nature biotechnology

Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases

Dirk Hockemeyer[1,4], Frank Soldner[1,4], Caroline Beard[1], Qing Gao[1], Maisam Mitalipova[1], Russell C DeKelver[2], George E Katibah[2], Ranier Amora[2], Elizabeth A Boydston[2], Bryan Zeitler[2], Xiangdong Meng[2], Jeffrey C Miller[2], Lei Zhang[2], Edward J Rebar[2], Philip D Gregory[2], Fyodor D Urnov[2] & Rudolf Jaenisch[1,3]

Realizing the full potential of human embryonic stem cells (hESCs) and induced pluripotent stem cells (hiPSCs) requires efficient methods for genetic modification. However, techniques to generate cell type–specific lineage reporters, as well as reliable tools to disrupt, repair or overexpress genes by gene targeting, are inefficient at best and thus are not routinely used. Here we report the highly efficient targeting of three genes in human pluripotent cells using zinc-finger nuclease (ZFN)–mediated genome editing. First, using ZFNs specific for the *OCT4* (*POU5F1*) locus, we generated OCT4-eGFP reporter cells to monitor the pluripotent state of hESCs. Second, we inserted a transgene into the *AAVS1* locus to generate a robust drug-inducible overexpression system in hESCs. Finally, we targeted the *PITX3* gene, demonstrating that ZFNs can be used to generate reporter cells by targeting non-expressed genes in hESCs and hiPSCs.

Gene targeting by homologous recombination in hESCs has proven difficult, and since the derivation of the first hESCs more than 10 years ago, only a few reports have described successful gene targeting[1–9]. These studies illustrate the utility of genetically modifying hESCs by gene targeting, but a general approach to manipulate the hESC genome is still lacking. Recently, a technique based on the introduction of DNA double-strand breaks by site-specific ZFNs to facilitate homologous recombination has been used to target endogenous genes in human cells[10,11]. A ZFN is generated by fusing the FokI nuclease domain to a DNA recognition domain composed of engineered C2H2 zinc-finger motifs that specify the genomic DNA binding site for the chimeric protein (**Fig. 1a**). Upon binding of two such fusion proteins at adjacent genomic sites, the nuclease domains dimerize, become active and cut the genomic DNA. When a donor DNA that is homologous to the target on both sides of the double-strand break is provided, the genomic site can be repaired by homology-directed repair, allowing the incorporation of exogenous sequences placed between the homologous regions[12,13]. This technique, also called 'genome editing', has been applied in systems not easily amenable to genetic modifications, such as zebrafish, plants and rats[14–19], and has also been used to edit the *CCR5* locus in hESCs[13].

We used the locus encoding the pluripotency-associated gene *OCT4* (also known as *POU5F1*)—one of the few genes that has been successfully targeted in hESCs[6]—to compare the efficiency of ZFN-mediated gene targeting in hESCs with that of conventional homologous recombination. We designed four ZFN pairs that recognize unique sequences in the first intron of the *OCT4* gene (**Fig. 1a**, **Supplementary Fig. 1** and **Supplementary Table 1**) and generated targeting donor constructs with short homology arms for the three most active ZFNs pairs. Correct targeting of these donor constructs containing a splice acceptor (SA) followed by an enhanced green fluorescent protein (eGFP)-2A-puromycin cassette (2A is a self-cleaving peptide sequence) results in the expression of two proteins—a fusion protein comprising the first 132 amino acids of human OCT4 fused to eGFP ($OCT4^{EX1\text{-}eGFP}$) and puromycin N-acetyltransferase—both under the control of the endogenous *OCT4* promoter. Southern blot analysis using external probes 3′ and 5′ to the donor homology regions and an internal probe against eGFP revealed that for ZFN pair #1, 40 out of 42 individual cell lines established from puromycin-resistant clones were correctly targeted (efficiency >94%; **Fig. 1b** and **Table 1**). ZFN pair #2 had a correct targeting frequency of 36–53% (**Table 1**). Of the remaining clones, most were correctly targeted in the *OCT4* locus but also carried additional nonhomologous integrations (**Fig. 1b** and **Table 1**). ZFN pair #3 had the lowest targeting efficiency and generated only a few puromycin-resistant clones, some of which were correctly targeted. $OCT4^{EX1\text{-}eGFP}$-targeted cells maintained a pluripotent state, as indicated by the expression of the pluripotency markers OCT4, NANOG, SOX2, Tra-1-60 and SSEA4 (**Fig. 1c**) and their ability to form cell types originating from all three developmental germ layers in teratoma-formation assays (**Fig. 1d**).

We detected expression of the $OCT4^{EX1\text{-}eGFP}$ fusion protein in hESCs by western blotting with antibodies against OCT4 and eGFP (**Fig. 1e**). When targeted hESCs were differentiated into fibroblasts, no $OCT4^{EX1\text{-}eGFP}$ protein was detected (**Fig. 1e**) and the cells regained puromycin sensitivity, demonstrating the validity of the reporter expression. Unexpectedly, clones targeted by ZFN pair #1 showed significantly lower $OCT4^{EX1\text{-}eGFP}$ protein

[1]The Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, USA. [3]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [2]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 8 June; accepted 10 August; published online 13 August 2009; doi:10.1038/nbt.1562

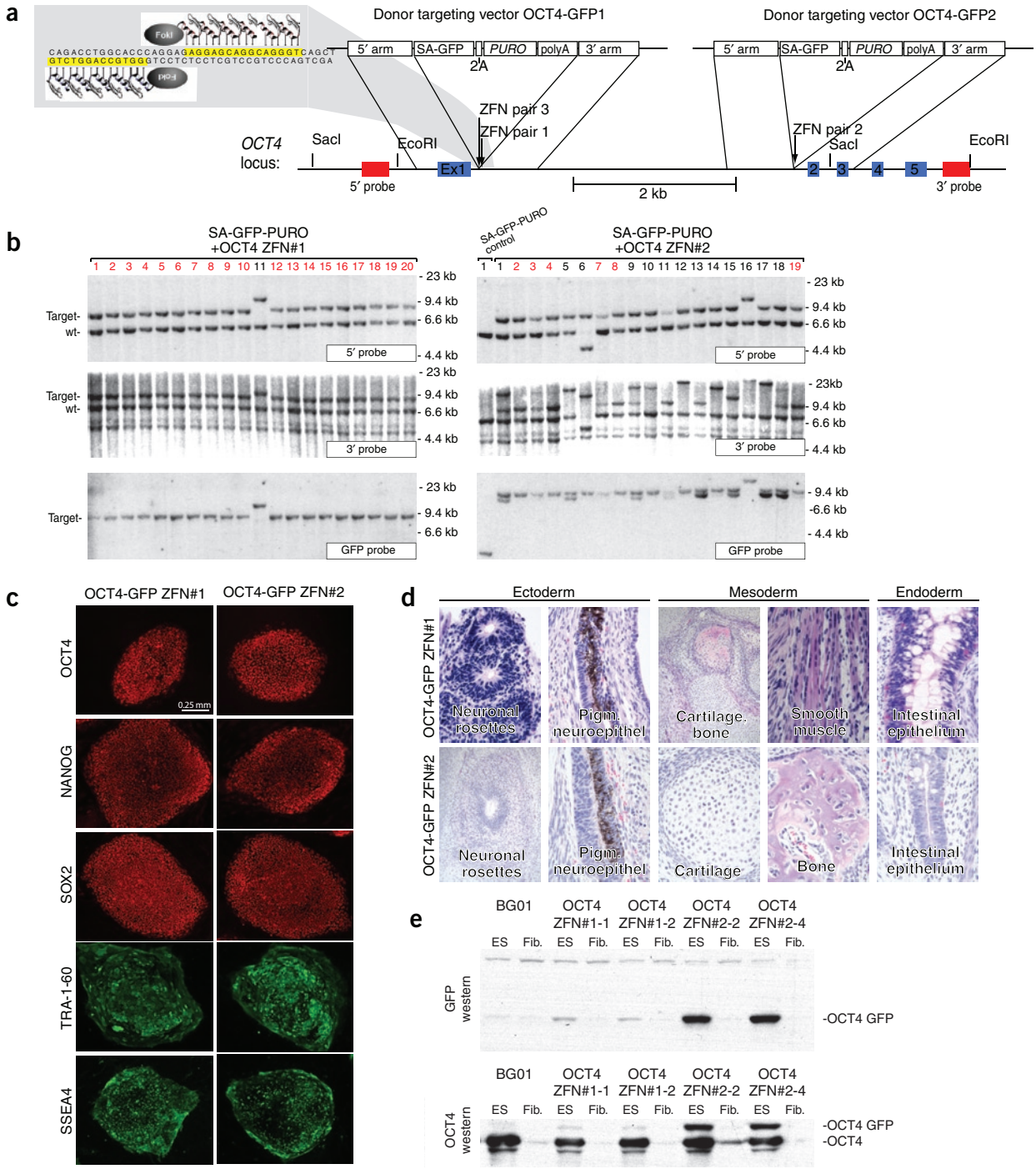

Figure 1 Targeting of *OCT4* in hESCs using ZFNs. (**a**) Schematic overview depicting the targeting strategy for the *OCT4* locus. Red boxes, probes used for Southern blot analysis; blue boxes, exons of *OCT4*; arrows, genomic site cut by the respective ZFN pair. Shown above is a schematic of the donor plasmid design. Donor plasmids were created corresponding to the cleavage location of the three ZFN pairs and carried roughly 700-bp regions of homology to the *OCT4* sequence. SA-eGFP, splice acceptor–eGFP sequence; 2A, self-cleaving peptide sequence; *PURO*, puromycin resistance gene; polyA, polyadenylation sequence. Inset at upper left is a cartoon of two ZFNs binding at a specific genomic site (yellow), leading to dimerization of the FokI nuclease domains. (**b**) Southern blot analysis of BG01 cells targeted with the indicated ZFN pairs using the corresponding donor plasmids. Genomic DNA was either digested with EcoRI and hybridized with the external 3′ probe or digested with SacI and hybridized with the external 5′ probe or internal eGFP probe. Correctly targeted clones without additional integrations are indicated in red. Fragment sizes: for 5′ probe and eGFP probe, wt, 6.4 kb, targeted, 8.4 kb; for 3′ probe, wt, 7.1 kb, targeted, 9.1 kb. (**c**) Immunofluorescence staining of BG01 cells targeted with the indicated ZFN pairs using the corresponding donor plasmids. Cells were stained for the pluripotency markers OCT4, NANOG, SOX2, Tra-1-60 and SSEA4. (**d**) Hematoxylin and eosin staining of teratoma sections generated from BG01 cells targeted with the indicated ZFN pairs and the corresponding donor plasmids. (**e**) Western blot analysis for the expression of OCT4 and eGFP in BG01 wild-type (wt) cells and BG01 cells targeted with the indicated ZFN pairs using the corresponding donor plasmids. Cell extracts were derived from either undifferentiated cells (ES) or *in vitro*–differentiated fibroblast-like cells (Fib.)

Table 1 Summary of targeting experiments

a Summary of targeting efficiencies for the *OCT4*, *AAVS1* and *PITX3* loci

OCT4

Cell line targeted	ZFN pair	Donor	Clones picked	Random integration[a]	Targeted + additional integrations[a]	Correct targeted clones: Het.[a]	Correct targeted clones: Homo.[a]	Correct targeted clones: Targeting efficiency (%)[a]
BG01	Control	OCT4-eGFP #1, #2, #3	2/1	2/1	0	0	0	0
BG01	ZFN #1 (2.5 μg)	OCT4-eGFP #1	4/21	0/1	0	4/20	0	100/95
BG01	ZFN #1 (10 μg)	OCT4-eGFP #1	17	1	0	16	0	94
BG01	ZFN #2 (2.5 μg)	OCT4-eGFP #2	15/22	0/1	7/13	8/8	0	53/36
BG01	ZFN #2 (10 μg)	OCT4-eGFP #2	31	1	18	12	0	39
BG01	ZFN #3 (2.5 μg)	OCT4-eGFP #3	2	1	0	1	0	50
	ZFN #3 (10 μg)	OCT4-eGFP #3	1	0	1	0	0	0

AAVS1

Cell line targeted	ZFN pair	Donor	Clones picked	Random integration	Targeted + additional integrations	Correct targeted clones: Het.	Correct targeted clones: Homo.	Correct targeted clones: Targeting efficiency (%)
BG01	Control	AAVS1/SA-PURO	10	10	0	0	0	0
BG01	AAVS1	AAVS1/SA-PURO	32	2	12	16	2	56
BG01	Control	AAVS1/PGK-PURO	36	36	0	0	0	0
BG01	AAVS1	AAVS1/PGK-PURO	35	13	5	16	1	49
BG01	AAVS1	AAVS1/TetO-eGFP FW	46	5	19	15	7	47
BG01	AAVS1	AAVS1/TetO-eGFP BW	35	0	21	10	4	40
iPS PD2^{1lox}-17-Puro-5	AAVS1	AAVS1/SA-PURO	23	1	8	11	3	61
iPS PD2^{1lox}-17-Puro-5	AAVS1	AAVS1/PGK-PURO	15	5	5	5	0	33
iPS PD2^{1lox}-17-Puro-10	AAVS1	AAVS1/PGK-PURO	37	9	9	15	4	51

PITX3

Cell line targeted	ZFN pair	Donor	Clones picked	Random integration	Targeted + additional integrations	Correct targeted clones: Het.	Correct targeted clones: Homo.	Correct targeted clones: Targeting efficiency (%)
BG01[b]	PITX3	PITX3-eGFP FW/BW	96/74	126	14/12	7/11	0	11
iPS PD2^{1lox}-17-Puro-10/ iPS PD2^{1lox}-21-Puro-20[c]	PITX3	PITX3-eGFP FW	30/20	23/18	4/1	3/1	0	8

[a]When two numbers are shown, this indicates the results from two independent experiments. Het., heterozygous; Homo., homozygous. [b]The first number indicates the result for targeting with the PITX3-GFP FW donor and the second indicates the result with the BW donor. PITX3-GFP FW and BW donor plasmids differed only in the orientation of the puromycin selection cassette. [c]The first number indicates the result for targeting iPS PD2^{1lox}-17-Puro-10 cell line, the second number indicates those for the PD2^{1lox}-21-Puro-20 cell line. The iPSCs were described in ref. 21.

b Summary of off-target analysis for the *OCT4*, *AAVS1* and *PITX3* targeting experiments

ZFN	NHEJ frequency of 'wt allele' in heterozygous clones	NHEJ at the 'top 10' off-target sites in correctly targeted heterozygous clones
OCT4 ZFN #1	1[d]/11	0/72
OCT4 ZFN #1	0/12	0/40
AAVS1	0/5	0/36
PITX3 ZFN #2	1[d]/18	1/36

[d]For *OCT4*, the mutated allele carried a deletion of 9 bp; for *PITX3*, the one mutated allele carried a deletion of 8 bp.

expression than clones targeted by ZFN pair #2, as indicated by western blot and FACS analysis (**Fig. 1e** and **Supplementary Fig. 2**). Additionally, hESCs targeted by ZFN pair #1 tolerated only puromycin concentrations of 0.5 μg/ml, whereas those targeted by ZFN pair #2 could be maintained in up to 2 μg/ml puromycin (data not shown), suggesting that the former had lower expression of puromycin N-acetyltransferase. One possible explanation of this finding is that the integration of the SA-eGFP-2A-puromycin cassette occurred very close to the splice donor of the first coding exon of *OCT4* (separated by 102 bp), perhaps impeding splicing efficiency and resulting in reduced OCT4$^{EX1\text{-}eGFP}$ expression.

Approaches to expressing transgenes in hESCs include random integration of expression vectors and the insertion of an expression construct with isogenic vector arms into the *ROSA26* locus[4] by homologous recombination. However, unpredictable position effects or low targeting frequencies have limited the utility of these methods. To develop a highly efficient and robust expression system, we used ZFN technology to target the *AAVS1* locus located on chromosome 19, encoding the *PPP1R12C* gene, which is ubiquitously expressed. This well-characterized locus has been previously shown to allow stable and long-term expression of transgenes in multiple cell types, including hESCs[20]. To target the first intron of *PPP1R12C*, we used a ZFN pair that generates a double-strand break at its target locus in hESCs (**Supplementary Fig. 1b** and **Supplementary Table 1**) and can be used to efficiently insert transgenes into this locus in transformed

human cell lines (R.C. DeKelver, V.M. Choi, E.A. Moehle, D.E. Paschon, J.C. Miller, F.D. Urnov *et al.*, unpublished results).

Two constructs with identical short homology arms were used to target the *AAVS1* locus: (i) a gene-trap vector for the *PPP1R12C* promoter containing a SA-2A-puromycin selection cassette and (ii) a puromycin selection cassette driven from the phosphoglycerol kinase (PGK) promoter (**Fig. 2a**). As indicated by Southern blot analysis, both vectors generated ~50% puromycin-resistant clones that had correctly targeted insertions on one or both alleles with no additional random integrations (**Fig. 2b**, **Table 1** and **Supplementary Fig. 3**). The ability to target genes with vectors using a selection cassette controlled by an exogenous promoter is important for targeting genes not expressed in hESCs, as demonstrated below for *PITX3*. All tested AAVS1-targeted hESCs, including homozygous targeted clones, retained a normal karyotype ($n = 4$) (**Supplementary Fig. 4a**) and remained pluripotent as determined by immunofluorescence staining for pluripotency markers (**Supplementary Fig. 4b**) and teratoma-formation assays (**Supplementary Fig. 4c**). hiPSCs[21] could be targeted with efficiency similar to that for hESCs (**Supplementary Fig. 5a,b** and **Table 1**)

To develop a transgenic overexpression system, we targeted to the *AAVS1* locus a donor plasmid expressing eGFP under the control of the constitutively active CAGGS promoter (**Fig. 2c**). HESCs targeted with this construct showed persistent and uniform eGFP expression (**Fig. 2d**). To generate an inducible expression system, an AAVS1 donor plasmid containing a minimal cytomegalovirus (CMV) promoter and a tetracycline response element driving the eGFP cDNA (TetO-eGFP) was targeted into the *AAVS1* locus in either the same or the opposite orientation as the *PPP1R12C* gene (**Fig. 2e**). ZFN-mediated targeting of BG01 cells with these donor plasmids yielded correctly targeted heterozygous (AAVS1-TetO-eGFP$^{+/-}$) and homozygous (AAVS1-TetO-eGFP$^{+/+}$) hESCs (**Fig. 2f,g**) with high efficiencies similar to those described above (TetO-FW, 47%; TetO-BW, 40%; **Table 1**). Correctly targeted hESCs were transduced with a lentivirus carrying the M2rtTA reverse transactivator to render the cells doxycycline (DOX) responsive. In agreement with previously reported lentiviral transduction efficiencies of hESCs, approximately 10% of the cells showed eGFP expression after the addition of DOX (**Fig. 2f**). To test DOX-inducible eGFP expression, we withdrew DOX from the cultures for varying lengths of time and found that eGFP fluorescence became undetectable 7 d after DOX withdrawal (**Supplementary Fig. 6**). We next established cell lines that showed limited silencing of the M2rtTA viral transgene by single-cell subcloning (**Fig. 2f**). FACS analysis of these cell lines, cultured under different DOX concentrations, revealed a dose-dependent relationship between DOX concentration and eGFP expression (**Fig. 2h**). These experiments showed that eGFP expression was dependent on DOX addition and on the presence of M2rtTA, indicating tight regulation of the eGFP expression cassette when integrated into the *AAVS1* locus. There were no apparent differences between the two orientations of the TetO-eGFP cassette.

Finally, we tested whether ZFNs could be used to modify genes that are not expressed in hESCs and hiPSCs by targeting the first exon of *PITX3*. PITX3 is a transcription factor expressed in some differentiated cell types, such as dopaminergic neurons, but not in hESCs. To generate a *PITX3* reporter, we designed a targeting construct in which the *PITX3* open reading frame was joined after amino acid 32 to the gene for eGFP followed by a polyadenylation signal and a *loxP*-flanked PGK-puromycin cassette (**Fig. 3a**). Coelectroporation of the donor plasmid with the PITX3 ZFN pair #2 (**Supplementary Fig. 7**) into hESCs or two hiPSC[21] lines resulted in correctly targeted clones with efficiencies of 11% for hESCs and 8% for iPSCs, as determined by Southern blot analysis (**Fig. 3b,c** and **Table 1**). All tested hESC and hiPSC lines targeted in the *PITX3* locus maintained a normal karyotype (hESC $n = 3$, hiPSC $n = 3$). The PGK-puromycin cassette was subsequently removed by transient expression of Cre recombinase (**Fig. 3b**). The relatively high targeting efficiency in hESCs and hiPSCs demonstrates that ZFN-mediated gene targeting is a robust tool for modifying genes not expressed in hECS so as to generate cell type–specific reporter systems. ZFN-mediated gene targeting therefore has the potential to overcome one of the main obstacles in hESC and hiPSC research.

A potential limitation of the ZFN targeting approach is off-target DNA breaks induced at related sequences elsewhere in the genome, which may cause unpredictable genotoxic effects. Although our Southern blot analysis using internal and external probes excluded additional integrations and confirmed the clonality of targeted clones, these analyses do not detect ZFN-mediated double-strand breaks and error-prone repair elsewhere in the genome. To examine off-target cleavage, we determined the DNA binding specificity for all ZFNs by SELEX (**Supplementary Table 2**), which made it possible to identify the most probable off-target cleavage sites on a genome-wide basis (**Supplementary Figs. 8a**, **9a**, **10a** and **11a**). Using the Cel-1 assay, we quantified the frequency of nonhomologous end joining (NHEJ)–mediated alterations in up to ten potential off-target sites in clones generated by four different ZFNs used in this study (**Table 1b** and **Supplementary Figs. 8b,c**, **9b,c**, **10b,c** and **11b,c**). In analyzing 46 genomic loci, we detected one NHEJ alteration at one genomic site in one of four PITX3 clones (**Table 1** and **Supplementary Fig. 9b–d**); all the other putative off-target sites for all the ZFNs in all the clones were wild type. Finally, we determined in heterozygously targeted clones the frequency of NHEJ alterations on the allele that did not carry an integrated transgene. This frequency was 1/18 for the PITX3 clones and 1/12 for the OCT4 clones carrying a disruption on the other allele (**Table 1b**); in all other cases, the other allele remained wild type. To ensure that a functional wild-type allele will be present in heterozygously targeted clones, the recognition sequence of the ZFNs can be designed to recognize intron sequences, as was done for the *OCT4* and *AAVS1* loci.

A recent study[22] described the use of ZFNs to disrupt the *PIGA* gene in hESCs and hiPSCs through the insertion of a drug-resistance marker; this work, together with our data on targeting five distinct loci in three different genes, demonstrates the utility of ZFNs for disrupting genes in human pluripotent stem cells. In addition, we have engineered both the *PITX3* locus, which is not expressed in hESCs, and the *OCT4* locus to report on cell fate and the *AAVS1* locus to be a 'safe harbor' for an inducible transgene, illuminating the range of genotypes that can be generated with ZFNs. We also note that, to ensure the uniqueness of our intended targets within the human genome, we relied on ZFNs containing 4–6 zinc fingers, which recognize composite sites of 24–36 bp.

ZFN-mediated gene targeting requires only relatively short targeting arms (for *AAVS1*, about 500 bp for each arm), facilitating the generation of targeting constructs to achieve site-specific integration of exogenous genes whose expression can be controlled by constitutively active, inducible or tissue-specific promoters. This is in contrast to targeting of genes such as *OCT4* by conventional homologous recombination, which has been shown to result in targeting efficiencies of 20% and 40% when isogenic targeting arms of 7.9 and 12.8 kb, respectively, were used[6]. Moreover, the use of short targeting arms combined with ZFN technology results in high targeting efficiencies even when the same donor plasmids are used in genetically different cell lines,

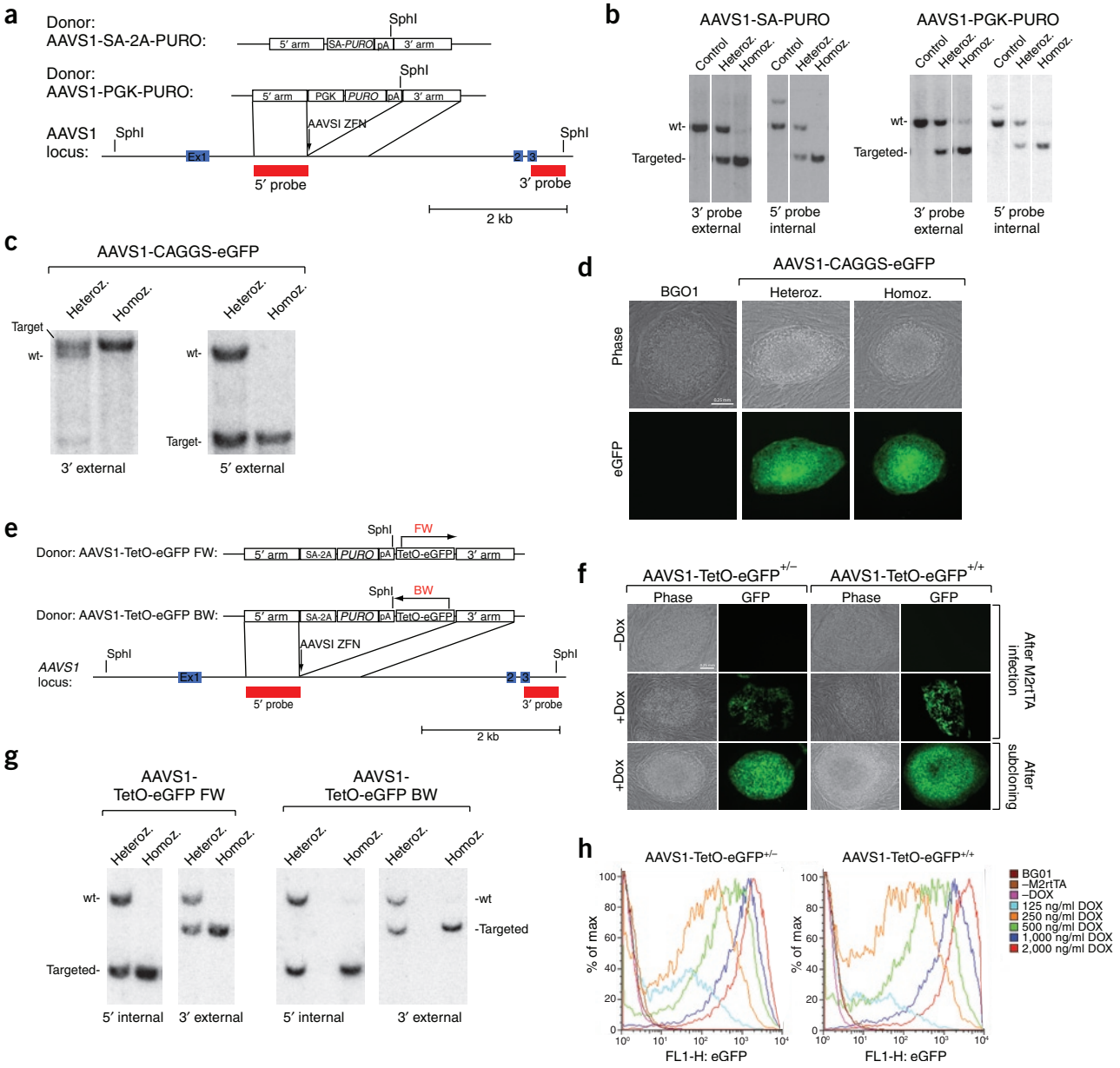

Figure 2 Targeting of the *AAVS1* locus using ZFNs. (**a**) Schematic overview depicting the targeting strategy for the *PPP1R12C* gene in the *AAVS1* locus. Red boxes, probes used for Southern blot analysis; blue boxes, first 3 exons of *PPP1R12C*; arrow, genomic site cut by the AAVS1 ZFNs. Donor plasmids used to target the locus are shown above. SA-PURO, splice acceptor sequence followed by a 2A self-cleaving peptide sequence and the puromycin resistance gene; pA, polyadenylation sequence; PGK, human phophoglycerol kinase promoter; *PURO*, puromycin resistance gene. (**b**) Southern blot analysis of BG01 cells targeted with the indicated donor plasmids using the AAVS1 ZFNs. Genomic DNA was digested with SphI and hybridized with the ^{32}P-labeled external 3′ probe or the internal 5′ probe. Fragment sizes for PGK-PURO: 5′ probe: wt, 6.5 kb, targeted, 4.2 kb; 3′ probe: wt, 6.5 kb, targeted, 3.7 kb. Fragment sizes for SA-PURO: 5′ probe: wt, 6.5 kb, targeted, 3.8 kb; 3′ probe: wt, 6.5 kb, targeted, 3.7 kb. (**c**) Southern blot analysis of BG01 cells targeted with an AAVS1 donor plasmid containing a CAGGS-driven eGFP cassette using the AAVS1 ZFNs. Genomic DNA was digested with SphI and hybridized with the ^{32}P-labeled external 3′ probe or the internal 5′ probe. Fragment sizes for CAGGS-GFP: 5′ probe: wt, 6.5 kb, targeted, 3.8 kb; 3′ probe: wt, 6.5 kb, targeted, 6.9 kb. (**d**) Phase-contrast picture and fluorescence imaging of eGFP in heterozygous or homozygous BG01 clones targeted with an AAVS1 donor plasmid containing a CAGGS-driven eGFP cassette and the AAVS1 ZFNs. (**e**) Schematic overview depicting the targeting strategy for the *PPP1R12C* gene in the *AAVS1* locus with a donor construct containing a DOX-inducible TetO-eGFP. Red boxes, probes used for Southern blot analysis; blue boxes, first three exons of the *PPP1R12C* gene in the *AAVS1* locus; arrows, genomic site cut by the ZFNs. Donor plasmids used to target the *AAVS1* locus are shown above. SA-PURO, splice acceptor sequence followed by a 2A self-cleaving peptide sequence and the puromycin resistance gene; pA, polyadenylation sequence; TetO, tetracycline response element. (**f**) Phase-contrast picture and fluorescence imaging of eGFP in BG01 cells either heterozygous (AAVS1 TetO-GFP$^{+/-}$) or homozygous (AAVS1-TetO-GFP$^{+/+}$) for the DOX-inducible eGFP cassette targeted to the *AAVS1* locus. Cells were transduced with a M2rtTA lentivirus to render them DOX responsive. Panel shows colonies before (top) and after FACS-assisted subcloning in the presence of DOX (bottom). (**g**) Southern blot analysis of BG01cells targeted with the indicated donor plasmids using the AAVS1 ZFNs. Genomic DNA was digested with SphI and hybridized with the ^{32}P-labeled external 3′ probe or the internal 5′ probe. Fragment sizes: 5′ probe: wt, 6.5 kb, targeted, 3.8 kb; 3′ probe: wt, 6.5 kb, targeted, 5.1 kb. (**h**) FACS analysis of AAVS1-TetO-GFP$^{+/-}$ and AAVS1-TetO-GFP$^{+/+}$ subclones for eGFP expression at different concentrations of DOX. BG01 cells, targeted cells before M2rtTA infection, and subcloned DOX-responsive cell lines cultured at different concentrations of DOX were analyzed. All cells were co-stained and analyzed for SSEA4 expression to exclude SSEA4-negative feeder cells from the analysis.

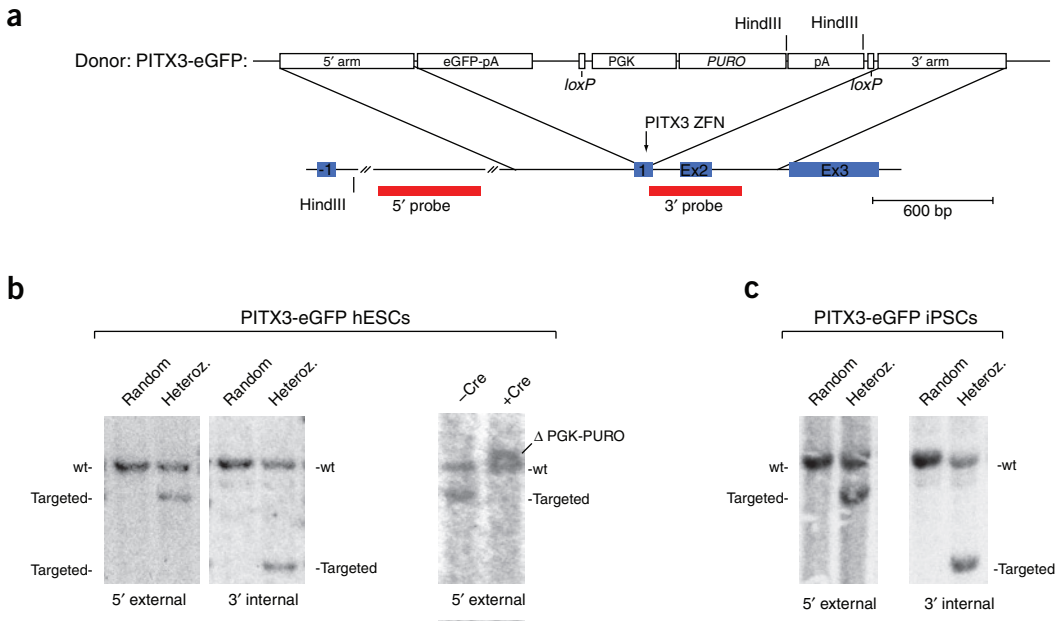

Figure 3 Targeting of *PITX3* in hESCs and hiPSCs using ZFNs. (**a**) Schematic overview depicting the targeting strategy for the *PITX3* gene. Red boxes, probes for Southern blot analysis; blue boxes, first exons of *PITX3*; arrows, genomic site cut by ZFN pair #2. Donor plasmids used to target the *PITX3* locus are shown above; these contained 5′ and 3′ homologous sequences of approximately 800 bp flanking the predicted ZFN pair #2 target site. eGFP, enhanced green fluorescent protein; PGK, human phophoglycerol kinase promoter; *PURO*, puromycin resistance gene, *loxP*, *loxP* sites; pA, polyadenylation sequence. Two constructs that differed only in the orientation of this selection cassette with respect to *PITX3* were successfully used to target *PITX3* (see also **Table 1**). (**b**) Southern blot analysis of BG01 cells targeted with the indicated donor plasmid using the PITX3 ZFNs. Right, Southern blot analysis of clones in which the PGK-PURO cassette was removed by transient expression of Cre recombinase. Genomic DNA was digested with HindIII and probed with ^{32}P-labeled external 5′ probe or the internal 3′ probe. Fragment sizes are: 5′ probe: wt, 8.8 kb, targeted, 7.4 kb, Δ-PGK-PURO, 10.5 kb; 3′ probe: wt, 8.8 kb, targeted, 4.3 kb. (**c**) Southern blot analysis of the hiPSCs targeted with the indicated donor plasmids using the PITX3 ZFNs. Genomic DNA was digested and probed as in **b**. Fragment sizes: 5′ probe: wt, 8.8 kb, targeted, 7.4 kb; 3′ probe: wt, 8.8 kb, targeted, 4.3 kb.

eliminating the need to construct isogenic targeting vectors. Although it is possible that not all nonexpressed genes can be targeted by ZFNs, the extraordinary flexibility available in designing zinc-finger binding motifs[23,24] will likely allow targeting of a substantial fraction of genes. Thus, this approach should be useful in generating genetic tools to study cell fate decisions and cell type–specific reporter systems to improve hESC differentiation protocols.

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

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturebiotechnology/.

ACKNOWLEDGMENTS

We thank R. Alagappan, P. Xu and E. Cook for technical support and J. Dausman, R. Flannery and D. Fu for their help with animal husbandry and processing of teratomas. We thank all the members of the Jaenisch laboratory for helpful discussions and comments on the manuscript. D.H. is a Merck Fellow of the Life Science Research Foundation. R.J. was supported by US National Institutes of Health grants R37-CA084198, RO1-CA087869 and RO1-HD045022: and by the Howard Hughes Medical Institute. Requests for ZFNs should be directed to F.D.U. (furnov@sangamo.com).

AUTHOR CONTRIBUTIONS

D.H., F.S. and R.J. designed the experiments and wrote the paper. C.B. provided assistance with construct design and Southern blot analysis. Q.G. analyzed all teratomas. M.M. provided hESCs. J.C.M. and L.Z. designed the ZFNs, which were assembled and tested by R.C.D., G.E.K. and R.A. X.M. performed the SELEX experiments. E.A.B. and B.Z. genotyped ZFN-edited clones for off-target effects. F.D.U., E.J.R. and P.D.G. supervised the ZFN design, characterization and off-target analysis and helped analyze the data and write the paper. D.H. and F.S. performed all other experiments.

COMPETING INTERESTS STATEMENT

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

Published online at http://www.nature.com/naturebiotechnology/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Thomson, J.A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
2. Urbach, A., Schuldiner, M. & Benvenisty, N. Modeling for Lesch-Nyhan disease by gene targeting in human embryonic stem cells. *Stem Cells* **22**, 635–641 (2004).
3. Costa, M. *et al.* A method for genetic modification of human embryonic stem cells using electroporation. *Nat. Protoc.* **2**, 792–796 (2007).
4. Irion, S. *et al.* Identification and targeting of the ROSA26 locus in human embryonic stem cells. *Nat. Biotechnol.* **25**, 1477–1482 (2007).
5. Suzuki, K. *et al.* Highly efficient transient gene expression and gene targeting in primate embryonic stem cells with helper-dependent adenoviral vectors. *Proc. Natl. Acad. Sci. USA* **105**, 13781–13786 (2008).
6. Zwaka, T.P. & Thomson, J.A. Homologous recombination in human embryonic stem cells. *Nat. Biotechnol.* **21**, 319–321 (2003).
7. Davis, R.P. *et al.* 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 (2008).
8. Xue, H. *et al.* A targeted neuroglial reporter line generated by homologous recombination in human embryonic stem cells. *Stem Cells* **27**, 1836–1846 (2009).
9. Ruby, K.M. & Zheng, B. Gene targeting in a HUES line of human embryonic stem cells via electroporation. *Stem Cells* **27**, 1496–1506 (2009).
10. Urnov, F.D. *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646–651 (2005).
11. Carroll, D. Progress and prospects: zinc-finger nucleases as gene therapy agents. *Gene Ther.* **15**, 1463–1468 (2008).
12. Moehle, E.A. *et al.* Targeted gene addition into a specified location in the human

genome using designed zinc finger nucleases. *Proc. Natl. Acad. Sci. USA* **104**, 3055–3060 (2007).
13. Lombardo, A. *et al.* Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat. Biotechnol.* **25**, 1298–1306 (2007).
14. Cai, C.Q. *et al.* Targeted transgene integration in plant cells using designed zinc finger nucleases. *Plant Mol. Biol.* **69**, 699–709 (2009).
15. Meng, X., Noyes, M.B., Zhu, L.J., Lawson, N.D. & Wolfe, S.A. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat. Biotechnol.* **26**, 695–701 (2008).
16. Doyon, Y. *et al.* Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat. Biotechnol.* **26**, 702–708 (2008).
17. Shukla, V.K. *et al.* Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature* **459**, 437–441 (2009).
18. Townsend, J.A. *et al.* High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature* **459**, 442–445 (2009).
19. Geurts, A.M. *et al.* Knockout rats via embryo microinjection of zinc-finger nucleases. *Science* **325**, 433 (2009).
20. Smith, J.R. *et al.* Robust, persistent transgene expression in human embryonic stem cells is achieved with AAVS1-targeted integration. *Stem Cells* **26**, 496–504 (2008).
21. Soldner, F. *et al.* Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. *Cell* **136**, 964–977 (2009).
22. Zou, J. *et al.* Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell* **5**, 97–110 (2009).
23. Pabo, C.O., Peisach, E. & Grant, R.A. Design and selection of novel Cys2His2 zinc finger proteins. *Annu. Rev. Biochem.* **70**, 313–340 (2001).
24. Klug, A. The discovery of zinc fingers and their development for practical applications in gene regulation. *Proc. Jpn. Acad.* **81**, 87–102 (2005).

ONLINE METHODS

Cell culture. Cell culture techniques have been described previously[21]. HiPSCs and the hESC line BG01 (NIH code BG01; BresaGen, Inc.) were maintained on mitomycin C–inactivated mouse embryonic fibroblast (MEF) feeder layers in hESC medium (DMEM/F12 (Invitrogen) supplemented with 15% fetal bovine serum (FBS) (HyClone), 5% KnockOut Serum Replacement (Invitrogen), 1 mM glutamine (Invitrogen), 1% nonessential amino acids (Invitrogen), 0.1 mM β-mercaptoethanol (Sigma) and 4 ng/ml FGF2 (R&D Systems)). Cultures were passaged every 5–7 d either manually or enzymatically with collagenase type IV (Invitrogen; 1.5 mg/ml). In order to perform FACS analysis of OCT4-eGFP clones in the absence of MEF feeder cells, hESCs were passaged onto Matrigel-coated plates in mTeSR medium (Stemcell Technologies). Karyotyping analysis was performed by Cell Line Genetics.

ZFN design and ZFN expression plasmids. ZFNs against the human *OCT4*, *AAVS1* and *PITX3* loci were designed using an archive of prevalidated two-finger modules exactly as described in published work[10,16,25]; complete sequences of the ZFNs, which carried obligate heterodimer forms of the FokI endonuclease[26], are provided in **Supplementary Table 1**. The ZFNs were designed and tested at Sangamo BioSciences for the purpose of disruption of their intended target loci by transient transfection into K562, HeLa and HEK293 cells, followed by Surveyor (Cel-1) endonuclease–based measurement of NHEJ at the target locus exactly as described[26,25] (primers used in Cel-1 analysis are provided in the **Supplementary Table 2**); the ZFN expression constructs were then provided to the Jaenisch laboratory.

Targeting of hESCs and hiPSCs using ZFN-mediated homologous recombination. hESCs and hiPSCs were cultured in rho kinase (ROCK) inhibitor (Calbiochem; Y-27632) 24 h before electroporation. Cells were harvested using 0.25% trypsin/EDTA solution (Invitrogen) and 1×10^7 cells resuspended in PBS were electroporated, if not otherwise indicated, with 40 μg of donor plasmids (designed and assembled by D.H. and F.S.) and 5 μg of each ZFN-encoding plasmid (Gene Pulser Xcell System, Bio-Rad; 250 V, 500 μF, 0.4-cm cuvettes[3]). Cells were subsequently plated on MEF feeder layers (DR4 MEFs for puromycin selection) in hESC medium supplemented with ROCK inhibitor for the first 24 h. Individual colonies were picked and expanded after puromycin selection (0.5 μg/ml) 10–14 d after electroporation.

Experimental genome-wide evaluation of ZFN action. A double-strand break persistently induced by designed ZFNs is repaired by NHEJ, an error-prone process[27] that generates small insertions and deletions at the site of the break[28]. This feature has been extensively used to profile the consequences of ZFN-driven editing on the target genome[15–17,25]. In the present work, such genotyping was performed essentially as described previously[16,25]. First, the consensus target for each ZFN was experimentally determined by SELEX as described[25] under conditions known to yield a biologically relevant consensus site for C2H2 ZFPs[29]. These studies yielded the targets provided in **Supplementary Table 3**. Next, the human genome was searched for candidate off-target sites that provided the best match to the experimentally determined base-frequency matrices obtained from the SELEX studies. In performing this step, we allowed ZFN site pairings with 5 or 6 bp between individual targets, in order to reflect the ability of our designed ZFNs to cleave equally well at these two spacings. Likewise, we also allowed an optional gap of 1 bp between the ninth and tenth bases of Oct-4 ZFN #1-R in order to reflect the binding characteristics of a longer flexible linker between the second and third fingers of this protein that allows binding to either target type. Finally, the experimentally determined base-frequency matrices for each ZFN pair were used to rank the potential off-target sites. For each ZFN target, the top-10-ranked off-target sites were then genotyped as follows. Four single cell–derived clones heterozygous for the transgene at the ZFN target site were chosen at random, genomic DNA was isolated, and every potential off-target site was amplified using 32 cycles of PCR with Accuprime Taq HiFi DNA polymerase (Invitrogen). The majority of the sites were then genotyped using the Surveyor endonuclease ('Cel-1'; Transgenomics) assay exactly as described[26], with the following modification: to address the potential for biallelic homozygous disruption (which would yield no Cel-1 signal), an equal amount of PCR product amplified from control cells was added to that from the ZFN-edited clone. Following a denaturation-renaturation step and treatment with Cel-1 to cleave heteroduplexes formed from wild-type and mutated DNA strands, the reaction was resolved by 10% nondenaturing PAGE (Bio-Rad) in 1× TBE. One putative off-target site was found to be heterozygous for a SNP, precluding the use of the Cel-1 assay; it was genotyped by cloning (TopoTA; Invitrogen) and Sanger sequencing. In addition to the analysis of putative off-target sites, for each heterozygous clone, the nontransgenic allele of the ZFN target locus was genotyped by PCR (using primers shown in **Supplementary Table 1**), cloning and sequencing.

Fibroblast differentiation of OCT4-eGFP hESCs. Embryoid body (EB)-induced differentiation was performed as previously described[30]. Briefly, hESC colonies were harvested using 1.5 mg/ml collagenase type IV (Invitrogen), separated from the MEF feeder cells by gravity, gently triturated and cultured for 7 d in nonadherent suspension culture dishes (Corning) in DMEM supplemented with 20% FBS. EBs were plated onto adherent tissue culture dishes and passaged according to primary fibroblast protocols using trypsin for at least four passages before the start of experiments.

Removal of PGK-PURO cassette by transient Cre-recombinase expression. HESCs targeted in the *PITX3* locus were cultured in ROCK inhibitor for 24 h prior to electroporation. Cells were harvested using 0.25% trypsin/EDTA solution (Invitrogen) and 1×10^7 cells resuspended in PBS were electroporated with pTurbo-Cre (40 μg; GenBank accession number AF334827) and pEGFP-N1 (10 μg; Clontech) as described previously[3] (Gene Pulser Xcell System, Bio-Rad; 250 V, 500 μF, 0.4-cm cuvettes). Cells were subsequently plated on MEF feeder layers in hESC medium supplemented with ROCK inhibitor. Cre recombinase–expressing cells were enriched by FACS sorting (FACS-Aria; BD Biosciences) of a single-cell suspension for eGFP expressing cells 60 h after electroporation followed by replating at a low density in ROCK inhibitor–containing hESC medium. Individual colonies were picked 10–14 d after electroporation.

Lentiviral infection of hESCs. The FUW-M2rtTA lentiviral vector has been described previously[30]. Vesicular stomatitis virus glycoprotein (VSVG)-coated lentiviruses were generated in 293 cells as described previously[31]. Briefly, culture medium was changed 12 h after transfection and virus-containing supernatant was collected 60–72 h after transfection. Viral supernatant was filtered through a 0.45-μm filter. This virus-containing supernatant was used to infect hESC aggregates that were separated from feeder cells by collagenase treatment and serial washes. Two consecutive infections in the presence of 2 μg/ml of polybrene were performed over a period of 12 h in suspension. hESC aggregates were replated after infection on feeder cells. Infection efficiencies were determined using FACS analysis for eGFP and SSEA4 (mouse monoclonal, Developmental Studies Hybridoma Bank) of cells cultured in the presence of DOX (Sigma-Aldrich; 2 μg/ml) for 2 d. To enrich for eGFP expressing cells, targeted and infected hESCs were FACS sorted as single-cell suspension 2 d after DOX induction in the presence of ROCK inhibitor (FACS-Aria; BD-Biosciences) and subsequently replated in the ROCK inhibitor–containing hESC medium. DOX-responsive GFP-expressing cell lines were isolated by manual picking of single colonies.

Teratoma formation and analysis. HESCs were collected by collagenase treatment (1.5 mg/ml) and separated from feeder cells by subsequent washes with medium and sedimentation by gravity. HESC aggregates were collected by centrifugation, resuspended in 250 μl of PBS and injected subcutaneously in the back of severe combined immunodeficient (SCID) mice (Taconic). Tumors generally developed within 4–8 weeks, and animals were killed before tumor size exceeded 1.5 cm in diameter. Teratomas were isolated and fixed in formalin. After sectioning, teratomas were diagnosed on the basis of hematoxylin and eosin staining.

Immunocytochemistry. Cells were fixed in 4% paraformaldehyde in PBS and immunostained according to standard protocols using the following antibodies: SSEA4 (mouse monoclonal, Developmental Studies Hybridoma Bank); Tra-1-60 (mouse monoclonal, Chemicon International); hSOX2 (goat polyclonal, R&D Systems); Oct-3/4 (mouse monoclonal, Santa Cruz Biotechnology); hNANOG (goat polyclonal R&D Systems) and appropriate Molecular Probes Alexa Fluor dye–conjugated secondary antibodies (Invitrogen).

Immunoblotting. HESCs were collected by collagenase treatment (1.5 mg/ml) and separated from feeder cells by subsequent washes with medium and sedimentation by gravity. hESC-derived fibroblasts were collected by trypsinization.

Cells were pelleted by centrifugation and washed with PBS and again collected by centrifugation. Cells were lysed in ice-cold buffer (50 mM Tris-HCl, pH 7.4, 20% glycerol, 1 mM EDTA, 150 mM NaCl, 0.5% Triton X-100, 0.02% SDS, 1 mM dithiothreitol, 2 mM phenylmethylsulfonyl fluoride, supplemented with proteinase inhibitor cocktail (Complete Mini, Roche). After 5 min on ice, 5 M NaCl was added to bring the final NaCl concentration to 400 mM. After 5 min on ice, an equal volume of ice-cold water was added and the lysate was mixed before immediate centrifugation in a microcentrifuge (14,000 rpm, 10 min). Protein concentration of the supernatant was determined by Bradford assay and 15 μg of protein was separated using 4–12% Bis-Tris gradient gels (Invitrogen). After transfer to PVDF membranes, blots were probed with antibodies to OCT4 (mouse monoclonal, Santa Cruz Biotechnology) or to GFP (Rbt pAB to GFP, Abcam ab290-50).

Southern blotting. Genomic DNA was separated on a 0.7% agarose gel after restriction digests with the appropriate enzymes, transferred to a nylon membrane (Amersham) and hybridized with ^{32}P-labeled random primer (Stratagene) probes.

25. Perez, E.E. *et al.* Establishment of HIV-1 resistance in CD4$^+$ T cells by genome editing using zinc-finger nucleases. *Nat. Biotechnol.* **26**, 808–816 (2008).
26. Miller, J.C. *et al.* An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.* **25**, 778–785 (2007).
27. Valerie, K. & Povirk, L.F. Regulation and mechanisms of mammalian doublestrand break repair. *Oncogene* **22**, 5792–5812 (2003).
28. Bibikova, M., Golic, M., Golic, K.G. & Carroll, D. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics* **161**, 1169–1175 (2002).
29. Phillips, C.M. *et al.* Identification of chromosome sequence motifs that mediate meiotic pairing and synapsis in *C. elegans*. *Nat. Cell. Biol.* **11**, 934–942 (2009).
30. Hockemeyer, D. *et al.* A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. *Cell Stem Cell* **3**, 346–353 (2008).
31. Brambrink, T. *et al.* Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* **2**, 151–159 (2008).

doi:10.1038/nbt.1562