

# Continuous directed evolution of DNA-binding proteins to improve TALEN specificity

Basil P Hubbard<sup>1</sup>, Ahmed H Badran<sup>1</sup>, John A Zuris<sup>1</sup>, John P Guilinger<sup>1</sup>, Kevin M Davis<sup>1</sup>, Liwei Chen<sup>1</sup>, Shengdar Q Tsai<sup>2–5</sup>, Jeffry D Sander<sup>2–5,7</sup>, J Keith Joung<sup>2–5</sup> & David R Liu<sup>1,6</sup>

**Nucleases containing programmable DNA-binding domains can alter the genomes of model organisms and have the potential to become human therapeutics. Here we present DNA-binding phage-assisted continuous evolution (DB-PACE) as a general approach for the laboratory evolution of DNA-binding activity and specificity. We used this system to generate transcription activator–like effectors nucleases (TALENs) with broadly improved DNA cleavage specificity, establishing DB-PACE as a versatile approach for improving the accuracy of genome-editing agents.**

Genome-editing tools are revolutionizing our understanding of how genotype influences phenotype and have the potential to serve as treatments for genetic diseases<sup>1,2</sup>. These tools include fusions of programmable DNA-binding domains (DBDs) such as zinc fingers (ZFs) and transcription activator–like effectors (TALEs) to functional domains including nucleases, recombinases and transposases<sup>2,3</sup>. ZFs are DBDs of approximately 30 amino acids that typically bind three DNA nucleotides along the major groove<sup>2,4</sup>. Several methods have been developed to generate ZF arrays with tailor-made DNA specificities<sup>5,6</sup>. TALE proteins consist of an N-terminal domain followed by a series of tandem repeats of 33–35 amino acids each, a nuclear localization sequence, a transcription activation domain and a C-terminal domain<sup>7,8</sup>. Two repeat variable diresidues at positions 12 and 13 within each repeat recognize and bind to a specific DNA base<sup>9,10</sup>, and altering the repeat variable diresidues allows TALE repeats to be programmed using a simple code<sup>7,11</sup>.

A limitation of TALEs is that the 5′ nucleotide of the target is specified to be T (ref. 9). Although promiscuous TALEs with no specificity at the 5′ position have been described<sup>12,13</sup>, no TALE variants that preferentially recognize 5′ A or 5′ C have been

described<sup>12,13</sup>. Moreover, TALEs bind appreciably to off-target sites within the genome, thereby limiting their potential as human therapeutics<sup>14</sup>. No method has been reported that enhances the specificity of a particular TALE array by decreasing its ability to bind to specific off-target DNA sequences found in a genome.

We recently developed a system, phage-assisted continuous evolution (PACE), that allows proteins to continuously evolve in the laboratory at a rate ~100-fold faster than that of conventional methods (Supplementary Fig. 1a and Supplementary Results)<sup>15</sup>. PACE has been used to rapidly evolve RNA polymerases and proteases with tailor-made properties<sup>15–17</sup>. We speculated that PACE could be adapted to continuously evolve DNA-binding domains with altered or improved DNA-binding specificity.

To develop a PACE-compatible DNA-binding selection, we linked a DBD of interest to a subunit of bacterial RNA polymerase III (RNAP). On the basis of previous one-hybrid systems<sup>18</sup>, we envisioned that binding of this fusion protein to operator sequences upstream of a minimal *lac* promoter would induce transcription of a downstream gene III-luciferase reporter through recruitment or stabilization of the RNAP holoenzyme (Fig. 1a). Using the DBD of Zif268 (residues 333–420)<sup>19</sup> fused to the ω subunit of RNAP, we established sequence-specific and binding-dependent induction of gene III and production of pIII, the phage protein that enables propagation during PACE (Supplementary Fig. 1b,c and Supplementary Results).

To port this selection system into PACE, we moved the DNA operator–gene III cassette to an accessory plasmid (AP) and moved the RNAP ω–Zif268 protein to a selection phage (SP) construct. We then performed plaque assays to establish activity-dependent phage propagation *in vitro* (Supplementary Figs. 2–4 and 5a and Supplementary Results) and validated continuous propagation of Zif268-SP phage in DNA-binding PACE (DB-PACE) (Supplementary Fig. 5b and Supplementary Results). Next, we performed a mock evolution to evolve DNA-binding activity starting from an inactive mutant Zif268 protein (Supplementary Fig. 5c,d and Supplementary Results).

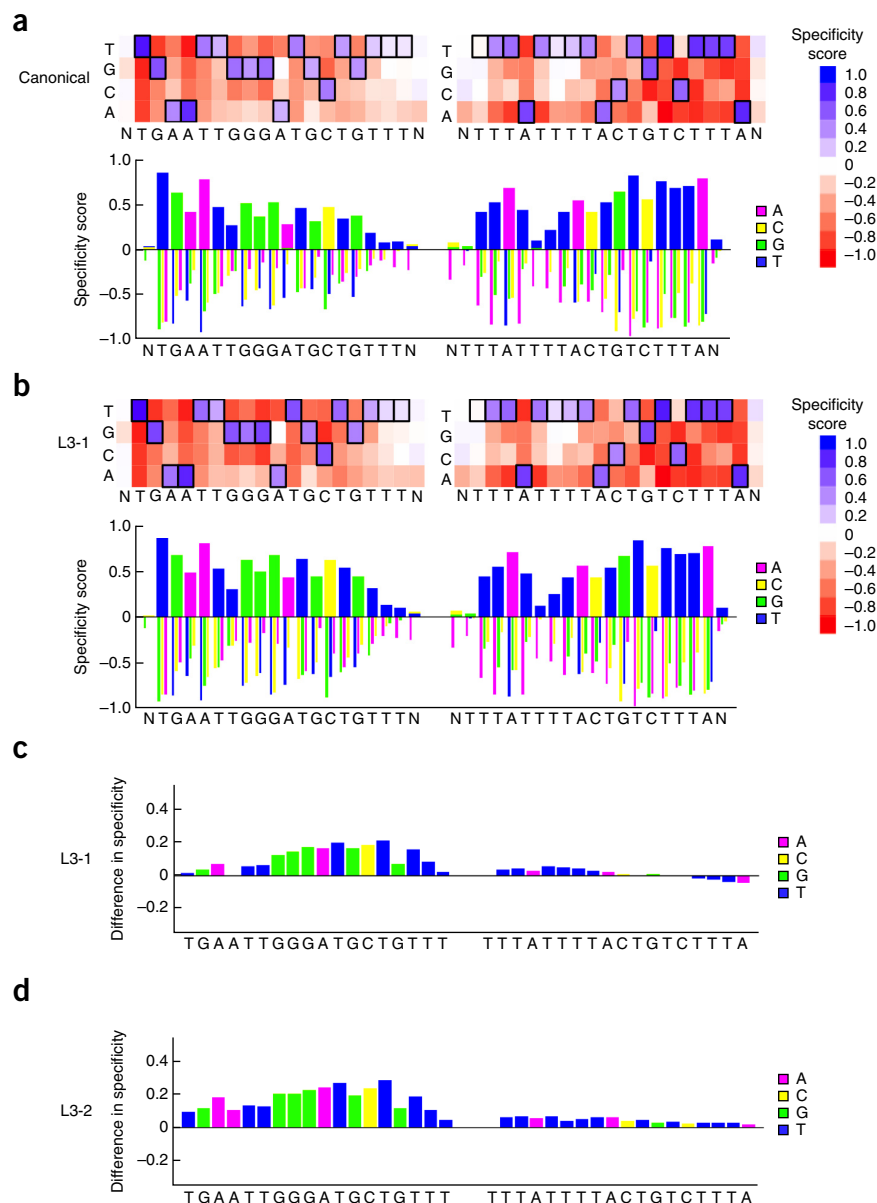
To apply this system to continuously evolve TALE proteins, we optimized the one-hybrid fusion architecture, and verified activity-dependent phage propagation *in vitro* and in PACE (Supplementary Figs. 6 and 7a,b and Supplementary Results). Next, we used DB-PACE to evolve a canonical CBX8-targeting TALE toward recognition of noncanonical 5′ nucleotides (5′ A, C and G). After 48 h of DB-PACE, we isolated phage with up to sixfold-higher activity on 5′ A relative to the canonical TALE protein, fivefold-higher activity on 5′ C and fivefold-higher activity on 5′ G target sequences (Supplementary Figs. 8 and 9

<sup>1</sup>Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts, USA. <sup>2</sup>Molecular Pathology Unit, Massachusetts General Hospital, Charlestown, Massachusetts, USA. <sup>3</sup>Center for Cancer Research, Massachusetts General Hospital, Charlestown, Massachusetts, USA. <sup>4</sup>Center for Computational and Integrative Biology, Massachusetts General Hospital, Charlestown, Massachusetts, USA. <sup>5</sup>Department of Pathology, Harvard Medical School, Boston, Massachusetts, USA. <sup>6</sup>Howard Hughes Medical Institute, Harvard University, Cambridge, Massachusetts, USA. <sup>7</sup>Present address: DuPont Pioneer, Johnston, Iowa, USA. Correspondence should be addressed to D.R.L. (drlu@fas.harvard.edu).

RECEIVED 14 JANUARY; ACCEPTED 26 JUNE; PUBLISHED ONLINE 10 AUGUST 2015; DOI:10.1038/NMETH.3515



**Figure 2** | High-throughput specificity profiling of canonical and evolved TALENs. Top, heat map showing DNA cleavage specificity scores across >10<sup>12</sup> off-target sequences for either canonical (a) or L3-1 evolved TALENs (b) targeting the *ATM* locus at each position in the left and right half-sites plus a single flanking position (N). Bottom, bar graph showing the quantitative specificity score for each nucleotide position. A score of 0 indicates no specificity, whereas a score of 1.0 corresponds to perfect specificity. (c) Bar graph indicating the quantitative difference in specificity score between the canonical and L3-1 evolved TALENs at each position in the target half-sites. A difference of 0 indicates no change in specificity. (d) Identical to c, except for the L3-2 evolved TALEN versus the canonical TALEN. For all heat maps, the cognate base for each position in the target sequence is boxed. For the right half-site, data for the sense strand are displayed.



found that TALEN pairs containing the evolved TALEs L3-1 and L3-2 showed a substantially lower ability to cleave a wide range of off-target sequences containing 4–9 mutations relative to the canonical TALEN (Supplementary Fig. 18 and Supplementary Table 2), indicating a broad specificity improvement. Moreover, TALEN pairs incorporating the evolved TALEs L3-1 or L3-2 displayed higher specificity than that of the canonical TALEN at nearly all positions in the left half-site of the *ATM* binding sequence, but there was no substantial change in specificity in the right half-site that was not used during DB-PACE (Fig. 2, Supplementary Figs. 19–24 and Supplementary Results).

Finally, we tested the behavior of our evolved TALENs in human cells. We found that whereas cleavage at the on-target *ATM* site was comparable for the canonical and evolved TALENs (Supplementary Tables 1 and 3), both evolved TALENs exhibited much lower (or undetectable) off-target activity than the canonical TALEN on all four genomic off-target loci assayed (Supplementary Table 3 and Supplementary Results). We confirmed that the improved DNA cleavage specificity of the evolved TALENs was independent of FokI architecture and cell line (Supplementary Table 3).

DB-PACE brings the power of continuous evolution to bear on improving the activity and specificity of a variety of DNA-binding proteins. Because DB-PACE does not require the use of targeted libraries that can constrain or bias evolutionary outcomes, it naturally supports the discovery of evolved solutions with desired properties that could not be rationalized a priori (Supplementary Discussion). Furthermore, DB-PACE coupled with *in vitro* specificity profiling represents a new systematic approach to removing specific off-target activities of TALENs, and it may be used to facilitate generation of highly specific genome engineering tools for therapeutic applications (Supplementary Fig. 25).

## METHODS

Methods and any associated references are available in the [online version of the paper](#).

**Accession codes.** NCBI Sequence Read Archive: [SRP055191](#) and [SRP053327](#).

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

## ACKNOWLEDGMENTS

This work was supported by Defense Advanced Research Projects Agency (DARPA) HR0011-11-2-0003, DARPA N66001-12-C-4207, a grant from the US National Institutes of Health (NIH)/NIGMS (R01 GM095501 and R01 GM065400) and the Howard Hughes Medical Institute. S.Q.T., J.D.S. and J.K.J. were supported by an NIH Director's Pioneer Award (DP1 GM105378). S.Q.T. was supported by NIH F32 GM105189. A.H.B. was supported by the Harvard Chemical Biology Program and a US National Science Foundation Graduate Research Fellowship.

## AUTHOR CONTRIBUTIONS

B.P.H. designed the research, performed experiments, analyzed data and wrote the manuscript. A.H.B. assisted in the design of the one-hybrid system,

and A.H.B. and K.M.D. contributed materials and performed experiments. J.A.Z., J.P.G. and L.C. performed experiments and data analysis. S.Q.T. prepared materials for TALEN cleavage analysis in cells. J.D.S. contributed experimentally validated TALE arrays. D.R.L. designed and supervised the research and wrote the manuscript. All of the authors contributed to editing the manuscript.

#### COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the [online version of the paper](#).

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Tebas, P. *et al. N. Engl. J. Med.* **370**, 901–910 (2014).
2. Gaj, T., Gersbach, C.A. & Barbas, C.F. III. *Trends Biotechnol.* **31**, 397–405 (2013).
3. Epinat, J.C. *et al. Nucleic Acids Res.* **31**, 2952–2962 (2003).
4. Klug, A. *Q. Rev. Biophys.* **43**, 1–21 (2010).
5. Maeder, M.L. *et al. Mol. Cell* **31**, 294–301 (2008).
6. Beerli, R.R., Segal, D.J., Dreier, B. & Barbas, C.F. III. *Proc. Natl. Acad. Sci. USA* **95**, 14628–14633 (1998).
7. Moscou, M.J. & Bogdanove, A.J. *Science* **326**, 1501 (2009).
8. Boch, J. *et al. Science* **326**, 1509–1512 (2009).
9. Deng, D. *et al. Science* **335**, 720–723 (2012).
10. Mak, A.N., Bradley, P., Cernadas, R.A., Bogdanove, A.J. & Stoddard, B.L. *Science* **335**, 716–719 (2012).
11. Cong, L., Zhou, R., Kuo, Y.C., Cunniff, M. & Zhang, F. *Nat. Commun.* **3**, 968 (2012).
12. Lamb, B.M., Mercer, A.C. & Barbas, C.F. III. *Nucleic Acids Res.* **41**, 9779–9785 (2013).
13. Tsuji, S., Futaki, S. & Imanishi, M. *Biochem. Biophys. Res. Commun.* **441**, 262–265 (2013).
14. Guilinger, J.P. *et al. Nat. Methods* **11**, 429–435 (2014).
15. Esvelt, K.M., Carlson, J.C. & Liu, D.R. *Nature* **472**, 499–503 (2011).
16. Carlson, J.C., Badran, A.H., Guggiana-Nilo, D.A. & Liu, D.R. *Nat. Chem. Biol.* **10**, 216–222 (2014).
17. Dickinson, B.C., Packer, M.S., Badran, A.H. & Liu, D.R. *Nat. Commun.* **5**, 5352 (2014).
18. Hu, J.C., Kornacker, M.G. & Hochschild, A. *Methods* **20**, 80–94 (2000).
19. Choo, Y. & Klug, A. *Proc. Natl. Acad. Sci. USA* **91**, 11163–11167 (1994).



## ONLINE METHODS

**Cloning and plasmid construction.** PCR fragments for pOH, pAP, pAPNeg, pJG and SP plasmids (see **Supplementary Note 1** for plasmid design specifics) were generated using either PfuTurbo Cx Hotstart (Agilent) or VeraSeq Ultra (Enzymatics) DNA polymerases and were assembled by USER cloning (NEB) according to the manufacturer's instructions. The Q5 site-directed mutagenesis kit (NEB) was used for all site-directed mutagenesis and to produce minimized pOH plasmids (pTet). DNA encoding TALEN cleavage sites was purchased as gBlocks (IDT) and inserted into pUC19 using XbaI and HindIII restriction enzymes. Representative primer sequences used for cloning are presented in **Supplementary Table 4**.

**Phage-assisted continuous evolution (PACE) of DNA-binding domains.** In general, PACE setup was as previously described<sup>16</sup>. *E. coli* cells were maintained in chemostats containing 200 mL of Davis' rich medium (DRM)<sup>16</sup> using typical flow rates of 1–1.5 vol/h. DRM was supplemented with appropriate antibiotics to select for transformed plasmids: APs (50 µg/mL carbenicillin), APNegs (75 µg/mL spectinomycin) and MPs (25 µg/mL chloramphenicol). Lagoon dilution rates were 1.3–2 vol/h. In all PACE experiments, S1030 cells carried a mutagenesis plasmid (MP), either the previously reported pJC184 (ref. 16), or a variant of this plasmid lacking RecA, pAB086a (see **Supplementary Note 1**). Mutagenesis was induced by continuously injecting arabinose (500 mM) at a rate of 1 mL/h into each 40-mL lagoon. Typical phage titers during each PACE experiment were 10<sup>6</sup>–10<sup>8</sup> p.f.u./mL. Specific parameters for each evolution experiment are detailed below.

**Reversion of Zif268-V24R.** A lagoon receiving host cell culture from a chemostat containing S1059 cells transformed with an MP was inoculated with Zif268-V24R phage. The lagoon flow rate during drift was 2 vol/h. After 24 h of drift, phage were isolated and used to inoculate a PACE experiment with S1030 host cells carrying pAPZif268 (see **Supplementary Notes 1 and 2**) and an MP. Evolved phage were isolated after 24 h and characterized using plaque assays. DNA and protein sequences for the ω-Zif268-DBD fusion protein are shown in **Supplementary Note 3**.

**Positive selection of TALEs with altered 5' preference (5' A, C, G).** Three parallel evolution experiments were performed to evolve phage with higher affinity for 5' A, 5' C or 5' G target sequences. For each experiment, two separate lagoons receiving culture from a chemostat containing S1030 cells transformed with the appropriate AP (pAPCBXTAL:5A, pAPCBXTAL:5C, pAPCBXTAL:5G) (see **Supplementary Note 1**) and an MP were inoculated with SPCBXTAL. PACE proceeded for 48 h at a lagoon dilution rate of 1.3 vol/h before harvest and analysis of the resultant phage pools. The protein sequence for the CBX8-directed TALE-ω fusion protein is shown in **Supplementary Note 4**.

**Negative selection to generate TALEs with 5' A specificity.** Two separate lagoons receiving culture from a chemostat containing a mixed population of S1030 cells were inoculated with evolved 5' A phage from the positive-selection experiment. This *E. coli* population consisted of a 1:1:1 mixture of host cells carrying an APNeg plasmid (pAPNegCBXTAL:5C, pAPNegCBXTAL:5G, or pAPNegCBXTAL:5T) together with pAPCBXTAL:5A and an MP. Over the course of a 6-d PACE experiment, an increasing dose of theophylline was added to

each lagoon at a rate of 1 mL/h to yield increasing final theophylline lagoon concentrations of 0.1 mM, 0.2 mM and 0.3 mM (+0.1 mM theophylline every 48 h).

**Positive selection and negative selection (OffA17) of ATM-L TALE.** Two separate lagoons receiving culture from a chemostat containing a S1030 cells transformed with pAPATMTAL, pAPNegATMTAL:OffA17 and an MP were inoculated with SPATMTAL phage (see **Supplementary Note 1**). The lagoon flow rate was 1.3 vol/h. Theophylline was added to each lagoon at increasing quantities (+0.1 mM every 24 h), from a starting dose of 0 mM to a final concentration of 0.4 mM; the injection rate into each lagoon was 1 mL/h. After 120 h of PACE, phage from both lagoons were pooled and subjected to an additional 24 h of PACE at a lagoon flow rate of 2 vol/h in the presence of 0.4 mM theophylline.

**Luciferase assay.** pOH plasmids were transformed by electroporation into S1030 cells (see **Supplementary Notes 1 and 2**), and grown overnight at 37 °C on LB agar plates supplemented with 50 µg/mL carbenicillin. Single colonies were used to inoculate cultures, which were allowed to grow for ~12 h at 37 °C in DRM supplemented with 50 µg/mL carbenicillin in a shaker. Cultures were diluted to an OD<sub>600</sub> of ~0.3 and allowed to grow for an additional 2 h at 37 °C. Next, each culture was diluted 1:15 into 300 µL of DRM supplemented with 50 µg/mL carbenicillin in the presence or absence of 200 ng/mL anhydrotetracycline and incubated in a 96-well plate for an additional 4–6 h (shaking). 200-µL aliquots of each sample were then transferred to 96-well opaque plates, and luminescence and OD<sub>600</sub> readings were taken using a Tecan Infinite Pro instrument. Luminescence data were normalized to cell density by dividing by the OD<sub>600</sub> value.

**Plaque assays.** S1030 cells were transformed with the appropriate plasmids via electroporation and grown in LB medium to an OD<sub>600</sub> of 0.8–1.0. Diluted phage stock samples were prepared (10<sup>-4</sup>, 10<sup>-5</sup>, 10<sup>-6</sup>, or 10<sup>-7</sup>-fold dilution) by adding purified phage stock to 250 µL of cells in Eppendorf tubes. Next, 750 µL of warm top agar (0.75% agar in LB, maintained at 55 °C until use) was added to each tube. After mixing by pipette, each 1-mL mixture was pipetted onto one quadrant of a quartered Petri plate that had previously been prepared with 2 mL of bottom agar (1.5% agar in LB). Following solidification of the top agar, plates were incubated overnight at 37 °C before analysis. Colorimetric plaque assays were performed in parallel with regular plaque assays using S2060 cells instead of S1030 cells and used S-Gal/LB agar blend (Sigma) in place of regular LB agar.

**High-throughput analysis of TALE mutations.** PCR fragments containing evolved phage with ~500 bp of flanking sequence on either end were amplified from minipreps (Qiagen) of cells infected with evolved phage pools using the following primers: HTSFwd, 5'-TGAAAATATTGTTGATGCGCTGGCAGTGTTTC-3'; HTSRev, 5'-TAGCAGCCTTTACAGAGAGAATAACATAAAA-3. HTS preparation was performed as previously reported<sup>20</sup> using a Nextera kit (Illumina). Briefly, 4 µL of amplified DNA (2.5 ng/µL), 5 µL TD buffer and 1 µL TDE1 were mixed together and heated at 55 °C for 5 min to perform 'tagmentation'. Following DNA clean up using a Zymo-Spin column (Zymo), samples were amplified with Illumina-supplied primers according to the manufacturer's

instructions. The resulting products were purified using AMPure XP beads (Agencourt), and the final concentration of DNA was quantified by qPCR using PicoGreen (Invitrogen). Samples were sequenced on a MiSeq Sequencer (Illumina) using  $2 \times 150$  paired-end runs according to the manufacturer's protocols. Analysis of mutation frequency was performed using Matlab as previously described<sup>20</sup>. Observed background mutation frequencies were subtracted from the mutation frequencies of each experimental sample to account for DNA sequencing errors<sup>20</sup>.

**YFP assay.** pTet plasmids were cotransformed with pAPNeg plasmids by electroporation into S1030 cells (see **Supplementary Notes 1 and 2**) and grown overnight at 37 °C on LB agar plates supplemented with 50 µg/mL carbenicillin and 100 µg/mL spectinomycin. Single colonies were used to inoculate cultures, which were allowed to grow for ~12 h in antibiotic-supplemented DRM in a bacterial shaker. Cultures were diluted to an OD<sub>600</sub> of ~0.3 and allowed to grow for an additional 2 h at 37 °C. Next, each culture was diluted 1:15 into 300 µL of DRM supplemented with antibiotics and 5 mM theophylline in the presence or absence of 50 ng/mL anhydrotetracycline and incubated in a 96-well deep-well plate for an additional 4–6 h (shaking). 200-µL aliquots of each sample were then transferred to 96-well opaque plates, and YFP fluorescence ( $\lambda_{\text{ex}} = 514$  nm,  $\lambda_{\text{em}} = 527$  nm) and OD<sub>600</sub> readings were taken using a Tecan Infinite Pro instrument. Fluorescence data were normalized to cell density by dividing by the OD<sub>600</sub> value.

**In vitro TALEN cleavage assay.** *In vitro* TALEN cleavage assays were performed as previously described with slight modifications to the procedure<sup>14</sup>. Briefly, 1 µg of each TALEN-encoding plasmid (pJG) was added individually to 20 µL of methionine-supplemented T7-TnT coupled transcription/translation system (Promega) lysate and incubated for 1.5 h at 30 °C. Determination of protein concentrations and preparation of linear DNA for TALEN cleavage was performed as previously reported<sup>14</sup>. Each reaction consisted of 50 ng of amplified DNA, 12 µL NEB buffer 3,

3 µL of each *in vitro*-transcribed/translated TALEN left and right monomers (corresponding to ~15 nM final TALEN concentration), and 6 µL of empty lysate brought up to a final volume of 120 µL in distilled water. The digestion reaction was allowed to proceed for 30 min at 37 °C (or 1 h where indicated), and then incubated with 1 µg/µL RNase A (Qiagen) for 2 min before being purified using a Minielute column (Qiagen). Reactions were subsequently run in a 5% TBE Criterion PAGE gel (Bio-Rad), and stained with 1× Sybr Gold (Invitrogen) for 10 min. Gels were imaged using a Syngene G:BOX Chemi XRQ, and densitometry was performed using GelEval 1.37 software.

**High-throughput specificity profiling assay.** High-throughput specificity profiling of canonical and evolved TALEN pairs and subsequent data analysis was performed as previously described<sup>14</sup>.

**TALEN cleavage in HEK 293 and U2OS cells.** pJG29 and pJG30 plasmids (see **Supplementary Note 1**) were transfected into HEK 293 cells (a cell line that has a high transfection efficiency; obtained from ATCC) using Lipofect (Signagen) according to the manufacturer's instructions. pJG51 and pJG52 plasmids were electroporated into U2OS cells as previously described<sup>14</sup>. For both sets of experiments, genomic DNA isolation was performed as previously reported<sup>14,21</sup>. Primers for amplifying on and off-target genomic sites are listed in **Supplementary Table 4**. Illumina adaptor ligation, AMPure XP bead cleanup (Agencourt), sequencing and post-analysis were performed as previously described<sup>14,21</sup>. The HEK 293 cells used in this study tested negative for mycoplasma at the time of purchase, and the U2OS cell line was previously authenticated and shown to be negative for mycoplasma contamination<sup>14</sup>.

20. Dickinson, B.C., Leconte, A.M., Allen, B., Esvelt, K.M. & Liu, D.R. *Proc. Natl. Acad. Sci. USA* **110**, 9007–9012 (2013).

21. Zuris, J.A. *et al. Nat. Biotechnol.* **33**, 73–80 (2015).