Genetic engineering of human pluripotent cells using TALE nucleases

Dirk Hockemeyer[1,4], Haoyi Wang[1,4], Samira Kiani[1], Christine S Lai[1,2], Qing Gao[1], John P Cassady[1,2], Gregory J Cost[3], Lei Zhang[3], Yolanda Santiago[3], Jeffrey C Miller[3], Bryan Zeitler[3], Jennifer M Cherone[3], Xiangdong Meng[3], Sarah J Hinkley[3], Edward J Rebar[3], Philip D Gregory[3], Fyodor D Urnov[3] & Rudolf Jaenisch[1,2]

© 2011 Nature America, Inc. All rights reserved.

Targeted genetic engineering of human pluripotent cells is a prerequisite for exploiting their full potential. Such genetic manipulations can be achieved using site-specific nucleases. Here we engineered transcription activator–like effector nucleases (TALENs) for five distinct genomic loci. At all loci tested we obtained human embryonic stem cell (ESC) and induced pluripotent stem cell (iPSC) clones carrying transgenic cassettes solely at the TALEN-specified location. Our data suggest that TALENs employing the specific architectures described here mediate site-specific genome modification in human pluripotent cells with similar efficiency and precision as do zinc-finger nucleases (ZFNs).

Gene targeting of human pluripotent cells by homologous recombination is inefficient, which has impeded the use of human ESCs and iPSCs in disease models. To overcome this limitation, we and others have shown that ZFNs can be used to modify the genomes of ESCs and iPSCs[1–3]. ZFNs can be engineered to induce a double-strand break precisely at a predetermined position in the genome[4]. The double-strand break can be repaired through nonhomologous end-joining to drive targeted gene disruption or through the homology-directed DNA repair pathway using an exogenous donor plasmid as a template. Depending on the donor design, this repair reaction can be used to generate large-scale deletions, gene disruptions, DNA addition[4] or single-nucleotide changes[5].

Recent work on transcription activator–like effectors (TALEs) provides an alternative approach to the design of site-specific nucleases[6]. Natural TALEs are transcription factors used by plant pathogens to subvert host genome regulatory networks[6]. The DNA-binding domain of TALEs is unusual. Multiple units of ~34 amino acids (called TALE repeats) are arranged in tandem, their sequence nearly identical except for two highly variable amino acids that establish the base-recognition specificity of each unit[7,8]. Each individual domain determines the specificity of binding to one DNA base pair in the TALE recognition sequence, and therefore arrays of four different repeat units are sufficient to generate TALEs with novel DNA recognition sites[7,8]. Nucleases based on such engineered TALE domains have been shown to target endogenous genes in transformed human cells[9,10]. Here we evaluate the use of TALENs for genetic engineering of endogenous loci in human ESCs and iPSCs.

We designed TALENs targeting the *PPP1R12C* (the AAVS1 locus), *OCT4* (also known as *POU5F1*) and *PITX3* genes at precisely the same positions as targeted earlier by ZFNs[2]. TALEN expression constructs and corresponding donor plasmids bearing homologous sequences were introduced into ESCs (line WIBR#3)[11] and iPSCs (line C1)[12] by electroporation (**Supplementary Fig. 1** and **Supplementary Tables 1** and **2**). Southern blot analysis was used to identify correctly targeted clones.

We targeted *PPP1R12C* with a gene trap approach (expressing puromycin (Puro) from the endogenous gene; **Fig. 1a**,**b** and **Table 1**) or with an autonomous selection cassette (puromycin expressed from the PGK-promoter (**Fig. 1**, **Table 1** and **Supplementary Figs. 1**–**3**). Targeting efficiency was high and similar to that with ZFNs[2]: 50% of the clones were targeted in one or both alleles and carried no randomly integrated transgenes (**Fig. 1b**, **Table 1** and **Supplementary Fig. 2**). Similarly, an SA-Puro-CAGGS-eGFP transgene was highly expressed from this locus (**Fig. 1a** and **Supplementary Fig. 3a**,**b**). Notably, such targeted cells remained pluripotent based on analysis of marker expression and of teratomas. (**Supplementary Fig. 3c**–**e**). Cells of all germ layers in teratomas expressed eGFP, indicating that TALEN- as well as ZFN-mediated targeting of *PPP1R12C* results in robust transgene expression in pluripotent as well as in differentiated cells (**Supplementary Fig. 3a**,**b**,**d**,**e**).

OCT4 was targeted using three different donor plasmids, resulting in expression of puromycin and an OCT4 exon1-eGFP fusion protein under control of the endogenous *OCT4* promoter. The first two donor plasmids were designed to integrate a splice acceptor-eGFP-2A-Puro cassette into the first intron of *OCT4*, whereas the third donor generated an in-frame fusion of exon 1 with the eGFP-2A-Puro cassette (**Supplementary Fig. 4**). Targeting efficiency in ESCs and iPSCs was 70–100% as determined by Southern blot analysis and DNA sequencing of single cell–derived clones (**Table 1** and **Supplementary Figs. 1** and **4**).

We also targeted the first exon of *PITX3*, which is not expressed in human pluripotent cells, and found that ~6% of drug-resistant clones carried the transgene solely at the *PITX3* locus as evaluated by Southern blot analysis (**Table 1** and **Supplementary Figs. 1** and **5**). Notably, in one clone the transgene had integrated into both alleles of *PITX3*.

To target the C terminus of OCT4, which has not been previously targeted by ZFNs, we generated TALENs directed against sequences flanking the stop codon of *OCT4* using two donor plasmids: the last *OCT4* codon was either fused in frame with an eGFP-PGK-Puro construct or with eGFP preceded by a 2A sequence (2A-eGFP-PGK-Puro; **Fig. 1c** and **Table 1**). After excision of the LoxP-flanked PGK-Puro cassettes, either a C-terminal OCT4-eGFP fusion protein or a separately translated eGFP protein will be expressed under the control of endogenous *OCT4* (ref. 13). Southern blot analysis showed that 2–10% of single cell–derived ESC and iPSC clones carried the transgene cassette at the *OCT4* locus (**Fig. 1d**,**e** and **Table 1**). Cre-mediated excision of the PGK-

[1]The Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, USA. [2]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.[3]Sangamo BioSciences, Inc., Richmond, California, USA. [4]These authors contributed equally to this work. Correspondence should be addressed to R.J. (jaenisch@wi.mit.edu).

Received 11 March; accepted 28 June; published online 7 July 2011; doi:10.1038/nbt.1927

© 2011 Nature America, Inc. All rights reserved.

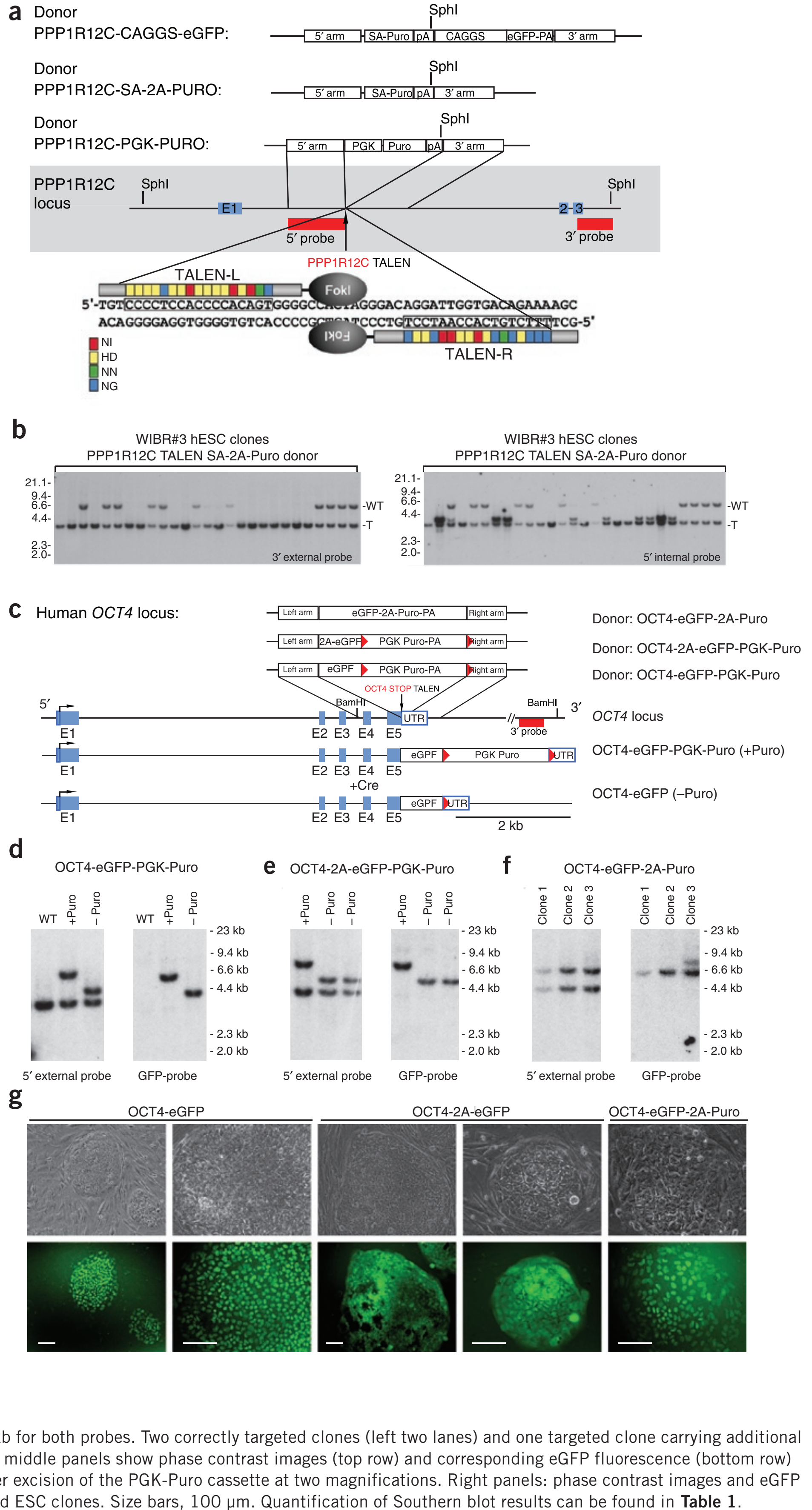

Figure 1 Genetic engineering of ESCs and iPSCs using TALENs. (**a**) Schematic overview depicting the targeting strategy for *PPP1R12C*. Southern blot probes are shown as red boxes, exons as blue boxes; the arrow indicates cut site by the TALENs. Donor plasmids: SA-2A-Puro, splice acceptor sequence followed by a 2A self-cleaving peptide sequence and the puromycin resistance gene; pA, polyadenylation sequence; PGK, phosphoglycerate kinase promoter; Puro, puromycin resistance gene; CAGGS, synthetic CAGGS promoter containing the actin enhancer and the cytomegalovirus early promoter; eGFP, enhanced green fluorescent protein. Below, scheme of PPP1R12C TALENs and their recognition sequence. TALE repeat domains are colored to indicate the identity of the repeat variable di-residue (RVD)[8]; each RVD is related to the cognate targeted DNA base by the following code (NI = A, HD = C, NN = G, NG = T)[7]. (**b**) Southern blot analysis of WIBR#3 ESCs targeted using PPP1R12C TALENs and the SA-2A-Puro donor plasmid. Genomic DNA was digested with SphI and hybridized with an ^{32}P-labeled external 3′-probe (left) or with an internal 5′-probe (right). 5′-probe detects a 6.5 kb WT and a 3.8 kb targeted fragment; 3′-probe a 6.5 kb WT and a 3.7 kb targeted fragment. WT, wild type; T, correctly targeted allele. (**c**) Schematic overview of the targeting strategy for the *OCT4* locus using the OCT4-STOP TALENs. Southern blot probes and exon of *OCT4* are colored as in **a** and the vertical arrow indicates the OCT4-STOP TALEN cut site. Shown above are the donor plasmid used to target the *OCT4* locus, loxP-sites are shown as red triangles. UTR, untranslated region of *OCT4*. (**d**) Southern blot analysis of the WIBR#3 ESCs targeted in the *OCT4* locus with the OCT4-eGFP-PGK-Puro donor plasmids. Genomic DNA was digested with BamHI and hybridized with the ^{32}P-labeled external 3′-probe or with the internal eGFP probe. Expected fragment size: WT = 4.2 kb, targeted = 6.8 kb for both probes. WT clone (left lane); clones before (middle lane) and after (right lane) excision of PGK-Puro cassette. (**e**) Southern blot analysis as in **d** of WIBR#3 ESCs targeted with 2A-eGFP-PGK-Puro donor plasmids. Expected fragment size as in **d**. A clone before (left lane) and two clones after (right lanes) Cre-mediated excision of PGK-Puro cassette. (**f**) Southern blot analysis as in **d** and **e** of WIBR#3 ESCs targeted with the eGFP-2A-Puro donor plasmids. Expected fragment size: WT = 4.2 kb, targeted = 5.6 kb for both probes. Two correctly targeted clones (left two lanes) and one targeted clone carrying additional aberrant integration (right lane). (**g**) Left and middle panels show phase contrast images (top row) and corresponding eGFP fluorescence (bottom row) of OCT4-eGFP and OCT4-2A-eGFP ESCs after excision of the PGK-Puro cassette at two magnifications. Right panels: phase contrast images and eGFP fluorescence of OCT4-eGFP-2A-Puro–targeted ESC clones. Size bars, 100 μm. Quantification of Southern blot results can be found in **Table 1**.

Table 1 Summary of targeting experiments using TALENs

OCT 4

Cell line targeted	TALEN pair	Donor	No. of clones analyzed	Random integration	Targeted + additional integration	Correctly targeted clones: Heterozygous	Homozygous	Targeting efficiency (%)
WIBR#3 ESC	OCT4 intron 1	OCT4-SA-eGFP#1	68	6	0	62	0	91
C1 iPSC	OCT4 intron 1	OCT4-SA-eGFP#1	7	0	0	7	0	100
WIBR#3 ESC	OCT4 intron 1	OCT4-SA-eGFP#3	35	0	0	34	0	97
C1 iPSC	OCT4 intron 1	OCT4-SA-eGFP#3	5	0	0	5	0	100
WIBR#3 ESC	OCT4 intron 1	OCT4-eGFP exon1 fusion	12	4	0	8	0	67
C1 iPSC	OCT4 intron 1	OCT4-eGFP exon1 fusion	1	0	0	1	0	100
WIBR#3 ESC	OCT4 STOP	OCT4-eGFP PGK-Puro (C-term fusion)	48	42	1	5	0	10
C1 iPSC	OCT4 STOP	OCT4-eGFP PGK-Puro (C-term fusion)	48	45	3	0	0	0
WIBR#3 ESC	OCT4 STOP	OCT4-2A-eGFP PGK-Puro (C-term fusion)	48	44	1	3	0	6
C1 iPSC	OCT4 STOP	OCT4-2A-eGFP PGK-Puro (C-term fusion)	48	46	1	1	0	2
WIBR#3 ESC	OCT4 STOP	OCT4-eGFP- 2A-Puro (genetrap)	48	0	26	22	0	46

PPP1R12C

Cell line targeted	TALEN pair	Donor	No. of clones analyzed	Random integration	Targeted + additional integration	Correctly targeted clones: Heterozygous	Homozygous	Targeting efficiency (%)
WIBR#3 ESC	PPP1R12C	PPP1R12C-SA-2A-Puro	96	4	35	34	21	57
WIBR#3 ESC	PPP1R12C	PPP1R12C-PGK-Puro	107	19	23	52	13	61
WIBR#3 ESC	PPP1R12C	PPP1R12C-CAGGS-eGFP	32	1	16	12	3	47
C1 iPSC	PPP1R12C	PPP1R12C-SA-2A-Puro	24	1	7	8	8	66
C1 iPSC	PPP1R12C	PPP1R12C-PGK-Puro	55	4	24	19	8	49
C1 iPSC	PPP1R12C	PPP1R12C-CAGGS-eGFP	23	1	12	5	5	43
WIBR#3 ESC	PPP1R12C eHiFi	PPP1R12C-SA-2A-Puro	22	4	5	14	3	77
C1 iPSC	PPP1R12C eHiFi	PPP1R12C-SA-2A-Puro	20	0	7	9	4	65

PITX3

Cell line targeted	TALEN pair	Donor	No. of clones analyzed	Random integration	Targeted + additional integration	Correctly targeted clones: Heterozygous	Homozygous	Targeting efficiency (%)
WIBR#3 ESC	PITX3 exon 1	PITX3-eGFP FW	96	62	22	11	1	13
C1 iPSC	PITX3 exon 1	PITX3-eGFP FW	96	88	4	4	0	4
C1 iPSC	PITX3 exon 1	PITX3-eGFP BW	96	84	11	1	0	1
WIBR#3 ESC	PITX3 STOP	PITX3-eGFP C-term fusion	48	20	16	11	1	23
C1 iPSC	PITX3 STOP	PITX3-eGFP C-term fusion	48	32	7	9	0	19
WIBR#3 ESC	PITX3 STOP	PITX3-eGFP C-term 2A fusion	96	32	44	19	1	21
C1 iPSC	PITX3 STOP	PITX3-eGFP C-term 2A fusion	48	24	13	9	2	23

The gene targeting efficiencies for *OCT4*, *PPP1R12C* and *PITX3* (last column) in either ESCs or iPSCs (first column) using the indicated TALEN pairs and donor plasmids and the number of clones analyzed, the number of clones found to carry additional unwanted integrations and the number of either homozygous or heterozygous correctly targeted clones. eHiFi indicates TALENs designed to function as obligatory heterodimers.

Puro cassette from OCT4-eGFP–targeted ESCs resulted in nuclear eGFP fluorescence, whereas clones targeted with OCT4-2A-eGFP showed pan-cellular eGFP fluorescence (**Fig. 1g**). Fluorescence was higher in OCT4-2A-eGFP–targeted clones than in OCT4-eGFP–targeted clones as determined by fluorescence-activated cell sorting (FACS) analysis, suggesting different protein stabilities of the OCT4-eGFP fusion protein and eGFP (**Supplementary Fig. 6**). A gene trap vector fusing an eGFP-2A-Puro cassette with the last *OCT4* codon had a 50% targeting efficiency (**Fig. 1c**,**f**,**g**, **Table 1** and **Supplementary Fig. 6**) similar to that of gene-trap vectors designed to target the first intron. As expected, eGFP expression became undetectable after differentiation into fibroblast-like cells (**Supplementary Fig. 7**), validating eGFP expression as a faithful reporter of OCT4 expression.

To illustrate the general utility of TALENs for generating such C-terminal fusion proteins, we designed TALENs that cut at the last coding exon of *PITX3* and generated ESCs and iPSCs with an in-frame fusion of 2A-eGFP or eGFP. Targeting was highly efficient, resulting in some clones carrying the transgene on both alleles (**Supplementary Fig. 8** and **Table 1**).

© 2011 Nature America, Inc. All rights reserved. npg

To assess the frequency of off-target modification not detected by Southern blot analysis, we determined the binding specificity of the *PPP1R12C* TALENs using systematic evolution of ligands by exponential enrichment. Genotyping of a panel of 19 of the most likely potential off-target sites revealed unintended cleavage at a low frequency (**Supplementary Figs. 9** and **10**, **Supplementary Tables 3** and **4** and **Supplementary Notes** for detailed description). A strategy to minimize potential off-target events is to design TALENs to function as obligatory heterodimers[14]. Such heterodimeric nucleases, fused to the TALE DNA binding domain, enabled high-efficiency targeting of the *PPP1R12C* locus (**Table 1**).

Individual TALE repeats can be joined to produce DNA binding domains capable of recognizing endogenous sequences in mammalian cells[9,10,15]. In the present work, we have built on the recent development of an efficient TALEN architecture[9] and evaluated the utility of TALENs to drive targeted gene modifications in human ESCs and iPSCs. At all five genomic sites tested, we obtained clones carrying transgenes solely at the TALEN-specified locus at a frequency between 67% and 100% for the first exon of *OCT4*, between 2% and 46% for the targeting of the *OCT4* STOP codon, of about 50% for the *AAVS1* locus, between 1% and 13% for the targeting of the ATG of *PITX3* and between 19% and 23% for the *PITX3* STOP codon. These efficiencies are similar to those observed with ZFNs[2]. As this approach couples a simple DNA recognition code with robust activity in human pluripotent stem cells, our data suggest that TALENs are a useful tool for investigator-specified targeting and genetic modification in human pluripotent cells.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

We thank R. Alagappan, P. Xu, S. Cristea, S. Lam and A. Vincent for expert technical assistance. We thank F. Soldner for helpful discussions on the manuscript. D.H. is a Merck Fellow of the Life Sciences Research Foundation. R.J. was supported by US National Institutes of Health grants R37-CA084198, RO1-CA087869 and RO1-HD045022 and by a grant from the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

D.H., H.W. and R.J. designed the targeting experiments and wrote the manuscript. H.W. and D.H. generated donor plasmids. D.H. preformed targeting experiments. S.K., C.S.L. and H.W. assisted with Southern blot analysis. Q.G. analyzed teratomas. J.P.C. and D.H. performed FACS analysis of targeted cells. L.Z. and J.C.M. designed the TALENs, S.J.H. assembled the TALENs, G.J.C. and Y.S. tested the TALENs, and B.Z., J.M.C. and X.M. performed the off-target analysis. D.H., R.J., L.Z., G.J.C., J.C.M., B.Z., X.M. and F.D.U. analyzed the data. E.J.R., P.D.G. and F.D.U. designed and supervised the design of the TALENs and contributed to writing the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/nbt/index.html.

Published online at http://www.nature.com/nbt/index.html.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Zou, J. *et al. Cell Stem Cell* **5**, 97–110 (2009).
2. Hockemeyer, D. *et al. Nat. Biotechnol.* **27**, 851–857 (2009).
3. Hockemeyer, D. & Jaenisch, R. *Cold Spring Harb. Symp. Quant. Biol.* **75**, 201–209 (2010).
4. Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. & Gregory, P.D. *Nat. Rev. Genet.* **11**, 636–646 (2010).
5. Soldner, F. *et al. Cell* (in the press).
6. Boch, J. & Bonas, U. *Annu. Rev. Phytopathol.* **48**, 419–436 (2010).
7. Boch, J. *et al. Science* **326**, 1509–1512 (2009).
8. Moscou, M.J. & Bogdanove, A.J. *Science* **326**, 1501 (2009).
9. Miller, J.C. *et al. Nat. Biotechnol.* **29**, 143–148 (2011).
10. Cermak, T. *et al.* published online, doi:10.1093/nar/gkr218 (14 April 2011).
11. Lengner, C.J. *et al. Cell* **141**, 872–883 (2010).
12. Hockemeyer, D. *et al. Cell Stem Cell* **3**, 346–353 (2008).
13. Zwaka, T. & Thomson, J. *Nat. Biotechnol.* **21**, 319–321 (2003).
14. Doyon, Y. *et al. Nat. Methods* **8**, 74–79 (2011).
15. Zhang, F. *et al. Nat. Biotechnol.* **29**, 149–153 (2011).

© 2011 Nature America, Inc. All rights reserved.
npg