nature biotechnology

Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription

Feng Zhang[1–3,5,7,8], Le Cong[2–4,8], Simona Lodato[5,6], Sriram Kosuri[2,3], George M Church[2,3] & Paola Arlotta[5,6]

The ability to direct functional proteins to specific DNA sequences is a long-sought goal in the study and engineering of biological processes. Transcription activator–like effectors (TALEs) from *Xanthomonas sp.* are site-specific DNA-binding proteins that can be readily designed to target new sequences. Because TALEs contain a large number of repeat domains, it can be difficult to synthesize new variants. Here we describe a method that overcomes this problem. We leverage codon degeneracy and type IIs restriction enzymes to generate orthogonal ligation linkers between individual repeat monomers, thus allowing full-length, customized, repeat domains to be constructed by hierarchical ligation. We synthesized 17 TALEs that are customized to recognize specific DNA-binding sites, and demonstrate that they can specifically modulate transcription of endogenous genes (*SOX2* and *KLF4*) in human cells.

Systematic interrogation and engineering of biological systems in normal and pathological states depend on the ability to manipulate the genome of target cells with efficiency and precision[1,2]. Some naturally occurring DNA-binding proteins have been engineered to enable sequence-specific DNA perturbation, including polydactyl zinc fingers[3–5] and meganucleases[6,7]. In particular, engineered zinc fingers can be attached to a wide variety of effector domains such as nucleases, transcription effectors and epigenetic modifying enzymes to carry out site-specific modifications near their DNA-binding sites. However, owing to the lack of a simple correspondence between amino acid sequence and DNA recognition, design and development of sequence-specific DNA-binding proteins based on zinc fingers and meganucleases remain difficult and expensive, often involving elaborate screening procedures and long development time on the order of several weeks. Here we describe methods for constructing alternative DNA-targeting protein domains based on the naturally occurring TALEs from the plant pathogen *Xanthomonas sp.*[8–11].

TALEs are natural effector proteins secreted by numerous species of *Xanthomonas* to modulate gene expression in host plants and to facilitate bacterial colonization and survival[9,11]. Recent studies of TALEs have revealed an elegant code linking the repetitive region of TALEs with their target DNA-binding site[8,10]. Common among the entire family of TALEs is a highly conserved and repetitive region within the middle of the protein, consisting of tandem repeats of mostly 33 or 34 amino acid segments (**Fig. 1a**). Repeat monomers differ from each other mainly in amino acid positions 12 and 13 (repeat variable di-residues), and recent computational and functional analyses[8,10] have revealed a strong correlation between unique pairs of amino acids at positions 12 and 13 and the corresponding nucleotide in the TALE-binding site (e.g., NI to A, HD to C, NG to T, and NN to G or A; **Fig. 1a**). The existence of this strong association suggests a potentially designable protein with sequence-specific DNA-binding capabilities, and the possibility of applying engineered TALEs to specify DNA binding in mammalian cells. However, our ability to test the modularity of the TALE DNA-binding code remains limited owing to the difficulty in constructing customized TALEs with specific arrangements of tandem repeat monomers. Early studies have tested the DNA-binding properties of TALEs[8,12–17], including two studies that tested artificial TALEs with tailored repeat regions[8,12,17,18].

A prerequisite for exploring the modularity of TALE repeat monomers is the ability to synthesize TALEs with customized repetitive DNA-binding domains. Although this has been recently shown to be possible[12,18,19], the repetitive nature of the TALE DNA-binding domains renders routine construction of novel TALEs difficult when using PCR-based gene assembly or serial DNA ligation, and such construction may not be amenable to high-throughput TALE synthesis. Furthermore, even though commercial services can be employed for the synthesis of novel TALE-binding domains[12], they present a cost-prohibitive option for large-scale TALE construction and testing. Hence a more robust protocol to construct large numbers of TALEs would enable ready perturbation of any genome target in many organisms.

To enable high-throughput construction of TALEs, we developed a hierarchical ligation–based strategy to overcome the difficulty of constructing TALE tandem repeat domains (**Fig. 1b** and **Supplementary Methods**). To reduce the amount of repetitive sequence in TALEs, thereby facilitating amplification using PCR, we first optimized the DNA sequence of the four repeat monomers (NI, HD, NN and NG) to minimize repetitiveness while preserving the amino acid sequence. To assemble the individual monomers in a specific order, we altered the DNA sequence at the junction between each pair of monomers, similar to

[1]Society of Fellows, Harvard University, Cambridge, Massachusetts, USA. [2]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. [3]Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, Massachusetts, USA. [4]Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, Massachusetts, USA. [5]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts, USA. [6]Center for Regenerative Medicine and Department of Neurosurgery, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA. [7]Present address: Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; McGovern Institute for Brain Research, MIT, Cambridge, Massachusetts, USA; and Department of Brain and Cognitive Sciences, MIT, Cambridge, Massachusetts, USA. [8]These authors contributed equally to this work. Correspondence should be addressed F.Z. (zhang_f@mit.edu) or P.A. (paola_arlotta@hms.harvard.edu).

Received 18 November 2010; accepted 12 January 2011; published online 19 January 2011; doi:10.1038/nbt1775

the Golden Gate cloning strategy for multipiece DNA ligation[20,21]. Using different codons to represent the junction between each pair of monomers (Gly-Leu), we designed unique 4-bp sticky-end ligation adapters for each junction (**Supplementary Methods**). Using this strategy, four monomers can be ligated simultaneously to form 4-mer tandem repeats. Three 4-mer repeats can be simultaneously ligated to form the desired 12-mer tandem repeat and subsequently ligated into a backbone vector containing a half-length repeat monomer specifying the 13th nucleotide of the binding site at the C terminus of the repeat domain, as well as the N- and C-terminal nonrepetitive regions from the *Xanthomonas campestris pv. armoraciae* TALE hax3 (**Fig. 1b**).

Using this method, we constructed 17 artificial TALEs with specific combinations of 12.5-mer repeats to target 14-bp DNA-binding sites. TALEs require the first letter of the binding site to be a T^8, meaning that the 12.5-mer repeat targets a 13-bp binding site. We analyzed two clones for each of the 17 TALEs by sequencing and found that all TALEs were accurately assembled. To reduce the frequency of false ligation products, we maximized the number of mismatches between different ligation adapters. We were able to construct the 17 TALEs in parallel in 3 d, substantially quicker than obtaining a similar number of TALEs using commercial DNA synthesis and at a fraction of the cost.

The DNA-binding code of TALEs was identified based on analysis of TALE-binding sites in plant genomes[8,10]. The binding specificities of TALEs have been analyzed using various *in vitro* and *in vivo* methods[12–17,19]. To determine whether this code can be used to target DNA in mammalian cells, we designed a fluorescence-based reporter system (**Supplementary Fig. 1**) by placing the DNA-binding site for each TALE upstream of a minimal cytomegalovirus (CMV) promoter driving the fluorescence reporter gene mCherry (**Fig. 1c**). To generate TALE transcription factors, we replaced the endogenous nuclear localization signal (NLS) and acidic transcription activation domain of wild-type hax3 with a mammalian NLS derived from the simian virus 40 large T antigen and the synthetic transcription activation domain VP64 (ref. 22) (**Fig. 1c**). To allow quantitative comparison of TALE

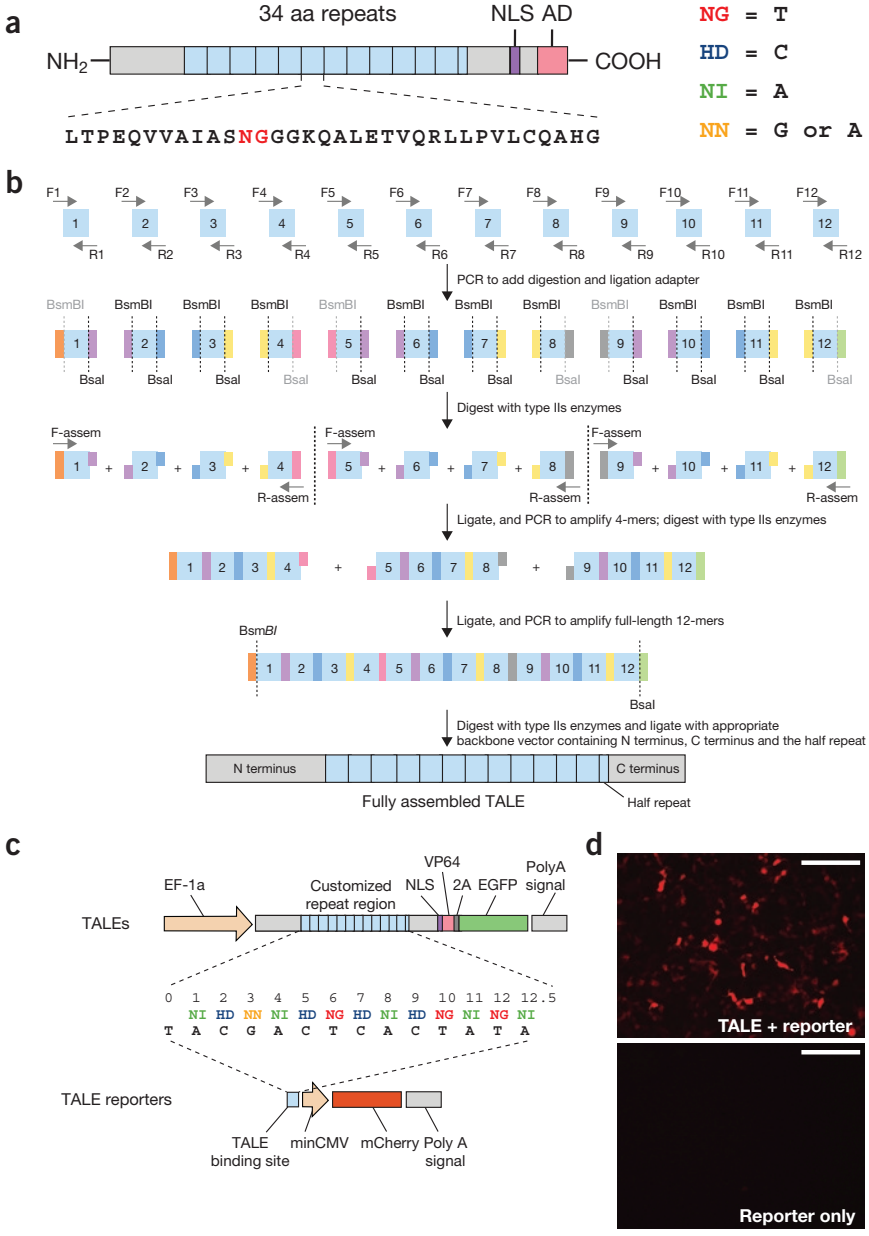

Figure 1 Design and construction of customized artificial TALEs for use in mammalian cells. (**a**) Schematic representation of the native TALE hax3 from *Xanthomonas campestris pv. armoraciae* depicting the tandem repeat domain and the two repeat variable di-residues (red) within each repeat monomer. These di-residues determine the base recognition specificity. The four most common naturally occurring di-residues used for the construction of customized artificial TAL effectors are listed together with their proposed major base specificity. NLS, nuclear localization signal; AD, activation domain of the native TAL effector. (**b**) Schematic of the hierarchical ligation assembly method for the construction of customized TALEs. Twelve separate PCRs are done for each of the four types of repeat monomers (NI, HD, NG and NN) to generate a set of 48 monomers to serve as assembly starting material. Each of the 12 PCR products for a given monomer type (e.g., NI) has a unique linker specifying its programmed position in the assembly (color-coded digestion and ligation adapters). After enzymatic digestion with a type IIs restriction endonuclease (e.g., BsaI), orthogonal overhangs are made by recoding each amino acid in the junction to use an alternative codon. The unique overhangs facilitate the positioning of each monomer in the ligation product. The ligation product was PCR amplified subsequently to yield the full-length repeat regions, which were then cloned into a backbone plasmid containing the N and C termini of the wild-type TALE hax3. (**c**) Schematic representation of the fluorescence reporter system for testing TALE-DNA recognition. The diagram illustrates the composition of the tandem repeat for a TALE and its corresponding 14-bp DNA-binding target in the fluorescent reporter plasmid. VP64, synthetic transcription activation domain; 2A, self-cleavage peptide. (**d**) 293FT cells co-transfected with a TALE plasmid and its corresponding reporter plasmid showed considerably greater mCherry expression compared with the reporter-only control. Scale bars, 200 μm.

activity, we also fused a self-cleaving green fluorescent protein (GFP) to the C terminus of each TALE, so that we could quantify the relative level of TALE expression using GFP fluorescence measurements.

Co-transfection of a TALE (TALE1) and its corresponding reporter plasmid in the human embryonic kidney cell line 293FT led to robust mCherry fluorescence (**Fig. 1d**). In contrast, transfection of 293FT cells with the reporter construct alone did not yield appreciable levels of fluorescence (**Fig. 1d**). Therefore, TALEs are capable of recognizing their target DNA sequences, as predicted by the TALE DNA-binding code, in mammalian cells. We quantified the level of reporter induction by measuring the ratio of total mCherry fluorescence intensity between cells co-transfected with a TALE and its corresponding reporter plasmid and cells transfected with the reporter plasmid alone. To account for differences in TALE expression level, we used the total GFP fluorescence from each TALE transfection as a normalization factor to assess the fold change of reporter induction compared to control.

Next we asked whether the DNA recognition code is sufficiently modular so that TALEs could be customized to target any DNA sequence of interest. We designed the 13 of the 17 TALEs in this study to target a range of DNA-binding sites with a wide range of GC content, from 8% to 80% (**Fig. 2a**) and found that 10 out of 13 TALEs (77%) drove robust mCherry expression (>tenfold) from their corresponding reporters. Three TALEs exhibited more than 50-fold reporter induction (TALE1, TALE4 and TALE8), and only 1 out of 13 TALEs (TALE11) generated less than fivefold induction of mCherry reporter (**Fig. 2a**). As a positive control, we constructed the artificial zinc finger–VP64 (ZF-VP64) fusion, where the zinc finger fusion has previously been shown to activate transcription from a binding site in the human *ERBB2* promoter[22]. ZF-VP64 protein was tested using

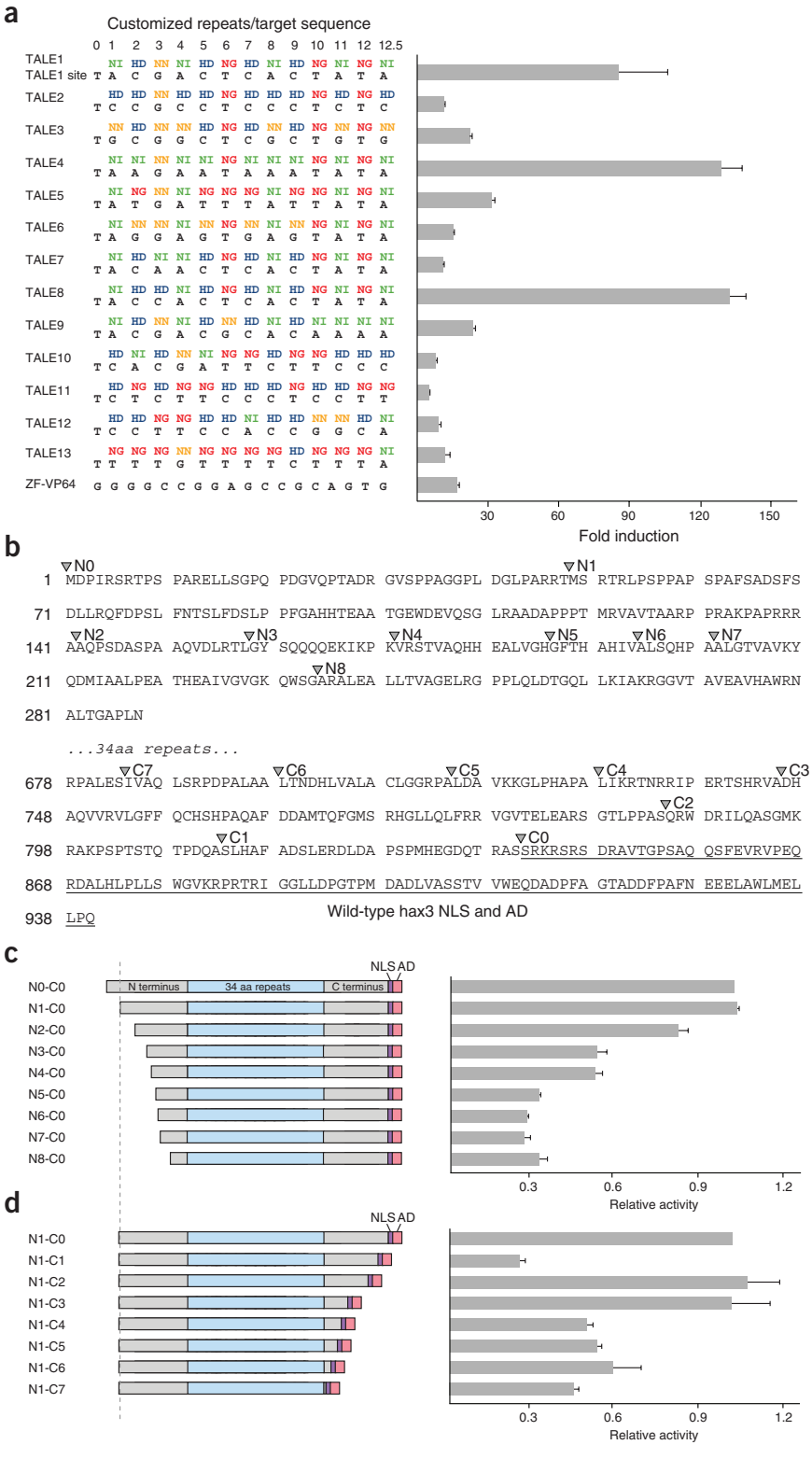

Figure 2 Functional characterization of the robustness of TALE-DNA recognition in mammalian cells and truncation analysis of TALE N- and C-termini. (**a**) Thirteen TALEs were tested with their corresponding reporter constructs. Customized repeat regions and binding site sequences are shown on the left. The activities of the TALEs on target gene expression are shown on the right as the fold induction of the mCherry reporter gene. Fold induction was determined by flow cytometry analysis of mCherry expression in transfected 293FT cells, and calculated as the ratio of the total mCherry fluorescence intensity of cells transfected with and without the specified TALE, normalized by the GFP fluorescence to control for transfection efficiency differences (Online Methods). (**b**) The N- and C-terminal amino acid sequence of wild-type TAL effector hax3 showing the positions of all N- and C-terminal truncation constructs tested in 293FT cells. N0 to N8 designates N-terminal truncation positions (N0 retains the full-length N terminus), and C0 to C7 designate C-terminal truncations. Amino acids representing the nuclear localization signal and the activation domain in the native hax3 protein are underlined. (**c**) Relative activity of each N-terminal TALE truncation construct compared to the TALE (N0-C0). TALE truncation positions are indicated in **b**. Error bars indicate s.e.m.; $n = 3$. TALE-TALE relative activity was calculated by dividing the fold induction of the construct by the fold induction of the reporter gene. Fold induction calculated as in **a**. (**d**) Relative activity of each C-terminal truncation compared to TALE(N1,C0).

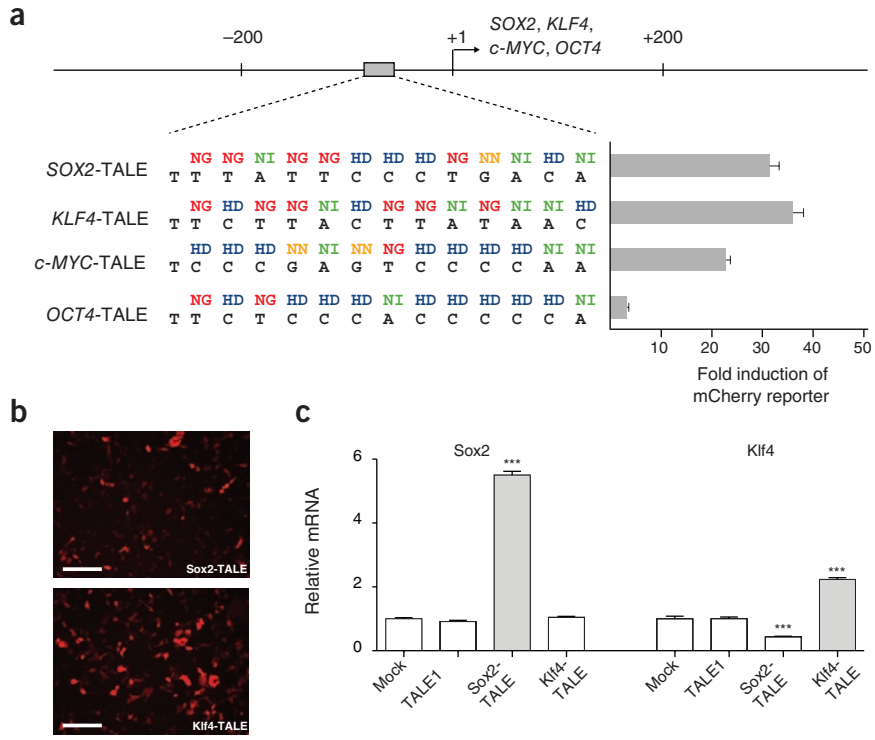

Figure 3 Activation of endogenous genes in human cells by TALEs. (**a**) TALEs designed to target the genes *SOX2*, *KLF4*, *c-MYC* and *OCT4* facilitate activation of mCherry reporter in 293FT cells. The target sites are selected from the 200-bp proximal promoter region. Fold induction was determined by flow cytometry analysis using the same methodology as in **Figure 2** and detailed in Online Methods. (**b**) Images of TALE-induced mCherry reporter expression in 293FT cells. Scale bar, 200 µm. (**c**) Levels of *SOX2* and *KLF4* mRNA in transfected 293FT cells, as determined by quantitative RT-PCR. Mock-treated cells received the transfection vehicle. TALE1 was used as a negative control. Error bars indicate s.e.m.; $n = 3$. *** indicates $P < 0.005$.

the same mCherry reporter assay and demonstrated ~16-fold increase in mCherry reporter expression (**Fig. 2a**).

These data indicate that sequence-specific TALEs can be designed and synthesized to target a wide spectrum of DNA-binding sites at a similar or greater level as an artificial ZF-VP64 transcription factor. Although most TALEs exhibited robust transcription activation in our reporter assay, the large range of observed activity suggests that other effects might contribute to TALE DNA-targeting efficacy. Possible causes might include differences in DNA-interacting capabilities such as binding strength of individual repeat types, context-dependence of monomer binding strength or complexities of mammalian transcription processes[8,17].

To further characterize the robustness of TALEs activities and their DNA-binding specificity, we altered the target nucleotides in the binding sites of TALE1 and TALE13 to test the impact of mismatch position and number on TALE activity. In general, we found that TALE activity is inversely correlated with the number of mismatches (**Supplementary Figs. 2** and **3**). However, the specific TALE recognition rules most likely depend on a combination of positional and contextual effects as well as the number of mismatches[8,13,14,17] and need to be characterized in greater detail.

Each fully assembled TALE has >800 amino acids. Therefore we sought to identify the minimal N- and C-terminal capping region necessary for DNA-binding activity. We used Protean (LASERGENE) to predict the secondary structure of the TALE N- and C-termini, and truncations were made at predicted loop regions. We first generated a series of N-terminal TALE1 truncation mutants and found that transcriptional activity is inversely correlated with the length of the N terminus (**Fig. 2b**,**c**). Deletion of 48 amino acids from the N terminus (truncation mutant N1-C0, **Fig. 2c**) retained the same level of transcription activity as TALE1 with the full-length N terminus, whereas deletion of 141 amino acids from the N terminus (truncation mutant N2-C0, **Fig. 2c**) retained ~80% of transcription activity. Therefore, given its full transcriptional activity, we chose to use truncation position N1 for all subsequent studies.

Similar truncation analysis in the C terminus revealed that a critical element for DNA binding resides within the first 68 amino acids (**Fig. 2b**,**d**). Truncation mutant N1-C3 retained the same level of transcriptional activity as the full C terminus, whereas truncation mutant N1-C4 reduces TALE1 activity by >50% (**Fig. 2d**). Therefore, to preserve the highest level of TALE activity, ~68 amino acids of the C terminus of hax3 should be preserved.

The modularity of the TALE code is ideal for designing artificial transcription factors for transcriptional manipulation from the mammalian genome. To test whether TALEs could be used to modulate transcription of endogenous genes, we designed four additional TALEs using the most active scaffold, N1-C0 (**Fig. 2c**), to directly activate transcription of *SOX2, KLF4, c-MYC* and *OCT4*. TALE-binding sites were selected from the proximal 200-bp promoter region of each gene (**Fig. 3a**). To assay the DNA-binding activity of the four new TALEs, we used the mCherry reporter assay as in previous experiments. Three out of four TALEs (SOX2-TALE, KLF4-TALE and c-MYC-TALE) exhibited >20-fold greater mCherry reporter activation (**Fig. 3a**,**b**).

To test the activity of TALEs on endogenous genes, we transfected each TALE into 293FT cells and quantified mRNA levels of each target gene using qRT-PCR. SOX2-TALE and KLF4-TALE were able to upregulate their respective target genes by 5.5 ± 0.1-fold and 2.2 ± 0.1-fold (**Fig. 3c**), providing a demonstration that TALE can be used to modulate transcription from the genome. To control for specificity of activation, we transfected 293FT cells in parallel with TALE1, which was not designed to target either *SOX2* or *KLF4*, and we found no change in the level of *SOX2* or *KLF4* expression relative to the mock control. Notably, we observed a statistically significant decrease in *KLF4* mRNA in 293FT cells transfected with SOX2-TALE (an approximately twofold reduction, $P < 0.001$, Student's *t*-test), potentially resulting from feedback activation and inhibition among pluripotency factors[23,24]. Finally, it is worth noting that c-MYC-TALE and OCT4-TALE did not upregulate their target genes (data not shown). This is not surprising as different genetic loci may not be equally accessible for activation, possibly due to epigenetic repression. Together, the data demonstrated that TALEs can be designed to bind and specifically activate transcription from the promoters of endogenous mammalian genes.

The modular nature of the TALE DNA recognition code provides an attractive solution for achieving sequence-specific DNA interaction in mammalian cells. Sequence-specific DNA-binding proteins

with predictable binding specificity can be generated economically in a matter of days, using molecular biology methods accessible to most laboratories. Nevertheless, several challenges remain for the widespread use of TALEs, including unknown sequence-specific effects on TALE binding, off-target activities and yet-to-be determined affinity for methylated DNA sequences. Future studies exploring the molecular basis of TALE-DNA interaction will likely extend the modular nature of the TALE code for increased precision, specificity and robustness. Given the ability of TALEs to efficiently anchor transcription effector modules to endogenous genomic targets, other functional modules, including nucleases[18,19,25,26], recombinases[27] and epigenetic modifying enzymes[2], can be similarly targeted to specific binding sites. The TALE toolbox will empower researchers, clinicians and technologists alike with a new repertoire of programmable and precise genome engineering technologies.

METHODS

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

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

ACKNOWLEDGEMENTS

This work was supported by the Harvard Society of Fellows (F.Z.), National Human Genome Research Institute Center for Excellence in Genomics Science (P50 HG003170, G.M.C.), Department of Energy Genomes to Life (DE-FG02-02ER63445, G.M.C.), Defense Advanced Research Projects Agency (W911NF-08-1-0254, G.M.C.), the Wyss Institute for Biologically Inspired Engineering (G.M.C.) and National Institutes of Health Transformative R01 (R01 NS073124-01, F.Z. and P.A.). S.L. was partially supported by a predoctoral fellowship from the European School of Molecular Medicine (S.E.M.M.). We thank the entire Church and Arlotta laboratories for discussion and support.

AUTHOR CONTRIBUTIONS

F.Z. and L.C. conceived the study. F.Z., L.C., S.L. and S.K. designed, performed and analyzed all experiments. P.A. supervised the work of S.L. and G.M.C. supervised the work of F.Z., L.C. and S.K. G.M.C., P.A. and F.Z. provided support for this study. F.Z., L.C. and P.A. wrote the manuscript with support from all authors. G.M.C and P.A. equally contributed to this work.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

1. Carr, P.A. & Church, G.M. Genome engineering. *Nat. Biotechnol.* **27**, 1151–1162 (2009).
2. Meister, G.E., Chandrasegaran, S. & Ostermeier, M. Heterodimeric DNA methyltransferases as a platform for creating designer zinc finger methyltransferases for targeted DNA methylation in cells. *Nucleic Acids Res.* **38**, 1749–1759 (2010).
3. Gonzalez, B. *et al.* Modular system for the construction of zinc-finger libraries and proteins. *Nat. Protoc.* **5**, 791–810 (2010).
4. Maeder, M.L., Thibodeau-Beganny, S., Sander, J.D., Voytas, D.F. & Joung, J.K. Oligomerized pool engineering (OPEN): an 'open-source' protocol for making customized zinc-finger arrays. *Nat. Protoc.* **4**, 1471–1501 (2009).
5. Blancafort, P., Magnenat, L. & Barbas, C.F., III. Scanning the human genome with combinatorial transcription factor libraries. *Nat. Biotechnol.* **21**, 269–274 (2003).
6. Rosen, L.E. *et al.* Homing endonuclease I-CreI derivatives with novel DNA target specificities. *Nucleic Acids Res.* **34**, 4791–4800 (2006).
7. Grizot, S. *et al.* Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. *Nucleic Acids Res.* **37**, 5405–5419 (2009).
8. Boch, J. *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509–1512 (2009).
9. Boch, J. & Bonas, U. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu. Rev. Phytopathol.* **48**, 419–436 (2010).
10. Moscou, M.J. & Bogdanove, A.J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501 (2009).
11. Bogdanove, A.J., Schornack, S. & Lahaye, T. TAL effectors: finding plant genes for disease and defense. *Curr. Opin. Plant Biol.* **13**, 394–401 (2010).
12. Morbitzer, R., Romer, P., Boch, J. & Lahaye, T. Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proc. Natl. Acad. Sci. USA* **107**, 21617–21622 (2010).
13. Kay, S., Hahn, S., Marois, E., Wieduwild, R. & Bonas, U. Detailed analysis of the DNA recognition motifs of the Xanthomonas type III effectors AvrBs3 and AvrBs3Deltarep16. *Plant J.* **59**, 859–871 (2009).
14. Romer, P. *et al.* Recognition of AvrBs3-like proteins is mediated by specific binding to promoters of matching pepper Bs3 alleles. *Plant Physiol.* **150**, 1697–1712 (2009).
15. Kay, S., Hahn, S., Marois, E., Hause, G. & Bonas, U. A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science* **318**, 648–651 (2007).
16. Romer, P. *et al.* Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. *Science* **318**, 645–648 (2007).
17. Scholze, H. & Boch, J. TAL effector-DNA specificity. *Virulence* **1**, 428–432 (2010).
18. Christian, M. *et al.* Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 757–761 (2010).
19. Miller, J.C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* published online, doi:10.1038/nbt.1755 (22 December 2010).
20. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS ONE* **4**, e5553 (2009).
21. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* **3**, e3647 (2008).
22. Beerli, R.R., Segal, D.J., Dreier, B. & Barbas, C.F., III. Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc. Natl. Acad. Sci. USA* **95**, 14628–14633 (1998).
23. Xu, N., Papagiannakopoulos, T., Pan, G., Thomson, J.A. & Kosik, K.S. MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells. *Cell* **137**, 647–658 (2009).
24. Wei, Z. *et al.* Klf4 interacts directly with Oct4 and Sox2 to promote reprogramming. *Stem Cells* **27**, 2969–2978 (2009).
25. Lee, H.J., Kim, E. & Kim, J.S. Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res.* **20**, 81–89 (2010).
26. Li, T. *et al.* TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.* **39**, 359–372 (2010).
27. Gordley, R.M., Gersbach, C.A. & Barbas, C.F., III. Synthesis of programmable integrases. *Proc. Natl. Acad. Sci. USA* **106**, 5053–5058 (2009).

ONLINE METHODS

Design and construction of TALEs and reporters. To simplify construction of TALEs, we synthesized (DNA2.0) a TALE backbone containing the N terminus, a single half-repeat region carrying the variable di-residue NI, and the C-terminus of hax3 and cloned it into a lentiviral expression vector containing the mammalian ubiquitous EF-1α promoter (pLECYT)[28]. To allow for insertion of customized repeat domains, we inserted a linker containing two Type IIs BsmBI sites between the N terminus and the half-repeat region. A DNA fragment containing a mammalian NLS, the transcription activation domain VP64 and 2A-GFP was assembled by PCR and fused to the C terminus of the synthesized TALE backbone (pLenti-EF1a-TALE(0.5 NI)-WPRE). HD, NG and NN versions of the backbone were generated by site-directed PCR mutagenesis using QuikChange II XL (Stratagene). The full nucleotide sequences for the four backbone vectors are available in **Supplementary Sequences**. Customized TALE repeat domains were synthesized by hierarchical ligation of individual repeat monomers (**Supplementary Methods**). To minimize repetitiveness of the final assembled tandem repeat domain, we optimized the DNA sequence for each type of repeat monomer (HD, NG, NI or NN) by altering the amino acid codons. The sequences for the optimized monomers are listed in **Supplementary Table 1**, and the assembly primers are listed in **Supplementary Table 2**. A step-by-step TALE assembly protocol as well as an optimized TALE assembly primer list can be found in **Supplementary Methods** and **Supplementary Table 3**, respectively. mCherry reporter plasmids carrying a TALE-binding site were generated by inserting sequences containing the binding site upstream of the minimal CMV promoter (**Supplementary Fig. 1**).

Cell culture and reporter activation assay. The human embryonic kidney cell line 293FT (Invitrogen) was maintained under 37 °C, 5% CO_2 using Dulbecco's modified Eagle's Medium supplemented with 10% FBS, 2 mM GlutaMAX (Invitrogen), 100 U/ml penicillin and 100 μg/ml streptomycin.

mCherry reporter activation was tested by co-transfecting 293FT cells with plasmids carrying TALEs and mCherry reporters. 293FT cells were seeded into 24- or 96-well plates the day before transfection at densities of 2×10^5 cells/well or 0.8×10^4 cells/well, respectively. Approximately 24 h after initial seeding, cells were transfected using Lipofectamine 2000 (Invitrogen). For 24-well plates we used 500 ng of TALE and 30 ng of reporter plasmids per well. For 96-well plates we used 100 ng of TALE and 7 ng of reporter plasmids per well. All transfection experiments were performed according to manufacturer's recommended protocol.

Flow cytometry. mCherry reporter activation was assayed by flow cytometry using a LSRFortessa cell analyzer (BD Biosciences). Cells were trypsinized from their culturing plates ~18 h after transfection and resuspended in 200 μl of media for flow cytometry analysis. The flow cytometry data were analyzed using BD FACSDiva (BD Biosciences). At least 25,000 events were analyzed for each transfection sample.

The fold induction of mCherry reporter gene by TALEs was determined by flow cytometry analysis of mCherry expression in transfected 293FT cells, and calculated as the ratio of the total mCherry fluorescence intensity of cells from transfections with and without the specified TALE. All fold-induction values were normalized to the expression level of TALE as determined by the total GFP fluorescence for each transfection.

Endogenous gene activation assay. 293FT cells were seeded in 6-well plates. 4 μg of TALE plasmid was transfected using Lipofectamine 2000 (Invitrogen). Transfected cells were cultured at 37 °C for 48 h, sorted for GFP-positive population using BD FACSAria (BD Biosciences) to obtain cells that were successfully transfected and expressing TALE. At least 1,000,000 cells were harvested and subsequently processed for total RNA extraction using the RNAeasy Mini Kit (Qiagen). cDNA was generated using the iScript cDNA Synthesis Kit (Bio-Rad) according to the manufacturer's recommended protocol. *OCT4*, *SOX2*, *c-MYC* and *KLF4* mRNA was detected using TaqMan Gene Expression Assays (Applied Biosystems).

28. Zhang, F. *et al*. Multimodal fast optical interrogation of neural circuitry. *Nature* **446**, 633–639 (2007).