HHS Public Access

Author manuscript

Nat Biotechnol. Author manuscript; available in PMC 2012 February 01.

Published in final edited form as:

Nat Biotechnol. ; 29(8): 731–734. doi:10.1038/nbt.1927.

Genetic engineering of human ES and iPS 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], and **Rudolf Jaenisch**[1,2]

[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

Abstract

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 Activation-Like Effector Nucleases (TALENs) for five distinct genomic loci. At all loci tested we obtained hESC and iPSC single-cell-derived clones carrying transgenic cassettes solely at the TALEN-specified location. Thus, TALENs mediate site-specific genome modifications in human pluripotent cells with comparable efficiency and precision as zinc finger nucleases (ZFNs).

Gene targeting of human pluripotent cells by homologous recombination is inefficient, which has impeded the use of human ES and iPS cells (hESCs and hiPSCs) in disease models. To overcome this limitation, we and others have shown that zinc finger nucleases can be used to modify the genomes of hESCs and iPSCs[1–3]. ZFNs can be engineered to induce a double strand break (DSB) precisely at a predetermined position in the genome[4]. This DSB can be repaired by end-joining to drive targeted gene disruption, or via the homology-directed DNA repair pathway using an ectopically provided 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) suggests[6] suggests an alternative approach to the design of site-specific nucleases. Natural TALEs are

[4]These authors contributed equally to this work

Author contributions

DH, HW, and RJ designed the targeting experiments and wrote the manuscript. HW and DH generated donor plasmids. DH preformed targeting experiments. SK and CSL, HW assisted with Southern blot analysis. QG analyzed teratomas. JPC and DH performed FACS analysis of targeted cells. LZ and JM designed the TALENs, SH assembled the TALENs, GC and YS tested the TALENs, and BZ, JC, and XM performed the off-target analysis. DH, RJ, LZ, GC, JM, BZ, XM, and FDU analyzed the data. ER, PG, and FDU designed and supervised the design of the TALENs and contributed to writing the manuscript.

transcription factors used by plant-pathogens to subvert host genome regulatory networks[6]. The DNA binding domain of TALEs is unusual: multiple ~34 amino acid units ("TALE repeats") are arranged in tandem, their sequence nearly identical except for two highly variable amino acids that establish the base recognition specificity for each unit[7, 8]. Each individual domain determines the specificity of binding to one DNA base pair in the TALE recognition sequence and therefore four different repeat units are sufficient to specify a novel site[7, 8]. Nucleases based on such engineered TALE domains can target endogenous genes in transformed human cells[9, 10]. Here, we assess whether TALENs can be used to genetically engineer endogenous loci in hESCs and iPSCs.

We designed TALENs targeting PPP1R12C (the AAVS1 locus), OCT4, PITX3 genes at precisely the same positions as targeted earlier by ZFNs[2]. TALEN expression constructs and corresponding homology bearing donor plasmids were electroporated into hESCs (line WIBR#3)[11] and iPSCs (line C1)[12] (see FigureS 1, TableS 1, 2). Southern blot analysis was used to identify correctly targeted clones.

We targeted the PPP1R12C gene with a gene trap approach (expressing puromycin (Puro) from the endogenous gene; Figure 1A,B; Table 1) or with an autonomous selection cassette (Puro expressed from the PGK-promoter (Figure 1, Table 1; Figure S1, 2, 3). Targeting efficiency was high and comparable to that with ZFNs[2] with 50% of the clones being targeted in one or both alleles carrying no randomly integrated transgenes (Figure 1B, Table 1; Figure S2). Similarly, a SA-Puro-eGFPeGFP transgene was highly expressed from this locus (Figure 1A; Figure S3A,B). Importantly, such targeted cells remained pluripotent based on analysis of marker expression and teratomas. Cells of all germ layers expressed eGFP indicating that TALEN as well as ZFN-mediated targeting of the PPP1R12C gene results in robust transgene expression in pluripotent as well as in differentiated cells (Figure S3C–E).

The OCT4 gene was targeted using three different donor plasmids resulting in expression of Puro 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-puromycin cassette (Figure S4). Targeting efficiency in ESCs and iPSCs was between 70% and 100% as determined by Southern blot analysis and DNA sequencing of single-cell-derived clones (Table 1, Figure S1,4).

We also targeted the first exon of the non-expressed PITX3 gene and found that ~6% of drug resistant clones carried the transgene solely at the PITX3 locus as evaluated by Southern blot analysis (Table 1, Figure S1, 5). Notably, in one clone the transgene had integrated into both alleles of the non-expressed PITX3 gene.

To target C' terminus in the OCT4 gene, not previously targeted by ZFNs, we generated TALENs directed against sequences flanking the stop codon of the OCT4 gene using two donor plasmids: the last OCT4 codon was either fused in frame with a GFP-PGK-Puro construct or a with an eGFP preceded by a 2A sequence (2A-GFP-PGK-Puro; Figure 1C, Table 1). Following excision of the LoxP flanked PGK-Puro cassettes either a C-terminal

OCT4-eGFP fusion protein or a separately translated eGFP proteins will be expressed under the control of the endogenous OCT4 gene[13]. Southern blot analysis showed that 2–10% of single cell derived hESC and hiPSC clones carried the transgene cassette at the OCT4 locus (Figure 1D, E; Table 1). Cre-mediated excision of the PGK-Puro cassette from OCT4-eGFP targeted hESCs resulted in nuclear eGFP fluorescence and from those targeted with OCT4-2A-eGFP -- in pan-cellular eGFP-fluorescence (Figure 1G). Fluorescence was higher in OCT4-2A-eGFP than in OCT4-eGFP targeted clones as determined by FACS analysis suggesting differential protein stabilities of the OCT4-eGFP fusion protein and eGFP (Figure S6). A gene trap vector fusing an eGFP-2A-Puro cassette with the last OCT4 codon had a 50% targeting efficiency (Figure 1C,F,G; Table 1; Figure S6) similar to that of gene trap-vectors designed to the first intron. As expected, eGFP expression became undetectable after differentiation into fibroblast-like cells (Figure S7) validating eGFP expression as a faithful reporter for OCT4 expression.

To illustrate the general utility of TALENs to generate such C-terminal fusion proteins, we designed TALENs to cut at the last coding exon of PITX3 and generated hESCs 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 (Figure S8; Table 1).

To assess the frequency of off-target modification not detected by Southern blot analysis we determined the binding specificity of the PPP1R12C TALENs using SELEX. When genotyping a panel of 19 maximal-likelihood potential off-target sites unintended cleavage was found only at low-frequency (See Figure S9, 10; Table S3, 4 and Supplemental material for detailed description). A strategy to minimize potential off-target events is to design TALENs to function as obligatory heterodimers[14]. As shown in Table 1, such heterodimeric nucleases in combination with the TALE DNA binding domain can yield high efficiency targeting of the PPP1R12C locus.

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 comparable to that 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 with efficiencies similar to those we have previously reported for ZFNs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank R. Alagappan, P. Xu, Sandra Cristea Stephen Lam and Anna Vincent for expert technical assistance. We thank Frank Soldner for helpful discussions on the manuscript. DH is a Merck Fellow of the Life Sciences Research

Foundation. RJ was supported by US National Institutes of Health grants R37-CA084198, RO1-CA087869 and RO1-HD045022 and by a grant from HHMI. RJ is an adviser to Stemgen and a cofounder of Fate Therapeutics. GJC, LZ, YS, JCM, BZ, JMC, XM, SJH, EJR, PDG and FDU are full-time employees of Sangamo BioSciences, Inc..

References

1. Zou J, et al. Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. Cell Stem Cell. 2009; 5:97–110. [PubMed: 19540188]
2. Hockemeyer D, et al. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat Biotechnol. 2009; 27:851–857. [PubMed: 19680244]
3. Hockemeyer, D.; Jaenisch, R. Gene Targeting in Human Pluripotent Cells; Cold Spring Harb Symp Quant Biol;
4. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with engineered zinc finger nucleases. Nat Rev Genet. 2010; 11:636–646. [PubMed: 20717154]
5. Soldner F, et al. Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell. 2011 *in press*.
6. Boch J, Bonas U. Xanthomonas AvrBs3 family-type III effectors: discovery and function. Annu Rev Phytopathol. 48:419–436. [PubMed: 19400638]
7. Boch J, et al. Breaking the code of DNA binding specificity of TAL-type III effectors. Science. 2009; 326:1509–1512. [PubMed: 19933107]
8. Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. Science. 2009; 326:1501. [PubMed: 19933106]
9. Miller JC, et al. A TALE nuclease architecture for efficient genome editing. Nat Biotechnol. 2011; 29:143–148. [PubMed: 21179091]
10. Cermak T, et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res.
11. Lengner CJ, et al. Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations. Cell. 141:872–883. [PubMed: 20471072]
12. Hockemeyer D, et al. A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. Cell Stem Cell. 2008; 3:346–353. [PubMed: 18786421]
13. Zwaka T, Thomson J. Homologous recombination in human embryonic stem cells. Nat Biotechnol. 2003; 21:319–321. [PubMed: 12577066]
14. Doyon Y, et al. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nat Methods. 8:74–79. [PubMed: 21131970]
15. Zhang F, et al. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol. 2011; 29:149–153. [PubMed: 21248753]

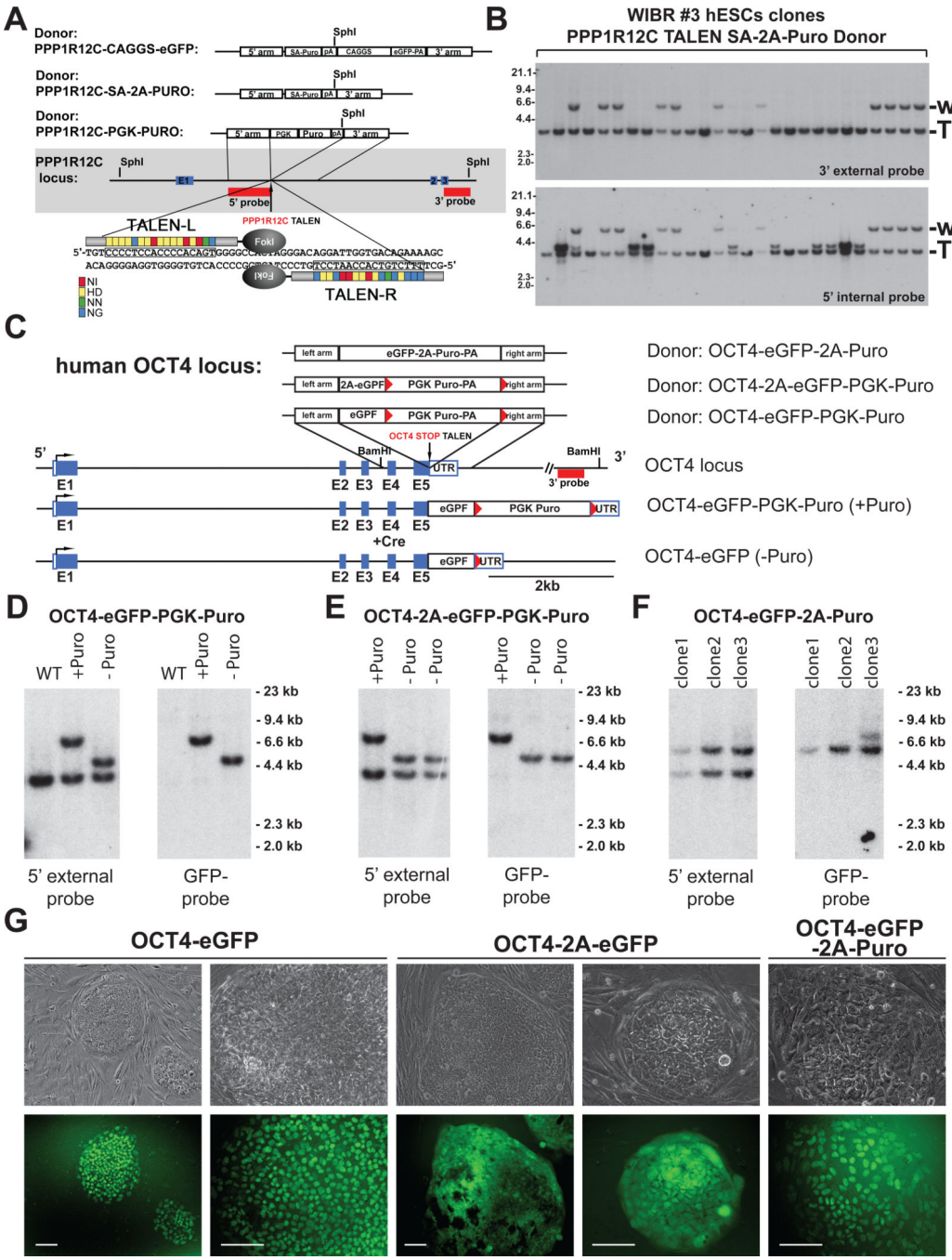

Figure 1. Genetic engineering of hESCs and iPSCs using TALENs

A. Schematic overview depicting the targeting strategy for the PPP1R12C gene. 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 CMV 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 hESCs 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 (Top) or with an internal 5'-probe (bottom). SA-2A-Puro 5'-probe detects a 6.5 kb wt and a 3.8 kb trageted fragment; 3'-probe a 6.5 kb wt and a 3.7kb targeted fragment. Wt = wild type and T = correctly targeted allele.

C. Schematic overview depicting the targeting strategy for the OCT4 locus using the OCT4-STOP TALENs. Southern blot probes and exon of OCT4 are colored as in panel 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 the OCT4 gene.

D. Southern blot analysis of the WIBR#3 hESCs 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. Two targeted clones are shown, one correctly targeted and one carrying a random integration. Expected fragment size: wt=4.2kb, targeted=6.8kb for both probes. Left lane: wt clone; middle lanes: clones before and after excision (right lane) of PGK-Puro cassette.

E. Southern blot analysis as in (D) of WIBR#3 hESCs 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 hESCs targeted with the eGFP-2A-Puro donor plasmids. Expected fragment size: wt 4.2kb, targeted 5.6kb for both probes. Two correctly targeted clones (left two lanes) and one targeted clone carriying 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 hESCs after excision of the PGK-Puro cassette at two magnifications. Right panels: phase contrast images and eGFP fluorescence of OCT4-eGFP-2A-Puro targeted hESCs clones. Size bars 100μm.

Nat Biotechnol. Author manuscript; available in PMC 2012 February 01.

Table 1

Summary of targeting experiments using TALENs OCT 4

						correctly targeted clones		
cell line targeted	**TALEN pair**	**donor**	**# of clones picked**	**random integration**	**targeted + additional integration**	**heterozygous**	**homozygous**	**targeting efficiency [%]**
WIBR#3 hESC	OCT4 intron1	OCT4-SAeGFP# 1	68	6	0	62	0	91
C1 iPSC	OCT4 intron1	OCT4-SAeGFP# 1	7	0	0	7	0	100
WIBR#3 hESC	OCT4 intron1	OCT4-SAeGFP# 3	35	0	0	34	0	97
C1 iPSC	OCT4 intron1	OCT4-SAeGFP# 3	5	0	0	5	0	100
WIBR#3 hESC	OCT4 intron1	OCT4-eGFP exon1 fusion	12	4	0	8	0	67
C1 iPSC	OCT4 intron1	OCT4-eGFP exon1 fusion	1	0	0	1	0	100
WIBR#3 hESC	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 hESC	OCT4 STOP	OCT4-2A-eGFP PGK-Puro (C-term fusion)	48	44	1	3	0	6
C1 iPSC	OCT4 STOP	OCT4-2Ae-GFP PGK-Puro (C-term fusion)	48	46	1	1	0	2
WIBR#3 hESC	OCT4 STOP	OCT4-eGFP-2A-Puro (genetrap)	48	0	26	22	0	46
PPP1R12C								
WIBR#3 hESC	PPP1R12C	PPP1R12C-SA-2A-Puro	96	4	35	34	21	57
WIBR#3 hESC	PPP1R12C	PPP1R12C-PGK-Puro	107	19	23	52	13	61
WIBR#3 hESC	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 hESC	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								
WIBR#3 hESC	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 hESC	PITX3 STOP	PITX3-eGFP C-term fusion	48	20	16	11	1	23

						correctly targeted clones		
cell line targeted	**TALEN pair**	**donor**	**# of clones picked**	**random integration**	**targeted + additional integration**	**heterozygous**	**homozygous**	**targeting efficiency [%]**
C1 iPSC	PITX3 STOP	PITX3-eGFP C-term fusion	48	32	7	9	0	19
WIBR#3 hESC	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