Technology Development in Mouse Genetics and Epigenetics

By

Chikdu Shakti Shivalila

B.S. Biology
The University of Pittsburgh, 2009

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2016

© 2016 Massachusetts Institute of Technology
All rights reserved

Signature of author: Signature redacted
Chikdu Shakti Shivalila
Department of Biology
January 8, 2016

Certified by: Signature redacted
Rudolf Jaenisch
Professor of Biology
Founding Member, Whitehead Institute
Thesis Supervisor

Accepted by: Signature redacted
Michael T. Hemann
Associate Professor of Biology
Co-Chair, Biology Graduate Committee
Technology Development in Mouse Genetics and Epigenetics

By

Chikdu Shakti Shivalila

Submitted to the Department of Biology
In partial fulfillment of the requirements for the degree of Doctor of Philosophy

Abstract

The importance and significance of a model organism in biological research cannot be overstated. The mouse in particular has been very useful in understanding questions in many areas of research such as developmental biology, cancer biology, neuroscience and genetics. However, even though the methods to make transgenic mice and gene knockin and knockouts have been successful, they are very inefficient, labor intensive and costly. Therefore, in this thesis we developed a novel methodology to rapidly and efficiently modify the mouse genome. Using CRISPR/Cas9, a novel genome-engineering technology developed from bacteria, we were able to genetically modify mouse embryonic stem cells and make mice that carried genetic modification by zygotic injections. Using CRISPR/Cas9 we were able to make mice in as little as three weeks that contained multiple gene knockouts, single nucleotide modifications, GFP and mCherry reporter alleles, epitope-tagged alleles, and conditional alleles.

Another interesting area of research in mouse genetics is epigenetic regulation, specifically how DNA methylation regulates development, gene expression, and cell state. Multiple studies have shown that this epigenetic modification plays an important regulatory role in these processes; however, the technology that has existed so far to investigate DNA methylation has only been able to look at snapshots of methylation patterns in fixed cell populations. In this thesis we have developed a novel technology named Reporter of Genomic Methylation (RGM), which allows for the investigation of methylation dynamics at single cell-resolution in vivo. The RGM technology was developed using a minimal synthetic secondary DMR promoter that drives the expression of a florescent protein. Using CRISPR/Cas9 the RGM reporter can be integrated into any genomic locus where it can report on the methylation state of its surroundings. We further show that the RGM reporter activity reflects the methylation state of non-coding regulatory elements such as promoters and enhancers. Furthermore, we show that the RGM technology allows for the dynamics of methylation and demethylation to be observed at these non-coding loci as cells transition between a pluripotent and differentiated state.

Thesis Supervisor: Rudolf Jaenisch
Title: Professor of Biology and Member of the Whitehead
Dedication

I dedicate this thesis to my family.
Acknowledgments

Nothing that is done is done alone. I think that science, like all creative pursuits, is not something that stands alone in isolation, but is rather the result of a temporal and spatial dynamic network of people that share a common interest. For this reason, there are to many people to thank, but I will try my best.

MIT (2011-2015)

First of all, I would like to thank my mentor Rudolf Jaenisch for giving me the opportunity to work in his laboratory for my PhD studies. Rudolf thank you for always giving me the freedom to research what I was interested in, for always giving me insightful advice on my research projects, and for always being patient with me. Rudolf you inspire in me the scientist I hope to one day become, your knowledge on so many vast areas of biology, your perpetual curiosity, your patience and understanding, your firm foundations in good experimentation, and your fortitude and flexibility in trying new things. I would also like to thank you for assembling a wonderful lab of great people, and for letting me be apart of it for my time at MIT. It has been an honor to work in your lab, and I will always have so much gratitude for the opportunities that you have given me.

I would also like to thank some amazing people I have had the chance to work with in the Jaenisch lab. I would like to thank Haoyi Wang, who when I first joined the Jaenisch lab, spent a lot of time mentoring me in basic experimental procedures, such as ES cell culture and TALEN genome engineering. As I adapted to become more of a self sufficient scientist we continued to work together. Thank you Haoyi for being an awesome mentor, friend, and colleague. I would also like to thank Hui Yang and Albert Wu Cheng for being good friends and also great collaborators. I would also like to thank Yonathan Stelzer for being a great friend and colleague, it has been really fun working with you. I have tremendously enjoyed all our talks about science and non-science related stuff. I would like to thank Frank Soldner, for being a good friend, a cynical but thought provoking science critique, and great person to go out for coffee and talk to about whatever. I would also like to thank Meelad Dawlaty, for being a good friend and for all the late night lab conversations. And also thanks to everyone else who has been in the Jaenisch for the last few years.

I would also like to thank my committee, Dr. David Bartel and Dr. Piyush Gupta. I really appreciate the time you have taken to give me advise on my research and for all the helpful feedback through my PhD studies. In addition, I really enjoyed your lectures that I attended during my first year here at MIT. Dr. Piyush you made computational biology interesting and fun. Dr. Bartel, I learned so much about RNA in your class, and that was the first time I ever started to think about CRISPR/Cas.

University of Pittsburgh (2007-2011)

I would like to thank Dr. Vernon Twombly, your lectures, lab courses and after class discussions had a big impact on me when I was deciding what to do after undergrad. In
addition, I am really thankful for your help in finding me an undergrad laboratory where I could do research while at the University of Pittsburgh. If it was not for your mentorship I probable would not have had the opportunities that I have had so far.

I would also like to thank Dr. Maria Teresa Saenz Robles. Dr. Robles I can’t even express how much gratitude I have for your mentorship. I learned so much from you, both how to pursue a scientific question and the technical skill to do so. You were always so patient and kind while teaching me all kinds of things. You were such a joy to work for. I will always keep the great memories of working with you. And I will always be really thankful for the opportunities you gave me.

I would also like to Thank Dr. James Pipas. Dr. Pipas, I appreciate so much the opportunity you gave me to work in your laboratory at the University of Pittsburgh. Working in your laboratory made me want to become a scientist, to do research and to discover new things. I think your excitement and fascination with biology was contagious and I got the bug.... Or virus. I will always have so much gratitude for everything you did for me.

University at Albany (2005-2007)

I would like to thank Keith Derbyshire for giving me my first opportunity to work in a biology lab

Adirondack community college (2004-2005)

I would like to thank professors Dave Hodgson (botany) and Joseph Eagan (zoology) for first introducing me to biology. Your lectures were amazing.

My family (1987-present)

I would like to thank my whole family, who I am today is because of every single one of you. Thank you to my Mom, Kamala, my Dad, Uli, my aunt Jammu, my uncle Nigel. My sisters: Roshi, Sodasi, Bumi and my Brothers: Osel, Pele, Hawazin, Ituri, Naya.

To my mom: Thank you for always being so kind, caring and loving. And thank you for raising me, teaching me, and always encouraging me to become whoever I wanted to become.

To my Dad: Thank you for teaching me to have independent thought, for all of the discussions about life, science, society etc.... Thank you for always being there when I needed you. Thank you for your patience, caring advice, and guidance.

To my brothers and sisters: Thank you all, we are always there for each other.

To my aunt Jammu and uncle Nigel: Thank you for all your support and kindness
Table of Contents

Abstract..................................................................................................................3
Acknowledgements..................................................................................................6
Table of Contents...................................................................................................9
Chapter 1. Introduction............................................................................................11
   Part 1. The mouse as a genetic model organism
      Early mouse development and pluripotent cells...........................................12
      Methods for generating genetically modified mice.......................................14
      Types of genetic modifications.......................................................................17
      Zygotic injections vs. Blastocyst injections...............................................20
   Part 2. The development of programmable nucleases for genome-engineering
      Zinc Finger Nucleases and TALENS.........................................................21
      CRISPR/Cas.................................................................................................23
      DNA repair machinery....................................................................................29
   Part 3. Epigenetics: DNA methylation
      DNA methylation............................................................................................32
      DNA methylation changes in mouse development.........................................35
      Methylation and gene regulation.....................................................................37
      DNA methylation and cancer..........................................................................38
      Technology available to study DNA methylation...........................................39
   Part 4. Thesis outline
      Locus-specific genome editing in the single-cell zygote...............................41
      A reporter for genomic methylation ..............................................................41

Chapter 2. One-step generation of mice carrying mutations in multiple genes
targeted by CRISPR-Cas mediated genome-engineering.....................................57

Chapter 3. One-step generation of mice carrying reporter and conditional
alleles by CRISPR/Cas-mediated genome-engineering......................................91

Chapter 4. Tracing dynamic changes of DNA methylation at single-cell
resolution................................................................................................................129

Chapter 5. Future directions
CRISPR/Cas9 genome-engineering in mice...........................................169
Further characterization and improvement of RGM.................................170
Application for RGM in imprinting, screening, and cancer..........................173
Curriculum vitae...................................................................................179
Chapter 1. Introduction

The mouse model has been an invaluable tool for understanding mammalian biology. The mouse shares 99% of its genes with humans, which has not only allowed the mouse to be used to understand basic mammalian biology in such research areas as neuroscience, epigenetics, and development, but also to model human diseases in vivo such as cancer and neurological disorders (Waterston 2002). One of the main reasons why the mouse has been used so extensively in biological research is because it has been well established as a genetic model organism (Paigen 2003).

Genetic research in mice has been facilitated by the development of methods and technology to manipulate the mouse genome. These methods consist of either transgene expression, mediated through random integration, or endogenous locus-specific genome editing, mediated through homologous recombination (Jaenisch et al., 1974; Smithies et al., 1985). However, the existing methods for mouse genome-engineering are still inefficient and can be improved by further technology development. Programmable nucleases, specifically CRISPR/Cas9, have the potential, when combined with existing methods and tools such as dsDNA-targeting vectors, ssDNA-oligos, and zygotic injection, to increase the efficiency of genome-engineering in the mouse model.

Epigenetics, specifically DNA methylation, has been shown to take part in many important processes in vertebrate biology such as: imprinting, gene regulation, development, and cancer (Holliday and Pugh, 1975; Riggs 1975). Multiple technologies have been developed to investigate DNA methylation. However, all of the current methods can only take a static snapshot of methylation patterns in a cell or tissue (Stelzer
and Jaenisch, 2015). Methods to Investigate DNA methylation is another area of research in mouse genetics that could be improved by further technology development. A technology that could report on the dynamics of locus-specific DNA methylation when combined with CRISPR/Cas9 genome engineering would allow for a better analysis of how this important epigenetic modification helps regulate non-coding regions such as promoters, enhancers, and imprinted regions in vivo.

Part 1. The mouse as a genetic model organism

Early mouse development and pluripotent cells

Upon fertilization the single-cell mouse zygote goes through a series divisions starting with the 2-and 4-cell-stage embryos where each cell is believed to be uniform in nature (Zernicka-Goetz et al., 2009). Upon division to the 8 and 16-cell stage, the embryo starts to specialize with cells in the center becoming the inner cell mass (ICM), and cells on the outside becoming the trophectoderm (TE). The compartmentalization of the ICM and TE is further completed in the 16-32-cell stage and early blastocyst (Zernicka-Goetz et al., 2009). Upon implantation of the late-stage blastocyst, the ICM will further differentiate to form the primitive endoderm (PE) and the epiblast (EPI). Further along the development of the embryo the EPI will form the three germ layers, the endoderm, mesoderm, and ectoderm. These three germ layers will eventually form all tissues in the adult organism (Zernicka-Goetz et al., 2009).

Mouse embryonic stem cells (mESCs) are designated as pluripotent cells that can give rise to the three germ layers, endoderm, ectoderm, and mesoderm. Therefore,
mESCs have the developmental potential to form every cell in the adult mouse but cannot give rise to the extra embryonic tissue (Jaenisch and Young, 2008). Dr. Martin Evans showed in 1981 that mouse pluripotent cells could be extracted from the ICM, and under the right culture conditions they could be propagated indefinitely in vitro (Martin et al., 1981). Further work showed that mESCs could contribute to form chimeric mice when injected into blastocyst-stage embryos, and that these cells could also contribute to the germline (Bradley et al., 1984). The maintenance of mESCs in vitro is facilitated by the extracellular signaling molecules LIF, Wnt, and activin/nodal which are thought to maintain the necessary microenvironment to keep mESCs in a pluripotent state (Smith et al., 1988; Ogawa et al., 2006). Furthermore, mESCs are characterized by the expression of a set of core transcription factors: Nanog, Sox2, and Oct4. These master transcription factors form an autoregulatory loop that maintains their own expression and represses other transcription factors necessary for differentiation (Boyer et al., 2005).

Although pluripotent mESCs are usually isolated from the ICM, work by Takahashi and Yamanaka showed that adult somatic cells could be reprogrammed into induced pluripotent stem cells (IPSCs) by viral transgenic expression of the transcription factors: Oct4, Sox2, c-Myc, and Klf4 (Takahashi et al., 2006). Further work showed that when the these IPSCs were selected for by endogenous Oct4 or Nanog expression they had an almost identical gene-expression profile compared with mESCs (Wernig et al., 2007). In addition, it was shown that when IPSCs are injected into blastocyst-stage embryos they contribute to form chimeric mice with chimerism occurring in both somatic and germline tissue (Okita et al., 2007, Maherali et al., 2007, Wernig et al., 2007). Since the initial work by Yamanaka, additional methods to induce reprogramming have been
established such as the use of doxycycline inducible polycistronic vectors, in vitro transcribed mRNAs of the reprogramming factors, and small-molecule inhibitors (Warren et al., 2010; Carey et al., 2009; Vidal et al., 2014).

**Methods for generating genetically modified mice**

The mouse was first adapted as a model organism for genetic studies in 1902 when Lucien Cuenot showed that coat color in mice followed Mendelian ratios (Cuenot 1902). However, the discipline of mouse genetics was not fully recognized until 1909 when C. C. Little established the first true inbred strains of mice so as to have reproducibility in genetic crosses (Paigen 2003). Early mouse genetics relied on spontaneous chemical- or radiation-induced mutations to investigate how specific genetic mutations caused phenotypic differences and abnormalities (Paigen 2003; Van der Weyden et al., 2011). This forward genetics approach was successful at isolating and cloning many important genes such as the c-kit tyrosine receptor \( W \) gene and its ligand *steel* (Chabot et al 1988; Brannan et al., 1991). However, the isolation of genes mutated by chemical mutagens such as ENU is cumbersome, and it was the development of transgenic technology that allowed for the isolation of genes for which when mutated affected development. In 1974 Rudolf Jaenisch and Beatrice Mintz developed the first transgenic mouse by injecting SV40 virus into mouse blastocyst-stage embryos, and then letting the blastocysts develop into pups by implanting them back into pseudo-pregnant female mice (Jaenisch et al., 1974). The integrated SV40 viral genome could be detected in multiple tissues from the adult mice that resulted from these injections. Jaenisch further showed that the Moloney Leukemia Virus (M-MuLV) could be transmitted to the
germline of mice when injected into pre-implantation stage embryos, and that the adult mice that developed from these embryos could pass on M-MuLV to their progeny (Jaenisch 1976). This early work was significant for two important reasons. First, it showed that foreign DNA could be integrated into the mouse genome. Second, this work showed that if a genetic modification occurred early enough in development then it could contribute to the germline and be propagated to the next generation.

Following on Dr. Jaenisch's work, multiple groups developed a system of generating transgenic mice by injecting plasmid DNA into the pro-nucleus of the mouse zygote (Gordon et al., 1980; Brinster et al., 1981; Costantini and Lacy, 1981; Harbers et al., 1981; Wagner et al., 1981). The first transgene to be successfully expressed in a transgenic mouse model was the viral TK gene (Gordon et al., 1980). Follow up work by Wagner showed that the complete rabbit Beta-globin gene could be integrated into the mouse genome, and that this transgene was expressed in a tissue specific manner (Wagner et al., 1981). This early work in mouse transgenics coincided with the development of plasmids that were able to express genes in mammalian cells. The synthesis of mammalian expression plasmids was facilitated by the isolation and cloning of mammalian promoters and polyA sequences that allowed for stable gene expression in mouse and human cells. (Doyle et al., 2012).

Furthermore, this early work in mouse transgenesis led to the development of other techniques that are still often used in mouse genetics. Lentiviruses were developed to express genes in mammalian cells and to create transgenic mice by infection of early blastocyst or zygote-stage embryos (Lois et al., 2002). Furthermore, gene-knockout mice were made by gene-trap experiments using lentivirus or other retroviral elements such as
the sleeping beauty or piggyback transposons (Perry et al., 1995; Luo et al., 1998; Dupuy et al., 2001). In addition, lentiviruses can be used to infect and express Cre-recombinase to knockout conditional alleles in adult tissues (DuPage et al., 2009).

Although transgenic mice proved to be very useful, this method relied on the random integration of a transgene. Random integration into the genome can lead to the disruption of a gene coding sequence, or can separate a gene from its endogenous cis-regulatory elements, which may result in epigenetic silencing. Furthermore, this method of making transgenic mice is not very efficient for knocking out an endogenous gene because the mutation cannot be targeted to a desired locus. This method rather relies on the isolation of knockout mutants by a forward genetics approach (Doyle et al., 2012).

The development of homologous recombination mediated genome editing by Capecchi and Smithies represents a major breakthrough as it allowed to predetermine the gene to be mutated. The method relies on targeting a specific genetic locus with an exogenous dsDNA-targeting vector that contains regions of homology to the locus of interest (Smithies et al., 1985; Thomas et al., 1986). The cell integrates the dsDNA-targeting vector, at a very low efficiency, into the designated locus through homologous recombination. The combination of homologous recombination and ES cell technology revolutionized our ability to generate mice with precise genetic modifications. To generate a mutant mESC clone, an exogenous dsDNA-targeting vector is introduced into the cells, and individual mESC colonies carrying the desired mutation are selected and verified initially by southern blot and later by PCR (Evans et al., 1981; Smithies et al., 1985; Thomas et al., 1986). Positive-selection antibiotic-resistant genes, such as puromycin, neomycin, and the negative-selection TK gene, greatly increase the isolation
of mESCs clones that contain the correct genetic modification. The antibiotic resistant cassette is added to the exogenous dsDNA-targeting vector and allows for selection against mESCs that did not become genetically modified (Doyle et al., 2012). After correctly genetically engineered mES cells have been isolated, they are injected into blastocyst-stage embryos to produce chimeric mice (Koller et al., 1989 Thompson et al., 1989). The mice are then mated to establish germline transmission of the modified ES cell clone.

This strategy of genetic engineering mESCs by homologous recombination has been the basis for further technical development in mouse genetics such as the generation of reporter alleles, specific gene knockouts, epitope-tagged alleles, and conditional alleles. However, one limitation of this method is that it is time consuming, requires multiple steps, and cannot be used in most other mammalian cell types where chimera competent ES cells have not been isolated. Furthermore, this method is very inefficient, with an estimated rate of recombination ranging from one in $1 \times 10^6$ to $1 \times 10^7$ (Kim and Kim, 204).

**Types of genetic modifications**

One of the first and most common genetic modifications made through homologous recombination in mESCs was locus-specific gene knockouts (Zijlstra et al., 1990; Donehower et al., 1992; Rudnicki et al., 1992). Knocking out a gene through homologous recombination can be accomplished through different mechanisms. The simplest way to cause a knockout is to introduce a stop codon or frameshift mutation into the coding region of a gene. This can be accomplished through the delivery of a dsDNA-
targeting vector that carries a stop codon or frameshift mutation, and which targets an exon of a gene (Sage et al., 2000). Another possible mechanism is to introduce a large fragment of DNA, such as a puromycin cassette, into the coding region of a gene, thereby terminating transcription (Xiong et al., 2012). Finally, a gene knockout can be made through deleting an entire exon(s) by having the dsDNA-targeting vector’s homology arms flank the region to be removed (Xiong et al., 2012). Because the efficiency of homologous recombination is so low, the sequential insertion of two different targeting vectors that express different resistant cassettes, such as hygromycin and neomycin, is required for deriving homozygous mutants by targeting both alleles. (Hudson et al., 1998)

Alternatively, heterozygous knockout mice can be made and then bred for homozygosity.

Single or multiple nucleotides can be exchanged through homologous recombination by using a dsDNA-targeting vector that contains the desired nucleotide differences (Wu et al., 1994). For example, this is done to change specific nucleotides that code for an amino acid that is important in the active site of an enzyme, or to introduce a mutation that activates an oncogene or deactivates a tumor suppressor. Because targeting efficiency is low, the targeting vector usually contains a selection cassette to allow for isolation of mESC colonies that have properly integrated the construct. The selection cassette can be flanked by two loxP sequences that allow for it to be removed after the mutation has been made. However, even with the removal of the selection cassette, this will not be a seamless mutation because the modified locus will still contain one loxP sequence.

Another important genetic modification is the introduction of a reporter allele into an endogenous gene to monitor the gene’s spatial and temporal expression (Bouabe et al.,
Reporter alleles can help address questions such as in what tissue is a specific gene expressed, and at what time during development does a specific gene turn on or become repressed. The most common reporter allele is the green fluorescent protein (GFP) isolated from the jellyfish *Aequorea victoria* or its many different synthetic variants that come in an array of fluorescent colors like red (mCherry), yellow (YFP), and blue (BFP) (Abe et al., 2013; Prasher et al., 1992, Srinivas et al., 2001; Tsien 1998). Reporter alleles can be introduced at the 5' or 3' terminus of a gene, so as to not disrupt its function, or they can be introduced in-frame of a coding exon, so as to both knockout the gene and report on its activity (Croxford et al., 2011). Alternatively, instead of a fluorescent reporter, an epitope-tag such as V5, FLAG, Myc, or Strep/Biotin can be inserted into the 5' or 3' coding region of an endogenous gene. Epitope-tags are short amino acid sequences that are recognized very strongly and specifically by their respective antibodies (Evan et al., 1985; Brizzard et al., Schmidt et al., 2007). Epitope-tags allow for the isolation, or relative quantification, of a protein when an antibody to that specific protein does not exist. Isolation or quantification of an epitope-tagged protein is important in many experimental protocols such as chromatin immunoprecipitation, mass spectrometry, ChIP-seq, western blot, biochemical assays, and immunofluorescence (Gavin et al., 2002; Ho et al., 2002, Kolodziej et al., 2009).

Sometimes it is impossible to make a knockout of a specific gene in a mouse because the mutation causes embryonic lethality. This prompted the development of conditional mutant alleles. This system takes advantage of the Cre-recombinase enzyme and its recognition sequence motif (loxP) isolated from the P1 bacteriophage (Sternberg et al 1981). In this system two loxP sequences are inserted on either side of an exon.
Upon expression of Cre-recombinase, the loxP-flanked (floxed) exon will recombine out resulting in a gene knockout (Dawlaty et al., 2011). Cre-recombinase can be expressed under a tissue specific promoter, so that the floxed-gene will only become deleted in the desired tissue (Orban et al., 1992). Furthermore, the Cre-recombination system can allow for cell lineage tracing. In this method, the expression of GFP is dependent on Cre-mediated recombination. When Cre is expressed under a tissue- or cell-specific promoter, only the progeny from cells that expressed Cre will be labeled by GFP (Mao et al., 2001).

**Zygotic injections vs. Blastocyst injections**

Mouse transgenics is a rapid method to genetically modify mice because direct zygotic injection of a DNA construct results in low chimerism and high germline contribution (Gordon et al., 1980). However, this method does not allow for locus-specific genetic modifications (Xiong et al., 2012). Homologous recombination mediated gene targeting in mES cells allows for locus-specific genome editing. However, modified mES cells are injected into blastocyst-stage embryos which results in high chimerism and low germline contribution. (Paigen 2003). Not all mES cell injections into blastocysts will result in chimeric mice, and the efficiency of germline contribution can vary depending on the mES cell line that is used (Guo et al., 2014). Furthermore, this method requires a long time to make genetically modified mice because of the multiple steps involved in the process. Ideally the most efficient way to genetically engineer a mouse would be locus-specific genome editing in the single-cell zygote-stage embryo because it would combine the expediency obtained by transgenics with the specificity obtained by
homologous recombination. The use of programmable nucleases has the potential to allow for locus-specific genome editing in the single-cell zygote.

Part 2. Programmable nucleases for gene targeting

Zinc Finger Nucleases and TALENS

Zinc Finger Nucleases (ZFNs) were the first programmable nucleases created for locus-specific genetic engineering. ZFNs are composed of two functional units, zinc finger proteins (ZFPs) and a FokI nuclease domain (Kim et al., 1996). ZFPs consist of a tandem array of C2H2 zinc fingers where each zinc finger can recognize a specific 3-bp DNA sequence motif (Tupler et al., 2001; Wolfe et al., 2000). To generate ZFNs that recognize and bind a specific DNA sequence multiple zinc fingers are arranged in one construct with the goal of creating a ZFP array that can bind to a 9-18 bps target sequence (Kim and Kim, 2014). The FokI nuclease is a type II restriction enzyme from \textit{Flavobacterium okeanokoites}, which contains two domains, a DNA-binding domain and a nuclease domain (Kim et al., 1996). To generate a programmable nuclease, the FokI nuclease domain is fused to the ZNP. Because the FokI nuclease domain has to dimerize to cleave dsDNA, two ZFNs need to be spaced 5-7 bp apart to cleave the target DNA (Kim and Kim, 2014). A complication of the ZNF technology is the difficulty of assembling ZFNs to target a specific sequence (Bae et al., 2003; Segal et al., 1999). A frequent problem is that the ZFN will not cause the desired dsDNA break because the ZFN can’t bind to and cut the target sequence, or that too many off target effects cause cytotoxicity (Kim and Kim, 2014). Nevertheless, ZFNs have been very useful in
genetically modifying hESCs and iPSCs (Soldner et al., 2011; Yusa et al., 2011). In addition, ZFNs have recently been used to genetically engineer hematopoietic stem and progenitor cells (Wang et al., 2015).

Transcription activator-like effector nucleases (TALENs) are similar to ZFNs in that both use the FokI nuclease domain as their functional unit to initiate a dsDNA break (Kim and Kim, 2014). However, unlike ZFNs, TALENs use a much more modular DNA binding domain that consists of transcription activator-like effectors (TALEs) (Mak et al., 2012; Deng et al., 2009; Boch et al., 2009). TALEs were discovered in the *Xanthomonas* species of bacteria and are composed of an array of 33-35 amino acid repeats (Kim and Kim, 2014). Each repeat can vary in the amino acids at position 12 and 13, which are called the repeat-variable diresidues (RVDs). Each repeat-domain depending on its RVD can recognize one of the 4 nucleotides: C, G, T, and A (Kim and Kim, 2014). Because of this single repeat to single nucleotide recognition, the repeats can be assembled to recognize any sequence of DNA. One TALEN is usually composed 12-20 RVDs that are attached to a FokI nuclease domain (Kim and Kim, 2014). Like ZFNs, a pair of TALENs must be used to target a specific genomic locus to cause a dsDNA break. Furthermore, even though TALENs are much easier to design and construct than ZFNs, they still require a complex cloning strategy called Golden Gate which is time consuming and not very user friendly (Ding et al., 2013). Despite the difficulty of assembly, TALENs have been used to efficiently edit the genome in human cells, zebrafish, *Oryza sativa*, *Caenorhabditis elegans*, and bovines (Carlson et al., 2012; Wood et al., 2011; Hockemeyer et al., 2011; Li et al., 2012).
CRISPR/Cas

In the 1990s, bioinformatic analysis discovered long repeats of short palindromic sequences in the genomes of some Bacteria and Archaea (Hermans et al., 1991; Bult et al., 1996; Hoe et al., 1999). The short hypervariable sequences between these palindromic repeats were shown to be homologous to sequences found in bacteriophage genomes and parasitic plasmids (Bolotin et al., 2005; Pourcel et al., 2005; Mojica et al., 2005). Further computational analysis led to the hypothesis that prokaryotes that carried these repetitive loci somehow used them as a way to protect themselves against bacteriophages or parasitic plasmids (Makarova et al., 2006). Research gave experimental validation that this repetitive locus, named Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), could act as an adaptive immune system when it was shown that that the CRISPR/Cas system in Streptococcus thermophilus provided resistance against phages that infect this species (Barrangou et al., 2007).

Further research showed that there are three phases of this adaptive immune response in Bacteria and Archaea, named adaption, expression, and interference (Figure 1.) (Bhya et al., 2011). Adaption or spacer acquisition occurs when a segment of DNA from a phage or parasitic plasmid is recognized as being foreign and is incorporated in between the palindromic repeats in the CRISPR locus (Garrett et al., 2010). These integrated segments of foreign DNA are called protospacers (Deveau et al., 2008). The second stage, expression, occurs when the pre-CRISPR RNA (pre-crRNA), which contains an array of multiple palindromic repeats and protospacers, is transcribed as a single long primary transcript. The pre-crRNA is then processed to form short CRISPR RNAs (crRNAs), with each crRNA only containing one protospacer sequence (Brouns et
al., 2008). In the third and final stage, interference, the short crRNA guides a protein or protein complex, which contains a functional DNA nuclease, to the invading phage or parasitic plasmid. Upon recognition, it causes degradation of the phage genome or parasitic plasmid by targeting the sequence that is complementary to the crRNA protospacer sequence (Deveau et al., 2010). An additional requirement for successful targeting is the presence of a protospacer adjacent motif (PAM) directly 3’ to the target protospacer sequence. Each CRISPR system recognizes a unique PAM motif, and even the same type of CRISPR system found in different species of bacteria can recognize different PAM motifs (Deveau et al., 2010). The CRISPR locus itself does not contain a PAM site directly 3’ to the incorporated protospacers. Therefore, the CRISPR system cannot target its own locus, but can only recognize protospacers that are flanked 3’ by a PAM motif in foreign DNA. However, the PAM motif is important for protospacer acquisition (Mojica et al., 2009).

There are three main subclasses of the CRISPR loci, Type-I, Type-II, and Type-III (Bhaya et al., 2011). All three subtypes share a common repetitive palindromic repeat sequence where new protospacers are acquired and expressed in the long primary pre-crRNA transcript (Deveau et al., 2010). Additionally, the loci of all three subtypes contain two genes which code for the proteins Cas1 and Cas2. These two proteins are important for protospacer acquisition (Bhaya et al., 2011). The main difference between the three subtypes is the protein or protein complex that is used to degrade the invading phage genome. The Type-I CRISPR locus contains multiple genes that code for a large multisubunit complex named CASCADE which binds the crRNA to target the phage genome (Jore et al., 2011). The Type-2 system uses a single protein called Cas9 and an
upstream non-coding RNA named the tracrRNA. Both of these components form a complex with the crRNA, which then functions to target and destroy the phage genome (Garneau et al., 2010). The type-3 subclass locus is even more complex and can use either a complex called the CMR, to target phage RNA, or a complex called the CSM, to target phage DNA (Bhaya et al., 2011).

![CRISPR/Cas adaptive immune response](image)

**Figure 1. CRISPR/Cas adaptive immune response.**
Bacteria and Achaia use CRISPR as an adaptive immune response to fight against invading phages or parasitic plasmids, this process is orchestrated through three main phases. (1) Acquisition is when a new protospacer from the invading phage is acquired and inserted into the CRISPR array. This new Protospcacer needs to be flanked directly 3' by a PAM sequence highlighted in red; the PAM sequence is not inserted into the CRISPR array. (2) Expression is when the Pre-crRNA is expressed as one long primary transcript with multiple crRNAs in tandem; this Pre-crRNA is then processed to form multiple functional crRNAs. (3) Interference is when the functional Cas nuclease and crRNA ribonucleoprotein complex cleaves its target sequence on either the invading phage or parasitic plasmid. Figure adopted from (Bhaya et al., 2011)
The type-II CRISPR locus is the simplest of the three subtypes. Along with the pre-crRNA and tracrRNA, the type-II locus contains only four protein coding genes, \textit{cas9, cas1, cas2} and either \textit{cas4} or \textit{csn2} (Bhaya et al., 2011). Work in bacteria showed that Cas9 along with a mature crRNA and tracrRNA, was necessary for resistance to phages (Garneau et al., 2010). Seminal in vitro work by Jennifer Doudna and Emmanuille Charpentier showed that a purified Cas9 protein from \textit{Streptococcus pyogenes}, along with a mature crRNA and tracrRNA, could initiate a dsDNA break in a plasmid that contains a PAM motif and a target sequence complementary to the crRNA protospacer sequence (Jinek et al., 2012).

This in vitro work showed that Cas9, along with a mature crRNA and tracrRNA, is sufficient to target and cleave dsDNA. Furthermore, it was revealed that the tracrRNA and crRNA form a stem-loop complex that binds Cas9. The crRNA then directs this ribonucleoprotein complex to bind its protospacer target sequence through Watson-Crick base complementarity (figure 2.) (Jinek et al., 2012). Upon recognition of its target sequence and PAM motif, Cas9, which contains two nuclease domains RuvC and HNH, will cleave dsDNA to form two blunt ends directly 3 nucleotides 5' to the PAM motif (Jinek et al., 2012). The PAM motif for \textit{S. Pyogenes} Cas9 is 5'-NGG-3' (Jinek et al., 2012). Furthermore, this work showed that the tracrRNA and crRNA could be combined together to form a synthetic chimeric RNA designated as “gRNA”. This chimeric gRNA forms the same stem-loop structure as the separate crRNA and tracrRNA, and along with Cas9 it can cleave dsDNA in vitro (Jinek et al., 2012). This work by Doudna and Charpentier suggested that because the Type-II CRISPR/Cas system is simple, consisting of only Cas9, a tracrRNA, and a programmable crRNA or gRNA, it could be developed
as a technology that could be used for genome-engineering in eukaryotic cells (Jinek et al., 2012).

Research by Fang Zhang and George Church further solidified CRISPR/Cas9 as a technology for genome-engineering when both labs simultaneously showed that this system could be used to edit genes in vivo, both in human and mouse cells (Cong et al., 2013; Mali et al., 2013). Both labs expressed a human codon-optimized version of *Streptococcus pyogenes* Cas9, and either a crRNA and tracrRNA or a chimeric gRNA, in mammalian cells, and showed that Cas9 could be directed to cleave an endogenous genomic locus that was complementary to the co-expressed crRNA or gRNA. In addition, this research indicated that CRISPR/Cas9 worked as well or even better than TALENs or Zinc Finger Nucleases when directed to the same locus (Cong et al., 2013). This in vivo work also showed that the dsDNA break caused by Cas9 could be repaired through homology-directed repair (HDR) by supplying either a ssDNA-oligo (to make a point mutations) or a dsDNA-template (to make a GFP reporter allele) (Cong et al., 2013; Mali et al., 2013). Furthermore, the CRISPR/Cas9 system only requires the expression of Cas9 and a gRNA that recognizes a 20nt genomic target sequence flanked directly 3’ by the 5’NGG-3’ PAM motif for successful genomic targeting. This fact makes the CRISPR/Cas9 system a much easier technology to design, construct, and implement than either TALENs or ZFNs. In addition, because Cas9 is the functional nuclease, it can be imagined that this system is very easy to multiplex to make multiple genetic modifications simultaneously by co-expressing different chimeric gRNAs along with Cas9 (Cong et al., 2013). The only limitation for CRISPR/Cas9 genome targeting is the presence of a PAM motif. Statistically, the 5’-NGG-3’ PAM motif for *S. pyogenes* Cas9
should occur every 16 nucleotides. However, this is not always the case in a genomic context. A potential solution is the recent development of CRISPR/Cas9 systems from different species of bacteria with alternative PAM motifs, which allows for a greater set of possible genomic targets (Zetsche et al., 2015; Kleinstiver et al., 2015).

Figure 2. In vitro and in vivo Cas9 target recognition.
Cas9 from the type-II CRISPR system forms a complex with both a crRNA and tracrRNA. The 20nt guide sequence on the crRNA recognizes its target sequence through base pairing to the 20nt complementary strand directly 5' to the NGG PAM motif. The constant region of the crRNA, highlighted in orange, forms a stem loop with the tracrRNA, highlighted in green. The stem loop conformation between the crRNA and tracrRNA and the stem loops 1,2,3 in the tracrRNA helps these RNAs bind Cas9. Upon target recognition, the Cas9 will cause a blunt end dsDNA break 3nt 5' of the PAM motif. Cas9 functions in vitro and in vivo when (a) the tracrRNA and crRNA are expressed separately or (b) when the tracrRNA and crRNA are fused together to form a chimeric sgRNA or gRNA. Figure adopted from (Kim and Kim, 2014)
DNA repair machinery

Locus-specific nucleases like CRISPR/Cas9, ZFNs, and TALENs increase the efficiency of genome-engineering because they cause a dsDNA break at the region of DNA to be modified. The cell is then forced to repair the dsDNA break. However, how the dsDNA break is repaired depends on various factors that are intrinsic to the cell, such as the time in the cell cycle when the dsDNA break is detected, and whether or not there is a segment of homologous DNA sequence present near the vicinity of the break (figure 3) (Kim and Kim, 2014). The two main repair pathways the cell uses to repair a dsDNA break are non-homologous-end joining (NHEJ) and homology-directed repair (HDR) (Kim and Kim, 2014).

NHEJ is the most common form of dsDNA break repair because it does not require the presence of a homologous DNA sequence, and it can occur in all phases of the cell cycle (Rothkamm et al., 2003). During NHEJ repair, when a dsDNA break is detected, a protein called KU binds very strongly (Kd $10^{-9}$) to each of the free DNA duplex ends (Falzon et al., 1993). KU then acts as a hub to recruit all of the different factors necessary for NHEJ repair. These include the Artemis:DNA-PKcs nuclease complex, the Polµ and Polλ DNA polymerases, and the XLF:XRCC:DNA ligase-IV complex (Ma et al., 2004). The Artemis:DNA-PKcs nuclease complex has both a 5' and 3' endonuclease activity, hairpin opening activity, and 5' exonuclease activity (Ma et al., 2002). These different functions allow the Artemis:DNA-PKcs nuclease complex to remove or modify any damaged single stranded ends that might result from a dsDNA break (Ma et al., 2005, Yannone et al., 2008). The Polµ and Polλ DNA polymerases are used to fill in single-stranded DNA ends. However, Polµ can also perform template-
independent synthesis (Lee et al., 2003; Nickmcelhinny et al., 2003). The XLF:XRCC:DNA ligase-IV complex is a very adaptable ligase complex that can ligate non-compatible ends, blunt ends, and ssDNA ends (Gu et al., 2007). There is not a specific order of recruitment of these complexes to KU, and different complexes can be recruited at the same time to the two separate KU bound DNA duplex ends (Ma et al., 2004). This fact is what leads to the random insertion or deletion of nucleotides during NHEJ repair and usually results in a random mutation once the two ends are joined (Lieber 2010). Because NHEJ randomly inserts or deletes nucleotides during the repair process, CRISPR/Cas9 can be used to knockout genes. If Cas9 is targeted to the coding region of a gene, any random insertion or deletion of 1n or 2n nucleotides will result in a frame-shift mutation and subsequent gene knockout if the exon is incorporated into the mRNA transcript after splicing.

A cell can also repair a dsDNA break through homology-directed repair (HDR) when the cell is in S/G2 phase and there is an adequate homologous dsDNA template present (Jasin et al., 2013). HDR is initiated when there is resection of the 5’ strand at each end of the dsDNA break which creates two 3’ ssDNA overhangs (Forget et al., 2010). In mammals, the resection of the 5’ strand is accomplished by the MRN complex along with the CtlP protein, and possibly other factors such as BRCA1, EXO1, and BLM (Jasin et al., 2013). The protein RAD51, which is recruited by BRCA2, binds to the 3’ strand and allows for strand invasion of the homologous dsDNA template (Sugawara et al., 2003). Upon invasion, the 3’ strand can act as a primer for homologous template-dependent synthesis by DNA polymerase (Forget et al., 2010). The resulting Holiday junction complex that forms can then be resolved by either non-crossover mechanism that
require RecQ and topoisomerase III or crossover mechanism that require Holiday 
junction Resolvase and DNA ligase (Jasin et al., 2013). Furthermore, there are alternative 
pathways of HDR called alt-NHEJ, MMEJ, or Single Strand Annealing (SSA) that 
require a 3’ ssDNA overhang and 5’ strand resection. However, rather than invasion, 
there is annealing across the break point (Jasin et al., 2013). HDR is the mechanism by 
which large pieces of DNA (such a GFP reporter alleles or Cre) can incorporated into the 
genome, and the alternative pathways are how ssDNA-oligos (such as epitope-tags or 
loxP sequences) are incorporated into the genome after a double stranded-break is created 
through CRISPR/Cas9 targeting.

**Figure 3. DNA repair pathways and genome-engineering**

Site-specific nucleases, like CRISPR/Cas, can be used for genome-engineering because 
they cause a dsDNA break that can be repaired either through either NHEJ or HDR. 
When no exogenous DNA template is supplied, the dsDNA break will be repaired 
through NHEJ, and will usually result in the random insertion or deletion of nucleotides 
at the break junction, this is ideal to knockout a gene by causing a frameshift mutation. If 
either ssDNA or dsDNA-targeting vector is supplied that has homology to both sides of 
the break junction, this DNA targeting vector will be incorporated through HDR. This 
ideal to make specific genetic modification, such specific point mutations, reporter 
alleles, conditional alleles, and epitope tagged alleles. Figure adopted from 
(Kim and Kim, 2014)
DNA methylation

DNA methylation is a highly conserved epigenetic modification that attaches a methyl group to the 5-carbon position of cytosine in DNA (Feng et al., 2010). The most prevalent DNA methylation in the genome occurs at CpG dinucleotides, but methylation can also occur at CpA dinucleotides (Arand et al., 2012). DNA methylation is a conserved epigenetic modification that is found in multiple species such as bacteria, flies, mice, and humans, where it is thought to play an essential role in gene regulation and development. One indication that CpG methylation plays an important regulatory role in the mammalian genome comes from the fact that globally the human genome is depleted of CpG dinucleotides (Smith and Meissner 2013). Furthermore, almost 60-80% of the 25 million CpG dinucleotides remain methylated in almost all adult tissues (Smith and Meissner 2013). However, a segment of these CpG dinucleotides occur in dense CpG regions called CpG islands, most of which are not methylated. Although, a few are methylated depending on cell type or developmental state (Smith and Meissner 2013; Schubeler 2015). Furthermore, CpGs are also found in low density CpG regions that are not CpG islands, but can become differentially methylated depending on cell or tissue type. These differently methylated regions (DMRs) sometimes overlap with promoters and enhancer elements which is a possible indication that methylation is playing a direct role in regulating these important cis-regulatory regions (Schubeler 2015).

DNMTs are a family of conserved enzymes that function to methylate DNA. In mice and humans there are 3 members: DNMT1, DNMT3A, and DNMT3B (Okano et
al., 1999; Li et al., 1992). In addition, there is another important non-enzymatic family member DNMT3L that is conserved between mice and humans (Chen et al., 2005). Each DNMT, except DNMT3L, can catalyze the methylation of cytosine in a CpG dinucleotide pair; however, each DNMT has a different function in this epigenetic process. DNMT1 is a maintenance methylase, which recognizes hemimethylated CpGs in dsDNA and then methylates the unmethylated cytosine to maintain the methylation pattern during DNA replication in mitosis (Avvakumov et al., 2008). DNMT1 interacts with PCNA and UHRF1, and this complex is recruited to and binds hemimethylated replicated DNA during S-phase of the cell cycle (Sharif et al., 2007; Chuang et al., 1997). DNMT3A and DNMT3B are de novo methylases and can methylate non-methylated DNA (Morgan et al., 2005). It has been shown that DNMT3A/B can interact with nucleosome remodeling complexes and histone methylases, which is exemplified during mESC differentiation when DNMT3A/B is recruited in a complex with G9A and LSH to the promoters of important ES-specific genes where it stably silences these promoters through methylation (Epsztejn-Litman et al., 2008; Myant et al., 2011). In addition, the factors TRIM28, SETDB1, and ZFP809 are known to be important for recruiting DMNT3A to silence LTRs in the genome (Wolf et al., 2009; Rowe et al., 2010; Wolf et al., 2007). DNMT3L cannot methylate DNA; however, it does play an important role of recruiting DNMT3A or DMTA3B to silence LINE1 and retroviral elements in the genome (Smith and Meissner 2013). The importance of these enzymes is highlighted by the fact that both Dnmt1 -/- and Dnmt3b -/- knockouts are embryonic lethal and dnmt3a -/- mice die shortly after birth (Li et al., 1992; Okano et al., 1999).
DNA demethylation is thought to occur through two different mechanisms. The first is passive DNA demethylation in which the maintenance DNMT1 is blocked and the genomic locus becomes diluted of the methyl mark through cell division (Kohli and Zhang, 2013). The second mechanism is active DNA demethylation in which the 5-methyl group on cytosine is removed by a series of enzymatic reactions without the need for cell division (Kohli and Zhang, 2013). Recently a family of enzymes called the Ten-Eleven Translocation (TET) enzymes, which consist of three family members, TET1, TET2, and TET3, were shown to be able to convert 5-methyl cytosine (5mc) to 5-hydroxymethylcytosine (5hmC) (Tahilani et al., 2009; Ito et al., 2010; Kriaucionis et al., 2009). The TET enzymes can further catalyze 5hmC to form 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Ito et al., 2011; He et al., 2011). TET enzymes could play an important function in either active or passive DNA demethylation. Converting 5mc to 5hmC, 5fc or 5caC could block DNMT1 from properly propagating the methyl mark during cell division. Alternatively, through an active process an enzyme called TDG, which is part of the BER DNA repair pathway, could remove these modified cytosines where they would then be replaced with non-methylated cytosines (Dalton and Bellacosa 2012). Besides the role of TET enzymes in either active or passive DNA demethylation, the modification of 5mc to 5hmC, 5fC, and 5caC is thought to play a role in gene regulation and other epigenetic processes (Kohli and Zhang, 2013). Furthermore, besides the TET family of proteins, other factors have been reported to play a role in active DNA demethylation such as the protein AID from the APOBEC family of enzymes (Bhutani et al., 2010; Popp et al., 2010).
Methylation changes in development

Methylation is a highly dynamic process during development, both globally and locally at specific genomic regions (Figure 4) (Kohli and Zhang, 2013). DNA methylation is thought to be important for committing cells to a specific lineage and making differentiation a unidirectional process (Messerschmidt et al., 2014). Upon fertilization, there is a global loss of methylation from both the paternal and maternal genomes. This epigenetic reprogramming event allows for the erasure of the germ cell-specific epigenetic signature (Santos and Dean, 2004). The rate of methylation loss from the maternal and paternal genomes is not the same. Loss of methylation occurs more rapid on the paternal genome possibly due to active demethylation mediated by the TET enzymes (Oswald et al., 2000; Kohli and Zhang, 2013). The loss of methylation from the maternal genome is the result of passive demethylation during DNA replication (Messerschmidt et al., 2014). After implantation a global wave of de novo methylation establishes the genomic methylation level that is characteristic of somatic cells (Smallwood and Kelsey, 2012). During this period, low-density CpG regions that overlap with certain developmental specific cis-regulatory elements become methylated or demethylated in a cell-type and tissue-specific manner allowing for specification and directionality of somatic differentiation (Messerschmidt et al., 2014; Schubeler, 2015). However, around embryonic day 7.25, a small fraction of cells migrates out of the epiblast and colonize the genital ridge, these cells establish the primordial germ cells (PGCs) which will further develop to form the sperm and oocytes after sex specification of the embryo (Smallwood and Kelsey, 2012). During PGC development there is another wave of global genomic demethylation, in all regions of the genome except LINE1 and
retrotransposable elements (Popp et al., 2010; Sasaki et al., 2008). This global demethylation during PGC development allows for the erasure of the somatic epigenetic signature and establishment of a new germ cell-specific signature (Messerschmidt et al., 2014).

There are a few very specific loci in the genome that do not become demethylated after fertilization. These regions are called imprinted germline differentially methylated regions (gDMRs). Maternal imprinted gDMRs are methylated in the oocyte and paternal imprinted gDMRs are methylated in the sperm (Smallwood and Kelsey, 2012). After fertilization these imprinted regions retain their parent-of-origin methylation pattern. These imprinted DMRs are CpG islands and are often imprint-control regions that control the imprinting of gene clusters next to them (Wutz et al. 1997; Thorvaldsen et al. 1998; Fitzpatrick et al. 2002; Lin et al. 2003). One example of this is the Prader-Willi/Angelman region which is a maternally imprinted gDMR. This region has a primary imprinted gDMR termed PWS-SRO which can regulate the imprinting of nearby secondary DMRs in the promoters of cis-proximal genes, such as SNRPN and UBE3 (Yang et al., 1998; Edwards and Ferguson-Smith 2007).
Figure 4. DNA methylation dynamics during development.
DNA methylation is a highly dynamic process during development. During development of the embryo primordial germ cells, PGCs, start to proliferate and migrate to form the genital ridge. As PGCs migrate, there is a progressive global loss of methylation from their genome. As the PGCs develop into either mature spermatogonia or oocytes, there is a wave of global genomic methylation. After fertilization of the sperm and egg and as the single-cell zygote develops to form the blastocyst, there is another loss of global genome methylation. However, a few loci in the genome do not become demethylated, these are called imprinted regions, and they retain their parent-of-origin methylation patterns. Finally, as the blastocyst develops to form the complete organism, there is further genomic methylation. This methylation event is tissue and cell-type specific, where it plays an important mechanism is gene regulation, development and unidirectional differentiation. Figure adopted from (Smallwood and Kelsey, 2012)

Methylation and gene regulation

Low CpG density promoters are regulated by methylation. In contrast, most CpG island-associated promoters are not regulated by DNA methylation; however, a subset of CpG islands associated promoters can be switched on and off by this epigenetic modification (Smith and Meissner, 2013). In addition, many repetitive elements, such as LINE1 and ERVs, are silenced through methylation of their respective promoters (Liang et al., 2002). Silencing of promoters through methylation coincides with changes in
histone modifications; specifically, the H3K9 methylation mark is highly associated with silenced promoters (Ayyanathan et al., 2003). The promoter for the Oct4 (pou5f1) gene, which is a master transcription factor for the pluripotent cell state, is a good example of a promoter that is regulated by DNA methylation and H3K9 methylation. (Athanasiadou et al., 2010). Upon differentiation and exit from pluripotency, the protein G9A is recruited to the Oct4 promoter where it initiates H3K9 methylation. This histone modification is followed by recruitment of the heterochromatin protein 1 (HP1). Finally, DNMT3A or DNMT3B are recruited to de novo methylate the promoter allowing for stable epigenetic silencing (Feldman et al., 2006; Athanasiadou et al., 2010). In addition to promoters, it is thought that the activity of other cis-regulatory regions, such as enhancers that overlap with low density CpG regions, can be regulated through DNA methylation (Schubeler 2015). Hypomethylation of enhancers has been correlated with gene expression changes. This is especially true in some cancer cell lines where it has been shown that enhancer methylation for certain genes is a better predictive indicator of a gene’s activity than promoter methylation (Aran et al., 2013, Sandovici et al., 2011, Messerchmidt et al., 2014).

**DNA methylation and cancer**

Aberrant methylation is associated with many different forms of cancer. Hypermethylation and subsequent silencing of genes that are important in DNA repair, cell-cycle regulation, apoptosis, and tumor cell invasion occur in many different types of cancer (Laird et al., 1995; Robertson 2005). Furthermore, the loss of imprinting (LOI) of some genes can cause or be associated with different cancer types. One important
example of this is the gene IGF2 which is an imprinted gene that is expressed from the paternal allele (Barlow et al. 1991). Loss of methylation on the maternal allele can cause overexpression of IGF2, and can lead to cancer in multiple tissue types such as the lung, liver, and colon (Moulton et al., 1994; Steenman et al. 1994). In addition, LOI can also occur when the normally expressed imprinted allele of a tumor suppressor gene becomes silenced (Robertson 2005). Examples include loss of expression of kinase inhibitor 1C, CDKN1C, in Wilms’ tumor, and RAS-related gene, DIRAS3 in colon cancer (Feinberg et al., 2002; Thompson et al., 1996). Besides the aberrant regulation of tumor suppressor genes and oncogenes by methylation, mutations in many of the core proteins that have a role in methylation dynamics are also found in cancer. A fusion of the MLL gene with TET1 is found in some patients with myeloid leukemia and mutations in the gene TDG occur in various cancer types (Ono et al., 2002; Dalton and Bellacosa 2012).

**Technology available to study DNA methylation**

One of the first experiments to show that DNA methylation occurred in eukaryotes made use of methylation-sensitive restriction enzymes (Bird and Southern, 1978; Cedar et al., 1979). Certain restriction enzymes cannot cut methylated DNA. Therefore, these enzymes can be used to distinguish if their respective sites are methylated in genomic DNA. This method was further improved when it was coupled with HPLC or mass spectrometry which allowed for relative total levels of methylated and non-methylated cytosine to be compared across different tissue or cell types (Gama-Sosa et al., 1983; Bestor et al., 1984). This method works well if the aim is to compare
the overall methylated cytosine content; however, it does not give any sequence specific-
information.

To investigate base-resolution methylation patterns, a method termed bisulfate
sequencing was developed (Herman et al., 1996; Frommer et al., 1992; Oakeley et al.,
1997). This method makes use of the chemical sodium bisulfate which specifically
deaminates non-methylated cytosine to form uracil; however, when the 5-carbon of
cytosine is methylated, this reaction does not proceed. Therefore, treating total genomic
DNA isolated from a cell or tissue in vitro with sodium bisulfate will convert all non-
methylated cytosine into uracil and all methylated cytosine will remain protected. Sodium
bisulfate treatment can be followed by PCR to gain information on the base-resolution
methylation pattern at any genomic locus that is able to be PCR amplified (Frommer et
al., 1992). However, primer design for bisulfate sequencing can be difficult because the
primers cannot contain CpG dinucleotides which would bias the reaction.

Finally, methods have been developed to investigate full-genome methylation
patterns. Methylated-cytosine-specific antibodies were combined with microarray
technology providing whole genome information on methylation albeit only with low
resolution (Gitan et al., 2002). New methods that couple bisulfate treatment with modern
sequencing technologies have now allowed for the establishment of full-genome base-
resolution methylation maps for many different species including human (Cokus et al.,
2008; Lister et al., 2009). Furthermore, a new initiative, the Roadmap Epigenomics
consortium, has recently published the full-genome methylation maps in multiple
different human and mouse cell types (Smith et al., 2014; Roadmap epigenomics
Consortium et al., 2015, Schulz et al., 2015).
Part 4: Thesis outline

Locus-specific genome editing in the single-cell zygote.

While the method of making knockin and knockout mice through homologous recombination has been very successful, the technique is time consuming, is inefficient, and is expensive. In the first aim of this thesis we investigate the possibility of developing a more efficient methodology for genetically engineering mice by combining the method of mouse zygotic injections, which was developed during the early period of transgenics, with that of programmable nucleases, specifically the CRISPR/Cas9 system. Because of its ability to cause locus-specific dsDNA breaks, the CRISPR/Cas9 system has been shown to increase the efficiency of making gene knockouts through NHEJ and genetic modification through HDR in human cells. The high efficiency of genome-engineering mediated by CRISPR/Cas9 could be used to make all of the classical genetic modifications, such as multiple gene knockouts, reporter alleles, epitope-tagged alleles, and conditional alleles, directly through zygotic injections. Cas9-mediated locus-specific genome-engineering through zygotic injections would allow for mice to be made in as little as three weeks at a much lower cost because it bypasses the intermediate steps of making genetically engineered mESCs though homologous recombination, blastocyst injections, and breeding chimeric mice to make pure F1s.

A reporter for genomic methylation

DNA methylation is a dynamic epigenetic modification that does not only play an important role in regulating normal biological processes, such as gene transcription, cell
identity, and development, but can also becomes highly deregulated in human diseases such as cancer. Although DNA methylation has been known to be a dynamic process for quite some time, present technology to capture DNA methylation patterns has been restricted to static snapshots in fixed cell populations. As of yet no technology exists to investigate the dynamics of DNA methylation in vivo. The second aim of this thesis is to establish a technology that can report on the DNA methylation state of an endogenous locus in vivo at single-cell resolution.
References


Ding, Q. et al. (2013) A TALEN genome-editing system for generating human stem cell-based disease models. Cell Stem Cell 12, 238–251


He YF. et al. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science 333, 1303–1307


Okano M, Bell DW, Haber DA, Li E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99, 247–257


Ono R. et al. (2002) LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). Cancer Res. 62, 4075–4080


Maherali, N. et al. (2007) Global epigenetic remodeling in directly reprogrammed fibroblasts. Cell Stem Cell 1, 55–70


Myant, K. et al. 2011. LSH and G9A/GLP complex are required for developmentally programmed DNA methylation. Genome Res. 21, 83–94


Soldner F. et al. (2011) Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell 146, 318–331


Wood AJ. et al. (2011) Targeted genome editing across species using ZFNs and TALENs. Science 333, 307


Chapter 2. One-step generation of mice carrying mutations in multiple genes targeted by CRISPR-Cas mediated genome-engineering

Haoyi Wang¹,⁷, Hui Yang¹,⁷, Chikdu S. Shivalila¹,²,⁷, Meelad M. Dawlaty¹, Albert W. Cheng¹,³, Feng Zhang⁵,⁶, Rudolf Jaenisch¹,³,⁸

1. Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, USA
2. Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
3. Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
5. Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA and
6. McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.
7. These authors contributed equally to this work.
8. Correspondence should be addressed to Rudolf Jaenisch [jaenisch@wi.mit.edu]

Published as:


*indicates equal contribution

HW, HY and CSS shared all experiments and analyses equally, except HY did all zygotic injections. MMD helped with 5hmC analysis. AWC helped with computational analysis of off targets. FZ proved CRISPR plasmids. HW, HY, CSS and RJ designed and conceived of experiments.
Mice carrying mutations in multiple genes are traditionally generated by sequential recombination in embryonic stem cells and/or time-consuming intercrossing of mice with a single mutation. The CRISPR/Cas system has been adapted as an efficient gene-targeting technology with the potential for multiplexed genome editing. We demonstrate that CRISPR/Cas-mediated gene editing allows the simultaneous disruption of five genes (Tet1, 2, 3, Sry, Uty - 8 alleles) in mouse embryonic stem (ES) cells with high efficiency. Coinjection of Cas9 mRNA and single-guide RNAs (sgRNAs) targeting Tet1 and Tet2 into zygotes generated mice with biallelic mutations in both genes with an efficiency of 80%. Finally, we show that coinjection of Cas9 mRNA/sgRNAs with mutant oligos generated precise point mutations simultaneously in two target genes. Thus, the CRISPR/Cas system allows the one-step generation of animals carrying mutations in multiple genes, an approach that will greatly accelerate the in vivo study of functionally redundant genes and of epistatic gene interactions.

Genetically modified mice represent a crucial tool for understanding gene function in development and disease. Mutant mice are conventionally generated by insertional mutagenesis (Copeland and Jenkins, 2010; Kool and Berns, 2009) or by gene-targeting methods (Capecchi, 2005). In conventional gene-targeting methods, mutations are introduced through homologous recombination in mouse embryonic stem (ES) cells. Targeted ES cells injected into wild-type (WT) blastocysts can contribute to the germline of chimeric animals, generating mice containing the targeted gene modification (Capecchi, 2005). It is costly and time consuming to produce single-gene knockout mice and even more so to make double-mutant mice. Moreover, in most other mammalian species, no established ES cell lines are available that contribute efficiently to chimeric animals, which greatly limits the genetic studies in many species.

Alternative methods have been developed to accelerate the process of genome modification by directly injecting DNA or mRNA of site-specific nucleases into the one-cell embryo to generate DNA double-strand break (DSB) at a specified locus in various species (Bogdanove and Voytas, 2011; Carroll et al., 2008; Urnov et al., 2010). DSBs
induced by these site-specific nucleases can then be repaired by error-prone nonhomologous end joining (NHEJ) resulting in mutant mice and rats carrying deletions or insertions at the cut site (Carbery et al., 2010; Geurts et al., 2009; Sung et al., 2013; Tesson et al., 2011). If a donor plasmid with homology to the ends flanking the DSB is coinjected, high-fidelity homologous recombination can produce animals with targeted integrations (Cui et al., 2011; Meyer et al., 2010). Because these methods require the complex designs of zinc finger nucleases (ZNFs) or Transcription activator-like effector nucleases (TALENs) for each target gene and because the efficiency of targeting may vary substantially, no multiplexed gene targeting in animals has been reported to date. To dissect the functions of gene family members with redundant functions or to analyze epistatic relationships in genetic pathways, mice with two or more mutated genes are required, prompting the development of efficient technology for the generation of animals carrying multiple mutated genes.

Recently, the type II bacterial CRISPR/Cas system has been demonstrated as an efficient gene-targeting technology with the potential for multiplexed genome editing. Bacteria and archaea have evolved an RNA-based adaptive immune system that uses CRISPR (clustered regularly interspaced short palindromic repeat) and Cas (CRISPR-associated) proteins to detect and destroy invading viruses and plasmids (Horvath and Barrangou, 2010; Wiedenheft et al., 2012). Cas proteins, CRISPR RNAs (crRNAs), and transactivating crRNA (tracrRNA) form ribonucleoprotein complexes, which target and degrade foreign nucleic acids, guided by crRNAs (Gasiunas et al., 2012; Jinek et al., 2012). It was shown that the Cas9 endonuclease from Streptococcus pyogenes type II CRISPR/Cas system can be programmed to produce sequence-specific DSB in vitro by
providing a synthetic single-guide RNA (sgRNA) consisting of a fusion of crRNA and tracrRNA (Jinek et al., 2012). More intriguingly, Cas9 and sgRNA are the only components necessary and sufficient for induction of targeted DNA cleavage in cultured human cells (Cho et al., 2013; Cong et al., 2013; Mali et al., 2013) as well as in zebrafish (Chang et al., 2013; Hwang et al., 2013). A recent report also demonstrated disruption of a GFP transgene in mice using the CRISPR/Cas system (Shen et al., 2013). The ease of design, construction, and delivery of multiple sgRNAs suggest the possibility of multiplexed genome editing in mammals. Indeed, one study demonstrated that two loci separated by 119 bp could be cleaved simultaneously in cultured human cells at a low efficiency (Cong et al., 2013). The extent of achievable multiplexed genome editing has yet to be demonstrated in stem cells as well as in animals. Here, we use the CRISPR/Cas system to drive both NHEJ-based gene disruption and homology directed repair (HDR)-based precise gene editing to achieve highly efficient and simultaneous targeting of multiple genes in stem cells and mice.

Results

Simultaneous targeting up to five genes in ES cells

To test the possibility of targeting functionally redundant genes from the same gene family, we designed sgRNAs targeting the Ten-eleven translocation (Tet) family members, Tet1, Tet2, and Tet3 (Figure 1A). Tet proteins (Tet1/2/3) convert 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) in various embryonic and adult tissues and mutant mice for each of these three genes have been produced by homologous recombination in ES cells (Dawlaty et al., 2011; Gu et al., 2011; Li et al.,
2011; Moran-Crusio et al., 2011). To test whether the CRISPR/Cas system could produce targeted cleavage in the mouse genome, we transfected plasmids expressing both the mammalian-codon-optimized Cas9 and a sgRNA targeting each gene (Cong et al., 2013; Mali et al., 2013) into mouse ES cells and determined the targeted cleavage efficiency by the Surveyor assay (Guschin et al., 2010). All three Cas9-sgRNA transfections produced cleavage at target loci with high efficiency of 36% at Tet1, 48% at Tet2, and 36% at Tet3 (Figure 1B). Because each target locus contains a restriction enzyme recognition site (Figure 1A), we PCR amplified an 500 bp fragment around each target site and digested the PCR products with the respective enzyme. A correctly targeted allele will lose the restriction site, which can be detected by failure to cleave upon enzyme treatment. Using this restriction fragment length polymorphism (RFLP) assay, we screened 48 ES cell clones from each single-targeting experiment. Consistent with the Surveyor analysis, a high percentage of ES cell clones were targeted, with a high probability of having both alleles mutated (Figure S1A available online). The results summarized in Table 1 demonstrate that between 65% and 81% of the tested ES cell clones carried mutations in the Tet genes with up to 77% having mutations in both alleles.

The high efficiency of single-gene modification prompted us to test the possibility of targeting all three genes simultaneously. For this we cotransfected ES cells with the constructs expressing Cas9 and three sgRNAs targeting Tet1, 2, and 3. Of 96 clones screened using the RFLP assay, 20 clones were identified as having mutations in all six alleles of the three genes (Figures 1C and S1B and Table 1). To exclude that a PCR bias could give false positive results, we performed Southern blot analysis and confirmed complete agreement with the RFLP results (Figure 1C). We subcloned and sequenced the
PCR products of Tet1-, Tet2-, and Tet3-targeted regions to verify that all of eight tested clones carried biallelic mutations in all three genes with most clones displaying two mutant alleles for each gene with small insertions or deletions (indels) at the target site (Figure 1D). To test whether these mutant alleles would abolish the function of Tet proteins, we compared the 5hmC level of targeted clones to WT ES cells. Previously, we reported a depletion of 5hmC in Tet1/Tet2 double-knockout ES cells derived using traditional gene-targeting methods (Dawlaty et al., 2013). As expected from loss of function alleles, we found a significant reduction of 5hmC levels in all clones carrying biallelic mutations in the three genes (Figure 1E).

To further test the potential of multiplexed gene targeting by CRISPR/Cas system, we designed sgRNAs targeting two Y-linked genes, Sry and Uty (Figure S1C). Short PCR products encoding sgRNAs targeting all five genes (Tet1, Tet2, Tet3, Sry, and Uty) were pooled and cotransfected with a Cas9 expressing plasmid and the PGK puroR cassette into ES cells. Of 96 clones that were screened using the RFLP assay, 10% carried mutations in all eight alleles of the five genes (Figure S1D and Table S1), demonstrating the capacity of the CRISP/Cas9 system for highly efficient multiplexed gene targeting.

**One-step generation of single-gene mutant mice by zygote injection**

We tested whether mutant mice could be generated in vivo by direct embryo manipulation. Capped polyadenylated Cas9 mRNA was produced by in vitro transcription and coinjected with sgRNAs. Initially, to determine the optimal concentration of Cas9 mRNA for targeting in vivo, we microinjected varying
amounts of Cas9-encoding mRNA with Tet1 targeting sgRNA at constant concentration (20 ng/ml) into pronuclear (PN) stage one-cell mouse embryos and assessed the frequency of altered alleles at the blastocyst stage using the RFLP assay. As expected, higher concentration of Cas9 mRNA led to more efficient gene disruption (Figure S2A). Nevertheless, even embryos injected with the highest amount of Cas9 mRNA (200 ng/ml) showed normal blastocyst development, suggesting low toxicity.

To investigate whether postnatal mice carrying targeted mutations could be generated, we coinjected sgRNAs targeting Tet1 or Tet2 with different concentrations of Cas9 mRNA. Blastocysts derived from the injected embryos were transplanted into foster mothers and newborn pups were obtained. As summarized in Table 2, about 10% of the transferred blastocysts developed to birth independent of the RNA concentrations used for injection suggesting low fetal toxicity of the Cas9 mRNA and sgRNA. RFLP, Southern blot, and sequencing analysis demonstrated that between 50 and 90% of the postnatal mice carried biallelic mutations in either target gene (Figures 2A, 2B, and 2C and Table 2).

Surprisingly, specific D9 Tet1 and specific D8 and D15 Tet2 mutant alleles were repeatedly recovered in independently derived mice. Preferential generation of these alleles is likely caused by a short sequence repeat flanking the DSB (see Figure S2B) consistent with previous reports demonstrating that perfect microhomology sequences flanking the cleavage sites can generate microhomology-mediated precise deletions by end repair mechanism (MMEJ) (McVey and Lee, 2008; Symington and Gautier, 2011)(Figure S2B). A similar observation was also made when TALEN mRNA was injected into one-cell rat embryos (Tesson et al., 2011).
We also derived blastocysts from zygotes injected with Cas9 mRNA and Tet3 sgRNA. Genotyping of the blastocysts demonstrated that of eight embryos three were homozygous and three were heterozygous Tet3 mutants (two failed to amplify) (Figure S2C). Some blastocysts were implanted into foster mothers and, upon C section, we readily identified multiple mice of smaller size (Figure S2D), many of which died soon after delivery. Genotyping shown in Figure S2E indicated that all pups with mutations in both Tet3 alleles died neonatally. Only 2 out of 15 mice survived that were either Tet3 heterozygous mutants or WT (Figure S2F). These results are consistent with the lethal neonatal phenotype of Tet3 knockout mice generated using traditional methods (Gu et al., 2011), although we have not yet established which of the Tet3 mutations produced loss of function rather than hypomorphic alleles.

**One-Step generation of double-gene mutant mice by zygote injection**

To test whether Tet1/Tet2 double-mutant mice could be produced from single embryos, we coinjected Tet1 and Tet2 sgRNAs with 20 or 100 ng/ml Cas9 mRNA into zygotes. A total of 28 pups were born from 144 embryos transferred into foster mothers (21% live-birth rate) that had been injected at the zygote stage with high concentrations of RNA (Cas9 mRNA at 100 ng/ml, sgRNAs at 50 ng/ml), consistent with low or no toxicity of the Cas9 mRNA and sgRNAs (Table 3). RFLP, Southern blot analysis, and sequencing identified 22 mice carrying targeted mutations at all four alleles of the Tet1 and Tet2 genes (Figures 2D and 2E) with the remaining mice carrying mutations in a subset of alleles (Table 3). Injection of zygotes with low concentration of RNA (Cas9 mRNA at 20 ng/ml, sgRNAs at 20 ng/ml) yielded 19 pups from 75 transferred embryos.
(25% live-birth rate), which is a higher survival rate than from embryos injected with 100 ng/ml of Cas9 RNA. Nevertheless, more than 50% of the pups were biallelic Tet1/Tet2 double mutants (Table 3). These results demonstrate that postnatal mice carrying biallelic mutations in two different genes can be generated within one month with high efficiency (Figure 2F).

Although the high live-birth rate and normal development of mutant mice suggest low toxicity of CRISPR/Cas9 system, we sought to determine the off-target effects in vivo. Previous work in vitro, in bacteria, and in cultured human cells suggested that the protospacer-adjacent motif (PAM) sequence NGG and the 8 to 12 base "seed sequence" at the 3' end of the sgRNA are most important for determining the DNA cleavage specificity (Cong et al., 2013; Jiang et al., 2013; Jinek et al., 2012). Based on this rule, only three and four potential off targets exist in mouse genome for Tet1 and Tet2 sgRNA, respectively (Table S2 and Experimental Procedures), with each of them perfectly matching the 12 bp seed sequence at the 3' end and the NGG PAM sequence of the sgRNA (there is no potential off-target site for Tet3 sgRNA using this prediction rule). From seven double-mutant mice produced from injection with high RNA concentration we PCR amplified 400 to 500 bp fragments from all seven potential off-target loci and found no cleavage in the Surveyor assay (Figure S3), suggesting a high specificity of CRISPR/Cas system.

**Multiplexed Precise HDR-mediated genome editing in vivo**

The NHEJ-mediated gene mutations described above produced mutant alleles with different and unpredictable insertions and deletions of variable size. We explored
the possibility of precise homology directed repair (HDR)-mediated genome editing by coinjecting Cas9 mRNA, sgRNAs, and single-stranded DNA oligos into one-cell embryos. For this we designed an oligo targeting Tet1 so as to change two base pairs of a SacI restriction site and creating instead an EcoRI site and a second oligo targeting Tet2 with two base pair changes that would convert an EcoRV site into an EcoRI site (Figure 3A). Blastocysts were derived from zygotes injected with Cas9 mRNA and sgRNAs and oligos targeting Tet1 or Tet2, respectively. DNA was isolated, amplified, and digested with EcoRI to detect oligo-mediated HDR events. Six out of nine Tet1-targeted embryos and 9 out of 15 Tet2-targeted embryos incorporated an EcoRI site at the respective target locus, with several embryos having both alleles modified (Figure S4A). When Cas9 mRNA, sgRNAs, and single-stranded DNA oligos targeting both Tet1 and Tet2 were coinjected into zygotes, out of 14 embryos, four were identified that were targeted with the oligo at the Tet1 locus, seven that were targeted with the oligo at the Tet2 locus and one embryo (2) that had one allele of each gene correctly modified (Figure S4B). All four alleles of embryo 2 were sequenced, confirming that one allele of each gene contained the 2 bp changes directed by the oligo, whereas the other alleles were disrupted by NHEJ-mediated deletion and insertion (Figure S4C).

Blastocysts with double oligo injections were implanted into foster mothers and a total of 10 pups were born from 48 embryos transferred (21% live-birth rate). Upon RFLP analysis using EcoRI, we identified seven mice containing EcoRI sites at the Tet1 locus and eight mice containing EcoRI sites at the Tet2 locus, with six mice containing EcoRI sites at both Tet1 and Tet2 loci (Figure 3B). We also applied RFLP analysis using SacI and EcoRV to Tet1 and Tet2 loci, respectively, showing that all alleles not targeted
by oligos contained disruptions, which is in consistent with the high biallelic mutation rate by Cas9 mRNA and sgRNAs injection. These results were confirmed by sequencing demonstrating mutations in all four alleles of mouse 5 and 7 (Figure 3C). Our results demonstrate that mice with HDR-mediated precise mutations in multiple genes can be generated in one step by CRISPR/Cas-mediated genome editing.

Discussion

The genetic manipulation of mice is a crucial approach for the study of development and disease. However, the generation of mice with specific mutations is labor intensive and involves gene targeting by homologous recombination in ES cells, the production of chimeric mice, and, after germ line transmission of the targeted ES cells, the interbreeding of heterozygous mice to produce the homozygous experimental animals, a process that may take 6 to 12 months or longer (Capecchi, 2005). To produce mice carrying mutations in several genes requires time-consuming intercrossing of single-mutant mice. Similarly, the generation of ES cells carrying homozygous mutations in several genes is usually achieved by sequential targeting, a process that is labor intensive, necessitating multiple consecutive cloning steps to target the genes and to delete the selectable markers.

As summarized in Figure 4, we have established three different approaches for the generation of mice carrying multiple genetic alterations. We demonstrate that CRISPR/Cas-mediated genome editing in ES cells can generate the simultaneous mutations of several genes with high efficiency, a single-step approach allowing the production of cells with mutations in five different genes (Figure 4A). We chose the three
Tet genes as targets because the respective mutant phenotypes have been well defined previously (Dawlaty et al., 2011, 2013; Gu et al., 2011). Cells mutant for Tet1, 2 and 3 were depleted of 5hmCas would be expected for loss of function mutations of the genes (Dawlaty et al., 2013). However, we have not as yet established, which of the Cas9-mediated gene mutations produced loss of function rather than hypomorphic alleles.

We also show that mouse embryos can be directly modified by injection of Cas9 mRNA and sgRNA into the fertilized egg resulting in the efficient production of mice carrying biallelic mutations in a given gene. More significantly, coinjection of Cas9 with Tet1 and Tet2 sgRNAs into zygotes produced mice that carried mutations in both genes (Figure 4B, upper). We found that up to 95% of newborn mice were biallelic mutant in the targeted gene when single sgRNA was injected and when coinjected with two different sgRNAs, up to 80% carried biallelic mutations in both targeted genes. Thus, mice carrying multiple mutations can be generated within 4 weeks, which is a much shorter time frame than can be achieved by conventional consecutive targeting of genes in ES cells and avoids time-consuming intercrossing of single-mutant mice.

The introduction of DSBs by CRISPR/Cas generates mutant alleles with varying deletions or insertions in contrast to designed precise mutations created by homologous recombination. The introduction of point mutations into human ES cells, cancer cell lines, and mouse by ZNF or TALEN along with DNA oligo has been demonstrated previously (Chen et al., 2011; Soldner et al., 2011; Wefers et al., 2013). We demonstrate that CRISPR/Cas-mediated targeting is useful to generate mutant alleles with predetermined alterations, and coinjection of single-stranded oligos can introduce designed point mutations into two target genes in one step, allowing for multiplexed gene
editing in a strictly controlled manner (Figure 4B, lower). It will be of great interest to assess whether this targeting system allows for the production of conditional alleles, or precise insertion of larger DNA fragments such as GFP markers so as to generate conditional knockout and reporter mice for specific genes.

There are several potential limitations of the CRISPR/Cas technology. First, the requirement for a NGG PAM sequence of S. pyogenes Cas9 limits the target space in the mouse genome. It has been shown that the Streptococcus thermophiles LMD-9 Cas9 using different PAM sequence can also induce targeted DNA cleavage in mammalian cells (Cong et al., 2013). Therefore, exploiting different Cas9 proteins may enable to target most of the mouse genome. Second, although the sgRNAs used here showed high targeting efficiency, much work is needed to elucidate the rules for designing sgRNAs with consistent high targeting efficiency, which is essential for multiplexed genome-engineering. Third, although our off-target analysis for the seven most likely off targets of Tet1 and Tet2 sgRNAs failed to detect mutations in these loci, it is possible that other mutations were induced following as yet unidentified rules. A more thorough sequencing analysis for a large number of sgRNAs will provide more information about the potential off-target cleavage of the CRISPR/Cas system and lead to a better prediction of potential off-target sites. Last, oligo-mediated repair allows for precise genome editing, but the other allele is often mutated through NHEJ (Figures 3B, 3C, and S4C). We have shown that using lower Cas9 mRNA concentration generates more mice with heterozygous mutations. Therefore, it maybe possible to optimize the system for more efficient generation of mice with only one oligo-modified allele. In addition, employment of Cas9
nickase will likely avoid this complication because it mainly induces DNA single-strand break, which is typically repaired through HDR (Cong et al., 2013; Mali et al., 2013).

It is likely that a much larger number of genomic loci than targeted in the present work can be modified simultaneously when pooled sgRNAs are introduced. The methods presented here open up the possibility of systematic genome-engineering in mice, facilitating the investigation of entire signaling pathways, of synthetic lethal phenotypes or of genes that have redundant functions. A particularly interesting application is the possibility to produce mice carrying multiple alterations in candidate loci that have been identified in GWAS studies to play a role in the genesis of multigenic diseases. In summary, CRISPR/Cas mediated genome editing makes possible the generation of ES cells and mice carrying multiple genetic alterations and will facilitate the genetic dissection of development and complex diseases.

**Experimental Procedures**

**Procedures for generating sgRNAs expressing vector**

Bicistronic expression vector expressing Cas9 and sgRNA (Cong et al., 2013) were digested with BbsI and treated with Antarctic Phosphatase, and the linearized vector was gel purified. A pair of oligos for each targeting site (Table S3) was annealed, phosphorylated, and ligated to linearized vector.

**Cell Culture and Transfection**

V6.5 ES cells (on a 129/Sv x C57BL/6 F1 hybrid background) were cultured on
gelatin-coated plates with standard ES cell culture conditions. Cells were transfected with a plasmid expressing mammalian-codon-optimized Cas9 and sgRNA (single targeting), three plasmids expressing Cas9 and sgRNAs targeting Tet1, Tet2, and Tet3 (triple targeting), or five PCR products each coding for sgRNA targeting Tet1, Tet2, Tet3, Sry, and Uty, along with a plasmid expressing PGK-puroR using FuGENE HD reagent (Promega) according to the manufacturer’s instructions. Twelve hours after transfection, ES cells were re-plated at a low density on DR4 MEF feeder layers. Puromycin (2mg/ml) was added 1 day after replating and taken off after 48 hr. After recovering for 4 to 6 days, individual colonies were picked and genotyped by RFLP and Southern blot analysis, and the leftover ES cells on plate were collected for Suveryor assay.

**Suveryor Assay and RFLP analysis for genome modification**

Suveryor assay was performed as described by (Guschin et al., 2010). Genomic DNA from treated and control ES cells or targeted and control mice was extracted. Mouse genomic DNA samples were prepared from tail biopsies. PCR was performed using Tet1-, 2-, and 3- specific primers (Table S3) under the following conditions: 95C for 5 min; 35x (95C for 30 s, 60C for 30 s, 68C for 40s); 68C for 2 min; hold at 4C. PCR products were then denatured, annealed, and treated with Suveryor nuclease (Transgenomic). DNA concentration of each band was measured on an ethidium bromide-stained 10% acrylamide Criterion TBE gel (BioRad) and quantified using ImageJ software. The same PCR products for Suveryor assay were used for RFLP analysis. Ten microliters of Tet1, Tet2, or Tet3 PCR product was digested with SacI, EcoRV, or XhoI, respectively.
Digested DNA was separated on an ethidium bromide-stained agarose gel (2%). For sequencing, PCR products were cloned using the Original TA Cloning Kit (Invitrogen), and mutations were identified by Sanger sequencing.

**Dot Blot**

DNA was extracted from ES cells following standard procedures. DNA was transferred to nylon membrane using BioRad slot blot vacuum manifold apparatus. Anti-5hmC (Active Motif 1:10,000) was used to detect 5hmC following manufacturer’s protocol.

**Production of Cas9 mRNA and sgRNA**

T7 promoter was added to Cas9 coding region by PCR amplification using primer Cas9 F and R (Table S3). T7-Cas9 PCR product was gel purified and used as the template for in vitro transcription (IVT) using mMMESSAGE mMACHINE T7 ULTRA kit (Life Technologies). T7 promoter was added to sgRNAs template by PCR amplification using primer Tet1 F and R, Tet2 F and R, and Tet3 F and R (Table S3). The T7-sgRNA PCR product was gel purified and used as the template for IVT using MEGAscript T7 kit (Life Technologies). Both the Cas9 mRNA and the sgRNAs were purified using MEGAclean kit (Life Technologies) and eluted in RNase-free water.

**One-Cell embryo injection**

All animal procedures were performed according to NIH guidelines and approved by the Committee on Animal Care at MIT. B6D2F1 (C57BL/6 X DBA2) female mice and ICR mouse strains were used as embryo donors and foster mothers, respectively.
Superovulated female B6D2F1 mice (7–8 weeks old) were mated to B6D2F1 stud males, and fertilized embryos were collected from oviducts. Cas9 mRNAs (from 20 ng/ml to 200 ng/ml) and sgRNA (from 20 ng/ml to 50 ng/ml) was injected into the cytoplasm of fertilized eggs with well recognized pronuclei in M2 medium (Sigma). For oligos injection, Cas mRNA (100 ng/ml), sgRNA (50 ng/ml), and donor oligos (100 ng/ml) were mixed and injected into zygotes at the pronuclei stage. The injected zygotes were cultured in KSOM with amino acids at 37°C under 5% CO2 in air until blastocyst stage by 3.5 days. Thereafter, 15–25 blastocysts were transferred into uterus of pseudopregnant ICR females at 2.5 dpc.

**Southern Blotting**

Genomic DNA was separated on a 0.8% agarose gel after restriction digests with the appropriate enzymes, transferred to a nylon membrane (Amersham) and hybridized with 32P random primer (Stratagene)-labeled probes.

**Prediction of Potential off targets**

Potential targets of CRISPR sgRNAs were found using the rules outline in (Mali et al., 2013). For a 20 nt sgRNA targeting sequence of nnnnn nnMMM MMMMM MMMMM, where M are the seed bases preceding the PAM sequence NGG, four search sequences (MMM MMMMM MMMMM AGG; MMM MMMMM MMMMM CGG; MMM MMMMM MMMMM GGG; MMM MMMMM MMMMM TGG) were generated. Exact matches to these search sequences in the mouse genome (mm9) were found using bowtie and reported as potential targets of the CRISPR sgRNA
Acknowledgments

We thank Ruth Flannery and Kibibi Ganz for support with animal care and experiments. We thank Jaenisch lab members for helpful discussions on the manuscript. We are also grateful to G. Grant Welstead and Daniel B. Dadon for the help of editing the manuscript.

M.M.D is a Damon Runyon Postdoctoral Fellow. A.W.C. is supported by a Croucher scholarship. R.J. is an adviser to Stemgent and a cofounder of Fate Therapeutics. This work was supported by NIH grants R37-HD045022 and R01-CA084198 to R.J.
Figure 1. Multiplexed Gene Targeting in mouse ES cells
(A) Schematic of the Cas9/sgRNA-targeting sites in Tet1, 2, and 3. The sgRNA-targeting sequence is underlined, and the protospacer-adjacent motif (PAM) sequence is labeled in green. The restriction sites at the target regions are bold and capitalized. Restriction enzymes used for RFLP and Southern blot analysis are shown, and the Southern blot probes are shown as orange boxes.

(B) Surveyor assay for Cas9-mediated cleavage at Tet1, 2, and 3 loci in ES cells.

(C) Genotyping of triple-targeted ES cells, clones 51, 52, and 53 are shown. Upper: RFLP analysis. Tet1 PCR products were digested with SacI, Tet2 PCR products were digested with EcoRV, and Tet3 PCR products were digested with Xhol. Lower: Southern blot analysis. For the Tet1 locus, SacI digested genomic DNA was hybridized with a 5' probe. Expected fragment size: WT = 5.8 kb, TM (targeted mutation) = 6.4 kb. For the Tet2 locus, SacI, and EcoRV double digested genomic DNA was hybridized with a 3' probe. Expected fragment size: WT = 4.3 kb, TM = 5.6 kb. For the Tet3 locus, BamHI and XhoI double-digested genomic DNA was hybridized with a 5' probe. Expected fragment size: WT = 3.2 kb, TM = 8.1 kb.

(D) The sequence of six mutant alleles in triple-targeted ES cell clone 14 and 41. PAM sequence is labeled in red.

(E) Analysis of 5hmC levels in DNA isolated from triple-targeted ES cell clones by dot blot assay using anti-5hmC antibody. A previously characterized DKO clone derived using traditional method is used as a control. See also Figure S1.
Table 1. CRISPR/Cas-Medlated Gene Targeting In V6.5 ES Cells

<table>
<thead>
<tr>
<th>Gene</th>
<th>Mutant Alleles per Clone / Total Clones Tested</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Tet1</td>
<td>27/48</td>
</tr>
<tr>
<td>Tet2</td>
<td>37/48</td>
</tr>
<tr>
<td>Tet3</td>
<td>32/48</td>
</tr>
<tr>
<td>Tet1 + Tet2 + Tet3</td>
<td>20/96</td>
</tr>
</tbody>
</table>

Table 1. Plasmids encoding Cas9 and sgRNAs targeting Tet1, Tet2, and Tet3 were transfected separately (single targeting) or in a pool (triple targeting) into ES cells. The number of total alleles mutated in each ES cell clone is listed from 0 to 2 for single-targeting experiment, and 0 to 6 for triple-targeting experiment. The number of clones containing each specific number of mutated alleles is shown in relation to the total number of clones screened in each experiment. See also Table S1.

Table 2. CRISPR/Cas-Mediated Single-Gene Targeting In BDF2 Mice

<table>
<thead>
<tr>
<th>Gene</th>
<th>Cast9/sgRNA (ng/μl)</th>
<th>Blastocysts/Injected Zygotes</th>
<th>Transferred Embryos (Recipients)</th>
<th>Newborns (Dead)</th>
<th>Mutant Alleles per Mouse/Total Mice Tested</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Tet1</td>
<td>20/20</td>
<td>38/50</td>
<td>19 (1)</td>
<td>2 (0)</td>
<td>2/2</td>
</tr>
<tr>
<td></td>
<td>100/20</td>
<td>50/80</td>
<td>26 (1)</td>
<td>3 (0)</td>
<td>2/3</td>
</tr>
<tr>
<td></td>
<td>50/20</td>
<td>40/50</td>
<td>40 (2)</td>
<td>8 (3)</td>
<td>4/7</td>
</tr>
<tr>
<td></td>
<td>100/50</td>
<td>167/198</td>
<td>90 (3)</td>
<td>12 (2)</td>
<td>9/11</td>
</tr>
<tr>
<td>Tet2</td>
<td>100/50</td>
<td>176/203</td>
<td>108 (5)</td>
<td>22 (3)</td>
<td>19/20</td>
</tr>
<tr>
<td></td>
<td>100/50</td>
<td>85/112</td>
<td>84 (4)</td>
<td>15 (13)</td>
<td>9/13</td>
</tr>
<tr>
<td>Tet3</td>
<td>100/50</td>
<td>167/198</td>
<td>90 (3)</td>
<td>12 (2)</td>
<td>9/11</td>
</tr>
</tbody>
</table>

Table 2. Cas9 mRNA and sgRNAs targeting Tet1, Tet2, or Tet3 were injected into fertilized eggs. The blastocysts derived from injected embryos were transplanted into foster mothers and newborn pups were obtained and genotyped. The number of total alleles mutated in each mouse is listed from 0 to 2. The number of mice containing each specific number of mutated alleles is shown in relation to the total number of mice screened in each experiment. See also Table S2. Some of the pups were cannibalized.

Table 3. CRISPR/Cas-Mediated Double-Gene Targeting In BDF2 Mice

<table>
<thead>
<tr>
<th>Gene</th>
<th>Cast9/sgRNA (ng/μl)</th>
<th>Blastocysts/Injected Zygotes</th>
<th>Transferred Embryos (Recipients)</th>
<th>Newborns (Dead)</th>
<th>Mutant Alleles per Mouse/Total Mice Tested</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Tet1 + Tet2</td>
<td>100 / 50</td>
<td>194/229</td>
<td>144(7)</td>
<td>31(8)</td>
<td>22/26</td>
</tr>
<tr>
<td></td>
<td>20 / 20</td>
<td>92/109</td>
<td>75(9)</td>
<td>19(3)</td>
<td>11/19</td>
</tr>
</tbody>
</table>

Table 3. Cas9 mRNA and sgRNAs targeting Tet1 and Tet2 were coinjected into fertilized eggs. The blastocysts derived from the injected embryos were transplanted into foster mothers and newborn pups were obtained and genotyped. The number of total alleles mutated in each mouse is listed from 0 to 4 for Tet1 and Tet2. The number of mice containing each specific number of mutated alleles is shown in relation to the number of total mice screened in each experiment. Some of the pups were cannibalized.
**A** Tet1 mutants

![Diagram](image1)

**B** Tet2 mutants

![Diagram](image2)

**C**

\[
\begin{align*}
\text{Tet1} & \quad \text{Tet2} \\
\text{WT} & \quad \text{TM} \\
9.4 \text{kb} & \quad 6.5 \text{kb} \\
6.5 \text{kb} & \quad 4.4 \text{kb} \\
4.4 \text{kb} & \quad 3.8 \text{kb} \\
6.8 \text{kb} & \quad 5.8 \text{kb} \\
6.5 \text{kb} & \quad 5.8 \text{kb} \\
5.6 \text{kb} & \quad 4.3 \text{kb} \\
\end{align*}
\]

**D**

![Diagram](image3)

**E**

![Diagram](image4)

**F**

![Diagram](image5)
Figure 2. Single and Double Gene Targeting In Vivo by Injection into Fertilized Eggs

(A) Genotyping of Tet1 single-targeted mice.

(B) Upper: genotyping of Tet2 single-targeted mice. RFLP analysis; lower: Southern blot analysis.

(C) The sequence of both alleles of targeted gene in Tet1 biallelic mutant mouse 2 and Tet2 biallelic mutant mouse 4.

(D) Genotyping of Tet1/Tet2 double-mutant mice. Analysis of mice 1 to 12 is shown. Upper: RFLP analysis; lower: southern blot analysis. The Tet1 locus is displayed on the left and the Tet2 locus on the right.

E) The sequence of four mutant alleles from double-mutant mouse 9 and 10. PAM sequences are labeled in red.

(F) Three-week-old double-mutant mice. All RFLP and Southern digestions and probes are the same as those used in Figure 1. See also Figures S2 and S3.
Figure 3. Multiplexed HDR-Mediated Genome Editing In Vivo

(A) Schematic of the oligo-targeting sites at Tet1 and Tet2 loci. The sgRNA-targeting sequence is underlined, and the PAM sequence is labeled in green.
Oligo targeting each gene is shown under the target site, with 2 bp changes labeled in red. Restriction enzyme sites used for RFLP analysis are bold and capitalized.

(B) RFLP analysis of double oligo injection mice with HDR-mediated targeting at the Tet1 and Tet2 loci.

(C) The sequences of both alleles of Tet1 and Tet2 in mouse 5 and 7 show simultaneously HDR-mediated targeting at one allele or two alleles of each gene, and NHEJ-mediated disruption at the other alleles. See also Figure S4.
A Multiple Gene targeting in ES cells

B One Step Generation of Mice With Mutiple Mutations

Targeted Mutations (Deletion / Insertion)

Predefined Precise Mutations

Figure 4. Mutiplexed Genome Editing in ES Cells and Mouse

(A) Multiple gene targeting in ES cells.

(B) One-step generation of mice with multiple mutations. Upper: multiple targeted mutations with random indels introduced through NHEJ. Lower: multiple predefined mutations introduced through HDR-mediated repair.
References


epigenetic abnormalities but is compatible with post-natal development. Dev. Cell 24, 310–323.


Supplementary figures

Table S1. CRISPR/Cas-Mediated Quintuple Gene Targeting in V6.5 mES cells,

Related to Table 1.

<table>
<thead>
<tr>
<th>Mutant genes</th>
<th>Tet1, 2, 3</th>
<th>Tet1, 2, 3 + Sry</th>
<th>Tet1, 2, 3 + Sry + Uty</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Mutant alleles</td>
<td>6 and more</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>No. Mutant clones / Total clones</td>
<td>54/96</td>
<td>37/96</td>
<td>10/96</td>
</tr>
</tbody>
</table>

Plasmids encoding Cas9 and five PCR products expressing sgRNAs targeting Tet1, Tet2, Tet3, Sry, and Uty were co-transfected into mES cells. The number of clones containing mutations in all six Tet alleles is listed in the Tet1, 2, 3 column; the number of clones containing mutations in all six Tet alleles and Sry allele is listed in the Tet1, 2, 3 + Sry column; the number of clones containing mutations in all six Tet alleles and both Sry and Uty allele is listed in the Tet1, 2, 3 + Sry + Uty column. The increased efficiency of generating Tet1, 2, 3 triple targeted mES clones in this quintuple targeting experiment, compared to the triple targeting experiment (Table 1), is likely due to the use of short PCR products instead of plasmids that express sgRNAs. The much smaller size of pooled PCR products may lead to more efficient delivery into transfected cells.
Table S2. Potential Off Targets of Tet1 and Tet2 sgRNAs, Related to Table 2

<table>
<thead>
<tr>
<th>MatchName</th>
<th>Coordinate (mm9)</th>
<th>Strand</th>
<th>SEEDPAM</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tet1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tet1_1_TGG_3</td>
<td>chr10:62296293-62296308</td>
<td>-</td>
<td>ggcgcgtGTCAAGGAGGCTCAGGG</td>
<td>Tet1</td>
</tr>
<tr>
<td>Tet1_1_AGG_1</td>
<td>chr16:8891779-8891794</td>
<td>+</td>
<td>ggtttggGTCAAGGAGGCTCAGGG</td>
<td>33kb 3' of 1810013L24Rik</td>
</tr>
<tr>
<td>Tet1_1_AGG_2</td>
<td>chr18:75130318-75130333</td>
<td>-</td>
<td>gggccagGTCAAGGAGGCTCAGGG</td>
<td>9.4kb 5' of Lipg</td>
</tr>
<tr>
<td>Tet1_1_GGG_4</td>
<td>chr2:36287584-36287599</td>
<td>+</td>
<td>gtttagGTCAAGGAGGCTCAGGG</td>
<td>9.8kb 3' of Chfr339</td>
</tr>
<tr>
<td>Tet2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tet2_1_AGG_2</td>
<td>chr3:133148617-133148632</td>
<td>-</td>
<td>gaaagggCCAACAGATATCAGGG</td>
<td>Tet2</td>
</tr>
<tr>
<td>Tet2_1_AGG_1</td>
<td>chr2:120696599-120696614</td>
<td>+</td>
<td>gcaaggaCCAACAGATATCAGGG</td>
<td>Intron of Ubr1</td>
</tr>
<tr>
<td>Tet2_1_CGG_3</td>
<td>chr10:95206326-95206341</td>
<td>+</td>
<td>aggtagggCCAACAGATATCAGGG</td>
<td>2.4kb 5' of AK169506</td>
</tr>
<tr>
<td>Tet2_1_TGG_4</td>
<td>chr19:39098539-39098554</td>
<td>+</td>
<td>ccacctggCCAACAGATATCCTGG</td>
<td>Intron of Cyp2c55</td>
</tr>
<tr>
<td>Tet2_1_TGG_5</td>
<td>chr15:59188892-59188907</td>
<td>+</td>
<td>gagataaCCAACAGATATCCTGG</td>
<td>Intron of E430025E21Rik</td>
</tr>
</tbody>
</table>
Table S3. Oligonucleotides Used in This Study, Related to Experimental Procedures
Oligonucleotides used for cloning sgRNA expression vector

<table>
<thead>
<tr>
<th>Gene target</th>
<th>Direction</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>CACCGGCTGCTGCTGCAGGGAGCTCA</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>AAACCTGAGCTCCCTGACAGCAGCC</td>
<td></td>
</tr>
<tr>
<td>Tet2</td>
<td>F</td>
<td>CACCGAAAGTGCCAACAGATATCC</td>
</tr>
<tr>
<td>R</td>
<td>AAACGGATATCTGTGTTGGCACTTTTC</td>
<td></td>
</tr>
<tr>
<td>Tet3</td>
<td>F</td>
<td>CACCGAAGGAGGGGAAGAGTTTCTCG</td>
</tr>
<tr>
<td>R</td>
<td>AAACCGAGAACTCTTCCCCTCCTTC</td>
<td></td>
</tr>
<tr>
<td>Sry</td>
<td>F</td>
<td>CACCGCATTTATGCTGCTGCTCCCG</td>
</tr>
<tr>
<td>R</td>
<td>AAACCGGGGACCACACCATAAATGC</td>
<td></td>
</tr>
<tr>
<td>Uty</td>
<td>F</td>
<td>CACCGTTTCTTTTCTCATTACCTA</td>
</tr>
<tr>
<td>R</td>
<td>AAACTAGGTAATGAGGAAAAGAAAC</td>
<td></td>
</tr>
</tbody>
</table>

Oligonucleotides used for Suveryor assay and RFLP analysis

<table>
<thead>
<tr>
<th>Gene target</th>
<th>Direction</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tet1</td>
<td>F</td>
<td>TTGTTCCTCTCTCTCTGACTGC</td>
</tr>
<tr>
<td>R</td>
<td>TGATTGATCAAAATAGGCGCTGC</td>
<td></td>
</tr>
<tr>
<td>Tet2</td>
<td>F</td>
<td>CAGATGCTTAGGCCAATCAAG</td>
</tr>
<tr>
<td>R</td>
<td>AGAAGCAACACACATGAAGATG</td>
<td></td>
</tr>
<tr>
<td>Tet3</td>
<td>F</td>
<td>CCACCTCTGAGCGCAGAGTG</td>
</tr>
<tr>
<td>R</td>
<td>GATGAACACAGTTCCTGACAG</td>
<td></td>
</tr>
<tr>
<td>Sry</td>
<td>F</td>
<td>GTCTGTCTTTGTCTGTCTGTC</td>
</tr>
<tr>
<td>R</td>
<td>GGGTATTTCTCTCTGTGTTAGG</td>
<td></td>
</tr>
<tr>
<td>Uty</td>
<td>F</td>
<td>GAGTTTCCTTTGCGTTTACC</td>
</tr>
<tr>
<td>R</td>
<td>AATGAGCACTTTTCAGAGTGGG</td>
<td></td>
</tr>
</tbody>
</table>
Oligonucleotides used for making template for in vitro transcription

<table>
<thead>
<tr>
<th>Template</th>
<th>Direction</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cas9</td>
<td>F</td>
<td>TAATACGACTCACTATAGGGAGAATGGGACTATAAGGACCACGAC</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>GCGAGCTCTAGGAAATTTTTCAC</td>
</tr>
<tr>
<td>Tet1 sgRNA</td>
<td>F</td>
<td>TTAATACGACTCACTATAGGGCTGCTGTCAGGGAGC</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>AAAACGACCGACTCGGTGC</td>
</tr>
<tr>
<td>Tet2 sgRNA</td>
<td>F</td>
<td>TTAATACGACTCACTATAGGAAAGTGCCAACAGATTTC</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>AAAACGACCGACTCGGTGC</td>
</tr>
<tr>
<td>Tet3 sgRNA</td>
<td>F</td>
<td>TTAATACGACTCACTATAGGAAGGGAGGGAAGAGTTCTCG</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>AAAACGACCGACTCGGTGC</td>
</tr>
</tbody>
</table>

Oligonucleotides used for HDR-mediated repair through embryo injection

<table>
<thead>
<tr>
<th>Gene target</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tet1</td>
<td>aaagaaaaagcccatatatatacacaccttcggcagcagcacaagtggctgctgctcagggAatTCatggagactagttggaactctcctcccctcacttcagccctcgagcttctgc</td>
</tr>
<tr>
<td>Tet2</td>
<td>tcaactgcgtgacttaaggctctgctctcaagtgctcagaaacacgtaggtgaccacacacGAtTCaggtgcagaacggagacaccgcagccpapctgaggcagctcaagcgaacacaaagcaca</td>
</tr>
</tbody>
</table>
Chapter 3. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome-engineering

Hui Yang\textsuperscript{1,4}, Haoyi Wang\textsuperscript{1,4}, Chikdu S. Shivalila\textsuperscript{1,2,4}, Albert W. Cheng\textsuperscript{1,3}, Linyu Shi\textsuperscript{1}, Rudolf Jaenisch\textsuperscript{1,3,*}

1. Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
2. Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA
3. Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, MA 02142, USA
4. These authors contributed equally to this work
* Correspondence: jaenisch@wi.mit.edu

Published as:


*indicates equal contribution

HW, HY and CSS shared all experiments and analyses equally, except HY did all zygotic injections. AWC helped with computational analysis of off targets. LS helped with cell\textsuperscript{11} assay analysis and in vitro RNA synthesis. HW, HY, CSS and RJ designed and conceived of experiments.
The type II bacterial CRISPR/Cas system is a novel genome-engineering technology with the ease of multiplexed gene targeting. Here, we created reporter and conditional mutant mice by coinjection of zygotes with Cas9 mRNA and different guide RNAs (sgRNAs) as well as DNA vectors of different sizes. Using this one-step procedure we generated mice carrying a tag or a fluorescent reporter construct in the Nanog, the Sox2, and the Oct4 gene as well as Mecp2 conditional mutant mice. In addition, using sgRNAs targeting two separate sites in the Mecp2 gene, we produced mice harboring the predicted deletions of about 700 bps. Finally, we analyzed potential off-targets of five sgRNAs in gene-modified mice and ESC lines and identified off-target mutations in only rare instances.

Mice with specific gene modification are valuable tools for studying development and disease. Traditional gene targeting in embryonic stem (ES) cells, although suitable for generating sophisticated genetic modifications in endogenous genes, is complex and time-consuming (Capecchi, 2005). The production of genetically modified mice and rats has been greatly accelerated by novel approaches using direct injection of DNA or mRNA of site-specific nucleases into the one-cell-stage embryo, generating DNA double-strand breaks (DSB) at specified sequences leading to targeted mutations (Carbery et al., 2010; Geurts et al., 2009; Shen et al., 2013; Sung et al., 2013; Tesson et al., 2011; Wang et al., 2013). Coinjection of a single-stranded or double-stranded DNA template containing homology to the sequences flanking the DSB can produce mutant alleles with precise point mutations or DNA inserts (Brown et al., 2013; Cui et al., 2011; Meyer et al., 2010; Wang et al., 2013; Wefers et al., 2013). Recently, pronuclear injection of two pairs of ZFNs and two double-stranded donor vectors into rat fertilized eggs produced rat containing loxP-flanked (floxed) alleles (Brown et al., 2013). However, the complex and time-consuming design and generation of ZFNs and double-stranded donor vectors limit the application of this method.
CRISPR (clustered regularly interspaced short palindromic repeat) and Cas (CRISPR-associated) proteins function as the RNA-based adaptive immune system in bacteria and archaea (Horvath and Barrangou, 2010; Wiedenheft et al., 2012). The type II bacterial CRISPR/Cas system has been demonstrated as an efficient gene-targeting technology that facilitates multiplexed gene targeting (Cong et al., 2013; Wang et al., 2013). Because the binding of Cas9 is guided by the simple base-pair complementarities between the engineered single-guide RNA (sgRNA) and a target genomic DNA sequence, it is possible to direct Cas9 to any genomic locus by providing the engineered sgRNA (Cho et al., 2013; Cong et al., 2013; Gilbert et al., 2013; Hwang et al., 2013; Jinek et al., 2012; Jinek et al., 2013; Mali et al., 2013b; Qi et al., 2013; Wang et al., 2013).

Previously, we used the type II bacterial CRISPR/Cas system as an efficient tool to generate mice carrying mutations in multiple genes in one step (Wang et al., 2013). However, this study left a number of issues unresolved. For example, neither the efficiency of using the CRISPR/Cas gene-editing approach for the insertion of DNA constructs into endogenous genes nor its utility to create conditional mutant mice was clarified. Here, we report the one-step generation of mice carrying reporter constructs in three different genes as well as the derivation of conditional mutant mice. In addition, we performed an extensive off-target cleavage analysis and show that off-target mutations are rare in targeted mice and ES cells derived from CRISPR/Cas zygote injection.

Results

Targeted Insertion of Short DNA Fragments

In previous work, we introduced precise base pair mutations into the Tet1 and
Tet2 genes through homology directed repair (HDR)-mediated genome editing following coinjection of single-stranded mutant DNA oligos, sgRNAs, and Cas9 mRNA (Wang et al., 2013). To test whether a larger DNA construct could be inserted at the same DSBs at Tet1 exon 4 and Tet2 exon 3, we designed oligos containing the 34 bp loxP site and a 6 bp EcoRI site flanked by 60 bps sequences on each side adjoining the DSBs (Figure S1A available online). We coinjected Cas9 mRNA, sgRNAs, and single-stranded DNA oligos targeting both Tet1 and Tet2 into zygotes. The restriction fragment length polymorphism (RFLP) assay shown in Figure S1B identified 6 out of 15 tested embryos carrying the loxP site at the Tet1 locus, 8 carrying the loxP site at the Tet2 locus, and 3 had at least one allele of each gene correctly modified. The correct integration of loxP sites was confirmed by sequencing (Figure S1C). These results demonstrate that HDR-mediated repair can introduce targeted integration of 40 bp DNA elements efficiently through CRISPR/Cas-mediated genome editing (summarized in Table 1).

Mice with Reporters in the Endogenous Nanog, Sox2, and Oct4 Genes

Because the study of many genes and their protein products are limited by the availability of high-quality antibodies, we explored the potential of fusing a short epitope tag to an endogenous gene. We designed a sgRNA targeting the stop codon of Sox2 and a corresponding oligo to fuse the 42 bp V5 tag into the last codon (Figure 1A). After injection of the sgRNA, Cas9 mRNA, and the oligo into zygotes, in vitro differentiated blastocysts were explanted into culture to derive ES cells. PCR genotyping and sequencing identified 7 out of 16 ES cell lines carrying a correctly targeted insert (Figures 1B and 1C). Western blot analysis revealed a protein band at the predicted size
using V5 antibody in targeted ES cells but not in the control cells (Figure 1D). As expected from a correctly targeted and functional allele, Sox2 expression was seen in targeted blastocysts and ES cells using V5 antibody (Figures 1E and 1F). Twelve of 35 E13.5 embryos and live-born mice derived from injected zygotes carried the V5 tag correctly targeted into the Sox2 gene as indicated by PCR genotyping and sequencing (data not shown, Table 1). To assess whether a marker transgene could be inserted into an endogenous locus, we coinjected Cas9 mRNA, sgRNA, and a double-stranded donor vector that was designed to fuse a p2A-mCherry reporter with the last codon of the Nanog gene (Figure 2A). A circular donor vector was used to minimize random integrations. To assess toxicity and to optimize the concentration of donor DNA, we microinjected different amounts of Nanog-2A-mCherry vector. Injection with a high concentration of donor DNA (500 ng/ml) yielded mCherry-positive embryos with high efficiency, with most blastocysts being retarded, whereas injection with a lower donor DNA concentration (10 ng/ml) yielded mostly healthy blastocysts, most of which were mCherry-negative. When 200 ng/ml donor DNA was used, 75% (936/1,262) of the injected zygotes developed to blastocysts, 9% (86/936) of which were mCherry-positive (Figure 2C; Table S1). mCherry was mainly expressed in the inner cell mass (ICM), consistent with targeted integration of the mCherry transgene into the Nanog gene. We derived six ES cell lines from mCherry-positive blastocysts, four of which uniformly expressed mCherry with the signal disappearing upon cellular differentiation (Figure 2C). The other two lines showed variegated mCherry expression, with some colonies being mCherry positive and others negative (Figure S2A, Table S2) consistent with mosaic donor embryos, which would be expected if transgene insertion occurred later than the
zygote stage, as has been previously observed with ZNF and TALEN-mediated targeting (Brown et al., 2013; Cui et al., 2011; Wefers et al., 2013). Correct transgene integration in ES cell lines was confirmed by Southern blot analysis (Figure 2B). We also generated mice from injected zygotes. Southern blot analysis (Figures S2B and S2C) revealed that 7 out of 86 E13.5 embryos and live-born mice carried the mCherry transgene in the Nanog locus. One targeted mouse was mosaic (Table S2), because the intensity of targeted allele was lower than the wild-type allele (Figure S2B, #6). Two of the mice carried an additional randomly integrated transgene (Figure S2C, #3). As summarized in Tables 1 and S1, the efficiency of targeted insertion of the transgene was about 10% in blastocysts and mice derived from injected zygotes.

Finally, we designed sgRNA targeting the Oct4 3’ UTR, which was co-injected with a published donor vector designed to integrate the 3 kb transgene cassette (IRES-eGFP-loxP-Neo-loxP; Figure 2D) at the 3’ end of the Oct4 gene (Lengner et al., 2007). Blastocysts were derived from injected zygotes, inspected for GFP expression, and explanted to derive ES cells. About 20% (47/254) of the blastocysts displayed uniform GFP expression in the ICM region. Three of nine derived ES cell lines expressed GFP (Figure 2E), including one showed mosaic expression (Table S2). Three out of ten live-born mice contained the targeted allele (Table 1). Correct targeting in mice and ES cell lines was confirmed by Southern blot analysis (Figure 2F).

Conventionally, transgenic mice are generated by pronuclear instead of cytoplasmic injection of DNA. To optimize the generation of CRISPR/Cas9-targeted embryos, we compared different concentrations of RNA and the Nanog-mCherry or the Oct4-GFP DNA vectors as well as three different delivery modes: (1) simultaneous
injection of all constructs into the cytoplasm, (2) simultaneous injection of the RNA and the DNA into the pronucleus, and (3) injection of Cas9/sgRNA into the cytoplasm followed 2 hr later by pronuclear injection of the DNA vector. Table S1 shows that simultaneous injection of all constructs into the cytoplasm at a concentration of 100 ng/ml Cas9 RNA, 50 ng/ml of sgRNA and 200 ng/ml of vector DNA was optimal, resulting in 9% (86/936) to 19% (47/254) of targeted blastocysts. Similarly, the simultaneous injection of 5 ng/ml Cas9 RNA, 2.5 ng/ml of sgRNA, and 10 ng/ml of DNA vector into the pronucleus yielded between 9% (7/75) and 18% (13/72) targeted blastocysts. In contrast, the two-step procedure with Cas9 and sgRNA simultaneous injected into the cytoplasm followed 2 hr later by pronuclear injection of different concentrations of DNA vector yielded no or at most 3% (1/34) positive blastocysts. Thus, our results suggest that simultaneous injection of RNA and DNA into the cytoplasm or nucleus is the most efficient procedure to achieve targeted insertion.

**Conditional Mecp2 Mutant Mice**

We investigated whether conditional mutant mice can be generated in one step by insertion of two loxP sites into the same allele of the Mecp2 gene. To derive conditional mutant mice similar to those previously described using traditional homologous recombination methods in ES cells (Chen et al., 2001), we designed two sgRNAs targeting Mecp2 intron 2 (L1, L2) and two sgRNAs targeting intron 3 (R1, R2), as well as the corresponding loxP site oligos with 60 bp homology to sequences on each side surrounding each sgRNA-mediated DSB (Figure S3A). To facilitate detection of correct insertions, the oligos targeting intron 2 were engineered to contain a NheI restriction site
and the oligos targeting intron 3 to contain an EcoRI site in addition to the loxP sequences (Figures 3A and S3A). To determine the efficiency of single loxP site integration at the Mecp2 locus, we injected Cas9 mRNA and each single sgRNA and corresponding oligo into zygotes, which were cultured to the blastocyst stage and genotyped by the RFLP assay. As shown in Figure S3B, the L2 and R1 sgRNAs were more efficient in integrating the oligos with four out of eight embryos carrying the L2 oligo and two out of six embryos carrying the R1 oligo. Therefore, L2 and R1 sgRNAs and the corresponding oligos were chosen for the generation of a floxed allele (Figure3A).

A total of 98 E13.5 embryos and mice were generated from zygotes injected with Cas9 mRNA, sgRNAs, and DNA oligos targeting the L2 and R1 sites. Genomic DNA was digested with both NheI and EcoRI, and analyzed by Southern blot using exon 3 and exon 4 probes (Figures 3A and 3B). The L2 and R1 oligos contained, in addition to the loxP site, different restriction sites (NheI or EcoRI). Thus, single loxP site integration at L2 or R1 will produce either a 3.9 kb or a 2 kb band, respectively, when hybridized with the exon3 probe (Figures 3A and B). We found that about 50% (45/98) of the embryos and mice carried a loxP site at the L2 site and about 25% (25/98) at the R1 site. Importantly, integration of both loxP sites on the same DNA molecule, generating a floxed allele, produces a 700 bp band as detected by exon 3 probe hybridization (Figures 3A and 3B). RFLP analysis, sequencing (Figures S4A and S4B) and Southern blot analysis (Figure 3B) showed that 16 out of the 98 mice tested contained two loxP sites flanking exon 3 on the same allele. Table 2 summarizes the frequency of all alleles and shows that the overall insertion frequency of an L2 or R1 insertion was slightly higher in
females (21/38) than in males (28/60) consistent with the higher copy number of the X-linked Mecp2 gene in females. To confirm that the floxed allele was functional, we used genomic DNA for in vitro Cre-mediated recombination. Upon Cre treatment, both the deletion and circular products were detected by PCR in targeted mice, but not in DNA from wild-type mice (Figure 3C). The PCR products were sequenced and confirmed the precise Cre-loxP-mediated recombination (Figure S4C).

We noticed that some pups carried large deletions but no loxP insertions, raising the possibility that two cleavage events may generate defined deletions. To confirm this notion, we coinjected Cas9 mRNA, Mecp2-L2, and R1 sgRNAs but without oligos. PCR genotyping and sequencing (Figures 3D and 3E) revealed that 8 out of 23 mice carried deletions of about 700 bp spanning the L2 and R1 sites removing exon 3. This was confirmed by Southern analysis (data not shown). Because DNA breaks are repaired through the nonhomologous end joining (NHEJ) pathway, the ends of the breaks are different in different deletion alleles (Figure 3E).

Mosaicism

As mentioned above, we noted that some animals were mosaic for the targeted insertion. We decided to characterize the frequency of mosaicism in Mecp2-targeted mice by Southern blot analysis. Because Mecp2 is an X-linked gene, in males more than one allele and in females more than two different alleles suggest mosaicism, which would be expected if integration occurred after the zygote stage. For example, as shown in Figure 3B, female mouse #2 contained three different alleles (one WT allele, one floxed allele, and one L2-loxP allele), and female mouse #4 contained four different alleles (one
WT allele, one floxed allele, one L2-loxP allele, and one R1-loxP allele). Male mouse #5 contained two different alleles, with each allele carrying a single loxP site (Figure 3B). We identified eight mosaics out of 16 mice containing a Mecp2 floxed allele. The frequency of mosaicism among 49 embryos and mice containing loxP site was about 40% (20/49) (Table S2). Because Southern blot analysis cannot detect small in-del mutations caused by NHEJ repair, it is possible that this underestimates the overall mosaicism frequency.

Off-Target Analysis

Two recent studies identified a high level of off-target cleavage in human cell lines using the CRISPR/Cas system, with Cas9-targeting specificity being shown to tolerate small numbers of mismatches between sgRNA and target DNA in a sequence and position-dependent manner (Fu et al., 2013; Hsu et al., 2013). Similarly, using a transcription-based method, Mali et al. (2013a) reported that Cas9/sgRNA binding could tolerate up to three mismatches.

We characterized potential off-target (OT) mutations in mice and ES cell lines derived from zygotes injected with Cas9 and sgRNAs targeting the Sox2, the Nanog, the Oct4, and the Mecp2 gene. We identified all genomic loci containing up to three or four base pair mismatches compared to the 20 bp sgRNA coding sequence (Table S3). We amplified all 13 potential OT sites of Sox2 sgRNA in six mice and four ES cell lines carrying the Sox2-V5 allele and tested for potential off target mutations using the Surveyor assay. No mutation was detected in any locus. When nine Nanog sgRNA potential OT sites (including seven genomic loci containing four base pair mismatches in
the PAM distal region) were tested in five correctly targeted mice and four targeted ES cell lines, mutations were found in seven samples at OT1 (Table S3). Since Nanog OT1 has only one base pair difference at the very 5' end of the sgRNA (position 20, numbered 1–20 in the 3' to 5' direction of sgRNA target site), it may not be surprising to find such a high frequency of mutations at this locus. In contrast, no off-target mutation was seen in any other Nanog OTs, which contain three or four base pair difference. For Oct4, we tested all 11 OT sites containing up to three base pair mismatches in three targeted mice and three targeted ES cell lines. Mutations were found in four out of six samples at OT1, which has only one base pair mismatch at position 19, whereas no off-target mutation was identified in any other Oct4 OTs, which contain two or three base pair mismatches (Table S3). Finally, four potential off-targets sites for Mecp2 L2, and ten sites for Mecp2 R1 were analyzed in ten mice carrying a Mecp2 floxed allele. Only one off-target mutation was identified in one mouse at the Mecp2 R1 OT2 (Table S3). In summary, we tested all potential off-target sites differing by up to three or four base pairs in 35 mice or ES cell lines and only identified mutations in one off-target site for Nanog (OT1 7/9 samples, one mismatch at position 20), Oct4 (OT1 4/6 samples, 1 mismatch at position 19), and Mecp2 (R1-OT2 1/10 samples, two mismatches at positions 7 and 20). No off target mutation was identified in any genomic locus containing three base pair mismatches. Thus, although the off-target mutation rate was lower than what had been observed in the previous studies using cultured human cancer cell lines, our results are consistent with the conclusion that two or more interspaced mismatches dramatically reduce Cas9 cleavage (Fu et al., 2013; Hsu et al., 2013).
DISCUSSION

In this study, we demonstrate that CRISPR/Cas technology can be used for efficient one-step insertions of a short epitope or longer fluorescent tags into precise genomic locations, which will facilitate the generation of mice carrying reporters in endogenous genes. Mice and embryos carrying reporter constructs in the Sox2, the Nanog and the Oct4 gene were derived from zygotes injected with Cas9 mRNA, sgRNAs, and DNA oligos or vectors encoding a tag or a fluorescent marker. Moreover, microinjection of two MeCP2-specific sgRNAs, Cas9 mRNA, and two different oligos encoding loxP sites into fertilized eggs allowed for the one-step generation of conditional mutant mice. In addition, we show that the introduction of two spaced sgRNAs targeting the MeCP2 gene can produce mice carrying defined deletions of about 700 bp. Though all RNA and DNA constructs were injected into the cytoplasm or nucleus of zygotes, the gene modification events could happen at the one-cell stage or later. Indeed, Southern analyses revealed mosaicism in 17% (1/6) to 40% (20/49) of the targeted mice and ES cell lines indicating that the insertion of the transgenes had occurred after the zygote stage (Table S2).

Our previous experiments (Wang et al., 2013) demonstrated an efficiency of CRISPR/Cas sgRNA-mediated cleavage that was high enough to allow for the one-step production of engineered mice up to 90% of which carried homozygous mutations in two genes (four mutant alleles). The results reported here show that the sgRNA-mediated DSBs occur at a significantly higher frequency than insertion of exogenous DNA sequences. Therefore, the allele not carrying the insert will likely be mutated as a consequence of NHEJ-based gene disruption. Thus, the reporter allele would need to be
segregated away from the mutant allele in order to produce mice carrying a reporter as well as a wild-type allele.

Two recent studies reported a high off-target mutation rate in CRISPR/Cas9-transfected human cell lines (Fu et al., 2013; Hsu et al., 2013). We analyzed the off-target rate for five different sgRNAs and identified cleavage of Nanog OT1, Oct4 OT1, and Mecp2 R1 OT2. Nanog OT1 has only one base pair difference from the targeting sequence at the extreme 5’ end (position 20), whereas Oct4 OT1 contains one base pair mismatch at position 19; and Mecp2 R1 OT2 has one base pair mismatch at position 20, and one mismatch at position 7. Thus, the only off-target mutations in 2 bp mismatch targets were seen when one of the mismatches was at the distal 5’ end. This result is consistent with previous findings that Cas9 can catalyze DNA cleavage in the presence of single-base mismatches in the PAM distal region (Cong et al., 2013; Hsu et al., 2013; Jiang et al., 2013; Jinek et al., 2012). No mutations were detected in 42 potential OTs of Sox2, Nanog, Oct4, or Mecp2 containing 3 or 4 bp mismatches in a total of 35 mice and ES cell lines tested, consistent with the observation that three or more interspaced mismatches dramatically reduce Cas9 cleavage (Hsu et al., 2013). Thus, for designing the most suitable target sequences for generating Cas9 cleavage-mediated, genetically modified mice, it is important to avoid targeting sites that have only one or two mismatches in other genomic loci. This is particularly important for mismatches that are at the PAM distal region.

We consider several possibilities to explain the lower off-target cleavage rate seen in animals derived from manipulated zygotes and the results reported for CRISPR/Cas-treated human cell lines (Fu et al., 2013; Hsu et al., 2013). (1) In our study, the off-target
mutagenesis was based on the analysis of a "clonal genome" in animals derived from a single manipulated zygote in contrast to the previous reports that analyzed heterogeneous cell populations. The surveyor assay, based upon extensive PCR amplification, may identify any mutation, even very rare alleles that may be present in the heterogeneous population. (2) The transformed human cell lines may have different DNA damage responses resulting in a different mutagenesis rate than the normal one-cell embryo. (3) In our experiments CRISPR/Cas was injecting as short-lived RNA in contrast to Fu et al. (2013) and Hsu et al. (2013), who used DNA plasmid transfection, which may express the Cas9/sgRNA for longer time periods leading to more extensive cleavage. Thus, our data suggest high specificity of the CRISPR/Cas9 system for gene editing in early embryos aimed at generating gene-modified mice. Never the less, future characterization of off-target mutagenesis of CRISPR/Cas system using whole-genome sequencing would be highly informative and may allow designing sgRNAs with higher specificity.

In summary, CRISPR/Cas-mediated genome editing represents an efficient and simple method of generating sophisticated genetic modifications in mice such as conditional alleles and endogenous reporters in one step. The principles described in this study could be directly adapted to other mammalian species, opening the possibility of sophisticated genome-engineering in many species where ES cells are not available.

EXPERIMENTAL PROCEDURES

Production of Cas9 mRNA and sgRNA

Bicistronic expression vector px330 expressing Cas9 and sgRNA (Cong et al., 2013) was digested with BbsI and treated with Antarctic Phosphatase, and the linearized vector was
gel purified. A pair of oligos (Table S4) for each targeting site was annealed, phosphorylated, and ligated to the linearized vector. T7 promoter was added to Cas9 coding region by PCR amplification using primer Cas9 F and R (Table S4). T7-Cas9 PCR product was gel purified and used as the template for in vitro transcription (IVT) using mMESSAGE mMACHINE T7 ULTRA kit (Life Technologies). T7 promoter was added to sgRNAs template by PCR amplification using primer listed in Table S4. The T7-sgRNA PCR product was gel purified and used as the template for IVT using MEGAscript T7 kit (Life Technologies). Both the Cas9 mRNA and the sgRNAs were purified using MEGAclear kit (Life Technologies) and eluted in RNase-free water.

**Single-Stranded and Double-Stranded DNA Donors**

All single-stranded oligos were ordered as Ultramer DNA oligos from Integrated DNA Technologies. Nanog-2A-mCherry vector was modified from previously published targeting vector Nanog-2A-mCherry-PGK-Neo (Faddah et al., 2013). Nanog-2A-mCherry-PGK-Neo was digested with PacI and Ascl to drop out the PGK-Neo cassette; the 9.7 kb fragment was gel purified and blunt-ended using T4 DNA polymerase (New England Biolabs) then self-ligated using T4 DNA ligase (New England Biolabs). Oct4-IRES-eGFP-PGK-Neo vector is previously published (Lengner et al., 2007).

**Suveryor Assay and RFLP Analysis for Genome Modification**

Suveryor assay was performed as described (Guschin et al., 2010). Genomic DNA from targeted and control mice or embryos was extracted and PCR was performed using gene-specific primers (Table S4) under the following conditions: 95C for 5 min; 35 x (95C for
30 s, 60C for 30 s, 68C for 40 s); 68C for 2 min; hold at 4C. PCR products were then denatured, annealed, and treated with Suveryor nuclease (Transgenomic). DNA concentration of each band was measured on an ethidium-bromide-stained 10% acrylamide Criterion TBE gel (BioRad) and quantified using ImageJ software. For RFLP analysis, 10ul of Tet1, Tet2, Mecp2-R1, R2 PCR product was digested with EcoRI, 10 ul of Mecp2-L1, and L2 PCR product was digested with Nhel. Digested DNA was separated on an ethidium-bromide-stained agarose gel (2%). For sequencing, PCR products were cloned using the Original TA Cloning Kit (Invitrogen), and mutations were identified by Sanger sequencing.

**One-Cell Embryo Injection**

All animal procedures were performed according to NIH guidelines and approved by the Committee on Animal Care at MIT. B6D2F1 (C57BL/6 X DBA2) female mice and ICR mouse strains were used as embryo donors and foster mothers, respectively. Superovulated female B6D2F1 mice (7–8 weeks old) were mated to B6D2F1 stud males, and fertilized embryos were collected from oviducts. Different concentrations of Cas mRNA, sgRNA, and oligos or plasmid vectors were mixed and injected into the cytoplasm or pronucleus of fertilized eggs with well-recognized pronuclei in M2 medium (Sigma). The injected zygotes were cultured in KSOM with amino acids at 37C under 5% CO2 in air until blastocyst stage by 3.5 days. Thereafter, 15–25 blastocysts were transferred into uterus of pseudopregnant ICR females at 2.5 dpc.

**Southern Blotting**
Genomic DNA was separated on a 0.8% agarose gel after restriction digests with the appropriate enzymes, transferred to a nylon membrane (Amersham) and hybridized with 32P random primer (Stratagene)-labeled probes. Between hybridizations, blots were stripped and checked for complete removal of radioactivity before rehybridization with a different probe.

**In Vitro Cre Recombination**

A 20 ul reaction containing 1mg of genomic DNA and 10 units of recombinant Cre recombinase (New England Biolabs) in 1x buffer was incubated at 37C for 1 hr. For all targets, 1ul of the Cre reaction mix was used as template for PCR reactions with gene-specific primers. For each target, primers DF and DR were used for detecting the deletion products, and primers CF and CR were used to detect the circle product. All products were sequenced.

**Immunostaining and Western Blot Analysis**

For immunostaining, cells in 24-well were fixed in PBS supplemented with 4% paraformaldehyde for 15 min at room temperature (RT). The cells were then permeabilized using 0.2% Triton X-100 in PBS for 15 min at RT. The cells were blocked for 30 min in 1% BSA in PBS. Primary antibody against V5 (ab9137, abcam) was diluted in the same blocking buffer and incubated with the samples overnight at 4C. The cells were treated with a fluorescently coupled secondary antibody and then incubated for 1 hr at RT. The nuclei were stained with Hoechst 33342 (Sigma) for 5 min at RT. For western blot, Cell pellets were lysed on ice in Laemmli buffer (62.5 mM Tris-HCl [pH 6.8], 2%
sodium dodecyl sulfate, 5% beta-mercaptoethanol, 10% glycerol, and 0.01% bromophenol blue) for 30 min in presence of protease inhibitors (Roche Diagnostics), boiled for 5–7 min at 100°C, and subjected to western blot analysis. Primary antibodies: V5 (1:1,000, ab9137, Abcam), beta-actin (1:2,000). Blots were probed with anti-goat, or anti-rabbit IgG-HRP secondary antibody (1:10,000) and visualized using ECL detection kit (GE Healthcare).

ESC Derivation and Differentiation

Morulas or blastocysts were selected to generate ES cell lines. The zona pellucida was removed using acid Tyrode solution. Each embryo was transferred into one well of a 96-well plate seeded with ICR embryonic fibroblast feeders in ESC medium supplemented with 20% knockout serum replacement, 1,500 U/ml leukemia inhibitory factor (LIF), and 50mM of the MEK1 inhibitor (PD098059). After 4–5 days in culture, the colonies were trypsinized and transferred to a 96-well plate with a fresh feeder layer in fresh medium. Clonal expansion of the ESCs proceeded from 48-well plates to 6-well plates with feeder cells and then to 6-well plates for routine culture. For ESC differentiation, cells were harvested by trypsinization and transferred to bacterial culture dishes in the ES medium without or LIF. After 3 days, aggregated cells were plated onto gelatin-coated tissue culture dishes and incubated for another 3 days.

Prediction of Potential Off-Targets

Potential off-targets were predicted by searching the mouse genome (mm9) for matches to the 20 nt sgRNA sequence allowing for up to four mismatches (Nanog) or three
mismatches (Sox2, Oct4, MeCP2-L2, and MeCP2-R1) followed by NGG PAM sequence. Matches were ranked first by ascending number of mismatches then by ascending distance from the PAM sequence.

ACKNOWLEDGMENTS

We thank Ruth Flannery and Kibibi Ganz for support with animal care and experiments. We thank Jaenisch lab members for helpful discussions on the manuscript. We are also grateful to Dina A. Faddah for help with editing the manuscript. A.W.C. is supported by a Croucher scholarship. R.J. is an adviser to Stemgent and a cofounder of Fate Therapeutics. This work was supported by NIH grants HD 045022 and R37CA084198 to R.J.
Figure 1. One-Step Generation of the Sox2-V5 Allele

(A) Schematic of the Cas9/sgRNA/oligo-targeting site at the Sox2 stop codon. The sgRNA coding sequence is underlined, capitalized, and labeled in red. The protospacer-adjacent motif (PAM) sequence is labeled in green. The stop codon of Sox2 is labeled in orange. The oligo contained 60 bp homologies on both sides flanking the DSB. In the oligo donor sequence, the V5 tag sequence is labeled as a green box. PCR primers (SF, V5F, and SR) used for PCR genotyping are shown as red arrowheads.
(B) Top: PCR genotyping using primers V5F and SR produced bands with correct size in targeted ES samples T1 to T5, but not in WT sample. Bottom: PCR genotyping using primers SF and SR produced slightly larger products, indicating the 42 bp V5 tag sequence was integrated. T1 only contain larger product, suggesting either both alleles were targeted, or one allele failed to amplify.

(C) PCR products using primers SF and SR were cloned into plasmid and sequenced. Sequence across the targeting region confirmed correct fusion of V5 tag to the last codon of Sox2.

(D) Western blot analysis identified Sox2-V5 protein using V5 antibody in ES cells containing Sox2-V5 allele. Beta-actin was shown as the loading control. Because beta-actin and Sox2-V5 run at the same size, the samples for the V5 signal and beta-actin were run in parallel on the same gel.

(E) Immunostaining of targeted blastocyst using V5 antibody showed signal in ICM. Scale bar, 50 mm.

(F) Immunostaining of targeted ES cells using V5 antibody showed uniform Sox2 expression. Scale bar, 100 mm.
Figure 2. One-Step Generation of an Endogenous Reporter Allele
(A) Schematic overview of strategy to generate a Nanog-mCherry knockin allele. The sgRNA coding sequence is underlined, capitalized, and labeled in red. The protospacer-adjacent motif (PAM) sequence is labeled in green. The stop codon of Nanog is labeled in orange. The homologous arms of the donor vector are indicated as HA-L (2 kb) and HA-R (3 kb). The restriction enzyme used for Southern blot analysis is shown, and the Southern blot probes are shown as red boxes.

(B) Southern analysis of Nanog-mCherry targeted allele. NcoI-digested genomic DNA was hybridized with 3' external probe. Expected fragment size: WT (wild-type) = 11.5 kb, T (targeted) = 5.6 kb. The blot was then stripped and hybridized with mCherry internal probe. Expected fragment size: WT = N/A, T = 6.6 kb.

(C) Nanog-mCherry targeted blastocysts showed expression in ICM. Mouse ES cell lines derived from targeted blastocysts remain mCherry positive, and the mCherry expression disappear upon differentiation. Scale bar, 100mm.

(D) Schematic overview of strategy to generate an Oct4-eGFP knockin allele. The sgRNA coding sequence is underlined, capitalized, and labeled in red. The PAM sequence is labeled in green. The homologous arms of the donor vector are indicated as HA-L (4.5 kb) and HA-R (2 kb). The IRES-eGFP transgene is indicated as a green box, and the PGK-Neo cassette is indicated as a gray box. The restriction enzyme used for Southern blot analysis is shown, and the Southern blot probes are shown as red boxes.

(E) Oct4-eGFP targeted blastocysts showed expression in ICM. Scale bar, 50mm. Mouse ES cell lines derived from targeted blastocysts remain GFP positive. Scale bar, 100mm.

(F) Southern analysis of Oct4-eGFP targeted allele. Southern analysis of Oct4-eGFP targeted allele. HindIII-digested genomic DNA was hybridized with 3’ external probe. Expected fragment size: WT = 9kb, Targeted = 7.2 kb. The blot was then stripped and hybridized with eGFP internal probe. Expected fragment size: WT = N/A, Targeted = 7.2 kb. See also Figure S2.
(A) Schematic of the Cas9/sgRNA/oligo targeting sites in Mecp2 intron 2 and intron 3. The sgRNA coding sequence is underlined, capitalized, and labeled in red. The PAM sequence is labeled in green. In the oligo donor sequence, the loxP site is indicated as an

Figure 3. One-Step Generation of a Mecp2 Floxed Allele

(A) Schematic of the Cas9/sgRNA/oligo targeting sites in Mecp2 intron 2 and intron 3. The sgRNA coding sequence is underlined, capitalized, and labeled in red. The PAM sequence is labeled in green. In the oligo donor sequence, the loxP site is indicated as an
orange box, and the restriction site sequences are in bold and capitalized. The oligo contained 60 bp homologies on both sides flanking the DSB. Restriction enzymes used for RFLP and Southern blot analysis are shown, and the Southern blot probes are shown as red boxes.

(B) Southern analysis of targeted alleles. Data of five mice are shown. EcoRI/NheI-digested genomic DNA was hybridized with the exon3 probe. Expected fragment size: WT = 5.2 kb, 2loxP = 0.7 kb, L2-loxP = 3.9 kb, and R1-loxP = 2 kb. The blot was then stripped and hybridized with the exon 4 probe. Expected fragment size: WT = 5.2 kb, 2loxP = 3.2 kb, L2-loxP = 3.9 kb, and R1-loxP = 3.2 kb. The sequence of the floxed allele is shown in Figure S4B.
Table 1. Mice with Reporters in the Endogenous Genes

<table>
<thead>
<tr>
<th>Donor</th>
<th>Blastocysts/Injected Zygotes</th>
<th>Targeted Blastocysts/Total</th>
<th>Targeted ESCs/Total</th>
<th>Transferred embryos (recipients)</th>
<th>Knockin pre- and postnatal mice/Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tet1-loxP + Tet2-loxP</td>
<td>65/69</td>
<td>6/15</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Tet2-loxP</td>
<td>5/15</td>
<td></td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soc2-V5</td>
<td>414/498</td>
<td>ND</td>
<td>7/16</td>
<td>1200</td>
<td>12/35</td>
</tr>
<tr>
<td>Nanog-mCherry</td>
<td>936/1262</td>
<td>ND</td>
<td>415 (21)</td>
<td>7/88</td>
<td></td>
</tr>
<tr>
<td>Oct4-GFP</td>
<td>254/345</td>
<td>47/254</td>
<td>3/9</td>
<td>1000 (4)</td>
<td>3/10</td>
</tr>
</tbody>
</table>

Table 1. Cas9 mRNA, sgRNAs targeting Tet1, Tet2, Sox2, Nanog, or Oct4, and single-stranded DNA oligos or double-stranded donor vectors were injected into fertilized eggs. Targeted blastocysts were identified by RFLP or fluorescence of reporters. The blastocysts derived from injected embryos were derived ES cell lines or transplanted into foster mothers and E13.5 embryos and postnatal mice were obtained and genotyped. ND, not determined.

Table 2. Conditional Mecp2 Mutant Mice

<table>
<thead>
<tr>
<th>Donor</th>
<th>Blastocysts/ Injected Zygotes</th>
<th>Transferred Embryos (Recipients)</th>
<th>Pre- and Postnatal Mice with loxP/Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mecp2-L2 + Mecp2-R1</td>
<td>367/451</td>
<td>360 (18)</td>
<td>Male</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>29/66</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>29/66</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>49/98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>59/98</td>
</tr>
</tbody>
</table>

Table 2. Cas9 mRNA, sgRNAs targeting Mecp2-L2 and Mecp2-R1, and single-stranded DNA oligos were injected into fertilized eggs. The blastocysts derived from the injected embryos were transplanted into foster mothers and pre- and postnatal mice were genotyped.

(a) Total mice containing loxP site integration in the genome.
(b) Mice containing loxP site integrated at L2 site.
(c) Mice containing loxP site integrated at R1 site.
(d) These male mice were mosaic.
REFERENCES


Table S1. Efficiency of Generation of Reporter Embryos by Cytoplasm and Pronuclear Injection, Related to Figure 2 and Table 1

<table>
<thead>
<tr>
<th>Donor</th>
<th>Dose of Cas9/sgRNA (ng/µl)</th>
<th>Dose of Donor vector (ng/µl)</th>
<th>Injected zygotes</th>
<th>Targeted Blastocysts / Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>One-step injection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nanog-mCherry</td>
<td>100/50 (Cyto)</td>
<td>500 (Cyto)</td>
<td>186</td>
<td>1/81</td>
</tr>
<tr>
<td>Nanog-mCherry</td>
<td>100/50 (Cyto)</td>
<td>200 (Cyto)</td>
<td>1262</td>
<td>86/936</td>
</tr>
<tr>
<td>Nanog-mCherry</td>
<td>100/50 (Cyto)</td>
<td>50 (Cyto)</td>
<td>402</td>
<td>7/308</td>
</tr>
<tr>
<td>Nanog-mCherry</td>
<td>100/50 (Cyto)</td>
<td>10 (Cyto)</td>
<td>333</td>
<td>1/278</td>
</tr>
<tr>
<td>Oct4-GFP</td>
<td>100/50 (Cyto)</td>
<td>200 (Cyto)</td>
<td>345</td>
<td>47/254</td>
</tr>
<tr>
<td>Nanog-mCherry</td>
<td>5/2.5 (Nuc)</td>
<td>10 (Nuc)</td>
<td>98</td>
<td>7/75</td>
</tr>
<tr>
<td>Oct4-GFP</td>
<td>5/2.5 (Nuc)</td>
<td>10 (Nuc)</td>
<td>105</td>
<td>13/72</td>
</tr>
<tr>
<td><strong>Two-step injection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nanog-mCherry</td>
<td>100/50 (Cyto)</td>
<td>50 (Nuc)</td>
<td>45</td>
<td>0/0</td>
</tr>
<tr>
<td>Nanog-mCherry</td>
<td>100/50 (Cyto)</td>
<td>10 (Nuc)</td>
<td>91</td>
<td>1/34</td>
</tr>
<tr>
<td>Nanog-mCherry</td>
<td>100/50 (Cyto)</td>
<td>2 (Nuc)</td>
<td>85</td>
<td>1/68</td>
</tr>
</tbody>
</table>

Cas9 mRNA, sgRNAs targeting Nanog, or Oct4, and double stranded donor vectors were injected into cytoplasm or pronuclei of zygotes. In one-step injection, the RNA and the DNA were simultaneously injected into the cytoplasm or pronucleus. In two-step injection, Cas9/sgRNA were injected into the cytoplasm followed 2 hours later by pronuclear injection of the DNA vector. Targeted blastocysts were identified by fluorescence of reporters. Cyto, cytoplasm; Nuc, nucleus.
Table S2. Mosaicism in Targeted Mice, Related to Figure S2 and Tables 1 and 2

<table>
<thead>
<tr>
<th>Donor</th>
<th>Mice/Mosaic</th>
<th>Total targeted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nanog-Cherry</td>
<td>1/7</td>
<td>3/13</td>
</tr>
<tr>
<td>ESCs</td>
<td>2/6</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>0/3</td>
</tr>
<tr>
<td>Oct4-EGFP</td>
<td>0/3</td>
<td>1/6</td>
</tr>
<tr>
<td>ESCs</td>
<td>1/3</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1/6</td>
</tr>
<tr>
<td>Mecp2-L2 +</td>
<td>Male 11/28</td>
<td>20/49*</td>
</tr>
<tr>
<td>Mecpe-R1</td>
<td>Female 9/21</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Targeted mice or ESCs were identified by RFLP, Southern bolt or Sequencing. The frequency of mosaicism in targeted mice was determined by fluorescent reporter or Southern blot analysis. (a) These 49 mice contain at least one loxP integration.
### Table S3. Off-Target Analysis, Related to Figures 1, 2, 3 and Tables 1 and 2

<table>
<thead>
<tr>
<th>Site name</th>
<th>Sequence</th>
<th>Indel mutation frequency (Mutant/Total)</th>
<th>Coordinate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target_Sox2_Stop</td>
<td>TGCCCCCTGTCACATGTGAGGG</td>
<td>/</td>
<td>chr3: 34560278-34550300</td>
</tr>
<tr>
<td>OT1_Sox2_Stop</td>
<td>TGGCCCTGTCACATGTGAGGG</td>
<td>0/10</td>
<td>chr4: 126636377-126636399</td>
</tr>
<tr>
<td>OT2_Sox2_Stop</td>
<td>TGCCCCCTGTCACATGTGAGGG</td>
<td>0/10</td>
<td>chr14: 58080951-58080963</td>
</tr>
<tr>
<td>OT3_Sox2_Stop</td>
<td>TGCCCCCTGTCACATGTGAGGG</td>
<td>0/10</td>
<td>chr1: 136641174-136641196</td>
</tr>
<tr>
<td>OT4_Sox2_Stop</td>
<td>TGCCCCCTGTCACATGTGAGGG</td>
<td>0/10</td>
<td>chr9: 69081989-69081994</td>
</tr>
<tr>
<td>OT5_Sox2_Stop</td>
<td>TGCCCCCTGTCACATGTGAGGG</td>
<td>0/10</td>
<td>chr1: 130633965-130633967</td>
</tr>
<tr>
<td>OT6_Sox2_Stop</td>
<td>TGCCCCCTGTCACATGTGAGGG</td>
<td>0/10</td>
<td>chr18: 61611640-61611662</td>
</tr>
<tr>
<td>OT7_Sox2_Stop</td>
<td>TGCCCCCTGTCACATGTGAGGG</td>
<td>0/10</td>
<td>chr5: 136841014-136841036</td>
</tr>
<tr>
<td>OT8_Sox2_Stop</td>
<td>TGCCCCCTGTCACATGTGAGGG</td>
<td>0/10</td>
<td>chr4: 141162434-141162456</td>
</tr>
<tr>
<td>OT9_Sox2_Stop</td>
<td>TGCCCCCTGTCACATGTGAGGG</td>
<td>0/10</td>
<td>chr3: 34560278-34550300</td>
</tr>
<tr>
<td>OT11_Sox2_Stop</td>
<td>TGCCCCCTGTCACATGTGAGGG</td>
<td>0/10</td>
<td>chr4: 141162434-141162456</td>
</tr>
<tr>
<td>OT12_Sox2_Stop</td>
<td>TGCCCCCTGTCACATGTGAGGG</td>
<td>0/10</td>
<td>chr5: 136841014-136841036</td>
</tr>
<tr>
<td>OT13_Sox2_Stop</td>
<td>TGCCCCCTGTCACATGTGAGGG</td>
<td>0/10</td>
<td>chr5: 136841014-136841036</td>
</tr>
<tr>
<td>Target_Nanog_Stop</td>
<td>CTTAGGGCTTACATTTCACTTG</td>
<td>7/9</td>
<td>chr6: 122663559-122663581</td>
</tr>
<tr>
<td>OT1_Nanog_Stop</td>
<td>CTTAGGGCTTACATTTCACTTG</td>
<td>3/9</td>
<td>chr6: 122663559-122663581</td>
</tr>
<tr>
<td>OT2_Nanog_Stop</td>
<td>CTTAGGGCTTACATTTCACTTG</td>
<td>3/9</td>
<td>chr6: 122663559-122663581</td>
</tr>
<tr>
<td>OT3_Nanog_Stop</td>
<td>CTTAGGGCTTACATTTCACTTG</td>
<td>3/9</td>
<td>chr6: 122663559-122663581</td>
</tr>
<tr>
<td>OT4_Nanog_Stop</td>
<td>CTTAGGGCTTACATTTCACTTG</td>
<td>3/9</td>
<td>chr6: 122663559-122663581</td>
</tr>
<tr>
<td>OT5_Nanog_Stop</td>
<td>CTTAGGGCTTACATTTCACTTG</td>
<td>3/9</td>
<td>chr6: 122663559-122663581</td>
</tr>
<tr>
<td>OT6_Nanog_Stop</td>
<td>CTTAGGGCTTACATTTCACTTG</td>
<td>3/9</td>
<td>chr6: 122663559-122663581</td>
</tr>
<tr>
<td>OT7_Nanog_Stop</td>
<td>CTTAGGGCTTACATTTCACTTG</td>
<td>3/9</td>
<td>chr6: 122663559-122663581</td>
</tr>
<tr>
<td>OT8_Nanog_Stop</td>
<td>CTTAGGGCTTACATTTCACTTG</td>
<td>3/9</td>
<td>chr6: 122663559-122663581</td>
</tr>
<tr>
<td>OT9_Nanog_Stop</td>
<td>CTTAGGGCTTACATTTCACTTG</td>
<td>3/9</td>
<td>chr6: 122663559-122663581</td>
</tr>
<tr>
<td>Target_Oct4_Stop</td>
<td>GCTAGGGCTTACATTTCACTTG</td>
<td>4/6</td>
<td>chr17: 35647634-35647656</td>
</tr>
<tr>
<td>OT1_Oct4_Stop</td>
<td>GCTAGGGCTTACATTTCACTTG</td>
<td>4/6</td>
<td>chr17: 35647634-35647656</td>
</tr>
<tr>
<td>OT2_Oct4_Stop</td>
<td>GCTAGGGCTTACATTTCACTTG</td>
<td>4/6</td>
<td>chr17: 35647634-35647656</td>
</tr>
<tr>
<td>OT3_Oct4_Stop</td>
<td>GCTAGGGCTTACATTTCACTTG</td>
<td>4/6</td>
<td>chr17: 35647634-35647656</td>
</tr>
<tr>
<td>OT4_Oct4_Stop</td>
<td>GCTAGGGCTTACATTTCACTTG</td>
<td>4/6</td>
<td>chr17: 35647634-35647656</td>
</tr>
<tr>
<td>OT5_Oct4_Stop</td>
<td>GCTAGGGCTTACATTTCACTTG</td>
<td>4/6</td>
<td>chr17: 35647634-35647656</td>
</tr>
<tr>
<td>OT6_Oct4_Stop</td>
<td>GCTAGGGCTTACATTTCACTTG</td>
<td>4/6</td>
<td>chr17: 35647634-35647656</td>
</tr>
<tr>
<td>OT7_Oct4_Stop</td>
<td>GCTAGGGCTTACATTTCACTTG</td>
<td>4/6</td>
<td>chr17: 35647634-35647656</td>
</tr>
<tr>
<td>OT8_Oct4_Stop</td>
<td>GCTAGGGCTTACATTTCACTTG</td>
<td>4/6</td>
<td>chr17: 35647634-35647656</td>
</tr>
<tr>
<td>OT9_Oct4_Stop</td>
<td>GCTAGGGCTTACATTTCACTTG</td>
<td>4/6</td>
<td>chr17: 35647634-35647656</td>
</tr>
<tr>
<td>OT10_Oct4_Stop</td>
<td>GCTAGGGCTTACATTTCACTTG</td>
<td>4/6</td>
<td>chr17: 35647634-35647656</td>
</tr>
<tr>
<td>OT11_Oct4_Stop</td>
<td>GCTAGGGCTTACATTTCACTTG</td>
<td>4/6</td>
<td>chr17: 35647634-35647656</td>
</tr>
</tbody>
</table>

**Coordinate**

- chr3: 34560278-34550300
- chr4: 126636377-126636399
- chr14: 58080951-58080963
- chr1: 136641174-136641196
- chr9: 69081989-69081994
- chr1: 130633965-130633967
- chr18: 61611640-61611662
- chr5: 136841014-136841036
- chr6: 122663559-122663581
- chr17: 35647634-35647656
- chr1: 136641174-136641196
- chr4: 141162434-141162456
- chr5: 136841014-136841036
- chr6: 122663559-122663581
- chr18: 61611640-61611662
- chr3: 34560278-34550300
- chr4: 141162434-141162456
- chr5: 136841014-136841036
- chr6: 122663559-122663581
- chr18: 61611640-61611662
- chr3: 34560278-34550300
- chr4: 141162434-141162456
- chr5: 136841014-136841036
- chr6: 122663559-122663581
- chr18: 61611640-61611662
- chr3: 34560278-34550300
- chr4: 141162434-141162456
- chr5: 136841014-136841036
- chr6: 122663559-122663581
- chr18: 61611640-61611662

### Mismatches from the on-target sequence are lower-case, boldface and underlined. Indel mutation frequencies in targeted mice or ESCs were calculated by Suveryor assay. Coordinate in which sites were located are shown. OT, off-target; /, not tested.

(a) Nanog OT1 and 2 contain 3bp mismatches; OT3 to 9 contain 4bp mismatches lying in PAM distal region.

(b) PCR products were cloned and sequenced to confirm off-target mutations.
Table S4. Oligonucleotides Used In This Study

Oligonucleotides used for cloning sgRNA

<table>
<thead>
<tr>
<th>Gene target</th>
<th>Direction</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tet1</td>
<td>F</td>
<td>caccggctgctgtcagggagctca</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>aaactgagctcccttgacagcagcc</td>
</tr>
<tr>
<td>Tet2</td>
<td>F</td>
<td>caccgaagtgcccaacagatatcc</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>aaacggatatatcttgtgacacttttc</td>
</tr>
<tr>
<td>Sox2</td>
<td>F</td>
<td>caccgtgccccctgtgacacatgtga</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>aaactcataatgtgacaggggac</td>
</tr>
<tr>
<td>Nanog</td>
<td>F</td>
<td>caccgcgtaaagtctcatatttcacc</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>aaacggtgaaaatagtgacatctgac</td>
</tr>
<tr>
<td>Oct4</td>
<td>F</td>
<td>caccgatcataacaggtgagcttac</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>aaacgtagtcataatgtgacatcagct</td>
</tr>
<tr>
<td>Mecp2</td>
<td>F</td>
<td>caccgtagcggcccccagcttgac</td>
</tr>
<tr>
<td>L1</td>
<td>R</td>
<td>aaactgggtgaatgtgacatctgac</td>
</tr>
<tr>
<td>Mecp2</td>
<td>F</td>
<td>caccgccaaggatacatgtatccta</td>
</tr>
<tr>
<td>L2</td>
<td>R</td>
<td>aaactgggataatcctgctgagcttggc</td>
</tr>
<tr>
<td>Mecp2</td>
<td>F</td>
<td>caccgaggagtggagtctagtacttt</td>
</tr>
<tr>
<td>R1</td>
<td>R</td>
<td>aaacaagtagcatgacatcactcactc</td>
</tr>
<tr>
<td>Mecp2</td>
<td>F</td>
<td>caccgtagcggcccccagcttgac</td>
</tr>
<tr>
<td>R2</td>
<td>R</td>
<td>aaacgacccatatcctacccacac</td>
</tr>
</tbody>
</table>

Oligonucleotides used for RFLP analysis and PCR genotyping.

<table>
<thead>
<tr>
<th>Gene target</th>
<th>Direction</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tet1</td>
<td>F</td>
<td>ttgttctctctctctgtgactgc</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>tgattgatcaaatagacctgc</td>
</tr>
<tr>
<td>Tet2</td>
<td>F</td>
<td>cagatgcctaggccaatcaag</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>agaagcacaacacatgagaatg</td>
</tr>
<tr>
<td>Sox2</td>
<td>SF</td>
<td>acatgtcagcagatctactcc</td>
</tr>
<tr>
<td>V5</td>
<td>V5F</td>
<td>taatttggatggatggtggtg</td>
</tr>
<tr>
<td>Mecp2</td>
<td>LF</td>
<td>aatgtgccacttttaacagcac</td>
</tr>
<tr>
<td>L1,L2</td>
<td>LR</td>
<td>ttcttgatgtttctgttggc</td>
</tr>
<tr>
<td>Mecp2</td>
<td>RF</td>
<td>aagcatgagccactacaacc</td>
</tr>
<tr>
<td>R1,R2</td>
<td>RR</td>
<td>cttgctcagaagccaaacag</td>
</tr>
</tbody>
</table>
Oligonucleotides used for making template for in vitro transcription

<table>
<thead>
<tr>
<th>Template</th>
<th>Direction</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cas9</td>
<td>F</td>
<td>ttaatacgactcactatagggaagatggactataaggagcacgac</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>ggcagctctaggaattttagc</td>
</tr>
<tr>
<td>Tet1</td>
<td>F</td>
<td>ttaatacgactcactatagggaagatggactataaggagcacgac</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>ggcagctctaggaattttagc</td>
</tr>
<tr>
<td>sgRNA</td>
<td>R</td>
<td>aaaaagccgcaacctgggtgcc</td>
</tr>
<tr>
<td>Tet2</td>
<td>F</td>
<td>ttaatacgactcactatagggaagatggactataaggagcacgac</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>ggcagctctaggaattttagc</td>
</tr>
<tr>
<td>Sox2</td>
<td>F</td>
<td>ttaatacgactcactatagggaagatggactataaggagcacgac</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>ggcagctctaggaattttagc</td>
</tr>
<tr>
<td>sgRNA</td>
<td>R</td>
<td>aaaaagccgcaacctgggtgcc</td>
</tr>
<tr>
<td>Nanog</td>
<td>F</td>
<td>ttaatacgactcactatagggaagatggactataaggagcacgac</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>ggcagctctaggaattttagc</td>
</tr>
<tr>
<td>sgRNA</td>
<td>R</td>
<td>aaaaagccgcaacctgggtgcc</td>
</tr>
<tr>
<td>Oct4</td>
<td>F</td>
<td>ttaatacgactcactatagggaagatggactataaggagcacgac</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>ggcagctctaggaattttagc</td>
</tr>
<tr>
<td>sgRNA</td>
<td>R</td>
<td>aaaaagccgcaacctgggtgcc</td>
</tr>
<tr>
<td>Mecp2-L1</td>
<td>F</td>
<td>ttaatacgactcactatagggaagatggactataaggagcacgac</td>
</tr>
<tr>
<td>sgRNA</td>
<td>R</td>
<td>ggcagctctaggaattttagc</td>
</tr>
<tr>
<td>Mecp2-L2</td>
<td>F</td>
<td>ttaatacgactcactatagggaagatggactataaggagcacgac</td>
</tr>
<tr>
<td>sgRNA</td>
<td>R</td>
<td>ggcagctctaggaattttagc</td>
</tr>
<tr>
<td>Mecp2-R1</td>
<td>F</td>
<td>ttaatacgactcactatagggaagatggactataaggagcacgac</td>
</tr>
<tr>
<td>sgRNA</td>
<td>R</td>
<td>ggcagctctaggaattttagc</td>
</tr>
<tr>
<td>Mecp2-R2</td>
<td>F</td>
<td>ttaatacgactcactatagggaagatggactataaggagcacgac</td>
</tr>
<tr>
<td>sgRNA</td>
<td>R</td>
<td>ggcagctctaggaattttagc</td>
</tr>
</tbody>
</table>

Oligonucleotides used for HDR-mediated repair through embryo injecti

<table>
<thead>
<tr>
<th>Gene target</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tet1-loxP</td>
<td>tcttgtgtgcattaagctgtgacatggactataaggagcacgac</td>
</tr>
<tr>
<td>Tet2-loxP</td>
<td>tcttgtgtgcattaagctgtgacatggactataaggagcacgac</td>
</tr>
<tr>
<td>Sox2-v5</td>
<td>taccagacggcaacctgggtgcc</td>
</tr>
<tr>
<td>Mecp2-L1</td>
<td>ccagccgcaacctgggtgcc</td>
</tr>
<tr>
<td></td>
<td>tctgtgtgcattaagctgtgacatggactataaggagcacgac</td>
</tr>
<tr>
<td>Mecp2-L2</td>
<td>ccagccgcaacctgggtgcc</td>
</tr>
<tr>
<td></td>
<td>tctgtgtgcattaagctgtgacatggactataaggagcacgac</td>
</tr>
<tr>
<td>Mecp2-R1</td>
<td>ccagccgcaacctgggtgcc</td>
</tr>
<tr>
<td></td>
<td>tctgtgtgcattaagctgtgacatggactataaggagcacgac</td>
</tr>
<tr>
<td>Mecp2-R2</td>
<td>ccagccgcaacctgggtgcc</td>
</tr>
<tr>
<td></td>
<td>tctgtgtgcattaagctgtgacatggactataaggagcacgac</td>
</tr>
</tbody>
</table>
Oligonucleotides used for off-targeted analysis

<table>
<thead>
<tr>
<th>Gene target</th>
<th>Direction</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>OT1_Sox2_Stop</td>
<td>F</td>
<td>atgacatgacccaagtaataaccc</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>ctccactctgtactagggac</td>
</tr>
<tr>
<td>OT2_Sox2_Stop</td>
<td>F</td>
<td>tgtggggtttttggtactgcc</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gacagatcatagatagaagaattg</td>
</tr>
<tr>
<td>OT3_Sox2_Stop</td>
<td>F</td>
<td>aaactgaggcagagtagtctg</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gtgacagccacactttgacc</td>
</tr>
<tr>
<td>OT4_Sox2_Stop</td>
<td>F</td>
<td>caccctaggttcatggccttc</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gatggtagctgattaaggac</td>
</tr>
<tr>
<td>OT5_Sox2_Stop</td>
<td>F</td>
<td>accatgtagggcattcaccc</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>catggagctcattactagtg</td>
</tr>
<tr>
<td>OT6_Sox2_Stop</td>
<td>F</td>
<td>ttctctgaagatgaaatgatt</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>cagttgcaagacttctgagag</td>
</tr>
<tr>
<td>OT7_Sox2_Stop</td>
<td>F</td>
<td>ttctctgctctgctcttc</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gcacatggttgacctgag</td>
</tr>
<tr>
<td>OT8_Sox2_Stop</td>
<td>F</td>
<td>gactcagatttccaaagcctg</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>acatctcatgctctttagac</td>
</tr>
<tr>
<td>OT9_Sox2_Stop</td>
<td>F</td>
<td>tgccatgtgctgttacctc</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>ttgatatttaagacaggtcctc</td>
</tr>
<tr>
<td>OT10_Sox2_Stop</td>
<td>F</td>
<td>gtaaggctagtaagaaccttgg</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>aattctcaactgaggaatactg</td>
</tr>
<tr>
<td>OT11_Sox2_Stop</td>
<td>F</td>
<td>ttctcagccagaaagcttg</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gactttagatgccagaggtag</td>
</tr>
<tr>
<td>OT12_Sox2_Stop</td>
<td>F</td>
<td>agctgacagaagacgtagag</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>taaaaccaagcaagttcatg</td>
</tr>
<tr>
<td>OT1_Nanog_Stop</td>
<td>F</td>
<td>gctgtggatgatgcacagtg</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gtctaaccctgcctatatgcaac</td>
</tr>
<tr>
<td>OT2_Nanog_STOP</td>
<td>F</td>
<td>agatccccattacggatggtg</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>ggcacactcaccatgattttgg</td>
</tr>
<tr>
<td>OT3_Nanog_Stop</td>
<td>F</td>
<td>tagattacttagtggctcacc</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>agtccccagtgcagagcacc</td>
</tr>
<tr>
<td>OT4_Nanog_Stop</td>
<td>F</td>
<td>gaccttttctaatggggcttg</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gttacagcaaggtagattaccc</td>
</tr>
<tr>
<td>OT5_Nanog_Stop</td>
<td>F</td>
<td>tcccatctaccctctgacctg</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gcctgaagaggaaaaaggttcc</td>
</tr>
<tr>
<td>OT6_Nanog_Stop</td>
<td>F</td>
<td>tctgaggtgaccaagcagt</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>acatccaccatgctctccgtg</td>
</tr>
<tr>
<td>OT7_Nanog_Stop</td>
<td>F</td>
<td>caatctttctcagtgaggtagg</td>
</tr>
<tr>
<td>OT8_Nanog_Stop</td>
<td>R</td>
<td>cttgttcagtgcatcttgctgc</td>
</tr>
<tr>
<td>----------------------</td>
<td>----</td>
<td>------------------------</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>tctctcgagaaagataggcc</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gttgcaatgcacttcttgag</td>
</tr>
<tr>
<td>OT9_Nanog_Stop</td>
<td>F</td>
<td>agtctagcatgtgagcttg</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>aacttcaagtggaaactgtctg</td>
</tr>
<tr>
<td>OT1_Mecp2_Left2</td>
<td>F</td>
<td>cacacacacactgaaataagt</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>agctggcttttgagccagac</td>
</tr>
<tr>
<td>OT2_Mecp2_Left2</td>
<td>F</td>
<td>tagtcacttagtttacttctc</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gtgatccagacttggcag</td>
</tr>
<tr>
<td>OT3_Mecp2_Left2</td>
<td>F</td>
<td>tcaactttccctcagtttcc</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>caatctatcattctctgaataag</td>
</tr>
<tr>
<td>OT4_Mecp2_Left2</td>
<td>F</td>
<td>gaaacctgtgacgagttcctc</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gacagacagctttgcttcccc</td>
</tr>
<tr>
<td>OT1_Mecp2_Right1</td>
<td>F</td>
<td>gcagcaccaggtgaaattaac</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gcctattatatgatctgccc</td>
</tr>
<tr>
<td>OT2_Mecp2_Right1</td>
<td>F</td>
<td>acagattgcaggctactcag</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gtccaagcctacttcctccac</td>
</tr>
<tr>
<td>OT3_Mecp2_Right1</td>
<td>F</td>
<td>tccgacaatgttttatgtctg</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>agatactagcagtgcagctg</td>
</tr>
<tr>
<td>OT4_Mecp2_Right1</td>
<td>F</td>
<td>gttcctgtgatgttttacttg</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>tagaccaattccacaccacag</td>
</tr>
<tr>
<td>OT5_Mecp2_Right1</td>
<td>F</td>
<td>tgctgtgaaactcagccttg</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>cttaagacacagccagaaag</td>
</tr>
<tr>
<td>OT6_Mecp2_Right1</td>
<td>F</td>
<td>cgccataaacctcccattag</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>ctctgtgattttgtaagccaac</td>
</tr>
<tr>
<td>OT7_Mecp2_Right1</td>
<td>F</td>
<td>ggcagcactaatattccccaaag</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>cttatattgtacatcattacc</td>
</tr>
<tr>
<td>OT8_Mecp2_Right1</td>
<td>F</td>
<td>ccatttgtaaagttggtatgc</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>ggtcgaatctctgtaaacaataac</td>
</tr>
<tr>
<td>OT9_Mecp2_Right1</td>
<td>F</td>
<td>gtcactttgactgtgagttg</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>ctagcaacttgggaagcaatag</td>
</tr>
<tr>
<td>OT10_Mecp2_Right1</td>
<td>F</td>
<td>ctaactcaactctcaagcttg</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>agagagcttccaaatttgtag</td>
</tr>
<tr>
<td>OT1_Oct4_Stop</td>
<td>F</td>
<td>accacactactcgataacctg</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>ggtatagcgtgagtggac</td>
</tr>
<tr>
<td>OT2_Oct4_Stop</td>
<td>F</td>
<td>atgtatgcaactaagggcaag</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>ccagtaatataacacctgcaatg</td>
</tr>
<tr>
<td>OT3_Oct4_Stop</td>
<td>F</td>
<td>gatccccatcctcttgagatc</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>aatccggcagctgacaaagcc</td>
</tr>
<tr>
<td>OT4_Oct4_Stop</td>
<td>F</td>
<td>tgtgcaatgcacagacactag</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>tttgatgcgtgcctcactgttc</td>
</tr>
<tr>
<td>OT5_Oct4_Stop</td>
<td>F</td>
<td>tatccctgggtatcataggttg</td>
</tr>
<tr>
<td>OT6_Oct4_Stop</td>
<td>F</td>
<td>acaagtgattactagaagggtg</td>
</tr>
<tr>
<td>OT6_Oct4_Stop</td>
<td>R</td>
<td>agtaggctttctatcaggatttc</td>
</tr>
<tr>
<td>OT7_Oct4_Stop</td>
<td>F</td>
<td>accatcttgatgattgagtagtag</td>
</tr>
<tr>
<td>OT7_Oct4_Stop</td>
<td>R</td>
<td>attccacactacaaaacttaatggc</td>
</tr>
<tr>
<td>OT8_Oct4_Stop</td>
<td>F</td>
<td>tggccggagctagccttgag</td>
</tr>
<tr>
<td>OT8_Oct4_Stop</td>
<td>R</td>
<td>cctccctgtgcaactaggaac</td>
</tr>
<tr>
<td>OT9_Oct4_Stop</td>
<td>F</td>
<td>ctttgattccattgtcaagg</td>
</tr>
<tr>
<td>OT9_Oct4_Stop</td>
<td>R</td>
<td>cagttcatcaccatctcttag</td>
</tr>
<tr>
<td>OT10_Oct4_Stop</td>
<td>F</td>
<td>aatcccaacttgaccatataagc</td>
</tr>
<tr>
<td>OT10_Oct4_Stop</td>
<td>R</td>
<td>cacccctttgccagctgtagac</td>
</tr>
<tr>
<td>OT11_Oct4_Stop</td>
<td>F</td>
<td>cactgcattgcataatagagag</td>
</tr>
<tr>
<td>OT11_Oct4_Stop</td>
<td>R</td>
<td>catggacttatgcacaaaggg</td>
</tr>
</tbody>
</table>
Chapter 4. Tracing dynamic changes of DNA methylation at single-cell resolution

Yonatan Stelzer\textsuperscript{1,3}, Chikdu Shakti Shivalila\textsuperscript{1,2,3}, Frank Soldner\textsuperscript{1}, Styliani Markoulaki\textsuperscript{1}, and Rudolf Jaenisch\textsuperscript{1,2,*}

1. Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
2. Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA
3. Co-first author
   *Correspondence: jaenisch@wi.mit.edu

Published as:
*indicates equal contribution

YS and CSS shared all experiments and analyses equally. FS helped with southern blot analysis. SM helped with blastocyst injections. YS, CSS and RJ designed and conceived of experiments.
Mammalian DNA methylation plays an essential role in development. To date, only snapshots of different mouse and human cell types have been generated, providing a static view on DNA methylation. To enable monitoring of methylation status as it changes over time, we establish a reporter of genomic methylation (RGM) that relies on a minimal imprinted gene promoter driving a fluorescent protein. We show that insertion of RGM proximal to promoter-associated CpG islands reports the gain or loss of DNA methylation. We further utilized RGM to report endogenous methylation dynamics of non-coding regulatory elements, such as the pluripotency-specific super enhancers of Sox2 and miR290. Loci-specific DNA methylation changes and their correlation with transcription were visualized during cell-state transition following differentiation of mouse embryonic stem cells and during reprogramming of somatic cells to pluripotency. RGM will allow the investigation of dynamic methylation changes during development and disease at single-cell resolution.

DNA methylation is recognized as a principal contributor to the stability and regulation of gene expression in development and maintenance of cellular identity (Bird, 2002; Cedar and Bergman, 2012; Jaenisch and Bird, 2003; Reik et al., 2001). Changes in DNA methylation are dynamic and it is still largely unknown how they dictate spatial and temporal gene expression programs (Smith and Meissner, 2013). Recent advancements in sequencing technologies enabled the establishment of methylation maps for multiple cell types in both human (Kundaje et al., 2015; Schultz et al., 2015; Smith et al., 2014; Ziller et al., 2013) and mouse (Hon et al., 2013), thus providing a framework for identifying key lineage-specific regulators (Rivera and Ren, 2013). DNA methylation is a dynamic process and current methods are only bulk and provide a static "snapshot" view of the methylation state during cell-state transitions. The difficulty in translating real-time epigenetic changes into a traceable readout is, to date, a limiting factor in our ability to follow the dynamics of DNA methylation. Therefore, a key challenge in the field is to generate tools that allow tracing changes in DNA methylation over time.
Here, we set out to generate a DNA methylation reporter system that is capable of visualizing genomic methylation states at single-cell resolution. The design of the reporter was based on two premises: (1) previous observations suggesting that CpG sites can serve as cis-acting signals, affecting the methylation state of adjacent CpGs (Brandeis et al., 1994; Mummaneni et al., 1995; Turker, 2002), and (2) a methylation-sensitive promoter that, when introduced in proximity to a CpG region of choice, may be utilized to report on methylation changes of the adjacent sequences. Thus, a key issue in establishing a DNA methylation reporter was identifying a methylation-sensitive promoter, which is not independently regulated by the DNA methylation machinery, but can be affected by methylation changes of adjacent sequences. Constitutively active genes usually contain hypomethylated high density CpG islands (CGIs) in their promoter regions and are not regulated by DNA methylation (Deaton and Bird, 2011) whereas gene promoters associated with low-density CGIs are activated and repressed in a tissue-specific manner. Because methylation of both classes of promoters is either not regulated by the DNA methylation machinery in all tissues or regulated in a tissue-dependent manner, these promoters cannot be utilized as DNA methylation reporters. In contrast, imprinted gene promoters exhibit inherent sensitivity to DNA methylation of adjacent genomic regions resulting in transcriptional activation or silencing. This mechanism has been established for a subgroup of germline-derived differentially methylated regions (DMRs) that affect in cis the methylation state of secondary regulatory promoter elements, which in turn control imprinted gene activity. Importantly, following their establishment, promoter-associated imprinted DMRs are not regulated by the DNA methylation machinery in a tissue-specific manner (Ferguson-Smith, 2011). We
hypothesized that these intrinsic characteristics of imprinted gene promoters make them attractive candidates for methylation sensors. Perhaps one of the best-studied examples is the Prader-Willi Angelman region, in which an imprinted DMR resides at the small nuclear ribonucleoprotein polypeptide N (Snrpn) gene promoter region controlling its parent-of-origin monoallelic expression (Buiting et al., 1995; Kantor et al., 2004). Furthermore, Snrpn is expressed in most of the tissues and thus serves as an attractive candidate to generate a DNA methylation reporter.

Changes in DNA methylation occur mostly at non-CGIs, some of which are associated with tissue-specific gene promoters (Jones, 2012). Nevertheless, a growing body of evidence suggests that the bulk of tissue-specific changes in DNA methylation is associated with non-coding sequences (Irizarry et al., 2009) such as distal regulatory elements, which include enhancers and transcription factor binding sites (Hon et al., 2013; Stadler et al., 2011; Ziller et al., 2013). Recent reports identified super-enhancers (SE) as clusters of TF and mediator-binding sites associated with bonafide enhancer chromatin marks to control the expression of key cell identity genes (Dowen et al., 2014; Hnisz et al., 2013; Whyte et al., 2013). Global genomic comparisons of tissue-specific DNA methylation and transcription factor (TF) chromatin immuno-precipitation sequencing (ChIP-seq) data correlated the chromatin with the methylation state (Xie et al., 2013). Thus, many tissue-specific enhancers are hypomethylated in tissues where the target genes are expressed, but are hypermethylated in tissues where the target genes are silent (Hon et al., 2013). In this paper, we establish a reporter of genomic methylation (RGM) that enables the visualization of changes in DNA methylation in live cells. We show that a minimal Snprn promoter can report on the DNA methylation state of
endogenous gene promoters. We also generated reporter cell lines for the pluripotency-specific miR290 and Sox2 SEs and demonstrate that RGM can be used to capture dynamic DNA methylation changes in distal non-coding regulatory regions. An attractive aspect of RGM is its utility to visualize DNA methylation changes in development and disease at single-cell resolution in the same sample.

RESULTS

A Methylation-Sensitive Reporter System Based on a Minimal Imprinted Promoter

To establish a methylation reporter, we generated a minimal Snrpn promoter that includes the conserved elements between human and mouse and contains the endogenous imprinted DMR region (Figure S1A). The minimal promoter region driving GFP was cloned into a sleeping beauty transposon vector (Ivics et al., 1997) to facilitate stable integration into the genome. Recent studies have demonstrated that different CGI vectors, when stably inserted into mouse embryonic stem cells (mESCs), adopt a methylation pattern that corresponds to the in vivo methylation pattern of the respective endogenous sequence (Sabag et al., 2014). To test whether DNA methylation can propagate into the Snrpn promoter region in vivo, we designed an experimental system in which the CGI regions of Gapdh and Dazl were cloned upstream of our reporter (Figure 1A). The promoter of Gapdh encompasses a hypomethylated CGI consistent with constitutive expression in all tissues. In contrast, the Dazl promoter-associated CGI is hypermethylated in all tissues excluding the germ cells (Hackett et al., 2013). Given the different expression and methylation patterns of both genes, upon stable integration of the two reporter vectors into mESCs the Gapdh CGI is expected to maintain its
hypomethylated state, while the Dazl CGI would be subjected to de novo methylation (Sabag et al., 2014). Figure 1B show that >95% of cells carrying the Gapdh reporter expressed GFP. In contrast, >30% of cells carrying the Dazl reporter were GFP-negative, corresponding to reporter silencing. The effect of the Dazl reporter becomes more robust upon continued passage, with >80% of the cells silencing their reporter within 4 weeks (Figure 1B).

To assess the DNA methylation levels of the Gapdh and Dazl reporters following introduction into mESCs, we sorted Gapdh GFP-positive and Dazl GFP-negative cell populations (Figure 1C). The GFP expression state was stable upon continuous culture and passaging of the two sorted cell populations for over 7 weeks (Figure 1C). DNA was extracted from both Gapdh GFP-positive and Dazl GFP-negative cells and subjected to bisulfite conversion and PCR sequencing. Figure 1D shows that Gapdh GFP-positive cells maintained the hypomethylated state at both Gapdh CGI and the Snrpn promoter regions, whereas Dazl GFP-negative cells became highly de novo methylated at the Dazl CGI region and its corresponding downstream Snrpn promoter (Figure 1E). These results are consistent with the hypothesis that DNA methylation can be propagated from the CGI into the Snrpn promoter region resulting in repression of transcriptional activity.

**RGM Is a Reporter for In Vivo Demethylation**

The experiments described above showed that RGM reports on de novo methylation imposed in vivo on the unmethylated Dazl CGI donor test sequence. Conversely, we were interested to assess whether a methylated and silent donor Snrpn promoter can be reactivated by means of demethylation acquired in vivo. For this, we
used the CpG methyltransferase M.Sssl to in vitro methylate both Gapdh and Dazi reporter constructs. Treatment of the plasmids with M.Sssl enzyme followed by bisulfite conversion, PCR amplification, and sequencing, confirmed the complete hypermethylation of both the CGI and Snrpn promoter regions (Figures 2A, S1B, and S1C). ESCs were transfected with either Gapdh or Dazi reporter and selected for cells carrying stably integrated vectors. Following 1 week of culture, we identified robust activation of GFP in virtually all cells carrying the integrated Gapdh reporter, whereas cells carrying the Dazl reporter remained GFP-negative (Figures 2B–2D). To assess the DNA methylation state of the Gapdh and Dazl CGI and the respective downstream Snrpn promoter regions, DNA was extracted from the two cell lines, subjected to bisulfite conversion, PCR amplification and sequencing. Figure 2E demonstrates that, consistent with high GFP expression, the Gapdh CGI and its downstream Snrpn promoter had become fully demethylated. In contrast, the Dazl CGI and its downstream Snrpn promoter sequences maintained the hypermethylated state in agreement with complete repression of the GFP signal (Figure 2F). Thus, our data support the hypothesis that a Snrpn promoter can report on in vivo demethylation of the CGI in its proximity.

**Dnmt1, Dnmt3a, and Dnmt3b Mediate Methylation and Reporter Activity**

We used ESCs deficient for the DNA methyltransferases Dnmt1, Dnmt3a, and Dnmt3b to gain mechanistic insights into demethylation and de novo methylation imposed on the Snrpn promoter in transfected ESCs. Figure 2G shows that introduction of an in vitro methylated Dazl Snrpn vector into Dnmt1 mutant cells resulted in ~80% GFP-positive cells by passage five, in contrast to no GFP-positive cells when inserted
into wild-type (WT) cells. In agreement with the role of Dnmt1 as being the maintenance DNA methyltransferase (Li et al., 1992), bisulfite sequencing analysis on the sorted GFP-positive cells confirmed that reactivation of the methylated Dazl reporter occurred by passive demethylation (Figure 2H). To clarify the mechanism of de novo methylation, we introduced an unmethylated version of both vectors into mESCs deficient for both de novo DNA methyltransferases Dnmt3a and Dnmt3b (Pawlak and Jaenisch, 2011). Figure 2I shows that the vast majority of cells carrying the Dazl or the Gapdh reporters were positive for GFP unlike Dazl reporter expression in control V6.5 cells (Figure 2I), which is consistent with Dnmt3a/b mediating de novo methylation and reporter silencing.

Recent studies have shown that culturing mESCs in 2i medium (inhibitors of MEK and GSK3), and leukemia inhibitory factor (LIF) results in downregulation of Dnmt3a and Dnmt3b, consequently leading to global hypomethylation (Lee et al., 2014). To assess whether these culture conditions affect reporter activity, we transfected the unmethylated Gapdh and Dazl reporters into WT mESCs cultured in 2i and LIF. Figure 2I shows that the great majority of the stably transfected cells were GFP-positive, consistent with 2i-mediated down regulation of the Dnmt3a and Dnmt3b.

**RGM Can Report on Methylation Associated with Endogenous Gene Promoters**

To test whether the Snrpn promoter could also report on DNA methylation levels associated with endogenous gene promoters, we utilized CRISP/Cas-mediated gene editing to target the endogenous CGIs located at the promoter regions of Gapdh and Dazl (Figures 3A, S2A, and S2B). Figure 3B shows 35/36 Dazl-vector-transfected clones were GFP-negative indicating robust silencing of the Dazl reporter whereas 20/21 Gapdh-
vector-transfected clones were GFP-positive (Figure 3B). FACS analysis of correctly targeted clones confirmed that Gapdh reporter cells were all GFP-positive with the CGI and Snrnp promoter unmethylated (Figures 3C and 3D) in contrast to Dazl GFP-negative clones with the corresponding sequences methylated (Figures 3E and 3F). Our results demonstrate that Snrnp reporter activity reports on the methylation state of its surrounding sequences and does not alter their methylation state. Furthermore, the endogenous targeting results suggested that the partial repression of the Dazl reporter (Figure 1B), observed at early passages of the transgene experiment, may be due to multiple genome integration and position effects.

**RGM Can Report on Methylation of Pluripotency-Specific Super-Enhancers**

Methylation of super enhancers (SEs) has been shown to change during differentiation. We tested whether RGM would report on the active and hypomethylated state of the pluripotency-specific SEs associated with the miR290 and Sox2 genes in mESCs and their methylated and inactive state in somatic cells (Figures 4A and S3A). In contrast to the CGIs located at gene promoters (Gapdh and Dazl), the SE regions of both Sox2 and miR290 represent low-density CpG sequences. Utilizing CRISP/Cas-mediated gene editing, we inserted a Snrnp tdTomato reporter into the endogenous miR290 and Sox2 enhancer (Figures 4B and S3B, respectively). As recipient cells, we used the previously established Oct4, Sox2, Klf4, and c-Myc (OSKM) polycistronic dox-inducible secondary reprogrammable mESCs (Carey et al., 2011), which also carried a GFP reporter knocked into the endogenous Nanog locus. Correct integration of the vector was validated by PCR and Southern analysis (Figure S3C). Figure 4C Shows that both
targeted ESC lines (miR290 #21 and Sox2#2) expressed tdTomato as well as Nanog-GFP. To assess whether the tdTomato expression correlated with hypomethylation of the inserted RGM, DNA extracted from the bulk mESCs population was bisulfite converted, amplified by PCR, and sequenced with the PCR amplification including both the SE CpG region and the downstream Snrpn promoter. As predicted from the methylation maps (Figures 4A and S3A), both endogenous miR290 and Sox2 CpG regions were mostly hypomethylated (Figure 4D). Importantly, the Snrpn promoter was also hypomethylated consistent with reporter expression. Of note, a few highly methylated alleles were detected (Figure 4D), possibly reflecting an inherent variation in the bulk population due to the presence of cells that carry an inactive reporter. To test this possibility, we analyzed the Sox2 SE region in the untargeted parental cell, which identified the presence of both methylated and unmethylated alleles at the same frequency as the targeted reporter cell line (Figure S3D). We conclude that RGM can report on the methylation state of distal genomic regulatory regions.

**Dynamic De Novo DNA Methylation during Differentiation**

To monitor real-time changes in genomic DNA methylation during in vitro differentiation, mESCs carrying the tdTomato reporters reflecting DNA methylation levels at the SE regions, were exposed to retinoic acid (RA), which induces a rapid exit from pluripotency, and cellular differentiation (Rhinn and Dolle, 2012). The presence of the Nanog-GFP reporter allowed monitoring exit from pluripotency by loss of GFP expression. Sorted double-positive (tdTomato+/GFP+) miR290 and Sox2 cells were plated on feeder-free gelatin coated plates, treated with 0.25 mM RA the following day
(Figure 5A) and analyzed at different times after addition of RA (Figures 5A and 5B). As expected, undifferentiated cells were double-positive (tdTomato+/GFP+). However, upon induction of differentiation a gradual reduction in the fraction of double-positive cells was observed with most disappearing over the time course of 7 days, resulting in a largely double-negative cell population (Figures 5B and 5C). This is in contrast to control Gapdh reporter cells that, as expected, appeared completely GFP-positive following 7 days of RA differentiation (Figure S4A). tdTomato and Nanog-GFP-positive cells disappeared with different kinetics: while singly tdTomato-positive cells (tdTomato+/GFP-) appeared after 2 days, only a few single Nanog-GFP-positive cells (tdTomato-/GFP+) were detected during differentiation (Figures 5B and 5C) suggesting that Nanog was silenced prior to methylation and silencing of the miR290 and Sox2 SEs.

To confirm that loss of the tdTomato signal correlated with accumulation of de novo methylation in both SE regions, we sorted the main populations at different time points during RA differentiation (Figure 5C). DNA was extracted from the different cell populations and subjected to bisulfite sequencing, thus allowing a comprehensive analysis of the methylation state in both the endogenous miR290 and Sox2 SE and their respective Snrpn promoter regions (Figures 5D, 5E, S4B, and S4C). In contrast to the bulk population of mESCs (Figure 4D), the sorted double-positive cells did not harbor completely methylated alleles, consistent with the notion that methylated alleles in the bulk population represent intrinsic variation. The methylation of both miR290 and Sox2 in single-positive cells (tdTomato+/GFP-) was low, consistent with tdTomato expression. The overall increased de novo methylation in the single-positive cells, compared with the double-positive cells, may suggest that DNA methylation mediated silencing was already
initiated in this intermediate cell population. Notably, our analysis identified completely methylated genomes in the Sox2 single-positive (tdTomato+/GFP-) cell population (Figure 5E). This suggests that during rapid changes of de novo methylation, the half-life of the fluorescent protein (FP) may lead to an over-estimation of cells that are still hypomethylated during cell-state transitions. Finally, in agreement with the silencing of tdTomato expression, the double-negative cells (tdTomato/GFP) exhibited robust hypermethylation on both endogenous SE regions and their respective Snrpn promoters (Figures 5D, 5E, S4B, and S4C). To test whether the targeted reporter allele correlated with the methylation levels of the untargeted allele (WT), we analyzed the WT allele in Sox2 reporter cells at different time points during differentiation. Figure S4D shows that similar to the reporter allele, the WT allele exhibited low levels of methylation in the sorted double-positive cells and high levels of methylation following 7 days of differentiation. We conclude that RGM allows dynamic monitoring de novo methylation events that are imposed on genomic sequences upon exiting from pluripotency. Our data suggest that the differentiation of ESCs induces silencing of Nanog prior to de novo methylation of the two miR290 and Sox2 SEs.

To test whether in vivo differentiation resulted in silencing of the tdTomato reporter in both miR290 and Sox2 SE regions, we analyzed 13.5 dpi chimeric embryos. As control, we injected ESCs harboring the Gapdh CGI reporter driving a GFP sequence, which had also been infected with lentiviruses resulting in constitutive expression of tdTomato. The robust expression of GFP in the Gapdh control embryos demonstrated the widespread expression signature of the Snrpn promoter throughout mouse tissues (Figure 6A). Unlike the Gapdh control, both miR290 and Sox2 embryos were completely
negative for both GFP and tdTomato, demonstrating robust repression of Nanog and the Snrpn promoter during in vivo differentiation (Figure 6A).

**DNA Demethylation during Cellular Reprogramming**

Reprogramming of somatic cells to iPS cells involves demethylation and activation of the pluripotency SEs Sox2 and miR290 (see Figures 4A and S3A). We investigated whether RGM could be used to capture demethylation events that are gradually acquired during cellular reprogramming. For this, we used secondary Dox-inducible reprogrammable mouse embryonic fibroblasts (MEFs) isolated from 13.5 dpi chimeric embryos that had been injected at the blastocyst stage with the OSKM DOX-inducible ESCs (Carey et al., 2011) carrying Nanog-GFP and the tdTomato reporter reflecting DNA methylation levels at the Sox2 or miR290 SE alleles (see Figure 6B). Culture of these MEFs in DOX induces the reprogramming factors while Nanog-GFP activation allows monitoring the course of reprogramming in the bulk somatic cell population (Buganim et al., 2012). As expected, MEFs isolated from 13.5 dpi embryos were negative for both GFP and tdTomato expression, as measured by fluorescent microscopy and fluorescence-activated cell sorting (FACS) analysis (Figures 6C and S5A). Importantly, consistent with tdTomato repression, both endogenous miR290 and Sox2 SE regions as well as their corresponding downstream Snrpn promoter regions were hypermethylated (Figure 6D). Further analysis of the WT allele in Sox2 MEF showed high correlation with the targeted reporter allele, demonstrating robust repression of the SE region in vivo (Figure S5B).
To test whether reprogramming-induced demethylation can be visualized by RGM, we treated the secondary MEFs with serum and LIF medium supplemented with 2mg/ml doxycycline (Dox). Both miR290 and Sox2 MEFs were successfully reprogrammed, resulting in double-positive cells (tdTomato+/GFP+ data not shown). It was recently shown that a combination of three chemicals, TGF-b antagonist ALK5 inhibitor II, GSK3b antagonist CHIR99021, and ascorbic acid, an enzymatic cofactor (from here on referred to as 3C), results in more efficient and synchronous reprogramming (Vidal et al., 2014). To achieve more synchronized and efficient reprogramming, both miR290 and Sox2 MEFs were subjected to 3C culture conditions and the dynamics of reporter activation was monitored by flow cytometry. While the first expression of tdTomato+ and GFP+ cells emerged at day16(Figure 6E), reporter activation of both miR290 and Sox2 occurred with different kinetics. Figure6E shows accumulation of miR290 reporter cells that activated both GFP and tdTomato (tdTomato+/GFP+) over time. A small population of single-positive GFP cells appeared in late stages of reprogramming consistent with a stochastic sequence of events in the reprogramming of the miR290 SE region. Unlike miR290 reporter cells, however, Sox2 cells showed a more robust and defined dynamic of activation of both reporters. By day 16, a population of single-positive GFP cells (tdTomato-/GFP+) had accumulated, which gradually shifted to become double-positive (tdTomato+/GFP+) over time (Figures 6E and S5C). To test whether the single-positive GFP cells give rise to double-positive cells, we sorted the single-positive GFP cells and replated them on feeders using Dox independent culture conditions. Consistent with the repression of the tdTomato signal, bisulfite sequencing confirmed that the single-positive GFP cells exhibit high levels of
methylation in the SE region, as well as in the downstream Snrpn promoter region (Figure S5D). Upon further culture, tdTomato-positive cells appeared demonstrating that single-positive GFP cells give rise to double-positive cells (Figure S5E).

Our results suggest that reprogramming of both miR290 and Sox2 SE regions are late events, with the Sox2 SE region being reprogrammed subsequently to the activation of endogenous Nanog. miR290 and Sox2 double-positive (tdTomato+/GFP+) cells invariably proceed to a Dox-independent iPS cell state (Figure 6F). To assess the methylation state of the Sox2 and miR290 SEs, we performed bisulfite sequencing on DNA extracted from sorted double-positive (tdTomato+/GFP+) iPS cells. As shown in Figure 6G, both miR290 and Sox2 SE regions and their corresponding downstream Snrpn promoters were demethylated. These results confirmed that RGM can visualize demethylation of regulatory genomic regions during reprogramming with single-cell resolution.

DISCUSSION

In this work, we have generated a DNA methylation reporter (RGM) that allows imaging of DNA methylation with single-cell resolution. The design of the reporter system took advantage of the intrinsic characteristics of imprinted gene promoters, for which the transcriptional activity reflects the DNA methylation state of adjacent sequences. Importantly, imprinted promoters are neutral to developmental or tissue-specific DNA methylation changes, with their activity strictly dependent on the methylation state of the adjacent regulatory elements. This is in contrast to CGI sequences such as Gapdh or tissue-specific elements such as the Dazl promoter-
associated sequences, which become demethylated or de novo methylated, respectively, when inserted into the genome of ESCs (Brandeis et al., 1994; Sabag et al., 2014). This indicates that methylation of these elements as opposed to imprinted promoters is sequence-dependent and subject to trans-acting signals and cell state-dependent regulation.

The RGM reporter system described here is based on the Snrpn minimal promoter that is not subjected to methylation changes by itself, and therefore GFP expression is solely dependent on the methylation state of surrounding sequences. Consistent with this premise, ES cells appeared GFP-positive when stably transfected with the methylated or unmethylated Gapdh/Snrpn-GFP vector, but were GFP-negative when transfected with the methylated or unmethylated Dazl/Snrpn-GFP reporter. This indicates that the Snrpn promoter region can be used as a faithful sensor for regional methylation changes of adjacent sequences.

To investigate whether RGM can report on the methylation state of endogenous loci, we targeted CGIs located at Gapdh and Dazl promoter regions, resulting in differential methylation and activity of the Snrpn reporter. Thus, the Snrpn promoter effectively reflects local methylation patterns without affecting the endogenous epigenetic state. As most of the tissue-specific DNA methylation changes occur in low-density CpG regulatory regions, we asked whether RGM could report on the methylation state of non-coding low-density CpG regions. We chose two pluripotency-specific SEs that are associated with the miR290 and Sox2 genes and are known to be active and unmethylated in ESCs but become methylated and inactive upon cellular differentiation. CRISPR/Cas-mediated insertion of the Snrpn-tdTomato reporter into ESCs resulted in
tdTomato-positive clones but tdTomato expression was silenced in mid-gestation chimeric embryos, which reflects the demethylation state of the SEs in pluripotent cells and their de novo methylation upon induction of differentiation. Consistent with this, MEFs isolated from chimeric embryos were tdTomato-negative with both elements highly methylated. Upon conversion of the MEFs into induced pluripotent stem cells (iPSCs), however, the cells became tdTomato-positive reflecting demethylation of the SEs during reprogramming to pluripotency. Our results establish that RGM reporter activity mirrors the changes of DNA methylation imposed on endogenous CGI and low-density CpG genomic elements during development, upon cellular differentiation, and during reprogramming. Extensive epigenomic analyses of multiple tissues and cell types in both human and mice, suggest that embryonic development and cell-type specification are associated with massive epigenomic remodeling at discrete enhancers (Hon et al., 2013; Kundaje et al., 2015; Schultz et al., 2015; Ziller et al., 2013). It will thus be of interest to test whether RGM can be utilized to report on the DNA methylation state associated with more discrete regulatory regions. Implementing the methylation reporter to tissue-specific DMRs holds the promise to further elucidate the link between DNA methylation and other epigenetic mechanisms, with cell-fate regulation.

Reprogramming of somatic cells into iPSCs involves extensive resetting of the epigenome (Buganim et al., 2013; Hanna et al., 2010), and coinciding with this notion, recent studies identified a key role for epigenetic modifiers during this process (Mansour et al., 2012; Rais et al., 2013; Soufi et al., 2012). However, the exact kinetics of these epigenetic changes during the reprogramming process are difficult to define because of cell heterogeneity and the stochastic nature of the reprogramming process. Here, we
followed the methylation changes of two SEs associated with Sox2 and miR290, demonstrating that demethylation of both regions is a late event in the reprogramming process. Simultaneous activation of endogenous Nanog and miR290 SE demethylation is consistent with Nanog directly regulating the expression of miR290 cluster during reprogramming to iPS cells (Gingold et al., 2014). The gradual activation of the Sox2 tdTomato reporter followed expression of endogenous Nanog consistent with demethylation of Sox2 SE being a late event in the process (Buganim et al., 2012). Systematic deletion of the Sox2 upstream SE region was recently shown to dramatically affect Sox2 expression in ESCs (Li et al., 2014; Zhou et al., 2014). Thus, the Sox2 SE methylation reporter cells provide a rigorous experimental system to investigate how DNA methylation changes at distal regulatory region influence the expression of downstream target genes.

Changes in DNA methylation during development, lineage commitment, and disease are dynamic, and studies of epigenetic changes are hampered by two experimental constraints that limit mechanistic studies of methylation and gene regulation: (1) current methodology provides only a static “snapshot” view of the methylation state during cell state transitions, and (2) current methylation analyses require the examination of multiple cells precluding assessment of epigenetic changes in single cells. Given the overwhelming evidence of cell-cell heterogeneity in embryos, cultured cells, or disease states such as cancer (Junker and van Oudenaarden, 2014), this is a serious limitation for a mechanistic understanding of the epigenetic state and gene expression during these complex processes. For example, monitoring the course of differentiation in both miR290 and Sox2 reporter cells confirmed the co-existence of cell
populations that harbor distinct epigenetic states. In contrast, commonly used bulk methodologies would not allow isolating and distinguishing the different cell populations. Thus, sorting and isolating different cell types according to their methylation states can be achieved only by using readout for methylation state at single-cell resolution. The RGM reporter system overcomes some of the limitations of conventional methylation analyses by providing real-time visualization of DNA methylation at single-cell resolution. As with any fluorescent protein-based reporter system, the accuracy to trace real-time changes depends on the half-life of the respective FP. Because the current version of the methylation reporter does not use a destabilized FP, silencing of the reporter after de novo methylation-induced repression of the Snrpn promoter is likely delayed. To generate a reporter that more rapidly responds to DNA methylation, changes would require the use of a destabilized FP. Targeting additional loci in future studies will allow us to further elucidate other possible limitations of the RGM reporter system, such as inhibition of the Snrpn transcriptional activity by chromatin conformation.

As RGM allows measuring dynamics of DNA methylation at single-cell resolution, it provides a framework for understanding epigenetic changes during cell state transition in heterogeneous cell populations. For example, replacing the fluorescent-based reporter system with Cre-Lox will enable the generation of epigenetic lineage tracing maps. Furthermore, utilizing RGM together with conventional gene expression reporters may offer detailed insights into the interplay between epigenetic cues and the execution of tissue-specific gene expression programs. The use of fluorescent reporters as readout for locus-specific methylation changes may also provide an effective screening platform
for the isolation of small molecule compounds that affect the methylation state of specific genomic regions.

EXPERIMENTAL PROCEDURES

mESCs Cell Culture

V6.5 mouse embryonic stem cells (mESCs) were cultured on irradiated mouse embryonic fibroblasts (MEFs) with standard ESCs medium: (500 ml) DMEM supplemented with 10% FBS (Hyclone), 10mg recombinant leukemia inhibitory factor (LIF), 0.1 mM beta-mercaptoethanol (Sigma-Aldrich), penicillin/streptomycin, 1 mM L-glutamine, and 1% nonessential amino acids (all from Invitrogen). For experiments in 2i culture conditions, mESCs were cultured on gelatin-coated plates with N2B27 + 2i + LIF medium containing: (500 ml), 240 ml DMEM/F12 (Invitrogen; 11320), 240 ml Neurobasal media (Invitrogen;21103), 5 ml N2 supplement (Invitrogen; 17502048), 10 ml B27 supplement (Invitrogen;17504044), 10mg recombinant LIF, 0.1 mM beta-mercaptoethanol (Sigma-Aldrich), penicillin/streptomycin, 1 mM L-glutamine, and 1% nonessential amino acids (all from Invitrogen), 50mg/ml BSA (Sigma), PD0325901 (Stemgent, 1mM), and CHIR99021 (Stemgent, 3mM).

Reporter Cell Lines

To generate stably integrated Gapdh and Dazl transgene reporter cell lines, either Gapdh-or Dazl-modified PiggyBac transposon (see Supplemental Experimental Procedures), and a helper plasmid expressing transposase, were transfected into mESCs cells using Xfect mESC Transfection Reagent (Clontech), according to the provider’s
protocol. Stably integrated reporter cells were selected with puromycin (2 mg/ml) for 4 days. To generate Dazl, Gapdh, miR290, and Sox2 SE reporter cell lines, targeting vectors, and CRISPR/Cas9 were transfected into mESCs using Xfect mESC Transfection Reagent (Clontech), according to the provider’s protocol. Forty-eight hours following transfection, cells were FACS-sorted for GFP or tdTomato expression (respectively) and plated on MEF feeder plates. Single colonies were further analyzed for proper and single integration by Southern blot and PCR analysis.

Flow Cytometry
To assess the proportion of GFP and tdTomato in the established reporter cell lines, a single-cell suspension was filtered and assessed on the LSR II SORP, LSRFortessa SORP, or FACSCanto II.

Retinoic Acid-Induced Differentiation
mESCs carrying the reporter for both miR290 and Sox2 SE regions were sorted for double-positive GFP and tdTomato expression and plated on gelatin-coated plates in ES cell medium (+LIF). The next day, cells were washed with PBS, resuspended in basal N2B27 medium (2i medium without LIF, insulin, and the two inhibitors), and supplemented with 0.25mM RA. Medium was replaced every other day.

Blastocyst Injections for the Generation of Chimeras and Secondary MEFs
Blastocyst injections were performed using (C57Bl/6xDBA) B6D2F2 host embryos. In brief, B6D2F1 females were hormone primed by an intraperitoneal (i.p.) injection of
pregnant mare serum gonadotropin (PMS, EMD Millipore) followed 46 hr later by an injection of human chorionic gonadotropin (hCG, VWR). Embryos were harvested at the morula stage and cultured in a CO2 incubator overnight. On the day of the injection, groups of embryos were placed in drops of M2 medium using a 16-um diameter injection pipet (Origio). Approximately ten cells were injected into the blastocoeel cavity of each embryo using a Piezo micromanipulator (Prime Tech). Approximately 20 blastocysts were subsequently transferred to each recipient female; the day of injection was considered as 2.5 days postcoitum (DPC). Fetuses were collected at 13.5 DPC for the extraction of embryonic fibroblasts as described before (Buganim et al., 2012).

**Southern Blots**

Genomic DNA (10–15 mg) was digested with appropriate restriction enzymes overnight. Subsequently, genomic DNA was separated on a 0.7% agarose gel, transferred to a nylon membrane (Amersham) and hybridized with 32P random primer (Stratagene)-labeled probes.

**Reprogramming to iPSCs**

MEFs isolated from miR290 and Sox2 fetuses were plated at density of 50,000 cells per 6-well in gelatin-coated plates with standard MEF medium (mESCs media without LIF). The following day MEF medium was replaced with mESCs medium containing 2 mg/ml doxycycline (Sigma). Alternatively, cells were grown in mESCs medium containing 2 mg/ml doxycycline and a combination of three compounds (TGF-b antagonist ALK5
inhibitor II, GSK3b antagonist CHIR99021, ascorbic acid) as described before (Vidal et al., 2014). Medium was replaced every other day during the course of reprogramming.

ACKNOWLEDGMENTS

We thank Thorold W. Theunissen, Patti Wisniewski, and Colin Zollo for FACS analyses and cell sorting, Denes Hnisz for providing ChIP-seq tracks, Kibibi Ganz for mouse injections, Huijing Yu for help in cloning, and Stefan Semrau for help with the RA differentiation and comments on the manuscript. This study was supported by NIH grant HD 045022. Y.S. is supported by a Human Frontier Postdoctoral Fellowship and. R.J. is co-founder of Fate Therapeutics and an adviser to Stemgent.
Figure 1. An Active Minimal Snrpn Promoter Can Be Repressed in cis by Means of Spreading of DNA Methylation into the Promoter Region

(A) Schematic representation of the sleeping-beauty-based vectors. Endogenous CpG Islands (CGI) of Dazl and Gapdh genes were cloned upstream of a minimal Snrpn promoter region-driving GFP. Open circle lollipops schematically represent individual unmethylated CpG.

(B) Flow cytometric analysis of V6.5 mESCs cultured in serum + LIF, following stable integration of unmethylated Gapdh and Dazl reporter vectors, demonstrating robust repression of GFP signal in the Dazl reporter cells over time. Shown are the mean percentages of GFP-negative cells ± STD of two biological replicates.

(C) Phase and fluorescence images of the sorted V6.5 mESCs, comprising stable integration of the Gapdh (left) and Dazl (right) vectors following prolonged culturing for 7 weeks.

(D and E) Bisulfite sequencing analysis of the stably transfected Gapdh (D) and Dazl (E) reporter cell lines was performed on the gene promoter-associated CGI (left) and the
downstream Snrpn promoter region (right). Open circles represent unmethylated CpGs; Filled circles, methylated CpGs. See also Figure S1.
Figure 2. An In Vitro Repressed Snrpn Promoter Can Be Reactivated in cis by Means of Spreading of DNA Demethylation into the Promoter Region

(A) Schematic representation of in vitro methylated sleeping-beauty-based vectors. Closed circle lollipops schematically represent individual methylated CpG.

(B) Phase and fluorescence images of the stably integrated V6.5 mESCs harboring Gapdh (left) and Dazl (right) in vitro methylated vectors, following 1 week of antibiotics selection.

(C and D) Flow cytometric analysis of the proportion of GFP-positive cells in V6.5 mESCs, stably integrated with either Gapdh (C) or Dazl (D) in vitro methylated vectors, following 2 weeks in culture.

(E and F) Bisulfite sequencing analysis of the stably transfected Gapdh (E) and Dazl (F) reporter cell lines, was performed on the gene promoter-associated CGI (left) and the downstream Snrpn promoter region (right).

(G) Flow cytometric analysis of the proportion of GFP-positive cells in V6.5 mESCs and Dnmt1 KO mESCs, stably integrated with in vitro methylated Dazl reporter vector.
(H) Bisulfite sequencing analysis of sorted GFP-positive Dnmt1 KO mESCs, stably integrated with in vitro methylated Dazl reporter vector.

(I) Flow cytometric analysis of the proportion of GFP-negative cells in control V6.5 mESCs, mESCs deficient for both Dnmt3a and Dnmt3b (Dnmt3abKO) and V6.5 mESCs cultured in 2i + LIF, which were stably integrated with unmethylated Gapdh (top) and Dazl (bottom) reporter vectors. See also Figure S1.
Figure 3. Generation of DNA Methylation Reporter Cell Lines for Endogenous Gene Promoters

(A) CRISPR/Cas-based strategy used to integrate the DNA methylation reporter into the endogenous promoter region of Gapdh and Dazl genes. TSS, transcription start site; green sequence, endogenous CGI region; black sequence, targeting CRISPR; red sequence, PAM recognition site.

(B) Flow cytometric analysis depicting the mean GFP intensity of randomly picked clones following antibiotic selection of both (top) Gapdh- and (bottom) Dazl-reporter-transfected V6.5 mESCs.

(C) Flow cytometric analysis of the proportion of GFP-positive cells in two representative clones correctly targeted with the methylation reporter at the promoter region of Gapdh.

(D) Bisulfite sequencing analysis was performed on mESCs harboring the DNA methylation reporter in Gapdh promoter region. For each cell line, the PCR amplicon
(marked with dashed line) includes both the endogenous CGI (left) and the downstream integrated Snrpn promoter region (right).

(E) Flow cytometric analysis of the proportion of GFP-positive cells in two representative clones correctly targeted with the methylation reporter at the promoter region of Dazl.

(F) Bisulfite sequencing analysis was performed on mESCs harboring the DNA methylation reporter in Dazl promoter region. For each cell line, the PCR amplicon (marked with dashed line) includes both the endogenous CGI (left) and the downstream integrated Snrpn promoter region (right). See also Figure S2.
Figure 4. Generation of DNA Methylation Reporter Cell Lines for the Pluripotent-Specific miR290 and Sox2 SE Regions

(A) Regional view depicting the DNA methylation (top) and chromatin (bottom) landscape of miR290 upstream pluripotent-specific SE. Shown are average methylation levels and enrichment of chromatin marks in mouse undifferentiated cells (green) and in adult tissues (gold), with respect to the genomic organization of the genes. DNA methylation varies from 1-hypermethylated to 0-hypomethylated. Characteristic clusters of typical enhancer marks and binding of tissue-specific TF determine the SE region (light blue).

(B) CRISPR/Cas-based strategy used to integrate the DNA methylation reporter into the endogenous SE region. HR, homologous recombination; green sequence, endogenous miR290 CpG region; black sequence, targeting CRISPR; red sequence, PAM recognition site.

(C) Phase and fluorescence images of correctly integrated DNA methylation reporter cell lines for miR290 (upper panel) and Sox2 (lower panel) endogenous SE regions. GFP marks endogenous expression levels of Nanog, whereas tdTomato reflects the endogenous DNA methylation levels at both miR290 and Sox2 SE regions.
(D) Bisulfite sequencing analysis was performed on undifferentiated mESCs harboring the DNA methylation reporter in either miR290 SE region (top) or Sox2 SE region (bottom). For each cell line, the PCR amplicon (marked with dashed line) includes both the endogenous CGI (left) and the downstream integrated Snrpn promoter region (right). See also Figure S3.
Figure 5. Dynamics of De Novo DNA Methylation of miR290 and Sox2 SE Regions upon In Vitro Differentiation

(A) Schematic representation of the RA-based differentiation protocol used on miR290 and Sox2 reporter cell lines. GFP marks endogenous expression levels of Nanog, whereas tdTomato reflects the endogenous DNA methylation levels at both miR290 and Sox2 SE regions.

(B) Flow cytometric analysis of the proportion of Nanog-GFP-positive cells (x axis) and tdTomato-positive cells (y axis) during 7 days of differentiation of miR290#21 (top) and Sox2#2 (bottom) reporter cell lines.

(C) Bar graph summarizing the proportion of the different cell populations during the course of 7 days RA differentiation for both miR290#21 (top) and Sox2#2 (bottom) reporter cell lines. Data represent two biological replicates. R, tdTomato; G, GFP.

(D and E) Bisulfite sequencing analysis on the three main cell populations sorted at 48 hr following initial treatment with RA. For both miR290 #21 (D) and Sox2#2 (E) cell lines, the PCR amplicon (marked with dashed line) includes the endogenous CGI (left) and the downstream integrated Snrpn promoter region (right). R, tdTomato; G, GFP. See also Figure S4.
Figure 6. Dynamics of DNA Demethylation of miR290 and Sox2 SE Regions during Cellular Reprogramming
(A) miR290 (top) and Sox2 (bottom) reporter chimeric experimental embryos (right embryo in each panel). As controls, Gapdh CGI reporter mESCs driving GFP and constitutively expressing tdTomato (Control, Gapdh-GFP, and tdTomato, respectively) were injected into host blastocyst (left embryo in each panel).

(B) Schematic representation of the experimental procedure to monitor the dynamics of demethylation during reprogramming of miR290 and Sox2 reporter cell lines. GFP marks endogenous expression levels of Nanog, whereas tdTomato reflects the endogenous DNA methylation levels at both miR290 and Sox2 SE regions.

C) Flow cytometric analysis of the proportion of GFP-positive cells (x axis) and tdTomato-positive cells (y axis) in P0 MEFs derived from miR290#21 (left) and Sox2#2 (right) chimeric embryos.

(D) Bisulfite sequencing analysis was performed on P0 MEFs derived from miR290#21 (top) and Sox2 #2 (bottom) chimeras. For each cell line, the PCR amplicon (marked with dashed line) includes both the endogenous CGI (left) and the downstream integrated Snrpn promoter region (right).

(E) Analysis of the proportion of GFP-positive cells (x axis) and tdTomato-positive cells (y axis) during the course of reprogramming of MEFs derived from miR290 #21 (upper panel) and Sox2#2 (lower panel) chimeras. Shown are flow cytometric data from different time points following addition of dox supplemented with 3C culture condition.

(F) Representative images of established miR290 and Sox2 iPSC lines, derived from sorted double-positive (tdTomato+/GFP+) colonies.

(G) Bisulfite sequencing analysis was performed on P2 iPSCs derived from miR290#21 (top) and Sox2#2 (bottom) MEFs. For each cell line, the PCR amplicon (marked with dashed line) includes both the endogenous CGI (left) and the downstream integrated Snrpn promoter region (right). See also Figure S5.
REFERENCES


Supplemental Experimental Procedures

Plasmid Cloning

To clone the PiggyBac-Insulator-GapdhCGI-Snrpn-GFP-polyA-PGK-PURO-sv40PolyA-Insulator construct, the minimal Snrpn promoter was PCR amplified using primers A1 and A2 (see complete primer list below). Snrpn PCR fragment was subsequently digested using MfeI and NheI restriction enzymes. GapdhCGI sequence was PCR amplified using primers A3 and A4, following digestion using SbfI and MfeI. A pCR2.1-TOPO-TA cloning vector (Life technologies) vector containing a GFP-PolyA-PGK-Puro cassette was digested using SbfI and NheI. Subsequently, these 3 DNA fragments were cloned using three-way ligation. The resulting GapdhCGI-Snrpn-GFP-PolyA-PGK-Puro cassette was then cloned into a PiggyBac transposon using the restriction enzymes SbfI and SacI to generate the PiggyBac-Insulator-GapdhCGI-Snrpn-GFP-polyA-PGK-PURO-sv40PolyA-Insulator vector. For the PiggyBac-Insulator-DazlCGI-Snrpn-GFP-polyA-PGK-PURO-sv40PolyA-Insulator construct, the same method was used, except that DazlCGI DNA fragment was PCR amplified using primers A5 and A6.

To Clone the mi290 super enhancer (SE) targeting vector, the 5' homology arm was PCR amplified using the primers B1 and B2, this DNA fragment was then digested using SbfI and MfeI restriction enzymes. The 3' homology arm was PCR amplified using the Primers B3 and B4, following digestion with AscI and FseI restriction enzymes. Both homology arms were subsequently ligated with Snrpn-tdTomato-PolyA-PGK-Puro fragment that had been digested with NheI and AscI restriction enzymes, and a pCR2.1-TOPO-TA cloning vector (Life technologies) backbone that had been digested with SbfI and FseI. To clone the Sox2 SE targeting vector, the same method was used except that 5' homology arm was amplified using primers C1 and C2, and the 3' homology arm was amplified using primers C3 and C4.

CRISPR oligonucleotides were ligated into px330 vector using BbsI restriction site as previously described (Wang et al., 2013). For Mi290 SE region oligonucleotides D3 and D4 were used and for the Sox2 SE region, the oligonucleotides D1 and D2 were used (see complete primer list below).

Bisulfite Conversion, PCR and Sequencing

Bisulfite conversion of DNA was established using the EpiTect Bisulfite Kit (Qiagen) following the manufacturer's instructions. The resulting modified DNA was amplified by first round of nested PCR, following a second round using loci specific PCR primers (see complete list of primers below). The first round of nested PCR was done as follows: 94 °C for 4 min ;55 °C for 2 min; 72 °C for 2 min; Repeat steps 1-3 1X; 94 °C for 1 min; 55 °C for 2 min; 72 °C for 2 min; Repeat steps 5-7 35X; 72 °C for 5 min; Hold 12°C. The second round of PCR was as follows: 95 °C for 4 min ;94 °C for 1 min; 55 °C for 2 min; 72 °C for 2 min; Repeat steps 2-4 35X; 72 °C for 5 min; Hold 12°C. The resulting amplified products were gel-purified, subcloned into A pCR2.1-TOPO-TA cloning vector (Life technologies), and sequenced.
Primer List – cloning

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 snrpnF-mfe</td>
<td>AATTAACAAATTGACGCTCAAATTTCCGCAGTAGG</td>
</tr>
<tr>
<td>A2 snrpnR-nhe</td>
<td>AATTAAGCTAGCAGAATCCACAAGGCCAGCTG</td>
</tr>
<tr>
<td>A3 gapdhF-sbf</td>
<td>AATTAACCTGAGGGGCTCCCTCTCCTCCCATTTCTC</td>
</tr>
<tr>
<td>A4 gapdhR-mfe</td>
<td>AATTAACATTGGAGAGAGGCCCAGCTACTCG</td>
</tr>
<tr>
<td>A5 dazlF-sbf</td>
<td>AATTAACCTGAGGGGCTCCCTCTCCTCCCATTTCTC</td>
</tr>
<tr>
<td>A6 dazlR-mfe</td>
<td>AATTAACATTGGAGAGAGGCCCAGCTACTCG</td>
</tr>
<tr>
<td>B1 miR290-5'arm F</td>
<td>AATTAACCTGAGGGGCTCCCTCTCCTCCCATTTCTC</td>
</tr>
<tr>
<td>B2 miR290-5'arm R</td>
<td>AATTAACATTGGAGAGAGGCCCAGCTACTCG</td>
</tr>
<tr>
<td>B3 miR290-3'arm F</td>
<td>AATTAACCTGAGGGGCTCCCTCTCCTCCCATTTCTC</td>
</tr>
<tr>
<td>B4 miR290-3'arm R</td>
<td>AATTAACATTGGAGAGAGGCCCAGCTACTCG</td>
</tr>
<tr>
<td>C1 Sox2-5'arm F</td>
<td>AATTAACCTGAGGGGCTCCCTCTCCTCCCATTTCTC</td>
</tr>
<tr>
<td>C2 Sox2-5'arm R</td>
<td>AATTAACATTGGAGAGAGGCCCAGCTACTCG</td>
</tr>
<tr>
<td>C3 Sox2-3'arm F</td>
<td>AATTAACCTGAGGGGCTCCCTCTCCTCCCATTTCTC</td>
</tr>
<tr>
<td>C4 Sox2-3'arm R</td>
<td>AATTAACATTGGAGAGAGGCCCAGCTACTCG</td>
</tr>
<tr>
<td>D1 SOX2-SE Crispr</td>
<td>CACCGCGTGCACATTTCAAAAATG</td>
</tr>
<tr>
<td>D2 SOX2-SE Crispr</td>
<td>AAACCTTTCAGCTTGACAAACCT</td>
</tr>
<tr>
<td>D3 miR290-EN2 Crispr</td>
<td>CACCGCGTGCACATTTCAAAAATG</td>
</tr>
<tr>
<td>D4 miR290-EN2 Crispr</td>
<td>AAACCTTTCAGCTTGACAAACCT</td>
</tr>
<tr>
<td>DazF-5' arm</td>
<td>AATTAACCTGAGGGGCTCCCTCTCCTCCCATTTCTC</td>
</tr>
<tr>
<td>DazR-5' arm</td>
<td>AATTAACATTGGAGAGAGGCCCAGCTACTCG</td>
</tr>
<tr>
<td>Dazl F - 3'arm</td>
<td>AATTAACATTGGAGAGAGGCCCAGCTACTCG</td>
</tr>
<tr>
<td>Dazl R - 3'arm</td>
<td>AATTAACATTGGAGAGAGGCCCAGCTACTCG</td>
</tr>
<tr>
<td>Dazl F Crispr</td>
<td>AATTAACATTGGAGAGAGGCCCAGCTACTCG</td>
</tr>
<tr>
<td>Dazl R Crispr</td>
<td>AATTAACATTGGAGAGAGGCCCAGCTACTCG</td>
</tr>
<tr>
<td>Gapdh F - 3'arm</td>
<td>AATTAACATTGGAGAGAGGCCCAGCTACTCG</td>
</tr>
<tr>
<td>Gapdh R - 3'arm</td>
<td>AATTAACATTGGAGAGAGGCCCAGCTACTCG</td>
</tr>
<tr>
<td>Dazi F - 3'arm</td>
<td>AATTAACATTGGAGAGAGGCCCAGCTACTCG</td>
</tr>
<tr>
<td>Dazi R - 3'arm</td>
<td>AATTAACATTGGAGAGAGGCCCAGCTACTCG</td>
</tr>
<tr>
<td>Dazi F Crispr</td>
<td>AATTAACATTGGAGAGAGGCCCAGCTACTCG</td>
</tr>
<tr>
<td>Dazi R Crispr</td>
<td>AATTAACATTGGAGAGAGGCCCAGCTACTCG</td>
</tr>
<tr>
<td>Gapdh F - 5'arm</td>
<td>AATTAACATTGGAGAGAGGCCCAGCTACTCG</td>
</tr>
<tr>
<td>Gapdh R - 5'arm</td>
<td>AATTAACATTGGAGAGAGGCCCAGCTACTCG</td>
</tr>
<tr>
<td>Gapdh F Crispr</td>
<td>AATTAACATTGGAGAGAGGCCCAGCTACTCG</td>
</tr>
<tr>
<td>Gapdh R Crispr</td>
<td>AATTAACATTGGAGAGAGGCCCAGCTACTCG</td>
</tr>
</tbody>
</table>

Primer List – Bisulfite

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP Nested R</td>
<td>CTCGACAAAAATAAAAACCCACCC</td>
</tr>
<tr>
<td>Dazl Nested F</td>
<td>GAAAGTTTTGTAATTAGTGTGTTGAG</td>
</tr>
<tr>
<td>Dazl F</td>
<td>CGATTAGAGAGTTGTTGTTGAG</td>
</tr>
<tr>
<td>Dazl R</td>
<td>CGTTACAAAAATCTCAACCC</td>
</tr>
<tr>
<td>Dazi-Snrpn F</td>
<td>CGATTAGAGAGTTGTTGTTGAG</td>
</tr>
<tr>
<td>Dazi-Snrpn R</td>
<td>CGTTACAAAAATCTCAACCC</td>
</tr>
<tr>
<td>Gapdh Nested F</td>
<td>GGTTGTAGGAGAGAGAGAG</td>
</tr>
<tr>
<td>Gapdh F</td>
<td>GGTTGTAGGAGAGAGAGAG</td>
</tr>
<tr>
<td>Gapdh R</td>
<td>GGTTGTAGGAGAGAGAGAG</td>
</tr>
<tr>
<td>Gapdh-Snrpn F</td>
<td>GGTTGTAGGAGAGAGAGAG</td>
</tr>
<tr>
<td>Gapdh-Snrpn R</td>
<td>GGTTGTAGGAGAGAGAGAG</td>
</tr>
<tr>
<td>miR290 Nested F</td>
<td>GGAGGATTTTTGGGAG</td>
</tr>
<tr>
<td>miR290 Nested R</td>
<td>GGAGGATTTTTGGGAG</td>
</tr>
<tr>
<td>miR290-Snrpn F</td>
<td>GGAGGATTTTTGGGAG</td>
</tr>
<tr>
<td>miR290-Snrpn R</td>
<td>GGAGGATTTTTGGGAG</td>
</tr>
<tr>
<td>Sox2 Nested F</td>
<td>GGTTGTAGGAGAGAGAGAG</td>
</tr>
<tr>
<td>Sox2 Nested R</td>
<td>GGTTGTAGGAGAGAGAGAG</td>
</tr>
<tr>
<td>Sox2-Snrpn F</td>
<td>GGTTGTAGGAGAGAGAGAG</td>
</tr>
<tr>
<td>Sox2-Snrpn R</td>
<td>GGTTGTAGGAGAGAGAGAG</td>
</tr>
</tbody>
</table>
Chapter 5. Future directions

CRISPR-Cas9 genome-engineering in mice.

In chapter 2 and 3 of this thesis we presented work that shows that CRISPR/Cas9 can be used to efficiently genetically engineer the mouse genome. Since this work was completed the technology has been substantially improved. One of the biggest questions that still needed to be addressed after the completion of this work was the potential off-target effects of using CRISPR/Cas9 for genetic engineering. However, recently full genome sequencing of CRISPR/Cas9 targeted human iPSCs and mice that were generated through Cas9 injected zygotes showed that there are very few off-target effects (Smith et al., 2014; Veres et al., 2014; Iyer et al., 2015). In addition, the CRISPR/Cas9 technology has been improved to further decrease off-target effects. Development of a Cas9 nickase, dCas9 fused to Fok1, and/or shorter gRNAs where shown to decrease unwanted off-target effects (Fujii et al., 2014; Ran et al., 2013; Tsai et al., 2014; Fu et al., 2014). Another area of our methodology that could be improved upon is the efficiency of HDR for larger dsDNA-targeting vectors. The efficiency of HDR has been increased by two different methods. The first method showed that the efficiency of HDR could be increased by injecting Cas9 protein into the zygote instead of mRNA (Aida et al., 2015). The second method showed that injecting an inhibitor SCR7 into the zygote, along with Cas9 mRNA, gRNA, and a dsDNA-targeting vector or ssDNA-oligos, could increase HDR (Maruyama et al., 2015). The SCR inhibits the DNA ligase-IV enzyme that is a critical component in the NHEJ pathway. Therefore, increasing the probability that the cell repairs a dsDNA break through HDR instead of NHEJ (Maruyama et al., 2015). Finally, one of the most difficult parts of our methodology for genome-engineering the
mouse by zygotic injections of CRISPR/Cas9, is the actual injection themselves. It requires a very skilled research scientist or technician to be able to inject Cas9 mRNA, gRNA, and DNA into the single-cell zygote. To make CRISPR/Cas9 genome-engineering in the mouse more accessible, a new delivery system was developed that showed that Cas9 mRNA, gRNA, and DNA could be electroporated into the single-cell zygote (Qin et al., 2015).

**Future directions for RGM**

In chapter 4 of this thesis we presented research on the creation of a new technology called Reporter for Genomic Methylation (RGM) that can be used to investigate the dynamics of methylation at an endogenous locus in vivo. We further showed that this reporter was able to accurately report on the dynamics of methylation at a few important cis-regulatory elements such as promoters and super-enhancers. However, there is still much work to be done, both in investigating how the tool works, improving the technology, and in understanding important biological questions that the technology can help address.

**Further characterization and improvement of RGM**

One important question that needs to be answered is the correct positional targeting of the RGM technology into an endogenous DMR. Is there a specific spatial distance where the RGM has to be integrated, does it have to be integrated into the DMR, very close to the DMR, or can it be inserted farther away? This most likely depends on how far methylation/demethylation can spread from the DMR and could be locus-specific
or be affected by other proximal sequence elements such as CTCF or SP1 binding sites. Research has shown that methylation can spread from DMRs, and indeed the fact that the RGM technology works proves that methylation and demethylation can spread from DMRs into cis-proximal promoters (Turker 2002; Irizarry et al., 2009). However, the distance that methylation and demethylation can spread is currently unknown. Furthermore, transcription factors such as SP1 have been shown to block the spread of methylation (Brandies et al 1994; Macleod et al., 1994). Also factors such as CTCF that act as insulators, and are important for the 3D chromatin architecture, could also impede the spreading of methylation and demethylation (Herold et al., 2012). These questions could be investigated by targeting an endogenous DMR, such as the DAZL DMR or the mi290 super-enhancer DMR, with the RGM technology. In each targeting event the RGM would be integrated progressively further away from the DMR boundary, this would allow for a better understanding into the spatial distance that methylation and demethylation can spread. Furthermore, repeating this at multiple endogenous DMRs, and bioinformatically aligning the genomic target sites with SP1, CTCF, other structural component, and transcription factor binding motifs, would allow for the formation of a set of parameters for targeting endogenous CpG islands/DMRs. This would indicate whether any specific sequence motifs effect the spread of methylation/demethylation.

Another important question that needs to be addressed is the on/off rate of our synthetic SNRP promoter. Is the methylation/demethylation of the synthetic promoter rapid or slow, does it vary at different loci, does it vary when the distance between the synthetic promoter and DMR it is reporting on changed? These are very complex questions to answer and control for, but they will have to be investigated so as to build a
framework for what the technology can be used for. One factor that will help in the investigation of the on/off rate will be to clone the synthetic SNRP promoter upstream of an unstable GFP or luciferase reporter. Because of the short half-life of these proteins, the promoter activity will be more efficiently correlated with the expression signal (Solberg and Krauss, 2013; Li et al., 1998).

Another interesting question that was not resolved in the initial study, is what CpGs in the synthetic promoter are important for controlling the on/off state when it is methylated/demethylated. The synthetic SNRP promoter is 284bp long and contains 16 CpGs dinucleotide. Therefore, this question could be addressed by doing site-directed mutagenesis of each of the CpGs, to see which are important for regulating the on/off state of this synthetic minimal promoter. Research on the HPRT promoter, which is also regulated by methylation, showed that only 3 CpGs were critical for turning off the promoter when methylated (Chen et al., 2001). This information could be used to potentially make a more sensitive variant of the synthetic SNRP promoter.

Finally, it will be interesting to investigate if the synthetic SNRP promoter can be used in other species besides the mouse model organism. The synthetic SNRP promoter was initially designed by taking the core homologous sequence elements between the human and mouse SNRP promoters. Therefore, it is highly probable that the RGM technology should work just as efficiently in human cells, especially because the methylation machinery is also highly conserved between humans and mice (Smith and Meissner, 2013). However, it will be interesting to see if the synthetic SNRP promoter will work in more evolutionary divergent model organisms, such as bacteria, zebrafish, fruit flies, and plants. Furthermore, if the synthetic SNRP promoter cannot work in any of
these model organisms, it will be interesting to investigate if alternative synthetic promoters can be made using similar design principles for these species.

**Applications for RGM in imprinting, screening and cancer**

There are multiple biological questions to answer using the RGM technology. However, for brevity, this discussion will look at a few of the more discernible ones. These include questions in imprinting, locus-specific methylation and demethylation, and how methylation regulates tumor suppressor and oncogene expression.

For imprinting, there are multiple questions that can be addressed using the RGM technology. First of all, it will be interesting to see if the RGM technology can be used to faithfully report on parent-of-origin imprints. This can be studied by making a mouse model that has the RGM reporter integrated into an imprinted DMR such as H19, IGF2, or DLK. In these mice the RGM reporter activity should depend on whether the allele is inherited from the mother or father (Smallwood and Kelsey, 2012). Furthermore, these mice could be used to investigate whether there is heterogeneous loss or gain of imprints in somatic tissues. Research on Macaque monkeys has indicated that there might be heterogeneous loss of imprinting at IGF2 and DLK in some somatic tissues (Cheong et al., 2015). In addition, it has been reported that there is loss of imprinting in mice at the DLK locus in niche astrocytes (Ferron et al., 2011). Another interesting question that can be addressed by using the RGM technology to target imprinted regions is whether there is heterogeneous loss or gain of methylation at these loci in mESCs and iPSCs. Some research indicates that ESCs and iPSCs heterogeneously lose their imprint at some loci and that this aberrant loss of methylation affects their developmental potential (Stadtfeld
et al., 2012; Dean et al., 1998; Sun et al., 2012). Heterogeneity of methylation at imprinted loci in mouse ES or IPS cell lines can be investigated by targeting both alleles of an imprinted locus, such as DLK-DIO3, with the RGM technology that expresses two different fluorescent markers. The ESC's or IPSC's developmental potential can be investigated by FACS sorting the negative, single-positive, and double-positive cells, and then subjecting these different populations of cells to tetraploid complementation or blastocyst injections.

Another interesting application for the RGM technology is to use it to genetically screen for factors that are important for locus-specific methylation during development. Although it is known that the DNMTs are the functional enzymes that methylate DNA, less is known about how these enzymes are recruited to specific loci such as super-enhancers or promoters during development. A search for the factors that recruit DNMTs to specific super-enhancers could be investigated by using the two cell lines, SOX2#2 and mi290#21, that were reported in chapter 4 of this thesis. A full genome CRISPR/Cas9 lentivirus library could be used to infect a population of these cells in the ES cell state (Shalem et al., 2014; Wang et al.; 2014). These cells could then be subjected to in vitro differentiation. As was reported in chapter 4, when these cells differentiate the mi290 and Sox2 super enhancer becomes methylated and the RGM reporter shuts off. However, the RGM reporter should remain active in any cells that have lost a factor that is necessary for recruiting DNMTs to the mi290 or Sox2 super-enhancer. After differentiation, the remaining fluorescent cells can be isolated by FACS, and the gene knockout of interest can be found by sequencing the integrated CRISPR gRNA.
As was noted in chapter 1 of this thesis, many genes in cancer become deregulated by changes that occur in the methylation state of their promoter elements. This includes loss of methylation at the promoters of oncogenes and the gain of methylation at the promoters of tumor suppressor genes. Furthermore, it was described how LOI at certain imprinted loci, such as the IGF2 locus, can also lead to cancer in multiple tissue types such as the lung and colon (Moulton et al., 1994; Steenman et al. 1994). It would be interesting to target the RGM technology to one of these important differentially methylated regions in a relevant cancer model/cell type. One locus of particular interest would be the imprinted DMR that regulates IGF2 in lung and colon cancer. If it were verified that there is LOI in this cancer model, it would be of a possible therapeutic benefit to use this RGM-targeted cancer cell model to screen for small molecules that restore normal imprinting at this locus.

Concluding remarks

The mouse model organism has had tremendous impact on biological research, allowing for the elucidation of many fundamental biological questions that are important not just for basic science but for human health. Future research using the mouse model organism will continue along this path. It is the hope of this author that the research presented in this thesis will help, even in a small way, future researchers in this endeavor.
References


Tsai SQ, Wyvekens N, Khayter C, Foden JA, Thapar V, Reyon D, Goodwin MJ,


Curriculum vitae

Chikdu Shakti Shivalila

Degree

- PhD. Biological Sciences : Massachusetts Institute of Technology
  Expected Graduation Date: December 2015

- B.S. Biological Sciences : University of Pittsburgh
  Graduation date: December 2009
  Undergraduate GPA: 3.92

Current Occupation

- Fourth year Graduate student in the Department of Biology at the Massachusetts Institute of Technology.
  August 2011-Present

Laboratory Experience

- **Position:** Graduate student in Dr. Rudolf Jaenisch’s lab at MIT/Whitehead Institute of Biomedical Sciences (May 2012-Present)
  **Research:** Genome engineering, Human and mouse stem cell biology, Epigenetics, Neuroscience

  In Dr. Rudolf Jaenisch’s Lab I have focused my research on technology development and synthetic biology, specifically in the area of genome engineering and epigenetic engineering. My earlier work consisted of developing a rapid way to genetically engineer the mouse model organism and mouse stem cells using Talens and Crispr/Cas. In addition, I have worked on building protocols to efficiently genetically engineer human stem cells using Crispr/Cas. More recently, I have been focused on developing a technology that can accurately report on DNA methylation/Demethylation dynamics at a single cell resolution. Furthermore, I have used this technology to understand DNA methylation dynamics at interesting genomic loci such as: super-enhancers, imprinted regions, and promoters.

- **Position:** Research specialist II/ Lab manager in Dr. James Pipas’s Lab at the University of Pittsburgh (January 2009 - August 2011)
  **Research:** Virology, Cancer Biology, Cell Biology, Biochemistry, Epigenetics

  In Dr. James Pipas’s laboratory my research was focused on SV40 Polyoma Virus, specifically the SV40 Large T Ag and how this oncoprotein causes cellular transformation through epigenetic deregulation.

Publications


**Teaching Experience**

- Teaching assistant: 7.01 (introductory biology) MIT. Fall 2012
- Teaching assistant: 7.02 (intro to Exp biology and comm) MIT. Spring 2015

**References**

Dr. Rudolf Jaenisch  
MIT/ Whitehead Institute of Biomedical Sciences  
9 Cambridge Center, Cambridge, MA 02142 :  
Lab phone: 617-258-5189

Dr. David Bartel  
MIT/ Whitehead Institute of Biomedical Sciences  
9 Cambridge Center, Cambridge, MA 02142 :  
Lab phone: 617-258-5287

Dr. Piyush Gupta  
MIT/ Whitehead Institute of Biomedical Sciences  
9 Cambridge Center, Cambridge, MA 02142 :  
Lab phone: 617-324-0086

Dr. James Pipas  
University of Pittsburgh, Department of Biological Sciences, 559B Crawford Hall 4249 Fifth Avenue, Pittsburgh, PA 15260 :  
Lab Phone: (412) 624-4691