Science, 2016. **351**(6271): p. 403-7.

133. Tabebordbar, M., et al., *In vivo gene editing in dystrophic mouse muscle and muscle stem cells.* Science, 2016. **351**(6271): p. 407-11.

134. Long, C., et al., *Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy.* Science, 2016. **351**(6271): p. 400-3.

135. Yang, Y., et al., *A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice.* Nat Biotechnol, 2016. **34**(3): p. 334-8.

136. Yin, H., et al., *Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components in vivo.* Nat Biotechnol, 2016. **34**(3): p. 328-33.

137. Savic, N. and G. Schwank, *Advances in therapeutic CRISPR/Cas9 genome editing.* Transl Res, 2016. **168**: p. 15-21.

138. Ding, Q., et al., *Permanent alteration of PCSK9 with in vivo CRISPR-Cas9 genome editing.* Circ Res, 2014. **115**(5): p. 488-92.

139. DeWitt, M.A., et al., *Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells.* Sci Transl Med, 2016. **8**(360): p. 360ra134.

140. Dever, D.P., et al., *CRISPR/Cas9 beta-globin gene targeting in human haematopoietic stem cells.* Nature, 2016. **539**(7629): p. 384-389.

141. Ye, L., et al., *Seamless modification of wild-type induced pluripotent stem cells to the natural CCR5Delta32 mutation confers resistance to HIV infection.* Proc Natl Acad Sci U S A, 2014. **111**(26): p. 9591-6.

142. Mandal, P.K., et al., *Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9.* Cell Stem Cell, 2014. **15**(5): p. 643-52.

143. Xu, L., et al., *CRISPR/Cas9-Mediated CCR5 Ablation in Human Hematopoietic Stem/Progenitor Cells Confers HIV-1 Resistance In Vivo.* Mol Ther, 2017.

144. Tebas, P., et al., *Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV.* N Engl J Med, 2014. **370**(10): p. 901-10.

145. Kaminski, R., et al., *Elimination of HIV-1 Genomes from Human T-lymphoid Cells by CRISPR/Cas9 Gene Editing.* Sci Rep, 2016. **6**: p. 22555.

146. Liao, H.K., et al., *Use of the CRISPR/Cas9 system as an intracellular defense against HIV-1 infection in human cells.* Nat Commun, 2015. **6**: p. 6413.

147. Yin, C., et al., *In Vivo Excision of HIV-1 Provirus by saCas9 and Multiplex Single-Guide RNAs in Animal Models.* Mol Ther, 2017. **25**(5): p. 1168-1186.

148. Cyranoski, D., *CRISPR gene-editing tested in a person for the first time.* Nature, 2016. **539**(7630): p. 479.

149. Chen, Z.H., et al., *Targeting genomic rearrangements in tumor cells through Cas9-mediated insertion of a suicide gene.* Nat Biotechnol, 2017.

150. Bikard, D., et al., *Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials.* Nat Biotechnol, 2014. **32**(11): p. 1146-50.

151. Citorik, R.J., M. Mimee, and T.K. Lu, *Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases.* Nat Biotechnol, 2014. **32**(11): p. 1141-5.

152. Yosef, I., et al., *Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria.* Proc Natl Acad Sci U S A, 2015. **112**(23): p. 7267-72.

153. Shalem, O., et al., *Genome-scale CRISPR-Cas9 knockout screening in human cells.* Science, 2014. **343**(6166): p. 84-7.

154. Wang, T., et al., *Genetic screens in human cells using the CRISPR-Cas9 system.* Science, 2014. **343**(6166): p. 80-4.

155. Sanjana, N.E., O. Shalem, and F. Zhang, *Improved vectors and genome-wide libraries for CRISPR screening.* Nat Methods, 2014. **11**(8): p. 783-4.

156. Koike-Yusa, H., et al., *Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library.* Nat Biotechnol, 2014. **32**(3): p. 267-73.

157. Chen, S., et al., *Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis.* Cell, 2015. **160**(6): p. 1246-60.

158. Zhu, S., et al., *Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library.* Nat Biotechnol, 2016. **34**(12): p. 1279-1286.

159. Shen, J.P., et al., *Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions.* Nat Methods, 2017.

160. Han, K., et al., *Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions.* Nat Biotechnol, 2017. **35**(5): p. 463-474.

161. Gantz, V.M. and E. Bier, *Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations.* Science, 2015. **348**(6233): p. 442-4.

162. Hammond, A., et al., *A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae.* Nat Biotechnol, 2016. **34**(1): p. 78-83.

163. Gantz, V.M., et al., *Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi.* Proc Natl Acad Sci U S A, 2015. **112**(49): p. E6736-43.

164. Yao, X., et al., *CRISPR/Cas9 - Mediated Precise Targeted Integration In Vivo Using a Double Cut Donor with Short Homology Arms.* EBioMedicine, 2017.

165. Dumanski, J.P., et al., *Smoking is associated with mosaic loss of chromosome Y.* Science, 2015. **347**(6217): p. 81-3.

166. Roper, R.J. and R.H. Reeves, *Understanding the basis for Down syndrome phenotypes.* PLoS Genet, 2006. **2**(3): p. e50.

167. Li, L.B., et al., *Trisomy correction in Down syndrome induced pluripotent stem cells.* Cell Stem Cell, 2012. **11**(5): p. 615-9.

168. Amano, T., et al., *Correction of Down syndrome and Edwards syndrome aneuploidies in human cell cultures.* DNA Res, 2015. **22**(5): p. 331-42.

169. Jiang, J., et al., *Translating dosage compensation to trisomy 21.* Nature, 2013. **500**(7462): p. 296-300.