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]

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

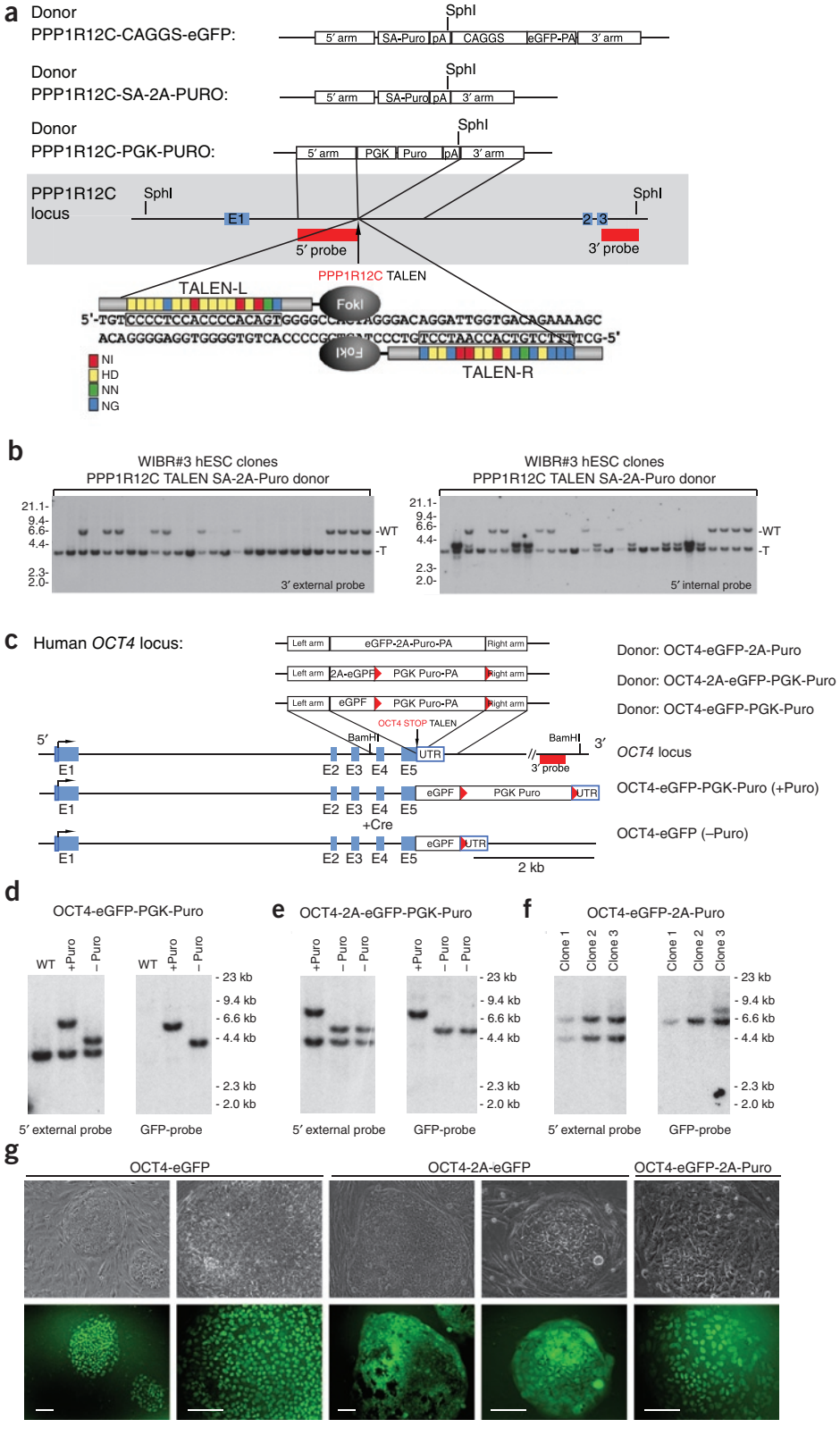


Figure 1 Genetic engineering of ESCs and iPSCs using TALENs. (**a**) Schematic overview depicting the targeting strategy for *PPP1R12C*. Southern blot probes are shown as red boxes, exons as blue boxes; the arrow indicates cut site by the TALENs. Donor plasmids: SA-2A-Puro, splice acceptor sequence followed 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**).

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).

SUPPLEMENTARY NOTES

EXPERIMENTAL PROCEDURES

Cell culture

Cell culture techniques have been described previously[1]. C1 hiPSCs[2] and the hESC lines WIBR#3[3] 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% KnockOutTM 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 to 7 days either manually or enzymatically with collagenase type IV (Invitrogen; 1.5 mg/ml). In order to partially remove feeder cells prior to FACS analysis, OCT4-eGFP targeted hESCs were passaged using collagenase. Cells were plated onto matrigel coated plates and cultured for 24-48 hours in mTESRI medium prior FACS analysis.

Genome editing with TALENs in transformed cells

TALENs were designed and assembled as described[4] except that the RVD in the C-terminal "half repeat" was not fixed as NG. Tandem arrays of TALE repeats were assembled from smaller fragments containing shorter tandem arrays of TALE repeats by PCR as described.[4]. All TALENs used the +63 truncation point[4]. The complete sequence of all TALENs used in this work is provided in supplementary information. K562 cells were obtained from the American Type Culture Collection (ATCC CCL-243) and grown and transfected with TALENs as described previously [4] . Four hundred thousand cells were transfected (estimated transfection efficiency ~90%), genomic DNA was extracted without selecting for transfected cells with Epicentre QuickExtract solution. Approximately 8,000 genome equivalents were used as input for each PCR. TALEN activity was

assayed via Surveyor nuclease (Transgenomic) using the following primers: PPP1R12C, Cel-I F2 (5'-ccc ctt acc tct cta gtc tgt gc-3') and Cel-I R1 (5'-ctc agg ttc tgg gag agg gta g-3'); PITX3, GJC 289F (5'-tta cga gga aac gct gct gg-3') and GJC 290R (5'-cca agc cag cgc ata ttc tc-3') along with GJC 291F (5'-aaa gcc aaa cag cac gcc tc-3') and GJC 292R (5'-gtc tgt gtg tag ggc cta gt-3'); OCT4, Group1F (5'-agg cgg ctt gga gac ctc tca-3') and Group1R (5'-tag gag atg tga gag acc ctg aca agg-3') along with GJC 295F (5'-ctt tct gtc att cac ttg cag-3') and GJC 296R (5'-act gaa gga caa acc aag at-3'). We used the primers indicated at a final concentration of 1 micromolar each in 50 microliter reactions with Accuprime Taq HiFi (Invitrogen) in buffer II, and ran the PCR reactions on a BioRad Tetrad as follows: an initial denaturation step (95 degrees, 3 min); next, 35 cycles of a denaturation step of 30 sec at 95 degrees, annealing step of 30 sec at 60 degrees, and an extension step of 30 sec at 72 degrees; a final extension step of 7 min at 72 degrees.

Targeting of hESCs and hiPSCs using TALEN mediated homologous recombination

hESCs and hiPSCs were cultured in Rho Kinase (ROCK)-inhibitor (Calbiochem; Y-27632) 24 hours prior to electroporation. Cell were harvested using 0.25% trypsin/EDTA solution (Invitrogen) and 1 x 10^7 cells resuspended in phosphate buffered saline (PBS) were electroporated if not otherwise indicated with 40 µg of donor plasmids (previously described in[5] or designed and assembled by D.H. and H.W.) and 5 µg of each TALEN encoding plasmid (Gene Pulser Xcell System, Bio-Rad: 250 V, 500µ F, 0.4 cm cuvettes[5]). Cells were subsequently plated on MEF feeder layers (DR4 MEFs for puromycin selection) in hESC medium supplemented with ROCK-inhibitor for the first 24 hours. Individual colonies were picked and expanded after puromycin selection (0.5 µg/ml) 10 to 14 days after electroporation.

Nature Biotechnology: doi:10.1038/nbt.1927

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 and Tra-1-81, (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).

FACS analysis

GFP-positive cells were measured using the BD LSR II flow cytometer and analyzed with FlowJo data analysis software.

Fibroblast differentiation of OCT4-eGFP hESCs

For EB induced differentiation, 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 days in non-adherent suspension culture dishes (Corning) in DMEM supplemented with 15% fetal bovine serum. EBs were plated onto adherent tissue culture dishes and passaged according to primary fibroblast protocols using trypsin.

Removal of PGK-Puro cassette by transient Cre-recombinase expression

hESCs targeted in the PITX3 locus were cultured in Rho Kinase (ROCK)-inhibitor (Calbiochem; Y-27632) 24 hours prior to electroporation. Cell were harvested using 0.25% trypsin/EDTA solution (Invitrogen) and 1 x 10^7 cells resuspended in PBS were electroporated with pTurbo-Cre (40 µg; Genbank Accession Number AF334827) and peGFP-N1 (10µg; Clontech) as described previously[5] (Gene Pulser Xcell System, Bio-Rad: 250 V, 500µF, 0.4 cm cuvettes). Cells were subsequently plated on MEF feeder layers (DR4 MEFs for puromycin selection) 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 hours after electroporation followed by replating at a low density in ROCK-inhibitor containing hESC medium. Individual colonies were picked 10 to 14 days after electroporation.

Teratoma formation and analysis:

hESCs were collected by collagenase treatment (1.5mg/ml) and separated from feeder cells by subsequent washes with medium and sedimentation by gravity. hESCs aggregates were collected by centrifugation, resuspended in 250µl of phosphate buffered saline (PBS) and injected subcutaneously in the back of SCID mice (Taconic). Tumors generally developed within 4-8 weeks and animals were sacrificed before tumor size exceeded 1.5 cm in diameter. Teratomas were isolated after sacrificing the mice and fixed in formalin. After sectioning, teratomas were diagnosed based on hematoxylin and eosin staining. Immunostaining of paraffin section was performed with standard techniques using a rabbit polyclonal anti-GFP antibody (Abcam 290).

Experimental determination of genome-wide specificity of action by the TALENs

The actual binding site for TALENs directed against the PPP1R12C locus was determined by in vitro SELEX exactly as described[6]; the SELEX data are shown in Supplemental Figure 9. The top ranked off-target sites were identified as previously described[6] using the SELEX-derived base frequency matrices in Supplemental Table 3. Capital letters indicate either (i) a match to the SELEX consensus or (ii) a non-targeted position within the TALEN heterodimer site. The genome searches for candidate off-target sites allowed for TALEN site pairings with 11 - 26 bp between individual targets. For logistical reasons, in this specific experiment we used TALENs carrying conventional FokI domains, but since TALENs carrying enhanced high-fidelity (eHiFi) obligate heterodimer FokI domains[7] directed against the PPP1R12C locus are robustly active both in transformed cells and hESCs/iPSCs (Table 1), we analyzed off-targets solely for TALEN heterodimers. Human ESCs were electroporated with TALEN expression

constructs exactly as described for ZFNs[5], and genomic DNA was harvested 72 hours later. Either 5µg of each TALEN and 40µg of a peGFP-N1 expression plasmid or, as a control, 50µg of the peGFP-N1 plasmid alone were electroporated. Illumina NextGen single-read 75 bp sequencing was used to interrogate the PPP1R12C on-target and top 20 SELEX-based off-target heterodimer sites for both control and TALEN-treated cells. Standard sequencing adapters were added via PCR in two steps using High-Fidelity Phusion Hot Start II Polymerase (NEB). In the first round, each site was amplified in a 20-cycle PCR using 200 ng genomic DNA and the primers listed in Supplemental Table 3, which also encoded the sequences 5'-CTACACTCTTTCCCTACACGACGCTCTTCCGATCT (For Seq) and 5'-CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT (Rev Seq) at the 5' ends. In the second round, an additional 20 cycles were run using 1 ul of product from the first PCR round and the primers SOLEXA-OUT-F1 (5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACG) and SOLEXA-OUT-R1 (5'-CAAGCAGAAGACGGCATA) to generate complete Illumina sequencing amplicons. The resulting PCR products were then gel-purified on 2% TBE-Agarose gels using Qiaquick Gel Extraction (Qiagen) and subsequently re-purified on Qiaquick PCR Purification columns (Qiagen). The samples were normalized to 20 nM each and combined in equal volumes so that no more than 11 sites were sequenced in a single Illumina lane. Control and TALEN-treated samples were processed in separate sequencing lanes. Sequences were binned by locus and then aligned to the appropriate wild-type sequence. Insertions and deletions (Indels) that occurred within a 10 bp window centered between the TALEN binding sites and contained 5 of fewer mismatched bases (excluding the InDel) were considered potential NHEJ events. This tended to yield a background InDel frequency of 0.04%-0.05%. To further reduce the background InDel frequency, InDels in the TALEN treated samples were compared to InDels in the corresponding control. Any InDel that had a normalized frequency at least 10-fold higher in the TALEN treated samples than in the corresponding control was considered for further analysis and background

corrected by subtracting the frequency (corrected for overall sample size) of the same InDel in the control samples. The converse correction was also performed (i.e. InDels that were observed at least 10-fold more frequently in the control than in the corresponding TALEN treated sample was considered for further analysis and corrected by subtracting the frequency of the same InDel observed in the corresponding TALEN treated sample). This correction yields a background level of approximately 0.0002% (i.e. one event for every 500,000 sequences). For unknown reasons, the control sample for off-target 19 yielded sequences for the incorrect locus, and while the TALEN-treated sample for that off-target yielded sequences in which InDels occurred at frequencies comparable to those where background correction demonstrates no off-target activity, lack of data from a control sample prevents us from fully analyzing that sample. In summary this analysis showed that in PPP1R12C TALEN-edited cells unintended cleavage occurs only at low-frequency: 17 loci remained wild-type, one of these sites was disrupted 169-fold less often than the intended target, and a second site was disrupted 1,140-fold less frequently (Figure S10 and Table S3 and 4).

Supplementary discussion and off-target analysis

Targeting the Stop Codon of OCT4- a position with redundant TALEN binding sites

Nucleases are routinely designed to target unique sites in the genome. In some rare cases, however, this is impossible; for instance, in editing the OCT4 locus, we found that a significant stretch of the DNA surrounding its stop codon occurs at 4 other unlinked positions in the genome within known OCT4 pseudogenes. The nucleases that cut at the stop, therefore, have 4 perfect matches elsewhere in the human genome (coordinates listed as per hg19 on genome.ucsc.edu):

% match	chr	strand	start	end
100.00%	6	-	31132351	31132405
100.00%	12	-	8286340	8286394
100.00%	1	+	155404027	155404080
98.20%	8	+	128429164	128429218
100.00%	10	-	69769937	69769988

Nature Biotechnology: doi:10.1038/nbt.1927

We genotyped all 4 of these other perfect target sites in 5 distinct single-cell-derived clones, all correctly edited at the OCT4 locus. The genotyping was done by PCR and Sanger sequencing, and was significantly complicated by the existence of a ~1.4 kb stretch of near-perfect homology between the sites ("ND" – no interpretable data obtained; "wt/?" – wild-type chromatids were obtained, but not at a number where a statistically significant argument could be made for wild-type homozygosity).

clone #	clone ID	chr 1	chr 8	chr 10	chr 12
1	7263	Δ16/?	N.D.	wt	Δ8/wt
2	7282	N.D.	wt	wt	N.D.
3	7111	wt/ Δ8	wt	wt	wt
4	7189	wt/?	Δ9/Δ5	wt	Δ3/Δ1
5	7187	wt/?	Δ8/wt	wt	Δ7/wt
6 (ctrl)	7464	wt	wt	wt	wt

One of the sites on chromosome 10 (chr 10) remained wild-type in all the clones. The site on chr 8 was wild-type in 2 clones, and carried small insertion or deletions (indels) in two others. For the site on chr 12 one clone remained wild-type, while 3 clones carried small indels. The site on chr 1 did not yield comprehensive data, but of note, clone #3 was wild-type for all off-target sites except an indel at one allele of that site. Since Southern blotting on these clones conclusively shows that the donor-specified insert was only transferred to the OCT4 locus, these data show that even in such rare circumstances (the occurrence of 4 perfect matches to the nuclease site), by relying on homology-directed repair it is possible to obtain cells that carry the insert at just one site, and with a minimal number of off-target mutations that do not affect cell phenotype in any measurable way. In the majority of cases, however, we design nucleases – both ZFNs and TALENs – to unique sites in the genome, and experimental measurement of nuclease action at off-target sites that inevitably carry multiple mismatches to the intended site has shown these to remain wild-

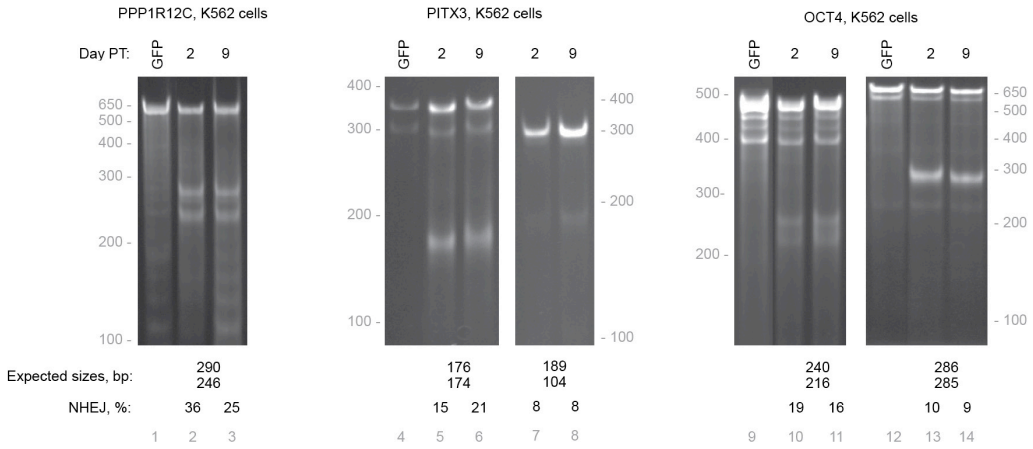
Nature Biotechnology: doi:10.1038/nbt.1927

type, with very rare exceptions (e.g., Perez et al, Nature Biotechnology 2008; Hockemeyer et al Nature Biotechnology 2009).

Nature Biotechnology: doi:10.1038/nbt.1927

Supplemental Figure 1

A

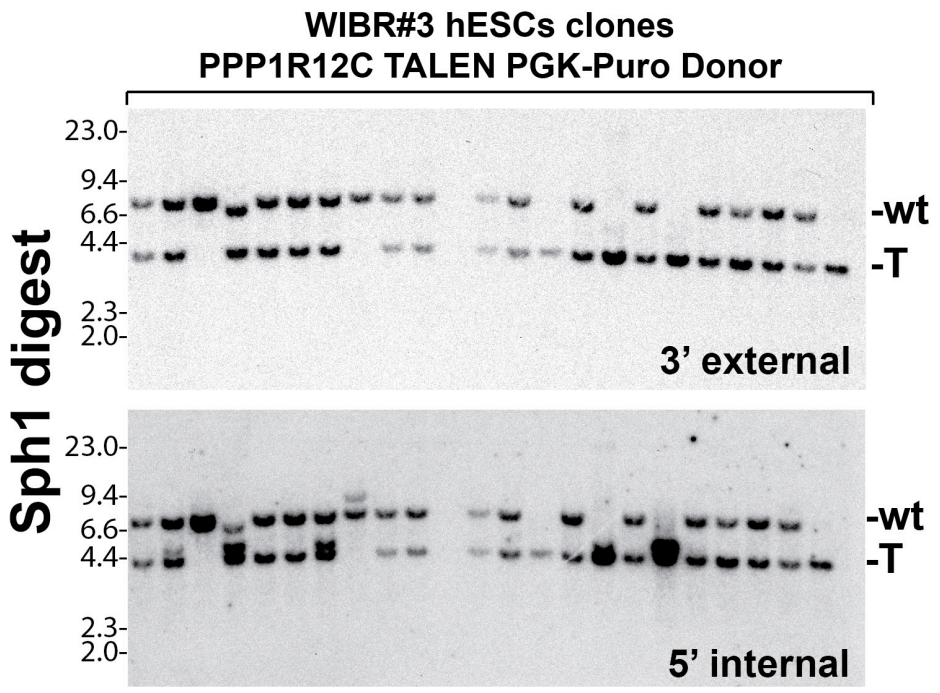


Assay of TALEN activity with the Surveyor nuclease. The AAVS1, PITX3, and OCT4 loci were PCR amplified and modification of the locus was measured using the Surveyor nuclease. The expected sizes of correctly digested fragments are shown below each panel. The degree of modification was quantified and is shown below each lane. GFP indicates GFP control transfection, 2 and 9 indicate days after transfection.

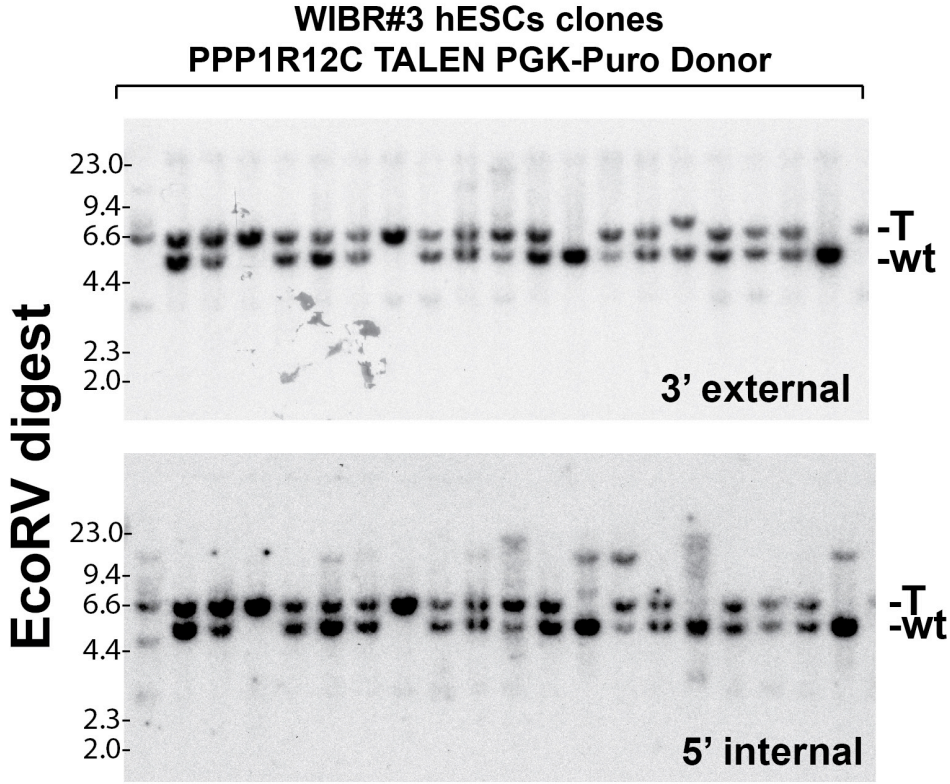
Nature Biotechnology: doi:10.1038/nbt.1927

Supplemental Figure 2

A



B



Nature Biotechnology: doi:10.1038/nbt.1927

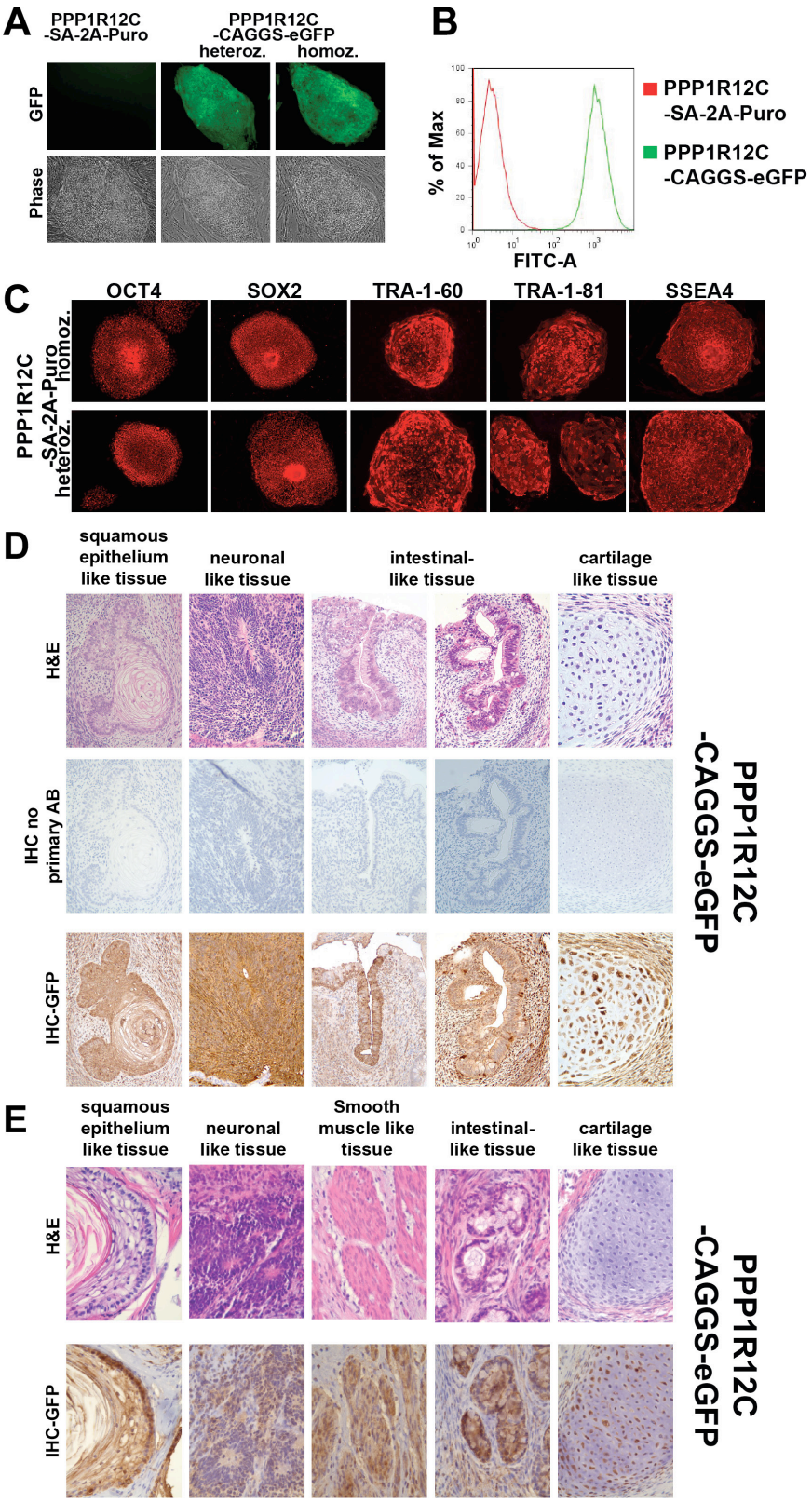
Targeting PPP1R12C with a PGK-Puro cassette using TALENs

A. Southern blot analysis of WIBR#3 hESCs cells targeted using the AAVS1-PGK-Puro donor plasmid and the PPP1R12C TALENs. Genomic DNA was digested with SphI and hybridized with the ^{32}P-labeled external 3'-probe or with the internal 5'-probe. Fragment sizes are: PGK-Puro: 5'-probe: wt=6.5 kb, targeted=4.2 kb; 3'-probe: wt=6.5 kb, targeted=3.7 kb.

B. Southern blot analysis of WIBR#3 hESCs cells targeted using the PPP1R12C-PGK-Puro donor plasmid and the PPP1R12C TALENs. Genomic DNA was digested with EcoRV and hybridized with the ^{32}P-labeled external 3'-probe or with the internal 5'-probe. As previously described for ZFN targeting of the PPP1R12C locus, a fraction of clones, although targeted, carried additional integrations[5]. Since EcoRV does not cut in the donor vector, the detection of higher than expected molecular weight restriction fragments in some of the targeted clones suggests that the additional integration detected with the internal 5'-probe might represent aberrant or multiple integrations of the donor plasmids in the PPP1R12C (AAVS1) locus rather than an integration caused by off-target effects of the PPP1R12C TALENs. These clones were considered as not correctly targeted and not analyzed further. The majority of clones obtained were correctly targeted on one or both TALEN-targeted alleles, and lacked randomly integrated DNA.

Nature Biotechnology: doi:10.1038/nbt.1927

Supplemental Figure 3



Nature Biotechnology: doi:10.1038/nbt.1927

Targeting PPP1R12C using TALENs

A. Phase contrast picture and fluorescence imaging of eGFP in heterozygous or homozygous WIBR#3 hESC clones targeted with the PPP1R12C-CAGGS-eGFP donor as shown in (A) using the PPP1R12C TALENs.

B. FACS analysis of eGFP fluorescence in WIBR#3 hESCs targeted in the PPP1R12C locus with the SA-2A-Puro donor or the CAGGS-eGFP donor.

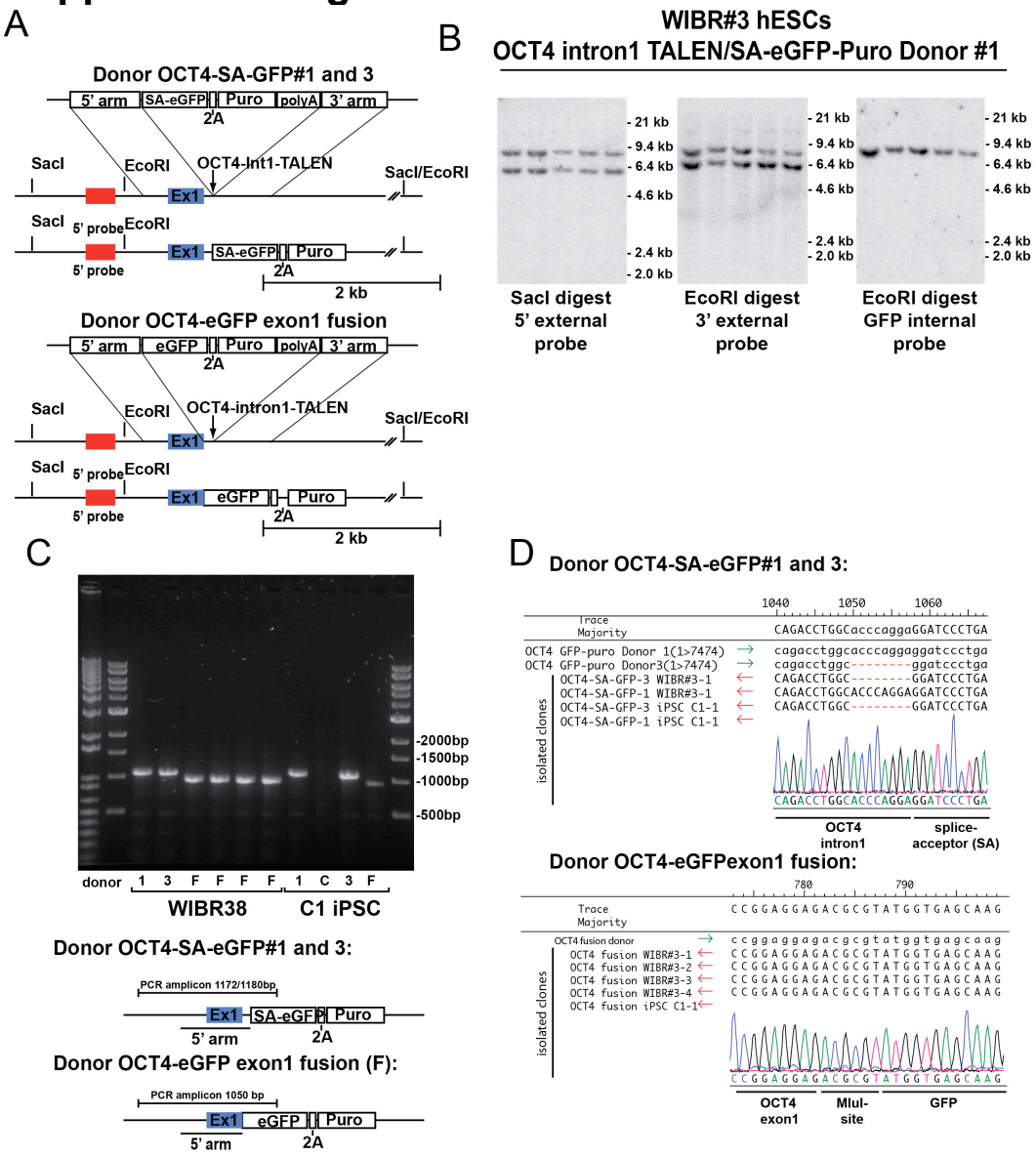
C. Immunofluorescence staining of WIBR#3 hESCs clones targeted with AAVS1 TALENs as shown in Figure 1A and B using the PPP1R12C-SA-2A-Puro donor plasmids. Cells were stained for the pluripotency markers OCT4, SOX2, Tra-1-60, Tra-1-81 and SSEA4.

D. Hematoxylin and eosin staining (H&E) of teratoma sections generated from WIBR#3 hESCs cells targeted with the PPP1R12C-CAGGS-eGFP donor plasmid and PPP1R12C **TALENs** (top row) and immunohistochemistry (IHC) staining for eGFP in neighboring sections (bottom row). As a control for staining specificity, neighboring sections were also stained omitting primary antibody incubation (middle row).

E. H&E of teratoma sections generated from WIBR#3 hESCs cells targeted with the PPP1R12C-CAGGS-eGFP donor plasmid and PPP1R12C **ZFN** (top row) and IHC staining for eGFP expression in neighboring sections (bottom row).

Nature Biotechnology: doi:10.1038/nbt.1927

Supplemental Figure 4



Targeting OCT4 to generate an OCT4-EXON1-GFP fusion protein

A. Schematic overview depicting the targeting strategy for the OCT4 locus using TALENs directed against the first intron of OCT4. We used three different donor plasmids to target the OCT4 locus. The first two donor plasmids (top schematic) (previously described as donor 1 and donor 3[5]) were designed to integrate a splice acceptor eGFP-2A-Puro cassette into the first intron of OCT4. The design of donor1 and 3 differed only in the positioning of the homology arms with respect to the TALEN cut. In donor 1, the 5' homology arm was comprised of homology sequences upstream of the 94 bp of the first OCT4 intron. In donor 3 this homology started at

Nature Biotechnology: doi:10.1038/nbt.1927

position 102 bp. The third donor – named OCT4 EXON1-GFP fusion-- (bottom panel) was engineered so that the (left) 5' homology arm starts with the last nucleotide of the first exon of OCT4. This targeting strategy requires that the DSB, generated by the TALENs, gets resected and single stranded DNA is generated to the start of the homology arm. Homology based DNA repair will result in a fusion of exon 1 to the reading frame of the eGFP-2A-puromycin cassette. Probes used for Southern blot analysis are shown as red boxes, exons of the OCT4 locus are shown as blue boxes and arrows indicate the genomic site cut of the OCT4 TALEN; SA-eGFP: splice acceptor eGFP sequence, 2A: self-cleaving peptide sequence, Puro: puromycin resistance gene, polyA: polyadenylation sequence.

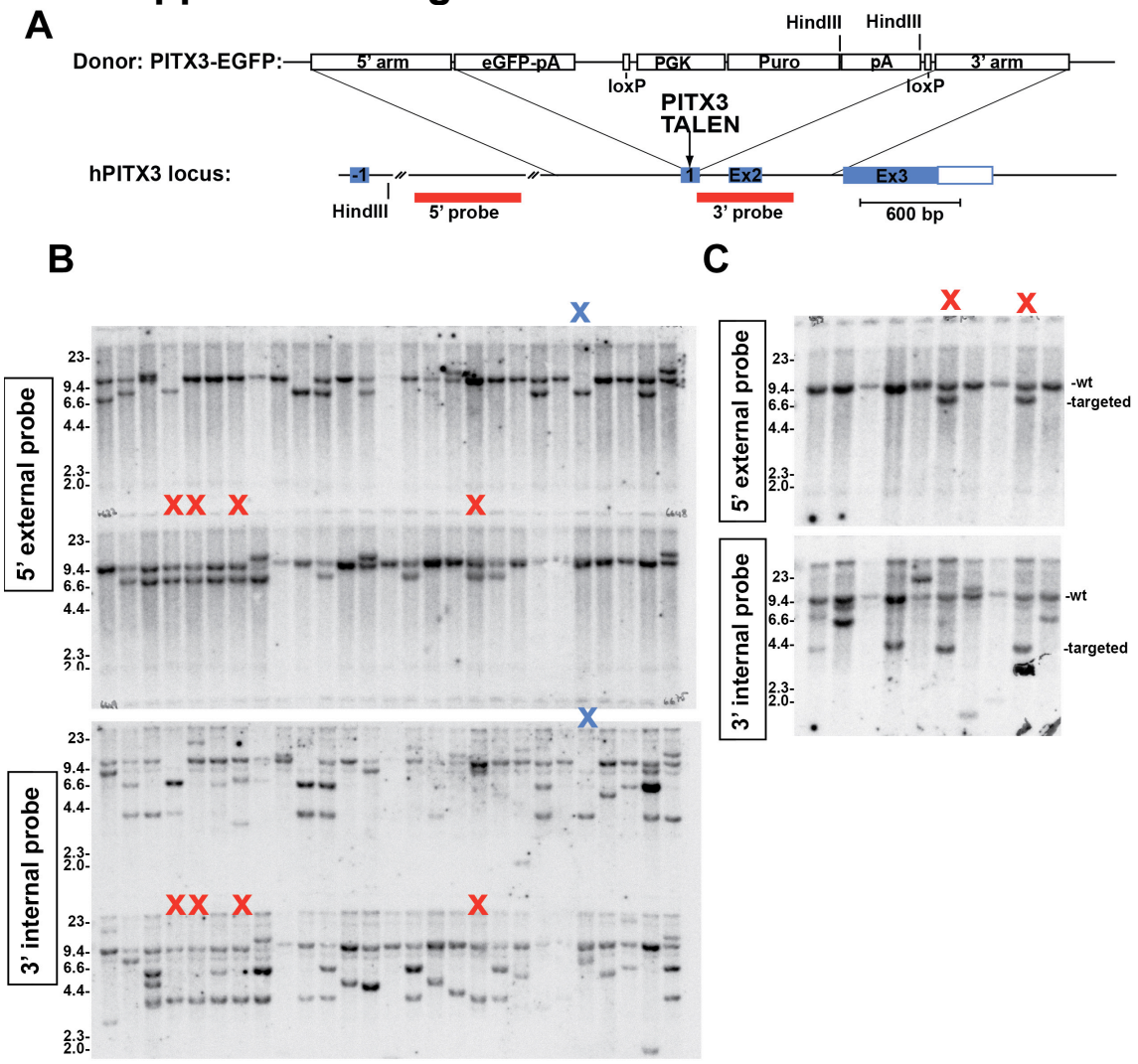
B. Southern blot analysis of WIBR#3 hESCs targeted with the indicated OCT4-Intron1 TALENs using the indicated donor plasmids. Genomic DNA was digested either with SacI and hybridized with the external 5'-probe (left panel) or digested with EcoRI and hybridized with the external 3'-probe (middle panel) or internal eGFP probe (right panel). Shown are three independent single cell derived clones for each donor.

C. Genotyping PCR of hESCs targeted using the indicated donors and OCT4-Intron1 TALENs as described in (A) and found to be correctly targeted based on Southern blot analysis as shown in (B). The PCR strategy is shown below. Primers were positioned in genomic regions 5' (outside) of the homology arms and in eGFP. Expected fragment size: Donor 1:1180 bp (lane labeled with 1), donor 3: 1172 bp (lane labeled with 3), donor fusion: 1050 bp (lane labeled with F). Primer sequences were: FW: GAGCAGAAGGATTGCTTTGG; BW: AAGTCGTGCTGCTTCATGTG. Size markers are 1kb+ (NEB) first lane and 1kb (NEB) second and last lane.

D. Sequencing analysis and sequence alignment of PCR products shown in (C). Sequencing confirmed the correct integration of the transgenes at the position specified by the donor plasmid. Top panel: sequence analysis and alignment for donor 1 and 3. Bottom panel: sequence analysis and alignment for OCT4 fusion donor. MluI indicates an in-frame restriction site used during donor design. This data illustrates that the position of transgene insertion using a given TALEN can be modified using alternative donor designs.

Nature Biotechnology: doi:10.1038/nbt.1927

Supplemental Figure 5



Targeting PITX3 using TALENs

A. Schematic overview depicting the targeting strategy for the PITX3 gene. Probes for Southern blot analysis are shown as red boxes, the first exons of the PITX3 locus are shown as blue boxes and arrows indicate the genomic site cut by the PITX3 TALENs. The donor plasmid used to target the PITX3 locus is shown above and contained 5' and 3' homologous sequences of approximately 800 bp flanking the predicted TALEN cut site. eGFP: enhanced green fluorescent protein, PGK: human phophoglycerol kinase promoter, Puro: puromycin resistance gene, loxP: loxP sites, pA: polyadenylation sequence. Two constructs (PITX3 donor FW and BW) that differed only in the orientation of this selection cassette with respect to the PITX3 gene were successfully used to target PITX3 (See also Table 1).

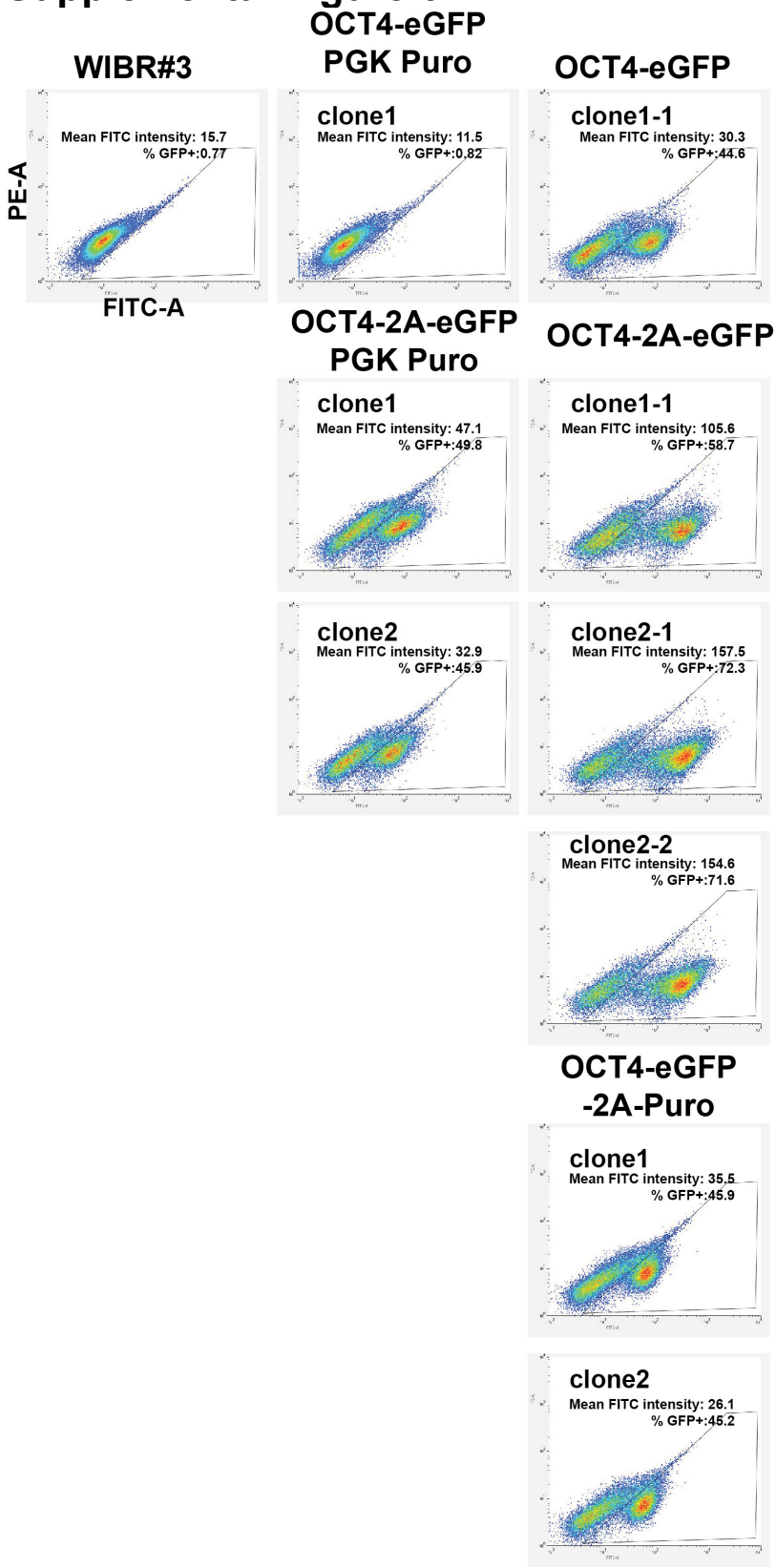
Nature Biotechnology: doi:10.1038/nbt.1927

B. Southern blot analysis of WIBR#3 hESCs targeted with the PITX3 FW donor plasmid using the PITX3 TALENs. Genomic DNA was digested with HindIII and probed with ^{32}P-labeled external 5'-probe (top panel) or with the internal 3'-probe (bottom panel). Fragment sizes are: 5' probe: wt=8.8 kb, targeted=7.4 kb; 3'-probe: wt=8.8 kb, targeted=4.3 kb. Heterozygous targeted clones are indicated in red and the homozygous targeted clone is indicated in blue.

C. Southern blot analysis of the C1 iPSCs targeted with PITX3 FW donor plasmids using the PITX3 TALENs. Genomic DNA was digested with HindIII and probed with ^{32}P-labeled external 5' probe (top panel) or with the internal 3'-probe (bottom panel). Fragment sizes are: 5' probe: wt=8.8 kb, targeted=7.4 kb; 3'-probe: wt=8.8 kb, targeted=4.3 kb. Heterozygous targeted clones are indicated in red.

Nature Biotechnology: doi:10.1038/nbt.1927

Supplemental Figure 6



Nature Biotechnology: doi:10.1038/nbt.1927

FACS analysis of OCT4-eGFP targeted cells

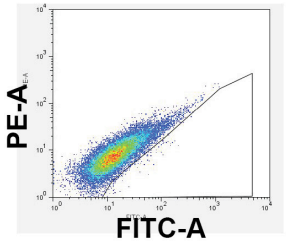
Graphs show FACS analysis of eGFP fluorescence (FITC-A, x axis) against PE-A auto-fluorescence (y-axis) in the indicated cell types. Left column shows wild-type WIBR#3 hESCs. Middle column shows one hESC clone targeted with OCT4-eGFP-PGK-Puro (top) and two clones targeted with OCT4-2A-eGFP-PGK-Puro (middle and bottom) before removal of the Puro selection cassette. eGFP fluorescence is low in these clones most likely due to the lack of an appropriate polyadenylation signal. The right column shows eGFP fluorescence for one clone after Cre-mediated removal of Puro (top), and for three clones expressing OCT4 and eGFP from a OCT4-2A-eGFP modified locus (middle). For all of these clones a significant increase in the percentage of eGFP positive cells (%GFP+) and the mean FITC intensity can be detected after eGFP is expressed using the natural OCT4 3' UTR. The bottom panels show the eGFP fluorescence of WIBR#3 clones targeted with the OCT4-eGFP-2A-Puro gene trap vector.

Nature Biotechnology: doi:10.1038/nbt.1927

Supplemental Figure 7

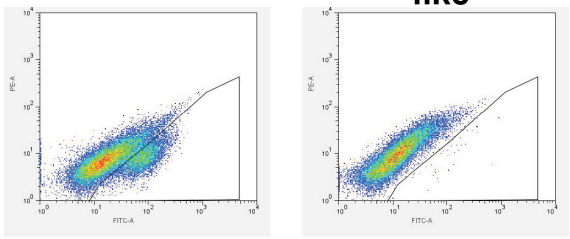
WIBR#3

hESCs



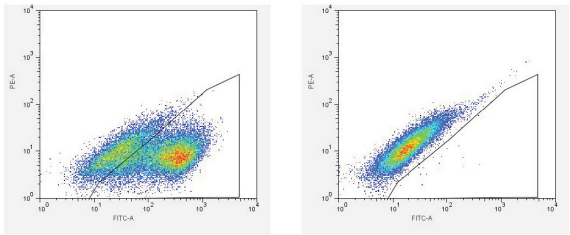
OCT4-eGFP

hESCs | **fibroblast -like**



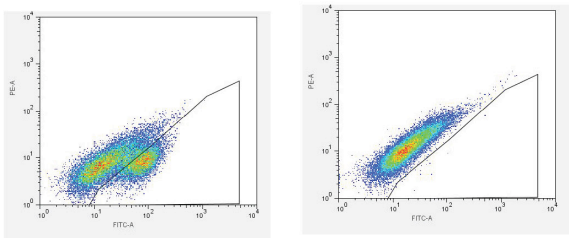
OCT4-2A-eGFP

hESCs | **fibroblast -like**



OCT4-eGFP -2A-Puro

hESCs | **fibroblast -like**



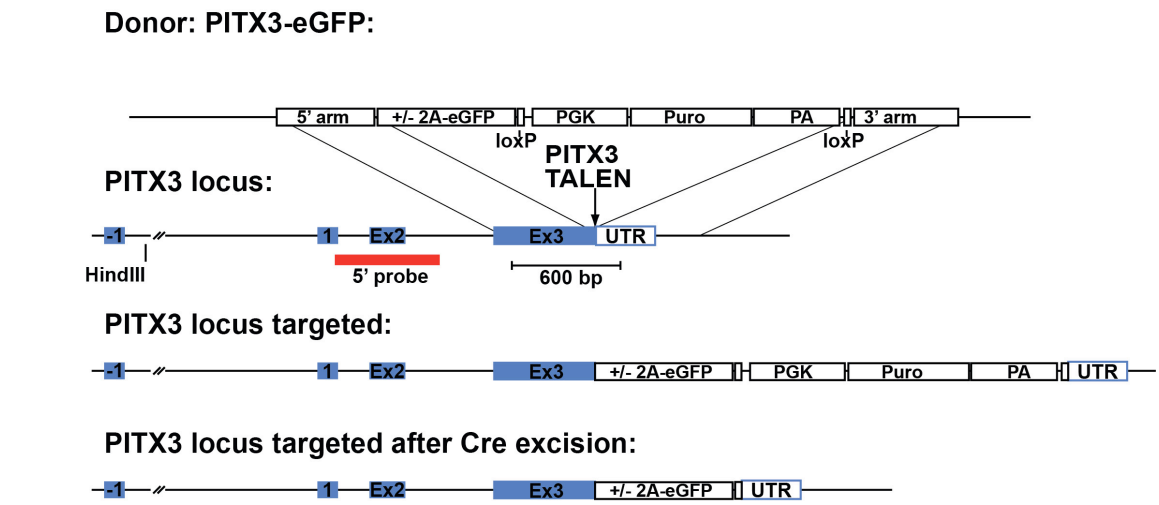
Nature Biotechnology: doi:10.1038/nbt.1927

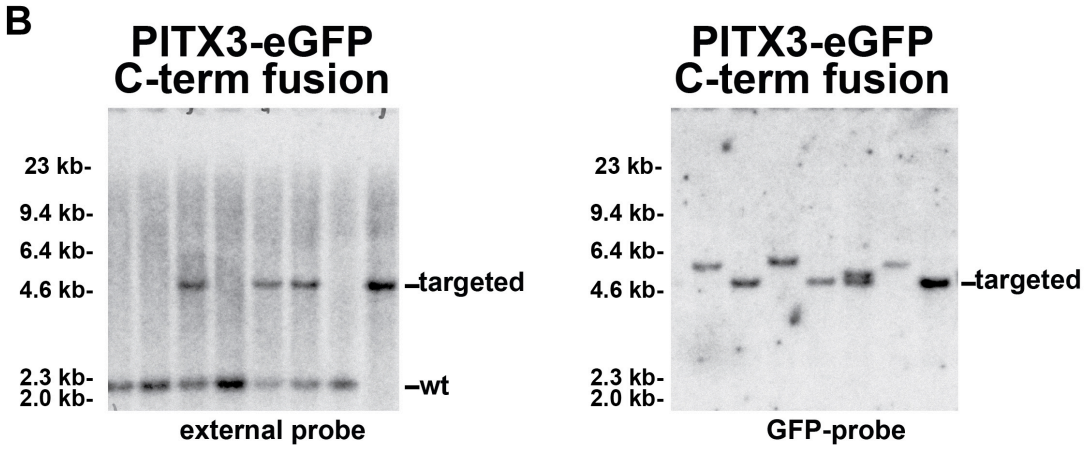
FACS analysis of OCT4-eGFP targeted cells after differentiation

Shown is FACS analysis of eGFP fluorescence (FITC-A, x axis) against PE-A auto-fluorescence (y-axis) in the indicated cell types. Top row shows unmodified wild-type hESCs. The graphs below show hESCs with the indicated modification at the OCT4-STOP codon. Shown on the left side are undifferentiated hESCs and on the right side the same cell line is shown after differentiation into fibroblast-like cells.

Nature Biotechnology: doi:10.1038/nbt.1927

Supplemental Figure 8





Using TALENs to generate a C-terminal fusion protein and eGFP-reporter line for PITX3

A. Schematic overview depicting the targeting strategy for the PITX3 locus using the PITX3-STOP TALENs. The external probe used for Southern blot analysis is indicated as a red box, exons of the PITX3 locus are shown as blue boxes and arrows indicate the genomic site cut by PITX3-STOP TALEN. Shown above is a schematic of the donor plasmid used to target the OCT4 locus (+/- 2A: indicates that two independent donor plasmids that only differ in this 2A-self cleaving peptide sequence were used), eGFP: enhanced green fluorescent protein, loxP-sites are indicated, PGK: human phophoglycerol kinase promotor, Puro: puromycin resistance gene, UTR: untranslated region of the PITX3 gene. Donor plasmids were constructed so that after the removal of the PGK-Puro cassette a PITX3-eGFP fusion protein or eGFP alone (from the 2A sequence) will be expressed using the natural 3' UTR of the PITX3 gene.

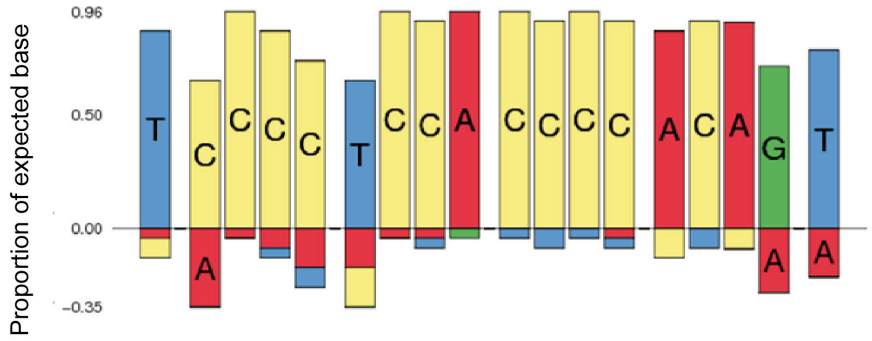
Nature Biotechnology: doi:10.1038/nbt.1927

B. Southern blot analysis of WIBR#3 hESCs targeted in the PITX3 locus to generate a C-terminal PITX3-eGFP fusion protein. Genomic DNA was digested with NcoI and hybridized with the ^{32}P-labeled external 5'-probe (left panel) or with the internal GFP probe (right panel).

Nature Biotechnology: doi:10.1038/nbt.1927

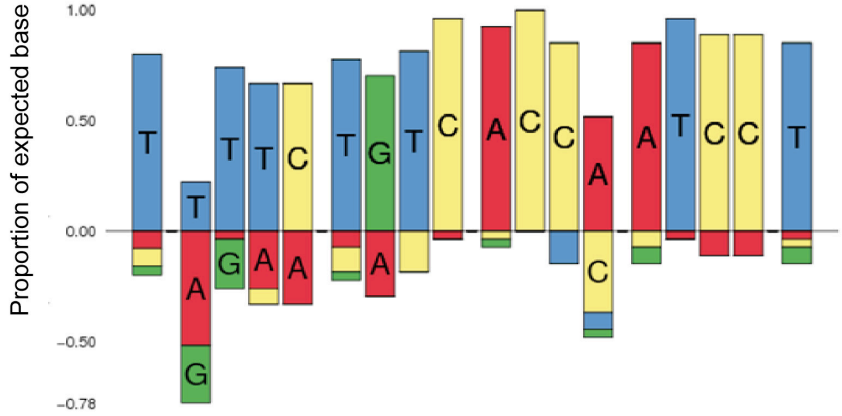
Supplemental Figure 9

A



PPP1R12C TALEN left

B



PPP1R12C TALEN right

Graphical depiction of a SELEX-derived base frequency matrix for the PPP1R12C TALENs.

At each matrix position, the frequency of the predicted target base is shown above the x-axis, whereas the remaining base frequencies are shown below the x-axis.

Nature Biotechnology: doi:10.1038/nbt.1927

Supplemental Figure 10

PPP1R12C deletions

locus	Number	Size	Loc.	Sequence
P1R12C				TATCTGTCCCCTCCACCCCACAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAG
P1R12C	1991	-1	52	TATCTGTCCCCTCCACCCCACAGTGGGGCCACT:GGGACAGGATTGGTGACAGAAAA
P1R12C	1855	-1	51	TATCTGTCCCCTCCACCCCACAGTGGGGCCAC:AGGGACAGGATTGGTGACAGAAAA
P1R12C	1174	-7	51	TATCTGTCCCCTCCACCCCACAGTGGGGCCAC:::::::AGGATTGGTGACAGAAAAGCCCCA
P1R12C	843	-10	46	TATCTGTCCCCTCCACCCCACAGTGGG:::::::::ACAGGATTGGTGACAGAAAAGCCCCATCC
P1R12C	671	-5	44	TATCTGTCCCCTCCACCCCACAGTG:::::CTACTAGGGACAGGATTGGTGACAGAAAAGC
P1R12C	622	-1	50	TATCTGTCCCCTCCACCCCACAGTGGGGCCA:TAGGGACAGGATTGGTGACAGAAAA
P1R12C	595	-2	51	TATCTGTCCCCTCCACCCCACAGTGGGGCCAC::GGGACAGGATTGGTGACAGAAAAG
P1R12C	571	-9	47	TATCTGTCCCCTCCACCCCACAGTGGGG:::::::::ACAGGATTGGTGACAGAAAAGCCCCATC
P1R12C	568	-9	43	TATCTGTCCCCTCCACCCCACAGT:::::::::AGGGACAGGATTGGTGACAGAAAAGCCCCATC
P1R12C	566	-17	53	TATCTGTCCCCTCCACCCCACAGTGGGGCCACTA:::::::::::::::::AGAAAAGCCCCATCCTTAG
P1R12C	565	-8	52	TATCTGTCCCCTCCACCCCACAGTGGGGCCACT::::::::GATTGGTGACAGAAAAGCCCCAT
P1R12C	564	-10	48	TATCTGTCCCCTCCACCCCACAGTGGGGC:::::::::AGGATTGGTGACAGAAAAGCCCCATCC
P1R12C	557	-9	41	TATCTGTCCCCTCCACCCCACA:::::::::CTAGGGACAGGATTGGTGACAGAAAAGCCCCATC
P1R12C	553	-9	50	TATCTGTCCCCTCCACCCCACAGTGGGGCCA:::::::::GGATTGGTGACAGAAAAGCCCCATC
P1R12C	551	-16	36	TATCTGTCCCCTCCACC:::::::::::::::AGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAG
P1R12C	549	-12	42	TATCTGTCCCCTCCACCCCACAG:::::::::::GGACAGGATTGGTGACAGAAAAGCCCCATCCTT
P1R12C	532	-18	32	TATCTGTCCCCTC:::::::::::::::::::CTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAG
P1R12C	519	-11	45	TATCTGTCCCCTCCACCCCACAGTGG:::::::::::ACAGGATTGGTGACAGAAAAGCCCCATCCT
P1R12C	457	-3	48	TATCTGTCCCCTCCACCCCACAGTGGGGC:::TAGGGACAGGATTGGTGACAGAAAAGC
P1R12C	430	-25	29	TATCTGTCCC:::::::::::::::::::::::::GGACAGGATTGGTGACAGAAAAGCCCCATCCTTAG
P1R12C	321	-23	33	TATCTGTCCCCTCC:::::::::::::::::::::::::ACAGGATTGGTGACAGAAAAGCCCCATCCTTAG
P1R12C	290	-6	50	TATCTGTCCCCTCCACCCCACAGTGGGGCCA::::::ACAGGATTGGTGACAGAAAAGCCCC
P1R12C	261	-4	49	TATCTGTCCCCTCCACCCCACAGTGGGGCC::::GGGACAGGATTGGTGACAGAAAAGCC
P1R12C	57	-9	39	TATCTGTCCCCTCCACCCCA:::::::::CACTAGGGACAGGATTGGTGACAGAAAAGCCCCAT
P1R12C	11	-7	52	TATCTGTCCCCTCCACCCCACAGTGGGGCCACT:::::::GGATTGGTGACAGAAAAGCCCCA

PPP1R12C Insertions

locus	Number	Size	Ins.	Sequence
P1R12C				TATCTGTCCCCTCCACCCCACAGTGGGGCCACTAGGGACAGGATTGGTGACAGAA
P1R12C	1694	+2	AC	TATCTGTCCCCTCCACCCCACAGTGGGGCCACACTAGGGACAGGATTGGTGACAGA
P1R12C	1558	+1	T	TATCTGTCCCCTCCACCCCACAGTGGGGCCACTTAGGGACAGGATTGGTGACAGAA
P1R12C	1402	+4	ACTA	TATCTGTCCCCTCCACCCCACAGTGGGGCCACTAACTAGGGACAGGATTGGTGACA
P1R12C	1183	+1	A	TATCTGTCCCCTCCACCCCACAGTGGGGCCACTAAGGGACAGGATTGGTGACAGAA
P1R12C	826	+3	CTA	TATCTGTCCCCTCCACCCCACAGTGGGGCCACTACTAGGGACAGGATTGGTGACAG
P1R12C	632	+3	TAA	TATCTGTCCCCTCCACCCCACAGTGGGGCCATAA::AGGGACAGGATTGGTGACAGAA
P1R12C	619	+4	GTGA	TATCTGTCCCCTCCACCCCACAGTGGGGCCAGTGACTAGGGACAGGATTGGTGACA
P1R12C	588	+2	TA	TATCTGTCCCCTCCACCCCACAGTGGGGCCACTATAGGGACAGGATTGGTGACAGA
P1R12C	573	+2	CT	TATCTGTCCCCTCCACCCCACAGTGGGGCCACTCTAGGGACAGGATTGGTGACAGA
P1R12C	553	+4	CTCC	TATCTGTCCCCTCCACCCCACAGTGGGGCCACCTCCTAGGGACAGGATTGGTGACA
P1R12C	481	+1	C	TATCTGTCCCCTCCACCCCACAGTGGGGCCACCTAGGGACAGGATTGGTGACAGAA
P1R12C	308	+3	GAC	TATCTGTCCCCTCCACCCCACAGTGGGGCCACGACTAGGGACAGGATTGGTGACAG
P1R12C	293	+4	CACT	TATCTGTCCCCTCCACCCCACAGTGGGGCCACTCACTAGGGACAGGATTGGTGACA
P1R12C	274	+4	TGCT	TATCTGTCCCCTCCACCCCACAGTGGGGCCACTTGCTAGGGACAGGATTGGTGACA
P1R12C	264	+5	TACTT	TATCTGTCCCCTCCACCCCACAGTGGGGCCACTTACTTAGGGACAGGATTGGTGAC
P1R12C	263	+4	CCAC	TATCTGTCCCCTCCACCCCACAGTGGGGCCACCCACTAGGGACAGGATTGGTGACA
P1R12C	17	+4	CTTA	TATCTGTCCCCTCCACCCCACAGTGGGGCCACTACTTAGGGACAGGATTGGTGACA
P1R12C	11	+6	AAGCTT	TATCTGTCCCCTCCACCCCACAGTGGGGCAAGCTTCACTAGGGACAGGATTGGTGA
P1R12C	10	+2	CC	TATCTGTCCCCTCCACCCCACAGTGGGGCCACCTAGGGACAGGATTGGTGACAGA

OT10 Deletions

locus	Number	Size	Loc.	Sequence
OT10				TGTTCTTCCACCCAGCCCTCAGTGCTCTGCTGAGGCAGGATGGGTGGCTGGAAAAATGGGGTCTTT
OT10	211	-5	45	TGTTCTTCCACCCAGCCCTCAGTGCT:::::GAGGCAGGATGGGTGGCTGGAAAAATGGGG
OT10	149	-3	48	TGTTCTTCCACCCAGCCCTCAGTGCTCTG:::AGGCAGGATGGGTGGCTGGAAAAATGG
OT10	3	-1	49	TGTTCTTCCACCCAGCCCTCAGTGCTCTGC:GAGGCAGGATGGGTGGCTGGAAAAAT
OT10	2	-1	45	TGTTCTTCCACCCAGCCCTCAGTGCT:TGCTGAGGCAGGATGGGTGGCTGGAAAAAT

OT14 Deletions

locus	Number	Size	Loc.	Sequence
OT14				AGTATTTTCTATAACTCATATTTTCACAAAACAATGTTTTTAGAAATTTTTGGGGTGG
OT14	51	-3	59	AGTATTTTCTATAACTCATATTTTCACAAAACAATGTTTTAA:::AAATTTTTGGGGTG

Nature Biotechnology: doi:10.1038/nbt.1927

Off-target analysis of the PPP1R12C locus

Insertion and deletions (InDels) observed in loci with InDel frequencies above background. Only InDels that occurred at least 10 times are shown for PPP1R12C and only InDels that occurred at least two times are shown for off target site (OT) 10 and OT14. Inserted bases are shown in red and deleted bases are indicated by the appropriate number of colons. The first column shows the locus, the second column shows the number of occurrences of each InDel, the third column shows the size of the InDel, the fourth column shows either the position of the 5' edge of the deletion or the identity of the inserted bases, and the fifth column shows the aligned sequence read. Many of these sequences can be annotated in multiple ways- only one annotation is shown. The wild-type sequence of each locus is shown above the corresponding InDels and the spacer between the TALEN binding sites is underlined.

Nature Biotechnology: doi:10.1038/nbt.1927

Supplemental Table 1:

PPP1R12C

101077

gt**CCCCTCCACCCCACAGT**ggggccactagggacaggattggtgacagaaaag
caggggaggtggggtgtcaccccggtgatccctg**TCCTAACCACTGTCTTT**tc

101079

PITX3

101146

gt**CAGACGCTGGCACT**ccgcacccccagctcccagagcacggctgcaagggccag
cagtctgcgaccgtgaggcgtgggggtcgagggtctcg**TGCCGACGTTCCCGG**tc

101148

PITX3 (flanking stop codon)

101236

gt**ACGCCGTGGAAAGGCC**cgta**tga**gcggccccgcccgtagatcatccccgag
catgcggcacctttccgggcatactcgccggggcggg**CATCTAGTAGGGGC**tc

101238

OCT4 (intron 1)

101123

gt**CACCTGCAGCTGCCCAG**acctggcacccaggagaggagcaggcagggtcag
cagtggacgtcgacgggtctggaccgtgggtcctc**TCCTCGTCCGTCCCAG**tc

101125

OCT4 (flanking stop codon)

101138

ct**CTGGGCTCTCCCAT**gcattcaaac**tga**ggtgcctgcccttctaggaatgggggac
gagacccgagagggtacgtaagtttgactccacggacg**GGAAGATCCTTACCCCC**tg

101142

Supplementary sequence information. TALEN nucleotide and amino acid sequences (RVDs underlined).

>101077-wt
ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCCCCCAAGAAGAAGAG
GAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGG
TCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCG
GCGCTTGGGACGGTGGCTGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACA
GTGGTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGC
TGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGCGCTCACCGGGGCCCCCTTGAACCTGACC
CCAGACCAGGTAGTCGCAATCGCGTCGCATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA
CCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCACATGACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAG
TTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAGCCATGATGGAGGGAAACAAGCATTGGAGACTGTCCAA
CGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGGCGGTAAGCAGGCGCT
GGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTCACCCCAGACCAGGTAGTCGCAATCGCGTCGAATGGCGGGG
GAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCA
TCACATGACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGT
TGTAGCGATTGCGAGCCATGATGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGA
CGCCTGCACAAGTGGTCGCCATCGCCTCCAATATTGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAG
GATCATGGGCTGACCCCAGACCAGGTAGTCGCAATCGCGTCGCATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCC
GGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCACATGACGGTGGCAAACAGGCTCTTGAGACGGTTC
AGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAGCCATGATGGAGGGAAACAAGCA

TTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGG
CGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGTTTGACCCCAGACCAGGTAGTCGCAATCG
CGTCGAACATTGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAA
GTCGTGGCCATTGCATCACATGACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCT
GACTCCCGATCAAGTTGTAGCGATTGCGAGCAACATCGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTC
AAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTG
CCTGTACTGTGCCAGGATCATGGCCTGACACCCGAACAGGTGGTCGCCATTGCTAGCAACGGGGGAGGACGGCCAGCCTTGGAGTCCAT
CGTAGCCCAATTGTCCAGGCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCG
CGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCCGAGAGAACTTCCCATCGA
GTCGCGGGATCCCAGCTGGTGAAGAGCGAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTACAT
CGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGACCGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACA
GGGGAAAGCACCTGGGCGGAAGCAGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGATTACGGCGTGATCGTGGACACA
AAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAGGCCGACGAGATGCAGAGATACGTGGAGGAGAACCAGACCCGGAATAAGCA
CATCAACCCCAACGAGTGGTGGAAGGTGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGCGGCCACTTCAAGGGCAACT
ACAAGGCCCAGCTGACCAGGCTGAACCACATCACCAACTGCAATGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGCGGCGAGATG
ATCAAAGCCGGCACCCTGACACTGGAGGAGGTGCGGCGCAAGTTCAACAACGGCGAGATCAACTTCAGATCTTGA

>101077-wt
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPA
ALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLT
PDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQ
RLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVAIA
SHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQ
DHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQA
LETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQ
VVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLL
PVLCQDHGLTPEQVVAIASNGGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHR
VAGSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDT
KAYSGGYNLPIGQADEMQRYVEENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEM
IKAGTLTLEEVRRKFNNGEINFRS.

>101079-wt
ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCCCCCAAGAAGAAGAG
GAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGG
TCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCG
GCGCTTGGGACGGTGGCTGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACA
GTGGTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGC
TGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGCGCTCACCGGGGCCCCCTTGAACCTGACC
CCAGACCAGGTAGTCGCAATCGCGTCGAATGGCGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA
CCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCAAACGGAGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAG
TTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAGCAATGGGGGAGGGAAACAAGCATTGGAGACTGTCCAA
CGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGGCGGTAAGCAGGCGCT
GGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTCACCCCAGACCAGGTAGTCGCAATCGCGTCGAATGGCGGGG
GAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGATCAAGTCGTGGCCATTGCA
AATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGT
TGTAGCGATTGCGAGCAATGGGGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGA
CGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAG
GATCATGGGCTGACCCCAGACCAGGTAGTCGCAATCGCGTCGAACATTGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCC
GGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCACATGACGGTGGCAAACAGGCTCTTGAGACGGTTC
AGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAGCCATGATGGAGGGAAACAAGCA
TTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCAATATTGG
CGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGTTTGACCCCAGACCAGGTAGTCGCAATCG
CGTCGAACATTGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAA
GTCGTGGCCATTGCATCAAACGGAGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCT
GACTCCCGATCAAGTTGTAGCGATTGCGAGCCATGATGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTC
AAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTG
CCTGTACTGTGCCAGGATCATGGCCTGACACCCGAACAGGTGGTCGCCATTGCTAGCAACGGGGGAGGACGGCCAGCCTTGGAGTCCAT
CGTAGCCCAATTGTCCAGGCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCG
CGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCCGAGAGAACTTCCCATCGA
GTCGCGGGATCCCAGCTGGTGAAGAGCGAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTACAT
CGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGACCGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACA
GGGGAAAGCACCTGGGCGGAAGCAGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGATTACGGCGTGATCGTGGACACA
AAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAGGCCGACGAGATGCAGAGATACGTGGAGGAGAACCAGACCCGGAATAAGCA
CATCAACCCCAACGAGTGGTGGAAGGTGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGCGGCCACTTCAAGGGCAACT
ACAAGGCCCAGCTGACCAGGCTGAACCACATCACCAACTGCAATGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGCGGCGAGATG
ATCAAAGCCGGCACCCTGACACTGGAGGAGGTGCGGCGCAAGTTCAACAACGGCGAGATCAACTTCAGATCTTGA

>101079-wt
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPA
ALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLT
PDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQ

Nature Biotechnology: doi:10.1038/nbt.1927

RLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIA
NNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQ
DHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQA
LETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQ
VVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLL
PVLCQDHGLTPEQVVAIASNGGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHR
VAGSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDT
KAYSGGYNLPIGQADEMQRYVEENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEM
IKAGTLTLEEVRRKFNNGEINFRS.

>101146-wt
ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCCCCCAAGAAGAAGAG
GAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGG
TCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCG
GCGCTTGGGACGGTGGCTGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACA
GTGGTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGC
TGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGCGCTCACCGGGGCCCCCTTGAACCTGACC
CCAGACCAGGTAGTCGCAATCGCGTCGCATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA
CCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCAAATATCGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAG
TTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAGCATTGGAGACTGTCCAA
CGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCAATATTGGCGGTAAGCAGGCGCT
GGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTCACCCCAGACCAGGTAGTCGCAATCGCGTCGCATGACGGGG
GAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGATCAAGTCGTGGCCATTGCA
AATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGT
TGTAGCGATTGCGAGCCATGATGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGA
CGCCTGCACAAGTGGTCGCCATCGCCTCCAACGGTGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAG
GATCATGGGCTGACCCCAGACCAGGTAGTCGCAATCGCCAACAATAACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCC
GGTCCTTTGTCAAGACCACGGCCTTACACCGGATCAAGTCGTGGCCATTGCAAATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTC
AGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAGCCATGATGGAGGGAAACAAGCA
TTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCAATATTGG
CGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGTTTGACCCCAGACCAGGTAGTCGCAATCG
CGTCGCATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTACTGTGCCAGGATCATGGCCTGACACCCGAACAG
GTGGTCGCCATTGCTAGCAACGGGGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCAATTGTCCAGGCCCGATCCCGCGTTGGCTGC
GTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCG
CATTGATCAAAAGAACCAACCGGCGGATTCCCGAGAGAACTTCCCATCGAGTCGCGGGATCCCAGCTGGTGAAGAGCGAGCTGGAGGAG
AAGAAGTCCGAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGACCG
CATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGAAAGCACCTGGGCGGAAGCAGAAAGCCTGACGGCG
CCATCTATACAGTGGGCAGCCCCATCGATTACGGCGTGATCGTGGACACAAAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAG
GCCGACGAGATGCAGAGATACGTGGAGGAGAACCAGACCCGGAATAAGCACATCAACCCCAACGAGTGGTGGAAGGTGTACCCTAGCAG
CGTGACCGAGTTCAAGTTCCTGTTCGTGAGCGGCCACTTCAAGGGCAACTACAAGGCCCAGCTGACCAGGCTGAACCACATCACCAACT
GCAATGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGCGGCGAGATGATCAAAGCCGGCACCCTGACACTGGAGGAGGTGCGGCGC
AAGTTCAACAACGGCGAGATCAACTTCAGATCTTGA

>101146-wt
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPA
ALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLT
PDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQ
RLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIA
NNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLPVLCQ
DHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQA
LETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQ
VVAIASNGGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGSQLVKSELEE
KKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQ
ADEMQRYVEENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRR
KFNNGEINFRS.

>101148-wt
ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCCCCCAAGAAGAAGAG
GAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGG
TCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCG
GCGCTTGGGACGGTGGCTGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACA
GTGGTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGC
TGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGCGCTCACCGGGGCCCCCTTGAACCTGACC
CCAGACCAGGTAGTCGCAATCGCCAACAATAACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA
CCACGGCCTTACACCGGATCAAGTCGTGGCCATTGCAAATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAG
TTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAGCCATGATGGAGGGAAACAAGCATTGGAGACTGTCCAA
CGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGGCGGTAAGCAGGCGCT
GGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTCACCCCAGACCAGGTAGTCGCAATCGCGTCGCATGACGGGG
GAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCA
TCAAACGGAGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGT
TGTAGCGATTGCGAGCAATGGGGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGA

Nature Biotechnology: doi:10.1038/nbt.1927

CGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAG
GATCATGGGCTGACCCCAGACCAGGTAGTCGCAATCGCGTCGCATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCC
GGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCAAATATCGGTGGCAAACAGGCTCTTGAGACGGTTC
AGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAGCA
TTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGG
CGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGTTTGACCCCAGACCAGGTAGTCGCAATCG
CGTCGCATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACGCCTGCACAA
GTGGTCGCCATCGCCAACAACAACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGCCT
GACACCCGAACAGGTGGTCGCCATTGCTAGCAACGGGGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCAATTGTCCAGGCCCGATC
CCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCGCGCTCGATGCAGTCAAAAAGGGTCTG
CCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCCGAGAGAACTTCCCATCGAGTCGCGGGATCCCAGCTGGTGAAGAG
CGAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCAGGAACA
GCACCCAGGACCGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGAAAGCACCTGGGCGGAAGCAGA
AAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGATTACGGCGTGATCGTGGACACAAAGGCCTACAGCGGCGGCTACAATCT
GCCTATCGGCCAGGCCGACGAGATGCAGAGATACGTGGAGGAGAACCAGACCCGGAATAAGCACATCAACCCCAACGAGTGGTGGAAGG
TGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGCGGCCACTTCAAGGGCAACTACAAGGCCCAGCTGACCAGGCTGAAC
CACATCACCAACTGCAATGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGCGGCGAGATGATCAAAGCCGGCACCCTGACACTGGA
GGAGGTGCGGCGCAAGTTCAACAACGGCGAGATCAACTTCAGATCTTGA

>101148-wt
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPA
ALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLT
PDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQ
RLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIA
SNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQ
DHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQA
LETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPAQ
VVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGL
PHAPALIKRTNRRIPERTSHRVAGSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSR
KPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLN
HITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINFRS.

>101236-wt
ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCCCCCAAGAAGAAGAG
GAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGG
TCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCG
GCGCTTGGGACGGTGGCTGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACA
GTGGTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGC
TGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGCGCTCACCGGGGCCCCCTTGAACCTGACC
CCAGACCAGGTAGTCGCAATCGCGTCGAACATTGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA
CCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCACATGACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAG
TTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAGCATTGGAGACTGTCCAA
CGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGGCGGTAAGCAGGCGCT
GGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTCACCCCAGACCAGGTAGTCGCAATCGCGTCGCATGACGGGG
GAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGATCAAGTCGTGGCCATTGCA
AATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGT
TGTAGCGATTGCGAGCAATGGGGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGA
CGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAG
GATCATGGGCTGACCCCAGACCAGGTAGTCGCAATCGCCAACAATAACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCC
GGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCAAATATCGGTGGCAAACAGGCTCTTGAGACGGTTC
AGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACGCCTGCACAAGTGGTCGCCATCGCCTCCAATATTGGCGGTAAGCAGGCG
CTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGTTTGACCCCAGACCAGGTAGTCGCAATCGCGTCGAACATTGG
GGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGATCAAGTCGTGGCCATTG
CAAATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTTTCTGTCAAGCCCACGGGCTGACTCCCGATCAA
GTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTT
GACGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCC
AGGATCATGGCCTGACACCCGAACAGGTGGTCGCCATTGCTAGCCACGATGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCAATTG
TCCAGGCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCGCGCTCGATGCAGT
CAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCCGAGAGAACTTCCCATCGAGTCGCGGGATCCC
AGCTGGTGAAGAGCGAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTACATCGAGCTGATCGAG
ATCGCCAGGAACAGCACCCAGGACCGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGAAAGCACCT
GGGCGGAAGCAGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGATTACGGCGTGATCGTGGACACAAAGGCCTACAGCG
GCGGCTACAATCTGCCTATCGGCCAGGCCGACGAGATGCAGAGATACGTGGAGGAGAACCAGACCCGGAATAAGCACATCAACCCCAAC
GAGTGGTGGAAGGTGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGCGGCCACTTCAAGGGCAACTACAAGGCCCAGCT
GACCAGGCTGAACCACATCACCAACTGCAATGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGCGGCGAGATGATCAAAGCCGGCA
CCCTGACACTGGAGGAGGTGCGGCGCAAGTTCAACAACGGCGAGATCAACTTCAGATCTTGA

>101236-wt
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPA
ALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLT
PDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQ

RLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIA
NNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQ
DHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQA
LETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVFCQAHGLTPDQ
VVAIANNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGRPALESIVAQL
SRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGSQLVKSELEEKKSELRHKLKYVPHEYIELIE
IARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRNKHINPN
EWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINFRS.

>101238-wt
ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCCCCCAAGAAGAAGAG
GAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGG
TCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCG
GCGCTTGGGACGGTGGCTGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACA
GTGGTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGC
TGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGCGCTCACCGGGGCCCCCTTGAACCTGACC
CCAGACCAGGTAGTCGCAATCGCGTCGCATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA
CCACGGCCTTACACCGGATCAAGTCGTGGCCATTGCAAATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAG
TTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAGCATTGGAGACTGTCCAA
CGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGCAGGCGCT
GGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTCACCCCAGACCAGGTAGTCGCAATCGCCAACAATAACGGGG
GAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCA
TCAAATATCGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGT
TGTAGCGATTGCGAGCAATGGGGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGA
CGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAG
GATCATGGGCTGACCCCAGACCAGGTAGTCGCAATCGCGTCGAACATTGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCC
GGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCAAACGGAGGTGGCAAACAGGCTCTTGAGACGGTTC
AGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAGCCATGATGGAGGGAAACAAGCA
TTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCAACGGTGG
CGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGTTTGACCCCAGACCAGGTAGTCGCAATCG
CGTCGAACATTGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTACTGTGCCAGGATCATGGCCTGACACCCGAACAG
GTGGTCGCCATTGCTAGCCACGATGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCAATTGTCCAGGCCCGATCCCGCGTTGGCTGC
GTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCG
CATTGATCAAAAGAACCAACCGGCGGATTCCCGAGAGAACTTCCCATCGAGTCGCGGGATCCCAGCTGGTGAAGAGCGAGCTGGAGGAG
AAGAAGTCCGAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGACCG
CATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGAAAGCACCTGGGCGGAAGCAGAAAGCCTGACGGCG
CCATCTATACAGTGGGCAGCCCCATCGATTACGGCGTGATCGTGGACACAAAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAG
GCCGACGAGATGCAGAGATACGTGGAGGAGAACCAGACCCGGAATAAGCACATCAACCCCAACGAGTGGTGGAAGGTGTACCCTAGCAG
CGTGACCGAGTTCAAGTTCCTGTTCGTGAGCGGCCACTTCAAGGGCAACTACAAGGCCCAGCTGACCAGGCTGAACCACATCACCAACT
GCAATGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGCGGCGAGATGATCAAAGCCGGCACCCTGACACTGGAGGAGGTGCGGCGC
AAGTTCAACAACGGCGAGATCAACTTCAGATCTTGA

>101238-wt
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPA
ALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLT
PDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQ
RLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIA
SNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQ
DHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQA
LETVQRLLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQ
VVAIASHDGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGSQLVKSELEE
KKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQ
ADEMQRYVEENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRR
KFNNGEINFRS.

>101123-wt
ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCCCCCAAGAAGAAGAG
GAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGG
TCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCG
GCGCTTGGGACGGTGGCTGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACA
GTGGTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGC
TGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGCGCTCACCGGGGCCCCCTTGAACCTGACC
CCAGACCAGGTAGTCGCAATCGCGTCGCATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA
CCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCAAATATCGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAG
TTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAGCCATGATGGAGGGAAACAAGCATTGGAGACTGTCCAA
CGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGGCGGTAAGCAGGCGCT
GGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTCACCGCAGACCAGGTAGTCGCAATCGCGTCGAATGGCGGGG
GAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGATCAAGTCGTGGCCATTGCA
AATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGT
TGTAGCGATTGCGAGCCATGATGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGA

CGCCTGCACAAGTGGTCGCCATCGCCTCCAATATTGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAG
GATCATGGGCTGACCCCAGACCAGGTAGTCGCAATCGCCAACAATAACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCC
GGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCACATGACGGTGGCAAACAGGCTCTTGAGACGGTTC
AGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAGCAATGGGGGAGGGAAACAAGCA
TTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGG
CGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGTTTGACCCCAGACCAGGTAGTCGCAATCG
CGTCGCATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAA
GTCGTGGCCATTGCATCACATGACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTTTCTGTCAAGCCCACGGGCT
GACTCCCGATCAAGTTGTAGCGATTGCGAGCCATGATGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTC
AAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCAATATTGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTG
CCTGTACTGTGCCAGGATCATGGCCTGACACCCGAACAGGTGGTCGCCATTGCTAGCAATAAAGGAGGACGGCCAGCCTTGGAGTCCAT
CGTAGCCCAATTGTCCAGGCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCG
CGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCCGAGAGAACTTCCCATCGA
GTCGCGGGATCCCAGCTGGTGAAGAGCGAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTACAT
CGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGACCGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACA
GGGGAAAGCACCTGGGCGGAAGCAGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGATTACGGCGTGATCGTGGACACA
AAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAGGCCGACGAGATGCAGAGATACGTGGAGGAGAACCAGACCCGGAATAAGCA
CATCAACCCCAACGAGTGGTGGAAGGTGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGCGGCCACTTCAAGGGCAACT
ACAAGGCCCAGCTGACCAGGCTGAACCACATCACCAACTGCAATGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGCGGCGAGATG
ATCAAAGCCGGCACCCTGACACTGGAGGAGGTGCGGCGCAAGTTCAACAACGGCGAGATCAACTTCAGATCTTGA

>101123-wt
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPA
ALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLT
PDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQ
RLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTADQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIA
NNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQ
DHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQA
LETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQ
VVAIASHDGGKQALETVQRLLPVFCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLL
PVLCQDHGLTPEQVVAIASNKGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHR
VAGSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDT
KAYSGGYNLPIGQADEMQRYVEENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEM
IKAGTLTLEEVRRKFNNGEINFRS.

>101125-wt
ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCCCCCAAGAAGAAGAG
GAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGG
TCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCG
GCGCTTGGGACGGTGGCTGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACA
GTGGTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGC
TGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGCGCTCACCGGGGCCCCCTTGAACCTGACC
CCAGACCAGGTAGTCGCAATCGCCAACAATAACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA
CCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCAAATATCGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAG
TTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAGCCATGATGGAGGGAAACAAGCATTGGAGACTGTCCAA
CGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGGCGGTAAGCAGGCGCT
GGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTCACCCCTGACCAGGTAGTCGCAATCGCGTCGCATGACGGGG
GAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCA
TCAAACGGAGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGT
TGTAGCGATTGCGAATAACAATGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGA
CGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAG
GATCATGGGCTGACCCCAGACCAGGTAGTCGCAATCGCGTCGCATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCC
GGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCAAACGGAGGTGGCAAACAGGCTCTTGAGACGGTTC
AGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAGCA
TTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGG
CGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGTTTGACCCCAGACCAGGTAGTCGCAATCG
CGTCGAATGGCGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAA
GTCGTGGCCATTGCATCACATGACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCT
GACGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCC
AGGATCATGGCCTGACACCCGAACAGGTGGTCGCCATTGCTAGCAACGGGGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCAATTG
TCCAGGCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCGCGCTCGATGCAGT
CAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCCGAGAGAACTTCCCATCGAGTCGCGGGATCCC
AGCTGGTGAAGAGCGAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTACATCGAGCTGATCGAG
ATCGCCAGGAACAGCACCCAGGACCGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGAAAGCACCT
GGGCGGAAGCAGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGATTACGGCGTGATCGTGGACACAAAGGCCTACAGCG
GCGGCTACAATCTGCCTATCGGCCAGGCCGACGAGATGCAGAGATACGTGGAGGAGAACCAGACCCGGAATAAGCACATCAACCCCAAC
GAGTGGTGGAAGGTGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGCGGCCACTTCAAGGGCAACTACAAGGCCCAGCT
GACCAGGCTGAACCACATCACCAACTGCAATGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGCGGCGAGATGATCAAAGCCGGCA
CCCTGACACTGGAGGAGGTGCGGCGCAAGTTCAACAACGGCGAGATCAACTTCAGATCTTGA

>101125-wt
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPA
ALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLT
PDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQ
RLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIA
SNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQ
DHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQA
LETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQ
VVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGRPALESIVAQL
SRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGSQLVKSELEEKKSELRHKLKYVPHEYIELIE
IARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRNKHINPN
EWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINFRS.

>101138-wt
ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCCCCCAAGAAGAAGAG
GAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGG
TCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCG
GCGCTTGGGACGGTGGCTGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACA
GTGGTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGC
TGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGCGCTCACCGGGGCCCCCTTGAACCTGACC
CCAGACCAGGTAGTCGCAATCGCGTCGCATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA
CCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCAAACGGAGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAG
TTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAGCATTGGAGACTGTCCAA
CGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGCAGGCGCT
GGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTCACCCCAGACCAGGTAGTCGCAATCGCCAACAATAACGGGG
GAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCA
TCACATGACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGT
TGTAGCGATTGCGAGCAATGGGGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGA
CGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAG
GATCATGGGCTGACCCCAGACCAGGTAGTCGCAATCGCGTCGAATGGCGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCC
GGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCACATGACGGTGGCAAACAGGCTCTTGAGACGGTTC
AGAGACTTCTCCCAGTTTTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAGCCATGATGGAGGGAAACAAGCA
TTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGG
CGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGTTTGACCCCAGACCAGGTAGTCGCAATCG
CGTCGAACATTGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTACTGTGCCAGGATCATGGCCTGACACCCGAACAG
GTGGTCGCCATTGCTAGCAACGGGGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCAATTGTCCAGGCCCGATCCCGCGTTGGCTGC
GTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCG
CATTGATCAAAAGAACCAACCGGCGGATTCCCGAGAGAACTTCCCATCGAGTCGCGGGATCCCAGCTGGTGAAGAGCGAGCTGGAGGAG
AAGAAGTCCGAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGACCG
CATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGAAAGCACCTGGGCGGAAGCAGAAAGCCTGACGGCG
CCATCTATACAGTGGGCAGCCCCATCGATTACGGCGTGATCGTGGACACAAAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAG
GCCGACGAGATGCAGAGATACGTGGAGGAGAACCAGACCCGGAATAAGCACATCAACCCCAACGAGTGGTGGAAGGTGTACCCTAGCAG
CGTGACCGAGTTCAAGTTCCTGTTCGTGAGCGGCCACTTCAAGGGCAACTACAAGGCCCAGCTGACCAGGCTGAACCACATCACCAACT
GCAATGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGCGGCGAGATGATCAAAGCCGGCACCCTGACACTGGAGGAGGTGCGGCGC
AAGTTCAACAACGGCGAGATCAACTTCAGATCTTGA

>101138-wt
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPA
ALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLT
PDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQ
RLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIA
SHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQ
DHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVFCQAHGLTPDQVVAIASHDGGKQA
LETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQ
VVAIASNGGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGSQLVKSELEE
KKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQ
ADEMQRYVEENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRR
KFNNGEINFRS.

>101142-wt
ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCCCCCAAGAAGAAGAG
GAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGG
TCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCG
GCGCTTGGGACGGTGGCTGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACA
GTGGTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGC
TGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGCGCTCACCGGGGCCCCCTTGAACCTGACC
CCAGACCAGGTAGTCGCAATCGCGTCGCATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA
CCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCACATGACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAG
TTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAGCCATGATGGAGGGAAACAAGCATTGGAGACTGTCCAA
CGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGGCGGTAAGCAGGCGCT
GGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTCACCCCAGACCAGGTAGTCGCAATCGCGTCGCATGACGGGG

GAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCA
TCAAATATCGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGATTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGT
TGTAGCGATTGCGAGCAATGGGGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGA
CGCCTGCACAAGTGGTCGCCATCGCCTCCAACGGTGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAG
GACCATGGGCTGACCCCAGACCAGGTAGTCGCAATCGCGTCGCATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCC
GGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCACATGACGGTGGCAAACAGGCTCTTGAGACGGTTC
AGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAGCAATGGGGGAGGGAAACAAGCA
TTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCAATATTGG
CGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGTTTGACCCCAGACCAGGTAGTCGCAATCG
CCAACAATAACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAA
GTCGTGGCCATTGCATCAAATATCGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCT
GACTCCCGATCAAGTTGTAGCGATTGCGAGCAACATCGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTC
AAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTG
CCTGTACTGTGCCAGGATCATGGCCTGACACCCGAACAGGTGGTCGCCATTGCTAGCAATAAAGGAGGACGGCCAGCCTTGGAGTCCAT
CGTAGCCCAATTGTCCAGGCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCG
CGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCCGAGAGAACTTCCCATCGA
GTCGCGGGATCCCAGCTGGTGAAGAGCGAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTACAT
CGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGACCGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACA
GGGGAAAGCACCTGGGCGGAAGCAGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGATTACGGCGTGATCGTGGACACA
AAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAGGCCGACGAGATGCAGAGATACGTGGAGGAGAACCAGACCCGGAATAAGCA
CATCAACCCCAACGAGTGGTGGAAGGTGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGCGGCCACTTCAAGGGCAACT
ACAAGGCCCAGCTGACCAGGCTGAACCACATCACCAACTGCAATGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGCGGCGAGATG
ATCAAAGCCGGCACCCTGACACTGGAGGAGGTGCGGCGCAAGTTCAACAACGGCGAGATCAACTTCAGATCTTGA

>101142-wt
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPA
ALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLT
PDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQ
RLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIA
SNIGGKQALETVQRFLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLPVLCQ
DHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQA
LETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQ
VVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLL
PVLCQDHGLTPEQVVAIASNKGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHR
VAGSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDT
KAYSGGYNLPIGQADEMQRYVEENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEM
IKAGTLTLEEVRRKFNNGEINFRS.

>101077-ELD
ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCCCCCAAGAAGAAGAG
GAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGG
TCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCG
GCGCTTGGGACGGTGGCTGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACA
GTGGTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGC
TGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGCGCTCACCGGGGCCCCCTTGAACCTGACC
CCAGACCAGGTAGTCGCAATCGCGTCGCATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA
CCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCACATGACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAG
TTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAGCCATGATGGAGGGAAACAAGCATTGGAGACTGTCCAA
CGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGGCGGTAAGCAGGCGCT
GGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTCACCCCAGACCAGGTAGTCGCAATCGCGTCGAATGGCGGGG
GAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCA
TCACATGACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGT
TGTAGCGATTGCGAGCCATGATGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGA
CGCCTGCACAAGTGGTCGCCATCGCCTCCAATATTGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAG
GATCATGGGCTGACCCCAGACCAGGTAGTCGCAATCGCGTCGCATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCC
GGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCACATGACGGTGGCAAACAGGCTCTTGAGACGGTTC
AGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAGCCATGATGGAGGGAAACAAGCA
TTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGG
CGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGTTTGACCCCAGACCAGGTAGTCGCAATCG
CGTCGAACATTGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAA
GTCGTGGCCATTGCATCACATGACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCT
GACTCCCGATCAAGTTGTAGCGATTGCGAGCAACATCGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTC
AAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTG
CCTGTACTGTGCCAGGATCATGGCCTGACACCCGAACAGGTGGTCGCCATTGCTAGCAACGGGGGAGGACGGCCAGCCTTGGAGTCCAT
CGTAGCCCAATTGTCCAGGCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCG
CGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCCGAGAGAACTTCCCATCGA
GTCGCGGGATCCCAGCTGGTGAAGAGCGAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTACAT
CGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGACCGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACA
GGGGAAAGCACCTGGGCGGAAGCAGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGATTACGGCGTGATCGTGGACACA
AAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAGGCCGACGAGATGGAGAGATACGTGGAGGAGAACCAGACCCGGGATAAGCA
CCTCAACCCCAACGAGTGGTGGAAGGTGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGCGGCCACTTCAAGGGCAACT

Nature Biotechnology: doi:10.1038/nbt.1927

ACAAGGCCCAGCTGACCAGGCTGAACCACATCACCAACTGCAATGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGCGGCGAGATG
ATCAAAGCCGGCACCCTGACACTGGAGGAGGTGCGGCGCAAGTTCAACAACGGCGAGATCAACTTCAGATCTTGA

>101077-ELD
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPA
ALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLT
PDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQ
RLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVAIA
SHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQ
DHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQA
LETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQ
VVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLL
PVLCQDHGLTPEQVVAIASNGGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHR
VAGSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDT
KAYSGGYNLPIGQADEMERYVEENQTRDKHLNPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEM
IKAGTLTLEEVRRKFNNGEINFRS.

>101079-KKR
ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCCCCCAAGAAGAAGAG
GAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCGCAACAGCAACAGGAGAAAATCAAGCCTAAGG
TCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCG
GCGCTTGGGACGGTGGCTGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACA
GTGGTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGC
TGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGCGCTCACCGGGGCCCCCTTGAACCTGACC
CCAGACCAGGTAGTCGCAATCGCGTCGAATGGCGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA
CCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCAAACGGAGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAG
TTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAGCAATGGGGGAGGGAAACAAGCATTGGAGACTGTCCAA
CGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGGCGGTAAGCAGGCGCT
GGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTCACCCCAGACCAGGTAGTCGCAATCGCGTCGAATGGCGGGG
GAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGATCAAGTCGTGGCCATTGCA
AATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGT
TGTAGCGATTGCGAGCAATGGGGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGA
CGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAG
GATCATGGGCTGACCCCAGACCAGGTAGTCGCAATCGCGTCGAACATTGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCC
GGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCACATGACGGTGGCAAACAGGCTCTTGAGACGGTTC
AGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAGCCATGATGGAGGGAAACAAGCA
TTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCAATATTGG
CGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGTTTGACCCCAGACCAGGTAGTCGCAATCG
CGTCGAACATTGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAA
GTCGTGGCCATTGCATCAAACGGAGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAGTTCTCTGTCAAGCCCACGGGCT
GACTCCCGATCAAGTTGTAGCGATTGCGAGCCATGATGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTC
AAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCCACGACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTG
CCTGTACTGTGCCAGGATCATGGCCTGACACCCGAACAGGTGGTCGCCATTGCTAGCAACGGGGGAGGACGGCCAGCCTTGGAGTCCAT
CGTAGCCCAATTGTCCAGGCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCG
CGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCCGAGAGAACTTCCCATCGA
GTCGCGGGATCCCAGCTGGTGAAGAGCGAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTGAAGTACGTGCCCCACGAGTACAT
CGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGACCGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACA
GGGGAAAGCACCTGGGCGGAAGCAGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGATTACGGCGTGATCGTGGACACA
AAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAGGCCGACGAGATGCAGAGATACGTGAAGGAGAACCAGACCCGGAATAAGCA
CATCAACCCCAACGAGTGGTGGAAGGTGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGCGGCCACTTCAAGGGCAACT
ACAAGGCCCAGCTGACCAGGCTGAACCGCAAAACCAACTGCAATGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGCGGCGAGATG
ATCAAAGCCGGCACCCTGACACTGGAGGAGGTGCGGCGCAAGTTCAACAACGGCGAGATCAACTTCTGA

>101079-KKR
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPA
ALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLT
PDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQ
RLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIA
NNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQ
DHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQA
LETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQ
VVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLL
PVLCQDHGLTPEQVVAIASNGGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHR
VAGSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDT
KAYSGGYNLPIGQADEMQRYVKENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNRKTNCNGAVLSVEELLIGGEM
IKAGTLTLEEVRRKFNNGEINF.

Binding sites of TALENs and TALEN nucleotide and amino acid sequences

Nature Biotechnology: doi:10.1038/nbt.1927

Supplemental Table 2:

Target	TALEN A	TALEN B	Activity	
PPP1R12C intron 1	GGGTACTTTTATCTGT	CCCTAGTGGCCCCACT	17%	
	CCCCTCCACCCCACAGT	TTTCTGTCACCAATCCT	36%	x
	GGTTCTGGGTACTTTT	CCCTAGTGGCCCCACT	7%	
	GGGTACTTTTATCTGT	TTTCTGTCACCAATCCT	10%	
OCT4 intron 1	GGGGAAGTGGTCACCt	GCTCCTCTCCTGGGT	2%	
	CACCTGCAGCTGCCCAG	GACCCTGCCTGCTCCT	9%	x
	GACCCTGCCTGCT	GACCCTGCCTGCTCCT	4%	
	GACCCTGCCTGCT	GCAGCTGCCCAGACCT	8%	
PITX3 exon 1	TGCAGCCGTGCTCT	GTCGCTGTCAGACGCT	16%	
	TGCAGCCGTGCTCT	CAGACGCTGGCACT	8%	
	GGCCCTTGCAGCCGT	GTCGCTGTCAGACGCT	15%	
	GGCCCTTGCAGCCGT	CAGACGCTGGCACT	21%	x

TALENs assembled according to the cipher previously described[4] and recognizing the DNA stretches indicated were tested for genome editing activity at the endogenous locus (indicated) using Surveyor endonuclease exactly as previously described [6]. The gene disruption activity is indicated; the pairs used for gene addition are indicated with an "x" in the last column.[4]

Nature Biotechnology: doi:10.1038/nbt.1927

Supplemental Table 3:

Site	Score	Chr	Location	Gene	Sequence	Site configuration TALEN-L / gap / TALEN-R	For Seq Primer	Rev Seq Primer
PPP1R12C	8.50E-06	19	55627106	PPP1R12C	CTtTTCTGTCACCAATCCTGTCCCTAGTGGCCCCACTGTGGGGTGGAGGGGAC	PPP1R12C-R / 15 / PPP1R12C-L	ctctggttctgggtacttttatctg	gtgtgtcaccagataaggaatctg
OTS 1	3.56E-11	1	206211317		TgGgTAaGTCACtCAaCCTCTCTGAATCATTTCAATTTACTCACCAtTGTGGGGTGGgGGGGAA	PPP1R12C-R / 26 / PPP1R12C-L	ccctgtattattttccttccc	agtacccaggagcagcaaga
OTS 2	2.75E-12	X	31479737	DMD	CTcTTCcaTCACtAATtCTGATGAACAGCTATCAAGCTATTCAtgGTaGGGTGGAGGGTAG	PPP1R12C-R / 23 / PPP1R12C-L	tttcattcattaactcttctcttcc	ccactcactcttcatattcaact
OTS 3	1.09E-12	1	24697472	C1orf201	TTCCCCaCCACaCCACAaTGCCCTGTCTCAAGaAactGTGACAGAATAA	PPP1R12C-L / 11 / PPP1R12C-R	ccttaaagctttgtacacaaattca	tgagccatcatgcccag
OTS 4	1.04E-12	18	60653602		GTATTCTGTCACtAtTCCTGGGACATTTCCTGAAGGACgaTGGGGcGtgGGGGgA	PPP1R12C-R / 17 / PPP1R12C-L	tccatccattgtcctttctc	gtcctccagttttcaaatttaa
OTS 5	9.91E-13	10	101836508	CPN1	CTCCCtcCCACCCCACctaCCATGAAGAATAGGATTGGgGgCAGgACtG	PPP1R12C-L / 11 / PPP1R12C-R	gctggcaggggttttcc	ttggcctgcttagatttgaa
OTS 6	7.02E-13	6	167286954		CccgTCaGTCACCCCcTCCTCTTGGGAGTGGGGCACaGTGGGGTGGAGtGGgT	PPP1R12C-R / 14 / PPP1R12C-L	cctggccagactcatgct	atctctggatcgtgtgcatt
OTS 7	6.43E-13	3	124702671	HEG1	ATCCCtTCCACCtCACcaaATAGTCCTCGCTAGCAGAAGGgTTGGgtACAGAACAA	PPP1R12C-L / 18 / PPP1R12C-R	cctgatgttttacacaactctatcc	gtcattttactcggtcttgactg
OTS 8	5.57E-13	9	135721616	AK8	GTATcCaGaCACCCAcCCTGGGCCTTCTGCAATGGAtgTGTGGGGTGGAtGGGgC	PPP1R12C-R / 17 / PPP1R12C-L	catctgtgggcccctct	tcaaatgttttgaggttggaaa
OTS 9	5.48E-13	22	51139502	SHANK3	GctCCCccCCACCCCcCAaaAAAATCCATTTCCGGGACtGGAgTGGgGACAGAAaAG	PPP1R12C-L / 18 / PPP1R12C-R	cccctccctaacctttaagtc	tctgccagaggccagagt
OTS 10	3.93E-13	15	67305689		CTtCCacCCAgCCCtCAGTGCTCTGCTGAGGCAGGATGGGTGgCtGgAaAA	PPP1R12C-L / 13 / PPP1R12C-R	gaatggatgaatgagtgaatgt	tcctgagttctcggttctttg
OTS 11	3.78E-13	4	148942809	ARHGAP10	GTtCCCaCCACCCCACAacCTTCTTACATCTGAACAAtGAcaGaTGACAGtAaAT	PPP1R12C-L / 17 / PPP1R12C-R	cctggactattatccacagctg	ccagtggagtcatatttcctg
OTS 12	3.10E-13	3	70995950		ATgCCCcCCACCCCAgAGTATATTCATTAGGACTCTTTGAttATaaGaGACAGAAaAT	PPP1R12C-L / 20 / PPP1R12C-R	cgtggttgcctttgtgaaat	tgcctggattttcaattatgtg
OTS 13	2.87E-13	6	45952064	CLIC5	CaGTaCTGTCACCtgTgCTATCCCTGGACTCACCAtgtTGGGGTGGAaGGGAC	PPP1R12C-R / 15 / PPP1R12C-L	tgctgtcttcaggtccgc	agctgaagctcccaaagtca
OTS 14	2.73E-13	5	165831361		ATtTTCTaTaACtCATatTTTCACAAAACAATGTTTTTAGAAAttTtTGGGGTGGAGGGGgT	PPP1R12C-R / 16 / PPP1R12C-L	aaaaattctgtttctaagaagt	ttcatccatcctatgaaaactca
OTS 15	2.42E-13	12	120531852	CCDC64	GTGgTCaGcCACCCcaCCTTGAGGGCAGCACAGGCACcacGGGGTGGAGGGGAG	PPP1R12C-L / 16 / PPP1R12C-R	aatggcaccagaggggtct	cagccactgcaggcttct
OTS 16	2.42E-13	2	74063519	STAMBP	CTCCCCTCCACCCCgtgGTGCCCGCGCTGCCCTCAAGGtgGGGTGgCtGAcCAC	PPP1R12C-L / 16 / PPP1R12C-R	aatggcaccagaggggtct	agctgttgcagggctctc
OTS 17	2.22E-13	3	110893534		CTGTTCcaTCgCCAcTgCTGCTCAAACCACTGAGGttTGgGGGGTGGgGGGTgG	PPP1R12C-R / 16 / PPP1R12C-L	cctgtctgtctgcatgctc	gtgtgctgggtactgaaacg
OTS 18	2.10E-13	12	12838163	GPR19	CcATcCcaTCACCCATCCTCACTCTTTAAATATGGTCTTTTTTTttTttTGGTGGgGGGGAG	PPP1R12C-R / 25 / PPP1R12C-L	catacctacctccctgtccc	ctgtggttttccccataagc
OTS 19	2.02E-13	19	6213603	MLLT1	AaGgaCTGTCcCtgcTCCTGCCCAGGAAGAGTCCAACTGTGGGGTGttGGGGgG	PPP1R12C-R / 16 / PPP1R12C-L	aggccacaccgggag	acaaggcggggaggctac
OTS 20	1.86E-13	9	141119479	FAM157B	AcACCCcCCACCCCcCAcaATTTGCTGTTAAAAGtGGAgTGtTGAtAGtATAG	PPP1R12C-L / 15 / PPP1R12C-R	ggaatattggttggtatggga	cacgaacagtggactctatttt

Location of SELEX-predicted PPP1R12C TALEN off-target sites and primer sequences for off-target site analysis

Nature Biotechnology: doi:10.1038/nbt.1927

Supplemental Table 4:

	TALENs			Negative Control		
Locus	total	InDels	%InDels	total	InDels	%InDels
PPP1R12C	2E+06	27409	1.8226	1464358	2	0.0001
OT1	2E+06	1	0.0001	1272038	2	0.0002
OT2	3E+06	4	0.0001	3228784	4	0.0001
OT3	5E+06	7	0.0002	3872171	8	0.0002
OT4	4E+06	3	0.0001	3948044	17	0.0004
OT5	3E+06	6	0.0002	2465847	3	0.0001
OT6	3E+06	6	0.0002	3092092	5	0.0002
OT7	2E+06	3	0.0001	2803754	4	0.0001
OT8	2E+06	0	0.0000	1424981	4	0.0003
OT9	2E+06	2	0.0001	1362994	3	0.0002
OT10	3E+06	366	0.0108	2871927	1	0.0000
OT11	3E+06	0	0.0000	4632791	17	0.0004
OT12	3E+06	3	0.0001	2963102	7	0.0002
OT13	3E+06	6	0.0002	4524613	6	0.0001
OT14	3E+06	53	0.0016	2597184	7	0.0003
OT15	3E+06	3	0.0001	3218486	5	0.0002
OT16	3E+06	7	0.0002	3020878	11	0.0004
OT17	2E+06	1	0.0001	1649244	0	0.0000
OT18	ND	ND	ND	ND	ND	ND
OT19	4E+06	3	0.0001	1862693	4	0.0002
OT20	4E+06	6	0.0002	3101727	7	0.0002

Summary of off-target analysis

The first column lists the predicted off-targets loci (OT) as described in Supplemental Table 3. The second column lists the total number of sequencing reads (total), the number of insertions and deletions (InDels) and their fraction to the total number of reads (in %) after electroporation with PPP1R12C TALENs. The third column shows the same values for eGFP control hESCs.

Nature Biotechnology: doi:10.1038/nbt.1927

Additional References

1. Soldner, F. et al. Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. *Cell* 136, 964-977 (2009).
2. Hockemeyer, D. et al. A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. *Cell Stem Cell* 3, 346-353 (2008).
3. Lengner, C.J. et al. Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations. *Cell* 141, 872-883.
4. Miller, J.C. et al. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* 29, 143-148 (2011).
5. Hockemeyer, D. et al. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol* 27, 851-857 (2009).
6. Perez, 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).
7. Doyon, Y. et al. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat Methods* 8, 74-79.