

Programmable Sequence-Specific Transcriptional Regulation of Mammalian Genome Using Designer TAL Effectors

Feng Zhang^{1,2,3,5,7*,±}, Le Cong^{2,3,4*}, Simona Lodato^{5,6}, Sriram Kosuri^{2,3}, George Church^{2,3,#,±}, and Paola Arlotta^{5,6,#,±}

¹ Society of Fellows

² Department of Genetics, Harvard Medical School

³ Wyss Institute for Biologically Inspired Engineering

⁴ Program in Biological and Biomedical Sciences, Harvard Medical School

⁵ Department of Stem Cell and Regenerative Biology Harvard University

⁶ Center for Regenerative Medicine and Department of Neurosurgery, Massachusetts General Hospital, Harvard Medical School

⁷ Present Address: Broad Institute of MIT and Harvard, Cambridge, MA

McGovern Institute for Brain Research, MIT, Cambridge, MA

Department of Brain and Cognitive Sciences, MIT Cambridge, MA

* These authors contributed equally to this work.

These authors contributed equally to this work.

± To whom correspondence should be addressed.

E-mail: zhang_f@mit.edu, gmc@harvard.edu, paola_arlotta@hms.harvard.edu

Feng Zhang, Ph.D.
Broad Institute of MIT and Harvard
7 Cambridge Center
Cambridge, MA 02142
E-mail: zhang_f@mit.edu

George Church, Ph.D.
Department of Genetics, Harvard Medical School
New Research Building, Rm 238
77 Avenue Louis Pasteur, Boston, MA 02115
E-mail: gmc@harvard.edu

Paola Arlotta, Ph.D.
Department of Stem Cell and Regenerative Biology, Harvard University
Center for Regenerative Medicine, Massachusetts General Hospital
Simches Building 4238
185 Cambridge Street
Boston, MA 02114
E-mail: paola_arlotta@hms.harvard.edu

The ability to direct functional domains to specific DNA sequences is a long sought-after goal for studying and engineering biological processes. Transcription activator like effectors (TALEs) from *Xanthomonas sp.* present a promising platform for designing sequence-specific DNA binding proteins. Here we describe a robust and rapid method for overcoming the difficulty of constructing TALE repeat domains. We synthesized 17 designer TALEs (dTALEs) that are customized to recognize specific DNA binding sites, and demonstrate that dTALEs can specifically modulate transcription of endogenous genes (Sox2 and Klf4) from the native genome in human cells. dTALEs provide a designable DNA targeting platform for the interrogation and engineering of biological systems.

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 designer polydactyl zinc finger (ZFs)³⁻⁵ and meganuclease^{6, 7} proteins. In particular, designer ZFs 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 site. However, due 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 designer ZFs and meganucleases remain difficult and expensive, often involving elaborate screening procedures and long development time on the order of several weeks. Here we developed an alternative DNA targeting platform based on the naturally occurring transcription activator-like effectors (TALEs) from *Xanthomonas sp.*⁸⁻¹¹

TALEs are natural effector proteins secreted by numerous species of *Xanthomonas* to modulate host gene expression and facilitate bacterial colonization and survival^{9, 11}. Recent studies of TALEs have revealed an elegant code linking the repetitive region of TALEs with its 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 (variable diresidues), and recent computational and functional analysis^{9, 11} 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 designer TALEs to specify DNA binding in mammalian cells. However, our ability to test the modularity of the TALE DNA binding code remains limited due to the difficulty in constructing custom TALEs with specific tandem repeat monomers. Early studies have tested the DNA binding properties of TALEs^{8, 12-16}, including two studies that tested artificial TALEs with customized repeat regions^{13, 14}.

A prerequisite for exploring the modularity TALE repeat monomers is the ability to synthesize designer TALEs with tailored repetitive DNA binding domains. While this has been recently shown to be possible^{13, 14, 17}, 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 may not be amenable to high-throughput TALE synthesis. Furthermore, even though commercial services can be employed for the synthesis of novel TALE binding domains¹⁴, they present a cost-prohibitive option for large scale TALE construction and testing. Hence a more robust protocol to construct large numbers of designer TALEs would enable ready perturbation of any genome target in many organisms.

To enable high-throughput construction of designer TALEs, we developed a reliable hierarchical ligation-based strategy to overcome the difficulty of constructing TALE tandem repeat domains (Fig. 1b and Supplementary Methods). To reduce the repetitiveness of designer TALEs and to facilitate amplification using PCR, we first optimized the DNA sequence of the four repeat monomers (NI, HD, NN, NG) to minimize repetitiveness while preserving the amino acid sequence. In order to assemble the individual monomers in a specific order, we altered the DNA sequence at the junction between each pair of monomers, similar to the Golden Gate cloning strategy for multi-piece DNA ligation^{18, 19}. Using different codons to represent the junction between each pair of monomers (Gly-Leu), we designed unique 4 base pair sticky-end ligation adapters for each junction (Supplementary Methods). Using this strategy, 4 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 0.5 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 non-repetitive regions from the *Xanthomonas campestris* pv. *armoraciae* TALE *hax3* (Fig. 1b).

Using this method, we attempted to construct 17 artificial TALEs with specific combinations of 12.5-mer repeats to target 14 base pair DNA binding sites – TALEs require the first letter of the binding site to be a T, and the 12.5-mer repeat targets a 13 bp binding site. We analyzed 2 clones for each of the 17 dTALEs via sequencing and found that all dTALEs were accurately assembled. Furthermore, we were able to construct 16 dTALEs in parallel in 3 days, a time substantially shorter than what is required for constructing a similar number of dTALEs via 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} and the binding specificity of TALEs have been analyzed using various *in vitro* and *in vivo* methods^{8, 14-17, 20-22}. In order 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 dTALE upstream of a minimal CMV promoter driving the fluorescence reporter gene mCherry (Fig. 1c). To generate dTALE transcription factors, we replaced the endogenous nuclear localization signal (NLS) and acidic transcription activation domain (AD) of wild type *hax3* with a mammalian NLS derived from the simian virus 40 large T-antigen and the synthetic transcription activation domain VP64 (Fig. 1c). To allow quantitative comparison of the dTALE activity, we also fused a self-cleaving green fluorescent protein (GFP) to the C-terminus of each dTALE, so that we could quantify the relative level of dTALE expression using GFP fluorescence measurements.

Co-transfection of a dTALE (dTAL1) 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, dTALEs 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 dTALE and its corresponding reporter plasmid, and cells transfected with the reporter plasmid alone. To account for differences in dTALE expression level, we use the total GFP fluorescence from each dTALE transfection as a normalization factor to assess the fold of reporter induction fold.

Next we asked whether the DNA recognition code is sufficiently modular so that dTALEs could be customized to target any DNA sequence of interest. We first synthesized 13 distinct

dTALEs targeting a range of DNA binding sites with diverse DNA sequence compositions (Fig. 2a) and found that 10 out of 13 dTALEs (77%) drove robust mCherry expression (> 10 folds) from their corresponding reporters. Three dTALEs exhibited more than 50 folds reporter induction (dTALE1, dTALE4 and dTALE8), and only one out of 13 dTALEs (dTALE11) generated less than 5 fold induction of mCherry reporter expression (Fig. 2a). As a positive control, we constructed an artificial zinc finger-VP64 (ZF-VP64) fusion, where the ZF has previously been shown to activate transcription from a binding site in the human *erbB-2* promoter²³. This artificial ZF-VP64 protein was tested using the same mCherry reporter assay and demonstrated approximately 16-fold mCherry reporter activation (Fig. 2a).

These data indicate that sequence-specific dTALEs can be designed and synthesized to target a wide spectrum of DNA binding sites at a similar or greater level as artificial ZF-VP64 transcription factors. While most dTALEs exhibited robust transcription activation in our reporter assay, the large range of observed activity suggests that other effects might contribute to dTALE 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 processes²².

To further characterize the robustness of dTALEs activities and their DNA binding specificity, we altered the target nucleotides in the binding sites of dTALE1 and dTALE13 to test the impact of mismatch position and number on dTALE activity. In general, we found that dTALE activity is inversely correlated with the number of mismatches (Supplementary Figs. 2 and 3). However, the specific dTALE recognition rules most likely depend on a combination of positional and contextual effects as well as the number of mismatches^{8, 15, 16, 22}, and need to be further characterized in greater detail.

Each fully assembled dTALE has more than 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 dTALE1 truncation mutants and found that transcriptional activity is inversely correlated with the N-terminus length (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 the full length N-term dTALE1, while deletion of 141 amino acids from the N-term (truncation mutant N2-C0, Fig. 2c) retained

approximately 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 dTALE1 activity by more than 50% (Fig. 2d). Therefore in order to preserve the highest level of dTALE activity, approximately 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. In order to test whether dTALE could be used to modulate transcription of endogenous genes, we designed 4 additional dTALEs to directly activate transcription of Sox2, Klf4, c-Myc, and Oct4. dTALE binding sites were selected from the proximal 200bp promoter region of each gene (Fig. 3a). To assay the DNA binding activity of the 4 new dTALEs, we used the mCherry reporter assay as in previous experiments. three out of four dTALEs (Sox2-dTALE, Klf4-dTALE, and cMyc-dTALE) exhibited greater than 20-fold of mCherry reporter activation (Fig. 3a,b).

To test the activity of dTALEs on endogenous genes, we transfected each dTALE into 293FT cells and quantified mRNA levels of each target gene using qRT-PCR. dTALE-Sox2 and dTALE-Klf4 were able to upregulate their respective target genes by 5.5 ± 0.1 and 2.2 ± 0.1 folds (Fig. 3c), providing a demonstration that dTALE can be used to modulate transcription from the genome. To control for specificity of activation, we transfected 293FT cells in parallel with dTALE1, which was not designed to target either Sox2 or Klf4, and found no change in the level of Sox2 or Klf4 expression relative to the mock control. Interestingly, we observed a statistically significant decrease in the level of Klf4 mRNA in 293FT cells transfected with Sox2-dTALE (approximately a 2-fold reduction). This is potentially due to secondary cross-regulation among reprogramming factors^{24, 25}. Finally, it is worth noticing that cMyc-dTALE and Oct4-dTALE 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 dTALE 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 a novel and attractive solution for achieving sequence-specific DNA interaction in mammalian cells. For the first time, sequence-specific DNA binding proteins with predictable binding specificity can be generated in a matter of days, economically using molecular biology methods accessible to most. 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 dTALEs to efficiently anchor transcription effector modules to endogenous genomic targets, other functional modules, including nucleases²⁶, recombinases²⁷, and epigenetic modifying enzymes², can be similarly targeted to specific binding sites. The designer TALE toolbox will empower researchers, clinicians, and technologists alike with a new repertoire of programmable precision 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 NIH Transformative R01 (R01 NS073124-01, for 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. and P.A. and F.Z. provided support for this study. F.Z. and L.C. wrote the manuscript with support from all authors.

Competing Financial Interests

The authors declare no competing financial interests.

References

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., 3rd 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. Li, T. et al. TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res* (2010).
13. Christian, M. et al. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 757-761 (2010).
14. 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 U S A* (2010).
15. 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).
16. 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).
17. Miller, J.C. et al. A TALE nuclease architecture for efficient genome editing. *Nat Biotech* **advance online publication** (2010).
18. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling: a one-pot DNA shuffling method based on type IIIs restriction enzymes. *PLoS One* **4**, e5553 (2009).
19. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3**, e3647 (2008).
20. 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).

21. Romer, P. et al. Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. *Science* **318**, 645-648 (2007).
22. Scholze, H. & Boch, J. TAL effector-DNA specificity. *Virulence* **1**, 428-432 (2010).
23. Beerli, R.R., Segal, D.J., Dreier, B. & Barbas, C.F., 3rd 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 U S A* **95**, 14628-14633 (1998).
24. 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).
25. Wei, Z. et al. Klf4 interacts directly with Oct4 and Sox2 to promote reprogramming. *Stem Cells* **27**, 2969-2978 (2009).
26. Lee, H.J., Kim, E. & Kim, J.S. Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res* **20**, 81-89 (2010).
27. Gordley, R.M., Gersbach, C.A. & Barbas, C.F., 3rd Synthesis of programmable integrases. *Proc Natl Acad Sci U S A* **106**, 5053-5058 (2009).

Online Methods

Design and construction of designer TALEs and reporters: To simplify construction of designer TALEs, a dTALE backbone containing the N-term, a single 0.5 repeat regions carrying the variable diresidue NI, and C-term of *hax3* was synthesized (DNA2.0) and cloned into a lentiviral expression vector containing the mammalian ubiquitous EF-1 α promoter (pLECYT)²⁸. To allow for insertion of customized repeat domains, a linker containing two Type IIs BsmBI sites are inserted between the N-term and the 0.5 repeat region. A DNA fragment containing a mammalian NLS, transcription activation domain VP64, and 2A-GFP was assembled via PCR assembly and fused to the C-term of the synthesized dTALE backbone (pLenti-EF1a-dTALE(0.5 NI)-WPRE). HD, NG, and NN versions of the backbone were generated via site directed PCR mutagenesis using QuikChange II XL (Stratagene). The full nucleotide sequences for the four backbone vectors are available in **Supplementary Information**. Customized dTALE repeat domains were synthesized via hierarchical ligation of individual repeat monomers (**Supplementary Methods**). To minimize repetitiveness of the final assembled tandem repeat domain, the DNA sequence for each type of repeat monomer (HD, NG, NI, or NN) has been optimized 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**.

mCherry reporter plasmids carrying dTALE 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 293 FT (Invitrogen) was maintained under 37°C, 5% CO₂ using Dulbecco's modified Eagle's Medium supplemented with 10% fetal bovine serum, 2mM GlutaMAX (Invitrogen), 100U/mL Penicillin, and 100 μ g/mL Streptomycin.

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

Flow cytometry: mCherry reporter activation was assayed via flow cytometry using a LSRTFortessa cell analyzer (BD Biosciences). Cells were trypsinized from their culturing plates approximately 18 hours after transfection and resuspended in 200ul of media for flow cytometry analysis. The flow cytometry data was 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 dTALEs was determined via 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 dTALE. All fold induction values were normalized to the expression level of dTALE as determined by the total GFP fluorescence for each transfection.

Endogenous gene activation assay: 293FT cells were seeded in 6 well plates. 4ug of dTALE plasmid was transfected using Lipofectamine 2000 (Invitrogen). Transfected cells were cultured at 37°C for 48 hours, sorted for GFP positive population using BD FACSAria (BD Biosciences) to obtain cells that were successfully transfected and expressing dTALE. 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, cMyc, and Klf4 mRNA were detected using TaqMan Gene Expression Assays (Applied Biosystems: Oct4 - Hs00999632_g1, Sox2 - Hs00602736_s1, cMyc - Hs00905030_m1, Klf4 - Hs01034973_g1).

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

Figure 1 | Design and construction of customized artificial transcription activator like effectors (dTALEs) 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 variable repeat diresidues (red) within each repeat monomer that specify the base recognition specificity. The four most common naturally occurring diresidues that were used for the construction of customized artificial TAL effectors are listed together with their proposed major base specificity. **b**, Schematic of the hierarchical ligation assembly method for the construction of customized dTALEs. 12 individual PCRs are performed for each of the 4 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 individual PCR products for a given monomer type (i.e. 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 cutter (e.g. BsaI), unique overhangs (generated by leveraging the alternate codons for each amino acid in the junction) are generated. 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 dTALE-DNA recognition. The diagram illustrates the composition of the tandem repeat for a dTALE and its corresponding 14bp DNA binding target in the fluorescent reporter plasmid. NLS, nuclear localization signal; AD, activation domain of the native TAL effector; VP64, synthetic transcription activation domain; 2A, self-cleavage peptide. **d**, 293FT cells co-transfected with a dTALE plasmid and its corresponding reporter plasmid exhibited significant level of mCherry expression compared to the reporter-only control. Scale bar, 200 μ m.

Figure 2 | Functional characterization of the robustness of dTALE-DNA recognition in mammalian cells and truncation analysis of TALE N- and C-termini. **a**, 13 dTALEs were tested with their corresponding reporter constructs. Customized repeat regions and binding site sequences are shown on the left. The activities of the dTALE to activate target gene expression are shown on the right as the fold induction of the mCherry reporter gene in a log scale. **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. The amino acids representing the nuclear localization signal and the activation domain in the native HAX3 protein were underlined. **c**, Sequential truncation at the N-

terminus of dTALE1 led to decreasing levels of reporter activity. Each truncation constructs is designated by its corresponding N-terminal and C-terminal truncation positions as indicated in panel b. Cartoon representations of all truncation constructs are shown on the left; the relative activity of each dTALE truncation construct compared to the dTALE(N0-C0) is demonstrated on the right. This relative activity is calculated from the fold induction of the reporter gene. **d**, Sequential truncation at the C-terminus of dTALE1 was used to characterize the optimal length of the C-terminus. Truncation position is designated in panel b. The relative activity of each truncation design compared to dTALE(N1,C0) is shown on the right. All error bars indicate s.e.m.; n=3. The fold induction was determined via 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 dTALE, normalized by the GFP fluorescence to control for transfection efficiency differences, as detailed in the **Online Methods**.

Figure 3 | Activation of endogenous pluripotency factors from the human genome by designer TALEs. **a**, dTALEs designed to target the pluripotency factors *Sox2*, *Klf4*, *c-Myc*, and *Oct4* facilitate activation of mCherry reporter in 293FT cells. The target sites are selected from the 200bp proximal promoter region. The fold induction was determined via flow cytometry analysis using the same methodology as stated in Fig. 2 and detailed in **Supplementary Methods**. **b**, Images of dTALE induced mCherry reporter expression in 293FT cells. Scale bar, 200 μ m. **c**, mRNA levels of *Sox2* and *Klf4* in 293FT cells transfected with mock, dTALE1, *Sox2*-dTALE and *Klf4*-dTALE. Bars represent the levels of *Sox2* or *Klf4* mRNA in the transfected cell as determined via quantitative RT-PCR. Mock consists of cells receiving the transfection vehicle and dTALE1 is used as a negative control. All error bars indicate s.e.m.; n=3. *** p < 0.005.

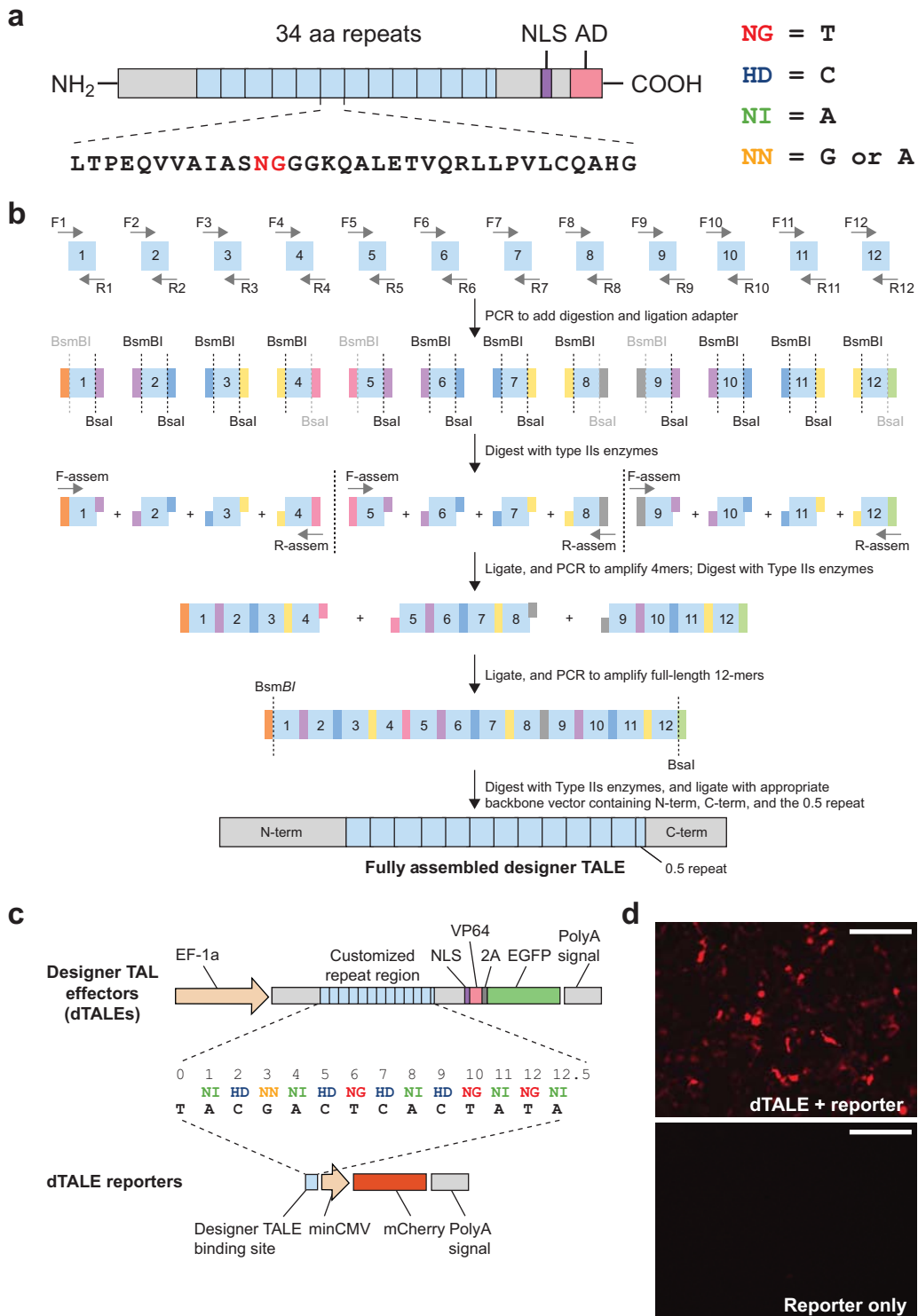


Figure 1

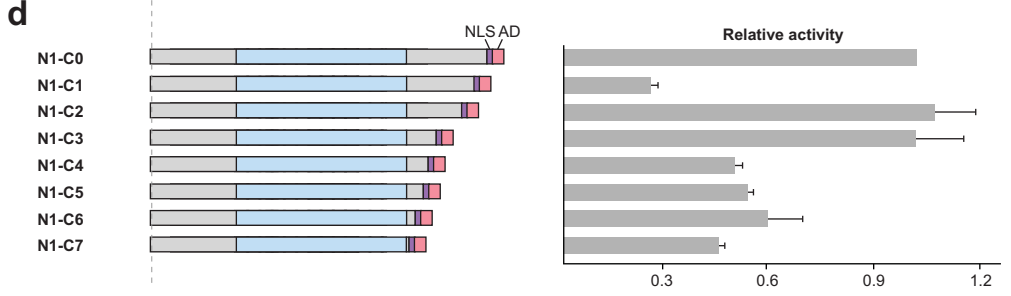
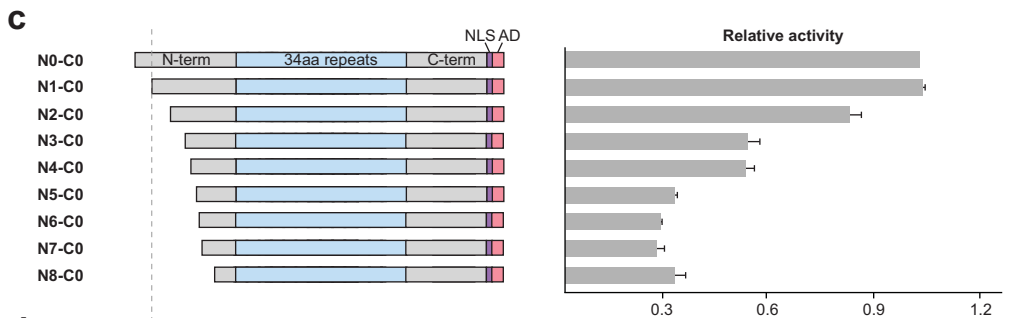
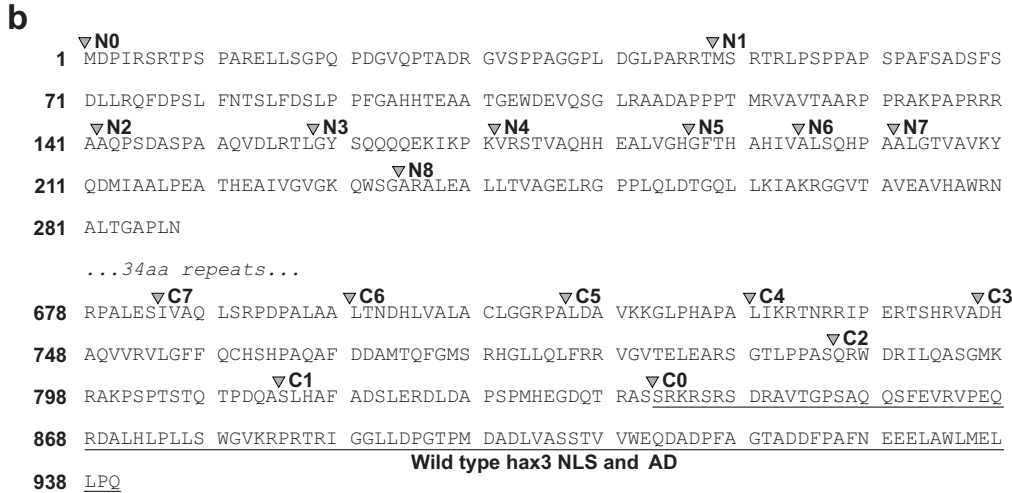
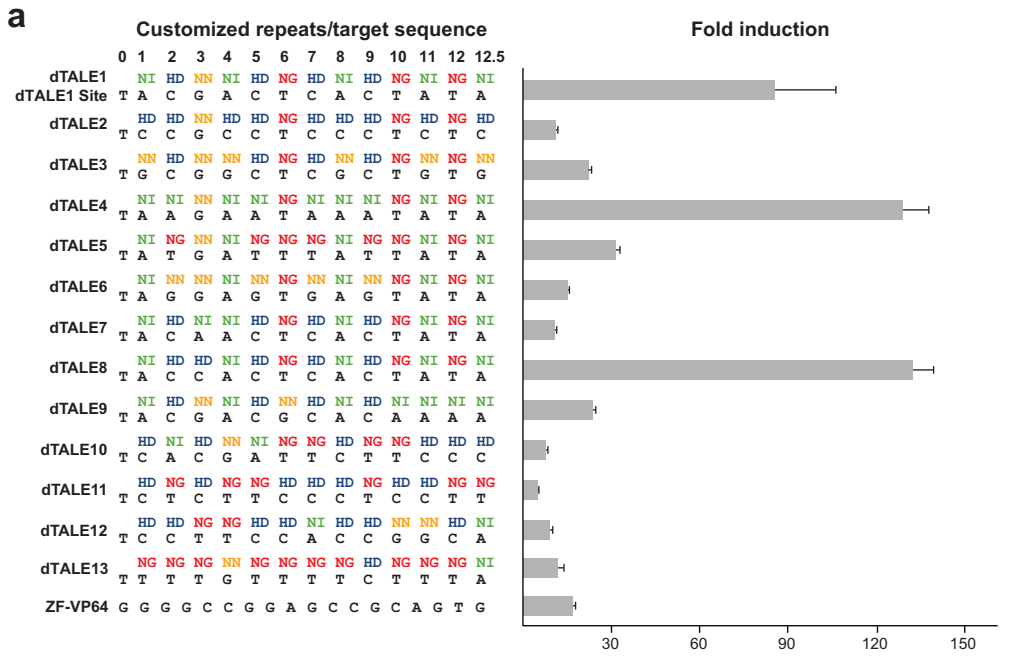


Figure 2

