

Gene Targeting in Human Pluripotent Cells

D. HOCKEMEYER¹ AND R. JAENISCH^{1,2}

*¹Whitehead Institute and Department of Biology, ^{1,2}Massachusetts Institute of Technology,
Cambridge, Massachusetts 02142*

Correspondence: jaenisch@wi.mit.edu

Mouse embryonic stem cells (mESCs) have the ability to differentiate into any cell type and can generate chimeric mice when transplanted into a host blastocyst. This remarkable potential, together with the development of robust gene targeting strategies in mESCs, were essential for establishing the mouse as the most widely used model organism in biomedical research. Recent advances have allowed the isolation of human embryonic stem cells and the derivation of induced pluripotent stem cells. Genetic tools similar to those proven routine in the mouse system are needed to realize the full potential of human pluripotent cells as disease models and putative therapeutics. Gene targeting in human cells, however, has proven to be more difficult, more time-consuming, and less robust than in mESCs. In this chapter, we discuss the strategies that have been used to allow specific genetic modifications in human pluripotent cells. We focus on the novel application of custom-engineered zinc-finger nucleases for gene targeting, which has promise to become a robust tool for efficient genetic manipulation of human pluripotent cells.

The isolation of the first human embryonic stem cells (hESCs) by Thomson et al. provided hope that these cells would not only be a useful tool to study early human development but could also be used to study human genetic diseases in vitro and derive human tissue for transplantation therapies (Thomson et al. 1998). Recent advances pioneered by Yamanaka and colleagues in 2006 have allowed the generation of human pluripotent cells from adult fibroblasts (Takahashi and Yamanaka 2006; Takahashi et al. 2007), limiting the need to harvest cells from human embryos. This process, called somatic-cell reprogramming, reverts a cell's epigenetic state to a developmentally earlier and therefore more potent state. These induced pluripotent stem cells (iPSCs) are attractive for biomedical research because they can be derived from patients who carry genetic predispositions for a given disease and from those who would benefit from cell transplantation. The realization of these applications for iPSC technology will require reliable tools for genetically modifying iPSCs. This would allow manipulation of the disease pathway of interest and potential genetic repair of disease-causing mutations in patient-derived cells. In addition, the ability to target genes in iPSCs would allow the generation of reporter cell lines that facilitate the identification and isolation of specific cell types in vitro. However, genetic manipulations of human cells by procedures that are standard in the mouse system have proven to be laborious and inefficient in hESCs and are not useful for routine applications. In this chapter, we discuss the basic principles of gene targeting and review the approaches that have been used thus far to manipulate genes in human cells. We then present a novel form of gene targeting in human pluripotent cells, using zinc-finger nucleases (ZFNs), that promises to provide a general tool to achieve gene targeting in human pluripotent cells.

GENE TARGETING IN MOUSE EMBRYONIC STEM CELLS

Before discussing gene targeting efforts in human cells, we first introduce the general principles and mechanisms of gene targeting as defined in the classic mouse system (Capecchi 1989; Koller and Smithies 1992). mESCs display high activity of the homologous recombination (HR) DNA repair pathway, a key feature that sets them apart from most other mammalian cells and allows efficient genetic manipulation by gene targeting. DNA repair by HR is one of two cellular pathways that normally repair double-strand DNA breaks (DSBs) (Helleday et al. 2007). Whereas HR repairs the DSB with high fidelity, the second pathway, the nonhomologous end joining (NHEJ) pathway, fuses the broken ends in an error-prone fashion (Valerie and Povirk 2003; Lieber 2010). The reason for the high fidelity of the HR pathway lies in its ability to use sequence information of the sister chromatid to repair DSB lesions. After the occurrence of a DSB, the DNA ends are resected, generating long single-strand 3' DNA ends that can invade the sister-chromatid-bearing homology to the site of the DSB. The 3' end of the broken DNA end is extended by DNA polymerases using the intact sister chromatid as a template. At this point, the sequence information of the sister chromatid is copied onto the broken strand, thereby restoring lost sequence information caused by the DSB. After this DNA synthesis step, the strand invasion is resolved and the initially broken chromosome ends are religated. In a case in which the sequence of the homologous template that is used in HR differs from the original sequence of the broken chromosome, HR can result in a gene conversion event. This gene conversion event is the underlying principle of gene targeting. During gene targeting, a cell aberrantly uses an ectopically provided DNA template

that mimics the sister chromatid by bearing homologous sequences to the targeted genomic region but is mutated to allow modifications of the endogenous sequence. In comparison to somatic cells, mESCs exhibit a higher propensity for the HR pathway over the NHEJ pathway, thereby facilitating precise alteration of the mouse genome with high efficiency.

It was originally presumed that similar strategies could be used for genetic manipulations in human cells; however, this did not turn out to be the case due to key differences in the biology of hESCs and mESCs. The exact reasons for the low efficiency of gene targeting by HR in hESCs are not well understood but may include reduced HR activity or a higher propensity to use the NHEJ pathway. Indeed, it was reported that hESCs preferentially use NHEJ to rejoin DSBs induced by ionizing radiation, in contrast to mESCs that show a higher propensity for HR (Banuelos et al. 2008). This difference could in part be due to the reduced expression of the NHEJ proteins Ku70/80 and Lig4 in mESCs and indicates that the higher rate of HR in mESCs is partially caused by their reduced ability to use the NHEJ pathway (Banuelos et al. 2008). In addition, the substantial biological differences between mESCs and hESCs, such as the ability to grow as single cells, may profoundly affect gene targeting efficiency.

The difficulties in targeting hESCs by HR is illustrated by the observation that, so far, <20 genes have been successfully targeted in hESCs (Thomson et al. 1998; Zwaka and Thomson 2003; Urbach et al. 2004; Costa et al. 2007; Irion et al. 2007; Davis et al. 2008a,b; Suzuki et al. 2008; Ruby and Zheng 2009; Xue et al. 2009; Song et al. 2010; see below). This is in contrast to the thousands of examples of successful gene targeting in mESCs. This limitation presents a major challenge that must be overcome to fully develop human pluripotent cells as a useful research and clinical tool.

GENE TARGETING STRATEGIES IN HUMAN PLURIPOTENT CELLS

Several independent systems have been developed to target genes in human cells. Here, we provide examples of the few successful conventional gene targetings in hESCs and illustrate some alternative strategies that have been used to enhance targeting efficiency in human pluripotent cells. Finally, we focus on one recently developed strategy based on ZFNs that is expected to substantially accelerate gene targeting in human pluripotent cells.

Conventional Gene Targeting in Human Pluripotent Cells

Although only a few successful examples of conventional gene targeting are available, they illustrate the variety of applications for gene targeting in pluripotent cells. The first report of using HR to manipulate hESCs by Zwaka and Thomson (2003) described the generation of a loss-of-function allele for the hypoxanthine-guanine phosphoribosyltransferase (*HPRT*) gene. Furthermore, Zwaka and Thomson (2003) used HR to knock in green

fluorescent protein (GFP) into the *OCT4* locus to report the transcriptional activity of this gene in hESCs. These cells were instrumental in the discovery of an alternative set of reprogramming factors, namely OCT4, KLF4, SOX2, NANOG, and LIN28, which were later used for the derivation of iPSCs from somatic human cells (Yu et al. 2007). The *OCT4* gene was targeted by a gene trap approach using a selection cassette without promoter that must therefore be expressed from the endogenous OCT4 promoter following correct insertion. Because the construct does not have a promoter, it is less likely to be expressed when aberrantly integrated into the genome. Therefore, the number of apparent random integrations is reduced and the identification of correctly targeted clones is simplified. Such gene trap approaches require active expression of the targeted gene, as is the case for *OCT4* in hESCs. Therefore, this approach cannot be used to target genes specific to differentiated cell types such as lineage markers in human pluripotent cells.

Conventional gene targeting approaches have also been used to generate cell lines that report in real time the specific cell-fate decisions of hESCs during differentiation (Davis et al. 2008b; Ruby and Zheng 2009; Xue et al. 2009). Targeting of the *Mixl1* (Davis et al. 2008b) and *Olig2* (Xue et al. 2009) loci exemplify this experimental approach. A GFP reporter was inserted into the *Mixl1* gene and used to identify and isolate cell populations enriched for primitive hematopoietic precursors. Similarly, *Olig2*, a transcription factor important for motor neuron and oligodendrocyte development, was targeted to generate a GFP reporter cell line. Following directed differentiation of *Olig2*-targeted hESCs, GFP expression was used to isolate cell types differentiating toward a neuronal or oligodendrocyte fate.

To establish a reliable system to ectopically express genes in hESCs, Irion et al. (2007) identified the human homolog of the mouse *ROSA26* locus and established a gene targeting strategy for this locus. *ROSA26* is frequently used as a "safe harbor" locus in the mouse because it can be targeted with high efficiency in mESCs and allows nearly ubiquitous expression without any reported detrimental effects when both alleles are disrupted (Zambrowicz et al. 1997). This locus has been successfully targeted in hESCs with an expression cassette for red fluorescent protein (RFP) (Irion et al. 2007). RFP was found in several different human cell types differentiated from these targeted hESCs, indicating that the *ROSA26* locus might be suitable for transgene expression in hESCs and hESC-derived cells. However, beyond this original report, the human *ROSA26* locus has thus far not been routinely used to express transgenes in hESCs.

Enhancing Gene Targeting by Alternative Forms of Homologous DNA Delivery

Gene targeting requires the delivery of ectopic DNA that can be used by cells as a template for HR. Generally, standard transfection and electroporation techniques are used to deliver targeting constructs in the form of plasmid DNA (for review, see Costa et al. 2007; Davis et al.

2008a). However, alternative strategies such as the use of viruses or bacterial artificial chromosomes (BACs) to deliver targeting constructs have been developed to facilitate gene targeting in human cells.

One approach is to use adeno-associated viruses (AAVs), which can be genetically engineered to be non-integrating and maintained as episomal particles in the cell. Such AAVs have been used to deliver DNA constructs for gene targeting in human cells and can increase gene targeting efficiency to up to 1% of the infected cells being correctly targeted by HR (Russell and Hirata 1998). Examples of genes targeted with this approach in hESCs were *HPRT* and *HMGAI* (high-mobility-group AT-hook) (Khan et al. 2010). However, AAVs are commonly used to deliver targeting constructs used for a gene trap approach, because a selection cassette that carries its own promoter would be constitutively expressed from the AAV episome. Therefore, it remains to be determined whether AAVs will also prove useful for gene targeting of nonexpressed genes in human pluripotent cells.

In addition to AAVs, BACs have also been used to deliver DNA to target genes in hESCs by HR (Song et al. 2010). Recombineering technology in bacteria can be used to genetically engineer BACs, which function similarly to conventional plasmids by carrying a selection cassette and homology arms. However, a key difference to regular targeting constructs is that BAC vectors can be generated with very long (>100 kb) homology arms. The increased size of the homology arms is thought to increase gene targeting efficiency in hESCs (Song et al. 2010). A BAC-based approach was used to disrupt both alleles of *ATM* or *p53* by sequential targeting in hESCs (Song et al. 2010). Further experiments are required to determine if BAC-mediated targeting can generally be applied to target genes in human pluripotent cells.

ZINC-FINGER NUCLEASES

In this section, we describe site-specific DNA nucleases (called ZFNs) that can be used to facilitate gene targeting in human cells in ways that offer several advantages over previous techniques.

An alternative strategy to enhance gene targeting efficiency came from the observation that the introduction of a DSB in a genomic region increased HR efficiency in the proximity of the break (Brenneman et al. 1996; Jasin 1996). This effect was initially described in experiments where a defined DSB was generated in cells using homing endonucleases. Homing endonucleases, such as the intron-encoded endonuclease I from *Saccharomyces cerevisiae* (SCE-I), are site-specific nucleases found in bacteria, phages, and lower eukaryotes. An important characteristic of these nucleases is that their recognition sequence is relatively large (the SCE-I site is 18 nucleotides) and therefore the mammalian genome contains only a limited number of such recognition sites. When SCE-I is expressed in cells that have been engineered to carry a single SCE-I recognition site, it binds to this site and cuts the DNA generating a DSB (Rouet et al. 1994). Cells that are provided with a “donor plasmid” that has homology to the

locus flanking the cut site can use this donor as a template to repair the DSB via a specialized form of the homologous DNA repair pathway called synthesis-dependent strand annealing (SDSA).

In SDSA (outlined in Fig. 1) (for review, see Helleday et al. 2007; Adelman and Boulton 2010), the DSB is resected by nucleases. The resulting single-strand DNA is coated by Rad51 and invades the provided donor plasmid at a homologous stretch of DNA. DNA repair polymerases then extend the single-strand DNA using the homologous strand as a template. During this step, lost DNA sequences are resynthesized or, if the exogenously applied donor contains additional sequence between the homology arms, then these get incorporated. Once DNA synthesis has generated a complementary sequence to the single-strand DNA on the other side of the DSB, the newly synthesized strand is displaced. Finally, the two strands of the DSB reanneal and synthesize DNA to fill in the gap, and DNA ligation rejoins the two strands. If the donor sequence diverges from the endogenous gene either by the addition of longer stretches of DNA or by a single point mutation, the resulting gene conversion can be detected and used to measure the frequency of gene targeting. In human cells, the rate of spontaneous gene targeting using such a construct in the absence of a DSB is low and usually occurs in less than one in one million cells. However, after SCE-I induces a DSB, targeting efficiency at the same locus and with the same donor construct increases by several orders of magnitude (Bibikova et al. 2003; Porteus and Baltimore 2003). This can result in gene targeting efficiencies of up to one in 10 cells undergoing a gene conversion (Bibikova et al. 2003; Porteus and Baltimore 2003).

After the nuclease has generated a DSB at the desired locus, the resection of the DSB is limited and results in approximately 30–100 bases of single-strand DNA (Elliott et al. 1998; Helleday et al. 2007; Moehle et al. 2007; Larocque and Jasin 2010). It is this single-strand region that is coated with RAD51 and used for homology search (Helleday et al. 2007; Adelman and Boulton 2010). As a result, only short homology arms are needed to mediate efficient gene targeting. However, the limited resection of the DSB after the nuclease cut also poses limitations because it leads to relatively short gene conversion tracts, and genetic manipulations are restricted to the region proximal to the DSB. This illustrates a fundamental difference between SDSA in human cells after a DSB and HR in mESCs, which allows genetic manipulations over long stretches of DNA.

The SCE-I nuclease has been instrumental in elucidating the details of this process; however, due its fixed 18-bp recognition site, SCE-I has limited use in gene targeting. To exploit the phenomenon that a DSB can facilitate the manipulation of genes within its proximity, site-specific nucleases called ZFNs were engineered (Bibikova et al. 2002; Porteus and Baltimore 2003). ZFNs are chimeric proteins composed of several Cys₂His₂ zinc fingers that form a zinc-finger DNA-binding domain, a structural motif found in a naturally occurring group of zinc-finger transcription factors (Kim and Chandrasegaran 1994; Kim et al. 1996, 1998). ZFNs are also comprised of a nuclease domain that is most commonly a

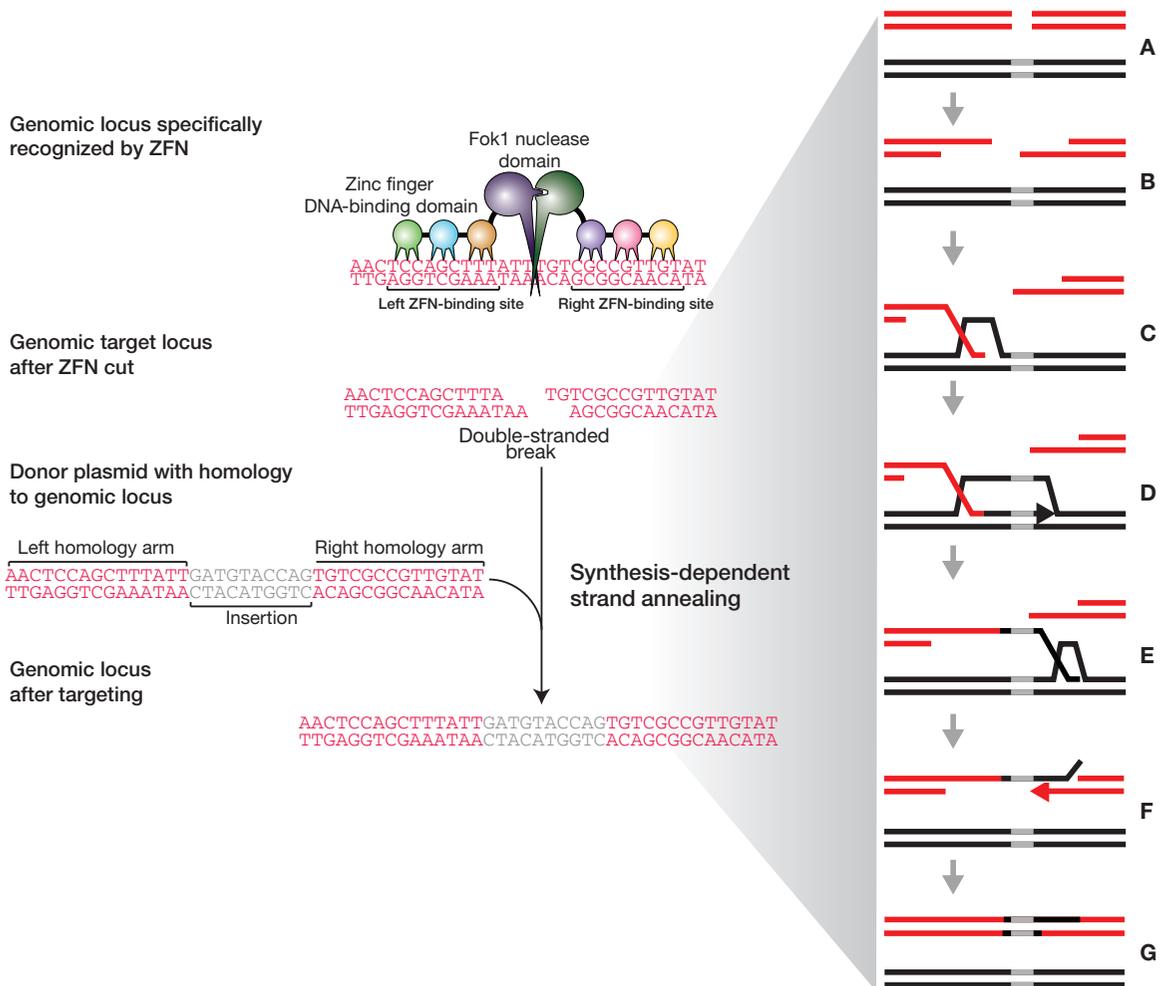


Figure 1. Overview of ZFN-mediated gene targeting. (*Left*) General concept of ZFN-mediated gene targeting. Two zinc-finger proteins are generated, which are designed to specifically recognize two adjacent DNA-binding sequences with a defined spacing. Following binding of the zinc fingers, their FOK1 nuclease domains dimerize, become active, and cut the DNA. If a donor plasmid carrying DNA homologous to the DSB is ectopically provided to the cell, it can be used to repair the DNA lesion. A donor plasmid can be designed so that it carries additional sequence between the homology arms. Following repair of the DSB with such a donor, the genomic locus will be altered to carry this additional sequence as an insertion at the site of the DSB. (*Right*) Underlying repair mechanism for this reaction is the SDSA (adapted from Helleday et al. 2007; Adelman and Boulton 2010). The sequence of events is as follows: After the ZFN generates a DSB (red) (*A*), the DNA is resected (*B*) to create single-stranded 3' overhangs on both sides of the DSB. The 3' single-strand DNA can strand-invade the donor plasmid (black) (*C*), thereby displacing one of the donor DNA strands and generating a D-loop structure. DNA synthesis extends the 3' end using the donor sequence as template (*D*). After this synthesis step, strand displacement and collapse of the D-loop structure leads to the release of the invading strand and restores the unchanged donor plasmid (*E*). The extended single strand can now anneal to the opposite side of the original break (*F*). DNA synthesis fills in the gaps and specific endonucleases prune the flap structures (*G*). Finally, DNA ligases rejoin the DNA strands (*G*). (Gray) Sequence divergence between the two homologous strands.

derivative of the FOK1 nuclease. Many aspects are considered in the design of a ZFN (Urnov et al. 2010) and different strategies have been developed (Greisman and Pabo 1997; Isalan et al. 2001; Hurt et al. 2003; Segal et al. 2003; Maeder et al. 2008; Meng et al. 2008; Foley et al. 2009; Sander et al. 2010). Briefly, each zinc finger in a zinc-finger DNA-binding domain specifically recognizes three base pairs while making only limited contact to the neighboring bases (Pavletich and Pabo 1993). This allows for the generation of combinatorial ZFN modules, a process called modular assembly (Segal et al. 2003), in which up to six zinc fingers with known binding activity are com-

bined so that they specifically recognize a stretch of ~18 consecutive nucleotides.

The other key component of a ZFN is the nuclease domain, which is designed so that its activity is dependent on the dimerization of two half-sites (Vanamee et al. 2001; Miller et al. 2007b; Szczepek et al. 2007). This restricted enzymatic activity acts as a safeguard to prevent nonspecific DNA cleavage. Because an individual zinc finger contains only one FOK1 half-site, two ZFNs must bind with a defined spacing in order for the nuclease domain to become active. This increases the specificity of the genomic site that is recognized by a ZFN pair to up to 36 nu-

cleotides in length. After the assembly of a specific ZFN pair, its activity and cytotoxicity are generally tested by appropriate assays (Miller et al. 2007a) before it is used in gene targeting experiments.

The ZFN pair and donor plasmid containing homology to the locus of interest are introduced into the cell using standard transfection methods suitable for the individual cell type. The donor plasmid is designed to carry ~500–750 bp of homologous sequence, flanking both sides of the ZFN recognition site. If desired, a single base-pair change can be introduced around the zinc-finger recognition site in order to generate a point mutation. This could be useful to introduce or repair a disease-specific mutation in hiPSCs. Alternatively, large cassettes carrying multiple transgenes and promoters can be placed between the homology arms. Such transgene integration is a required step in the generation of reporter cell lines, where a reporter such as GFP is inserted into the genome at a specific location.

It is worth noting that ZFNs were successfully used to genetically modify model organisms that could not easily be genetically targeted with conventional techniques. Gene correction has been achieved in *Drosophila* (Bibikova et al. 2003; Beumer et al. 2006, 2008), *Ara-bidopsis* (Zhang et al. 2010), tobacco plants (Townsend et al. 2009), and mice (Goldberg et al. 2010). Precise gene insertion at the site of the ZFN cut was performed in corn (Shukla et al. 2009) and tobacco plants (Cai et al. 2009). Furthermore, ZFN-mediated gene disruption has been successful in zebrafish (Doyon et al. 2008; Meng et al. 2008; Foley et al. 2009), rats (Geurts et al. 2009; Mashimo et al. 2010), and *Drosophila* (Bibikova et al. 2002). In contrast to ZFN-mediated gene addition, which relies on HR with an exogenous donor, ZFN-mediated gene disruption results from DNA repair by the NHEJ pathway when no donor is supplied. As mentioned above, the NHEJ pathway is imprecise and rejoins chromosome ends without the need for homology, regenerating lost sequence information, and/or the removal of aberrantly added nucleotides. Therefore, sporadic gene disruption can occur when a ZFN cut is made in the absence of a donor plasmid, leading to loss-of-function alleles. If present in germ cells, this approach can be used to generate knockout animals (Geurts et al. 2009; Mashimo et al. 2010).

ZFN-MEDIATED TARGETING IN HUMAN SOMATIC CELLS

Several genes have been targeted in human somatic cells using ZFNs (Table 1). In a seminal paper, Urnov et al. (2005) described correction of a disease-relevant mutation in the *IL2R γ* gene in human cells using ZFNs (Urnov et al. 2005). Mutations in this gene are responsible for X-linked severe combined immune deficiency (SCID). In a set of elegant experiments, these authors used ZFNs to introduce single out-of-frame point mutations into exon 5 of *IL2R γ* . In a second targeting step, ZFNs were used to correct these point mutations and restore the protein to wild-type levels (Urnov et al. 2005).

In addition, *IL2R γ* and other loci have been used to demonstrate that ZFNs can mediate transgene integration

in a large number of human cell types, including primary cells (Lombardo et al. 2007; Moehle et al. 2007). This approach allows the standardized generation of human cell lines that express transgenes and short hairpin RNAs from defined genetic loci. ZFN-mediated transgene insertion can thereby eliminate some of the unpredictable complications generating transgenic cell lines by random integration, such as position effects, differences in copy number, or differential epigenetic silencing. We and other investigators have also recently adopted ZFN technology to genetically modify hESCs and hiPSCs (Lombardo et al. 2007; Brunet et al. 2009; Hockemeyer et al. 2009; Zou et al. 2009).

TARGETING *OCT4* AND *PITX3* LOCI TO GENERATE REPORTER CELL LINES IN HUMAN PLURIPOTENT CELLS

As mentioned above, *OCT4* had been previously targeted by conventional HR in hESCs (Zwaka and Thomson 2003). Using isogenic vector arms, targeting efficiencies were 20% with 7.9-kb homologous sequences and 40% with 12.8-kb homologous sequences. In contrast, the ZFN targeting approach we used required only short homology arms of 500–750 bp and achieved gene targeting efficiencies of up to 94% correctly targeted clones (Hockemeyer et al. 2009). In these experiments, correct targeting resulted in a reporter system that allowed the identification of *OCT4*-positive pluripotent cells by either expression of GFP or the puromycin-resistance gene. This demonstrates the advantages of ZFN-mediated targeting over conventional targeting: ZFNs require only short targeting arms but nevertheless result in substantially higher targeting efficiency. This is relevant, because a major concern and uncertainty in conventional gene targeting is the extent to which small nucleotide polymorphisms (SNPs) in targeting constructs reduce gene targeting efficiency. Examples of successful conventional targeting experiments did not explicitly use isogenic DNA to generate their targeting constructs, but a comprehensive analysis in hESCs of the impact of SNPs in targeting constructs is lacking. In addition, the optimal length of homology arms must be considered in the design of the targeting strategy. Targeting arms such as those used to conventionally target hESCs were relatively long, and this length might have been crucial to targeting success. However, to generate such long arms isogenically and isochromosomally, in order to avoid SNPs, is relatively difficult and time-consuming. ZFN-mediated targeting requires only very short targeting arms and thereby largely eliminates these concerns.

In addition to modifying the *OCT4* locus, we also used ZFNs to target *PITX3* (Hockemeyer et al. 2009), a gene encoding a transcription factor that is expressed in differentiated cells, such as dopaminergic neurons, but not in hESCs. This targeting strategy was designed to generate a GFP reporter for the transcriptional state of *PITX3* by fusing the first exon of *PITX3* with the open reading frame (ORF) of GFP. Targeting of *PITX3* is significant because it illustrates that ZFNs can be used to target genes that are not expressed in hESCs.

Table 1. Gene Targeting Experiments Using ZFN in Human Cells

Cell type	Genetic modification	Gene	Clonal targeting efficiency	Reference				
Tumor cells	Gene disruption	<i>TP73</i>		Doyon et al. (2010)				
		<i>MAP3K14</i>						
		<i>EP300</i>						
		<i>BTK</i>						
		<i>TSC2</i>						
		<i>KDR</i>						
		<i>NR3C1</i>						
		<i>HOXB13</i>						
		<i>CXCR4</i>						
		<i>CFTR</i>						
Tumor cells	Gene addition	<i>PPP1R12C (AAVSI)</i>		Maeder et al. (2008)				
		<i>VEGF</i>						
		<i>CCR5</i>						
		<i>CXCR4</i>						
		<i>IL2RG</i>						
		Tumor cells	Gene correction		<i>IL2RG</i>		Urnov et al. (2005)	
					Gene disruption			
		Primary human cells	Gene disruption, correction		<i>CCR5</i>		Urnov et al. (2005)	
					<i>IL2RG</i>			
Gene addition	<i>PPP1R12C (AAVSI)</i>							
hESC/iPSCs	Gene disruption	<i>CCR5</i>		Lombardo et al. (2007)				
		<i>IL2RG</i>						
		<i>CCR5</i>						
		<i>PIGA</i>						
		<i>PIGA</i>						
	Gene addition	<i>IL2RG</i>		Lombardo et al. (2007)				
		<i>CCR5</i>						
		<i>PIGA</i>						
		<i>OCT4</i>	up to ~90%					
		<i>PITX3</i>	8%–10%					
	<i>PPP1R12C (AAVSI)</i>	~50%	Brunet et al. (2009) Hockemeyer et al. (2009) Dekelver et al. (2010)					

Adapted from Urnov et al. 2010.

ESTABLISHING *AAVSI* AS A SAFE-HARBOR LOCUS IN hESCs

AAVSI on chromosome 19 is a well-characterized locus that has been used to stably express transgenes in multiple transformed and primary cell lines without transgene silencing (Smith et al. 2008; Dekelver et al. 2010). *AAVSI* is the viral integration site for AAVs, which, following integration, disrupts the gene encoding the regulatory subunit 12C of protein phosphatase 1 (*PPP1R12C*). hESCs targeted in this locus using AAV-mediated gene integration exhibit long-term transgene expression and maintain a pluripotent state (Smith et al. 2008). To establish a robust overexpression system suitable for hESC cultures, we used a ZFN pair to target the first intron of *PPP1R12C*. Following the targeting of this locus with a constitutive expression cassette containing enhanced GFP, we observed long-term expression in hESCs, validating the use of ZFNs for modifying this locus. As mentioned above, an alternative approach for establishing controlled transgene

expression is the isogenic targeting of the *ROSA26* locus. The targeting frequency of the *ROSA26* locus was relatively low, with only 2.3% of the drug-resistant clones being correctly targeted (Irion et al. 2007), whereas targeting efficiencies of up to 50% were achieved by the ZFN-mediated approach (Hockemeyer et al. 2009).

In addition to expressing enhanced GFP constitutively from the *AAVSI* locus, we have also introduced a doxycycline (DOX)-inducible expression cassette into the *AAVSI* locus using ZFNs (Hockemeyer et al. 2009). Using this system, we found that a transgene can be expressed in a quantitative and temporally controlled manner in response to different DOX concentration in the medium. This inducible system requires a DOX-responsive promoter that drives transgene expression and expression of the reverse tetracycline-controlled transactivator (rtTA). Using simultaneous targeting of both alleles of the *AAVSI* locus, we have established a one-step targeting strategy in which rtTA is inserted into one allele and the DOX-responsive promoter, and the transgene of interest is inserted into the

other allele. Moreover, we have recently found that the *AAVSI* locus can also be used to express short hairpin RNAs to confer stable, long-term gene knockdown in hESCs (Dekelver et al. 2010).

OFF-TARGET SITE ALTERATIONS

ZFN technology has the potential to be a robust and reliable tool for gene targeting in human pluripotent cells, but only a few genes have been targeted using ZFNs thus far. Nevertheless, these examples indicate that current strategies for zinc-finger design (Pabo et al. 2001; Klug 2005) might allow the generation of selective and active ZFNs for most genes in the human genome. However, the introduction of active nucleases into human cells necessitates some caution and consideration regarding potential off-target and genotoxic effects of ZFNs. For basic research, several tools are at hand to ensure specificity and quality of ZFNs. Southern blotting using internal and external probes should be used to exclude ectopic integrations and confirm clonality of targeted clones, although such analyses do not detect ZFN-mediated DSBs and subsequent error-prone repair elsewhere in the genome.

To address these concerns, we investigated the frequency of off-target DNA cleavage and subsequent NHEJ-mediated repair for all currently used ZFNs (Hockemeyer et al. 2009). We first analyzed the DNA-binding specificity for all ZFNs with SELEX (systematic evolution of ligands by exponential enrichment) (Perez et al. 2008; Phillips et al. 2009), which allowed identification of the most likely off-target cleavage sites on a genome-wide scale. Using the Cel-1 assay (Miller et al. 2007a), we then quantified the actual frequency of NHEJ-mediated alterations in identified potential off-target sites. We analyzed 46 genomic loci and detected only one NHEJ alteration at one genomic site in a *PITX3*-targeted clone. All other putative off-target sites for all of the ZFNs tested were wild type. In addition, we determined the frequency of NHEJ alterations in heterozygous-targeted clones on the allele, which did not carry an integrated transgene. This frequency varied from 1 in 18 to 1 in 12 for *PITX3* and *OCT4* loci, respectively, whereas in all other cases, the other allele remained unaltered. To further ensure a functional wild-type allele in heterozygous-targeted clones, the recognition sequence of the ZFNs can be designed to recognize intron sequences, as was done for the *OCT4* and *AAVSI* loci. In summary, the use of ZFN can in rare cases lead to off-target effects, but these can usually be detected and minimized. The use of multiple, independent ZFN-mediated manipulations to confirm cellular phenotypes should also alleviate these concerns.

CONCLUDING REMARKS

The combination of patient-specific hiPSCs together with ZFN technology for performing genetic modifications of the human genome will allow the generation of specific cell types carrying disease-relevant alterations. This would be a significant step toward the generation of better in vitro models for human disease and the develop-

ment of potential therapeutics. Thus far, only a handful genes in human pluripotent cells have been targeted using ZFNs. Although targeting efficiency has been very high in the genes targeted so far, it is not clear what fraction of the human genome can be engineered in this way and how the chromatin conformation of a given gene may influence targeting efficiency. It will be important to resolve these issues in future work to assess the general utility of this technology for genetically manipulating the genome of human pluripotent cells.

ACKNOWLEDGMENTS

We thank Helen Bateup for critical comments on this manuscript. D.H. is a Merck Fellow of the Life Science Research Foundation. R.J. was supported by National Institutes of Science grants R37-CA084198, RO1-CA087869, and RO1-HD045022. R.J. is an adviser to Stemgen and a cofounder of Fate Therapeutics.

REFERENCES

- Adelman CA, Boulton SJ. 2010. Metabolism of postsynaptic recombination intermediates. *FEBS Lett* **584**: 3709–3716.
- Banuelos CA, Banath JP, MacPhail SH, Zhao J, Eaves CA, O'Connor MD, Lansdorp PM, Olive PL. 2008. Mouse but not human embryonic stem cells are deficient in rejoining of ionizing radiation-induced DNA double-strand breaks. *DNA Repair* **7**: 1471–1483.
- Beumer K, Bhattacharyya G, Bibikova M, Trautman JK, Carroll D. 2006. Efficient gene targeting in *Drosophila* with zinc-finger nucleases. *Genetics* **172**: 2391–2403.
- Beumer KJ, Trautman JK, Bozas A, Liu JL, Rutter J, Gall JG, Carroll D. 2008. Efficient gene targeting in *Drosophila* by direct embryo injection with zinc-finger nucleases. *Proc Natl Acad Sci* **105**: 19821–19826.
- Bibikova M, Golic M, Golic K, Carroll D. 2002. Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics* **161**: 1169–1175.
- Bibikova M, Beumer K, Trautman JK, Carroll D. 2003. Enhancing gene targeting with designed zinc finger nucleases. *Science* **300**: 764.
- Brenneman M, Gimble FS, Wilson JH. 1996. Stimulation of intrachromosomal homologous recombination in human cells by electroporation with site-specific endonucleases. *Proc Natl Acad Sci* **93**: 3608–3612.
- Brunet E, Simsek D, Tomishima M, DeKelver R, Choi VM, Gregory P, Urnov F, Weinstock DM, Jasin M. 2009. Chromosomal translocations induced at specified loci in human stem cells. *Proc Natl Acad Sci* **106**: 10620–10625.
- Cai C, Doyon Y, Ainley W, Miller J, Dekelver R, Moehle E, Rock J, Lee Y, Garrison R, Schulenberg L, et al. 2009. Targeted transgene integration in plant cells using designed zinc finger nucleases. *Plant Mol Biol* **69**: 699–709.
- Capecchi MR. 1989. Altering the genome by homologous recombination. *Science* **244**: 1288–1292.
- Costa M, Dottori M, Sourris K, Jamshidi P, Hatzistavrou T, Davis R, Azzola L, Jackson S, Lim S, Pera M, et al. 2007. A method for genetic modification of human embryonic stem cells using electroporation. *Nat Protoc* **2**: 792–796.
- Davis R, Costa M, Grandela C, Holland A, Hatzistavrou T, Micallef S, Li X, Goulburn A, Azzola L, Elefanta A, et al. 2008a. A protocol for removal of antibiotic resistance cassettes from human embryonic stem cells genetically modified by homologous recombination or transgenesis. *Nat Protoc* **3**: 1550–1558.
- Davis R, Ng E, Costa M, Mossman A, Sourris K, Elefanta A, Stanley E. 2008b. Targeting a GFP reporter gene to the *MIXL1* locus of human embryonic stem cells identifies human primitive

- streak-like cells and enables isolation of primitive hematopoietic precursors. *Blood* **111**: 1876–1884.
- Dekelver RC, Choi VM, Moehle EA, Paschon DE, Hockemeyer D, Meijnsing SH, Sancak Y, Cui X, Steine EJ, Miller JC, et al. 2010. Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. *Genome Res* **20**: 1133–1142.
- Doyon Y, McCammon J, Miller J, Faraji F, Ngo C, Katibah G, Amora R, Hocking T, Zhang L, Rebar E, et al. 2008. Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat Biotechnol* **26**: 702–708.
- Doyon Y, Choi VM, Xia DF, Vo TD, Gregory PD, Holmes MC. 2010. Transient cold shock enhances zinc-finger nuclease-mediated gene disruption. *Nat Methods* **7**: 459–460.
- Elliott B, Richardson C, Winderbaum J, Nickoloff JA, Jasin M. 1998. Gene conversion tracts from double-strand break repair in mammalian cells. *Mol Cell Biol* **18**: 93–101.
- Foley JE, Yeh J-RJ, Maeder ML, Reyon D, Sander JD, Peterson RT, Joung JK. 2009. Rapid mutation of endogenous zebrafish genes using zinc finger nucleases made by Oligomerized Pool Engineering (OPEN). *PLoS ONE* **4**: e4348.
- Geurts A, Cost G, Freyvert Y, Zeitler B, Miller J, Choi V, Jenkins S, Wood A, Cui X, Meng X, et al. 2009. Knockout rats via embryo microinjection of zinc-finger nucleases. *Science* **325**: 433.
- Goldberg AD, Banaszynski LA, Noh K-M, Lewis PW, Elsaesser SJ, Stadler S, Dewell S, Law M, Guo X, Li X, et al. 2010. Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* **140**: 678–691.
- Greisman HA, Pabo CO. 1997. A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* **275**: 657–661.
- Helleday T, Lo J, van Gent DC, Engelward BP. 2007. DNA double-strand break repair: From mechanistic understanding to cancer treatment. *DNA Repair* **6**: 923–935.
- Hockemeyer D, Soldner J, Beard C, Gao Q, Mitalipova M, Dekelver RC, Katibah GE, Amora R, Boydston EA, Zeitler B, et al. 2009. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol* **27**: 851–857.
- Hurt JA, Thibodeau SA, Hirsh AS, Pabo CO, Joung JK. 2003. Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc Natl Acad Sci* **100**: 12271–12276.
- Irion S, Luche H, Gadue P, Fehling H, Kennedy M, Keller G. 2007. Identification and targeting of the *ROSA26* locus in human embryonic stem cells. *Nat Biotechnol* **25**: 1477–1482.
- Isalan M, Klug A, Choo Y. 2001. A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nat Biotechnol* **19**: 656–660.
- Jasin M. 1996. Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet* **12**: 224–228.
- Khan IF, Hirata RK, Wang P-R, Li Y, Kho J, Nelson A, Huo Y, Zavaljevski M, Ware C, Russell DW. 2010. Engineering of human pluripotent stem cells by AAV-mediated gene targeting. *Mol Ther* **18**: 1192–1199.
- Kim YG, Chandrasegaran S. 1994. Chimeric restriction endonuclease. *Proc Natl Acad Sci* **91**: 883–887.
- Kim YG, Cha J, Chandrasegaran S. 1996. Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci* **93**: 1156–1160.
- Kim YG, Smith J, Durgesha M, Chandrasegaran S. 1998. Chimeric restriction enzyme: Gal4 fusion to Fok I cleavage domain. *Biol Chem* **379**: 489–495.
- Klug A. 2005. The discovery of zinc fingers and their development for practical applications in gene regulation. *Proc Jpn Acad Ser B* **81**: 87–102.
- Koller BH, Smithies O. 1992. Altering genes in animals by gene targeting. *Annu Rev Immunol* **10**: 705–730.
- Larocque JR, Jasin M. 2010. Mechanisms of recombination between diverged sequences in wild-type and BLM-deficient mouse and human cells. *Mol Cell Biol* **30**: 1887–1897.
- Lieber MR. 2010. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem* **79**: 181–211.
- Lombardo A, Genovese P, Beausejour C, Colleoni S, Lee Y, Kim K, Ando D, Urnov F, Galli C, Gregory P, et al. 2007. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol* **25**: 1298–1306.
- Maeder ML, Thibodeau-Beganny S, Osiaik A, Wright DA, Anthony RM, Eichinger M, Jiang T, Foley JE, Winfrey RJ, Townsend JA, et al. 2008. Rapid “open-source” engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell* **31**: 294–301.
- Mashimo T, Takizawa A, Voigt B, Yoshimi K, Hiai H, Kuramoto T, Serikawa T. 2010. Generation of knockout rats with X-linked severe combined immunodeficiency (X-SCID) using zinc-finger nucleases. *PLoS ONE* **5**: e8870.
- Meng X, Noyes M, Zhu L, Lawson N, Wolfe S. 2008. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol* **26**: 695–701.
- Miller J, Holmes M, Wang J, Guschin D, Lee Y, Rupniewski I, Beausejour C, Waite A, Wang N, Kim K, et al. 2007a. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol* **25**: 778–785.
- Miller JC, Holmes MC, Wang J, Guschin DY, Lee YL, Rupniewski I, Beausejour CM, Waite AJ, Wang NS, Kim KA, et al. 2007b. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol* **25**: 778–785.
- Moehle E, Rock J, Lee Y, Jouvenot Y, DeKelver R, Gregory P, Urnov F, Holmes M. 2007. Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc Natl Acad Sci* **104**: 3055–3060.
- Pabo C, Peisach E, Grant R. 2001. Design and selection of novel Cys₂His₂ zinc finger proteins. *Annu Rev Biochem* **70**: 313–340.
- Pavletich NP, Pabo CO. 1993. Crystal structure of a five-finger GLI-DNA complex: New perspectives on zinc fingers. *Science* **261**: 1701–1707.
- Perez E, Wang J, Miller J, Jouvenot Y, Kim K, Liu O, Wang N, Lee G, Bartsevich V, Lee Y, et al. 2008. Establishment of HIV-1 resistance in CD4⁺ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol* **26**: 808–816.
- Phillips CM, Meng X, Zhang L, Chretien JH, Urnov FD, Dernburg AF. 2009. Identification of chromosome sequence motifs that mediate meiotic pairing and synapsis in *C. elegans*. *Nat Cell Biol* **11**: 934–942.
- Porteus MH, Baltimore D. 2003. Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**: 763.
- Rouet P, Smith F, Jasin M. 1994. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol* **14**: 8096–8106.
- Ruby K, Zheng B. 2009. Gene targeting in a HUES line of human embryonic stem cells via electroporation. *Stem Cells* **27**: 1496–1506.
- Russell DW, Hirata RK. 1998. Human gene targeting by viral vectors. *Nat Genet* **18**: 325–330.
- Sander JD, Maeder ML, Reyon D, Voytas DF, Joung JK, Dobbs D. 2010. ZiFiT (zinc finger targeter): An updated zinc finger engineering tool. *Nucleic Acids Res* (suppl.) **38**: W462–W468.
- Segal DJ, Beerli RR, Blancafort P, Dreier B, Effertz K, Huber A, Koksche B, Lund CV, Magnenat L, Valente D, et al. 2003. Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry* **42**: 2137–2148.
- Shukla V, Doyon Y, Miller J, DeKelver R, Moehle E, Worden S, Mitchell J, Arnold N, Gopalan S, Meng X, et al. 2009. Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature* **459**: 437–441.
- Smith J, Maguire S, Davis L, Alexander M, Yang F, Chandran S, French-Constant C, Pedersen R. 2008. Robust, persistent transgene expression in human embryonic stem cells is achieved with AAVS1-targeted integration. *Stem Cells* **26**: 496–504.
- Song H, Chung S-K, Xu Y. 2010. Modeling disease in human ESCs using an efficient BAC-based homologous recombination system. *Cell Stem Cell* **6**: 80–89.

- Suzuki K, Mitsui K, Aizawa E, Hasegawa K, Kawase E, Yamagishi T, Shimizu Y, Suemori H, Nakatsuji N, Mitani K. 2008. Highly efficient transient gene expression and gene targeting in primate embryonic stem cells with helper-dependent adenoviral vectors. *Proc Natl Acad Sci* **105**: 13781–13786.
- Szczepek M, Brondani V, Buchel J, Serrano L, Segal DJ, Cathomen T. 2007. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat Biotechnol* **25**: 786–793.
- Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**: 663–676.
- Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**: 861–872.
- Thomson J, Itskovitz-Eldor J, Shapiro S, Waknitz M, Swiergiel J, Marshall V, Jones J. 1998. Embryonic stem cell lines derived from human blastocysts. *Science* **282**: 1145–1147.
- Townsend J, Wright D, Winfrey R, Fu F, Maeder M, Joung J, Voytas D. 2009. High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature* **459**: 442–445.
- Urbach A, Schuldiner M, Benvenisty N. 2004. Modeling for Lesch-Nyhan disease by gene targeting in human embryonic stem cells. *Stem Cells* **22**: 635–641.
- Urnov F, Miller J, Lee Y, Beausejour C, Rock J, Augustus S, Jamieson A, Porteus M, Gregory P, Holmes M. 2005. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**: 646–651.
- Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. 2010. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* **11**: 636–646.
- Valerie K, Povirk L. 2003. Regulation and mechanisms of mammalian double-strand break repair. *Oncogene* **22**: 5792–5812.
- Vanamee ES, Santagata S, Aggarwal AK. 2001. FokI requires two specific DNA sites for cleavage. *J Mol Biol* **309**: 69–78.
- Xue H, Wu S, Papadeas S, Spusta S, Swistowska A, Macarthur C, Mattson M, Maragakis N, Capecchi M, Rao M, et al. 2009. A targeted neuroglial reporter line generated by homologous recombination in human embryonic stem cells. *Stem Cells* **27**: 1836–1846.
- Yu J, Vodyanik M, Smuga-Otto K, Antosiewicz-Bourget J, Frane J, Tian S, Nie J, Jonsdottir G, Ruotti V, Stewart R, et al. 2007. Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**: 1917–1920.
- Zambrowicz BP, Imamoto A, Fiering S, Herzenberg LA, Kerr WG, Soriano P. 1997. Disruption of overlapping transcripts in the ROSA β geo 26 gene trap strain leads to widespread expression of β -galactosidase in mouse embryos and hematopoietic cells. *Proc Natl Acad Sci* **94**: 3789–3794.
- Zhang F, Maeder ML, Unger-Wallace E, Hoshaw JP, Reyon D, Christian M, Li X, Pierick CJ, Dobbs D, Peterson T, et al. 2010. High frequency targeted mutagenesis in *Arabidopsis thaliana* using zinc finger nucleases. *Proc Natl Acad Sci* **107**: 12028–12033.
- Zou J, Maeder M, Mali P, Pruett-Miller S, Thibodeau-Beganny S, Chou B, Chen G, Ye Z, Park I, Daley G, et al. 2009. Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell* **5**: 97–110.
- Zwaka T, Thomson J. 2003. Homologous recombination in human embryonic stem cells. *Nat Biotechnol* **21**: 319–321.