

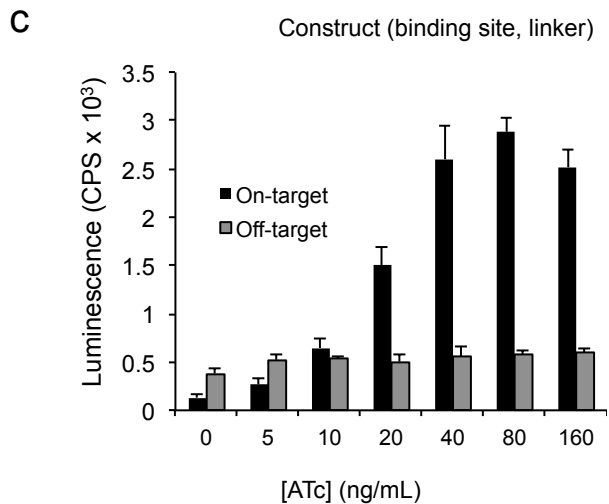
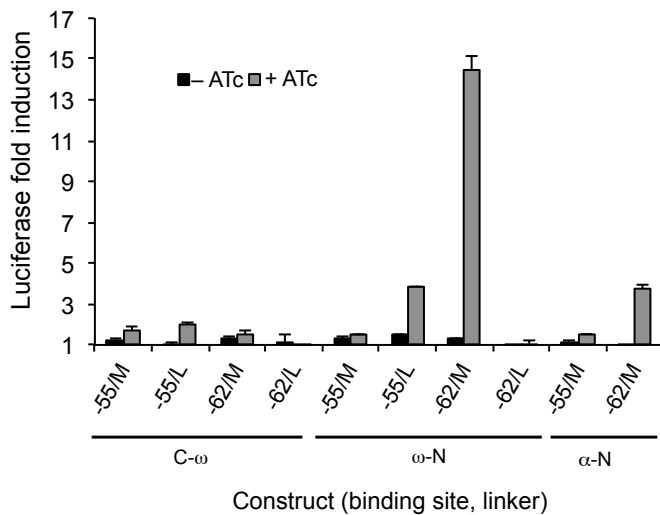
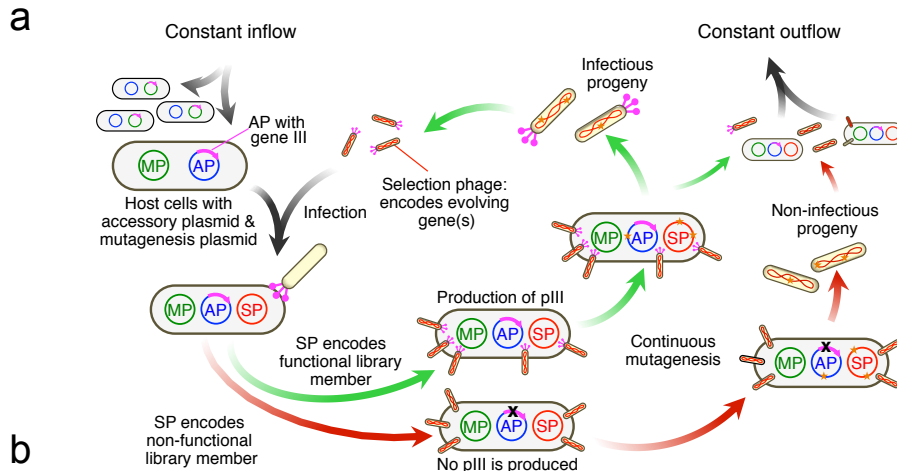
Continuous directed evolution of DNA-binding proteins to improve TALEN specificity

Basil P. Hubbard, Ahmed H. Badran, John A. Zuris, John P. Guillinger, Kevin M. Davis, Liwei Chen, Shengdar Q. Tsai, Jeffry D. Sander, J. Keith Joung, and David R. Liu

SUPPLEMENTARY INFORMATION

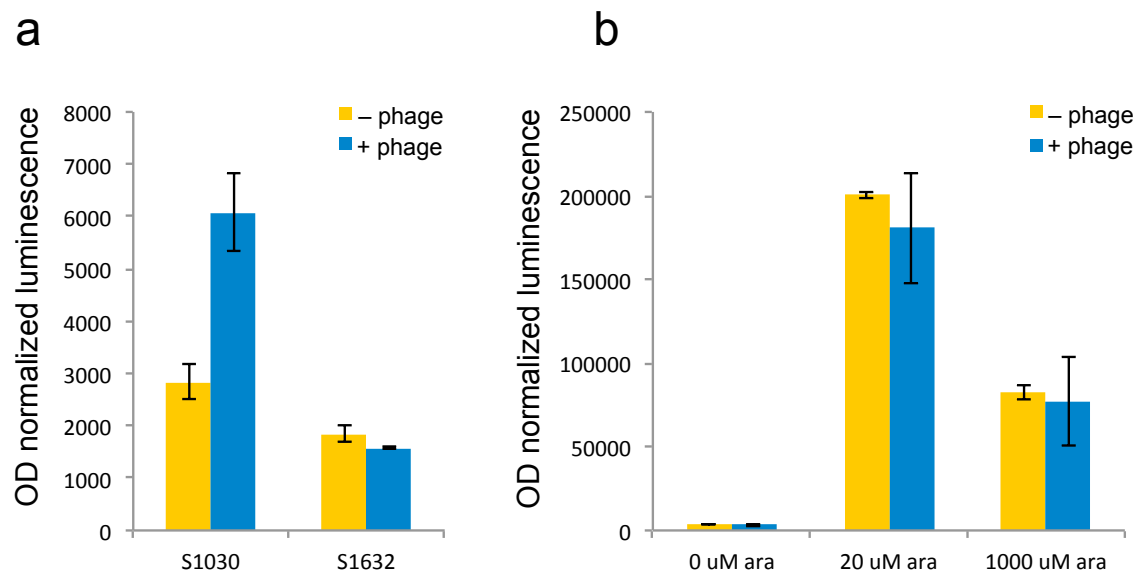
| | |
|--------------------------------|--|
| Supplementary Figure 1 | Diagram of PACE and optimization of a one-hybrid architecture for PACE |
| Supplementary Figure 2 | Chromosomal <i>pspBC</i> deletion enables small-molecule control of the phage shock promoter response |
| Supplementary Figure 3 | Generation of mutant PSP variants with altered dynamic range |
| Supplementary Figure 4 | Generation of S2060, a bacterial strain for chaperone overexpression and robust visualization of phage plaques |
| Supplementary Figure 5 | <i>In vitro</i> and continuous propagation of Zif268 SP, and reversion of an inactive Zif268 mutant to wild-type |
| Supplementary Figure 6 | Optimization of a TALE one-hybrid architecture for PACE |
| Supplementary Figure 7 | <i>In vitro</i> and continuous propagation of <i>CBX8</i> -targeting TALE SP, and assay of 5' specificity |
| Supplementary Figure 8 | Scheme of <i>CBX8</i> -targeting TALE and cognate binding sequence |
| Supplementary Figure 9 | Evolution of <i>CBX8</i> -TALE variants with increased activity towards 5' A, C, or G sequences |
| Supplementary Figure 10 | High-throughput sequence analysis of phage populations evolved to bind target sequences beginning with 5' A, C, or G |
| Supplementary Figure 11 | Characterization of mutations arising from evolution towards 5' A, C, or G target sequence binding |
| Supplementary Figure 12 | Specificity of phage evolved to recognize 5' A, C, or G, and negative selection validation |
| Supplementary Figure 13 | Genotypes and phenotypes of evolved clones following negative selection of 5' A-evolved phage against 5' C, G, and T binding |
| Supplementary Figure 14 | Characterization of mutations arising from negative selection PACE against target sequences beginning with 5' C, G, or T |
| Supplementary Figure 15 | Comparison of on-target cleavage efficiency of canonical and evolved L3-2 TALENs |

| | |
|---------------------------------|--|
| Supplementary Figure 16 | Characterization of evolved <i>ATM-L</i> TALEs following positive and negative selection PACE |
| Supplementary Figure 17 | Characterization of mutations identified in positive and negative selection <i>ATM-L</i> TALE PACE and evolved TALEN specificity |
| Supplementary Figure 18 | Global analysis of <i>in vitro</i> TALEN specificity |
| Supplementary Figure 19 | Specificity profile heat maps for the canonical <i>ATM</i> TALEN pair as a function of concentration |
| Supplementary Figure 20 | Specificity profile bar graphs of the canonical <i>ATM</i> TALEN pair as a function of concentration |
| Supplementary Figure 21 | Specificity profile heat maps of L2-1, L3-1, and L3-2 <i>ATM</i> TALEN pairs |
| Supplementary Figure 22 | Specificity profile bar graphs of L2-1, L3-1, and L3-2 <i>ATM</i> TALEN pairs |
| Supplementary Figure 23 | Specificity profile difference as a function of TALEN concentration for canonical and L3-1 <i>ATM</i> TALEN pairs |
| Supplementary Figure 24 | Bar graph showing difference in specificity of the canonical TALEN pair versus the L2-1, L3-1 and L3-2 TALEN pairs |
| Supplementary Figure 25 | Specificity of a <i>CCR5</i> -targeting TALE in the DB-PACE one-hybrid system |
| Supplementary Table 1 | Full target sequences used to study <i>ATM</i> TALENs |
| Supplementary Table 2 | Statistics of sequences selected by TALEN digestion |
| Supplementary Table 3 | Cellular modification rates, sample sizes and <i>P</i> values for high-throughput sequencing of TALEN cleavage in U2OS and 293 cells |
| Supplementary Table 4 | Primers sequences used for plasmid cloning and for high-throughput sequencing |
| Supplementary Results | |
| Supplementary Discussion | |
| Supplementary Note 1 | Summary of plasmid constructs |
| Supplementary Note 2 | Genotypes of bacterial strains used |
| Supplementary Note 3 | DNA and protein coding sequences for the ω -Zif268-DBD fusion protein |
| Supplementary Note 4 | Coding sequences for the <i>CBX8</i> - and <i>ATM-L</i> -directed TALE- ω fusion proteins |
| Supplementary References | |

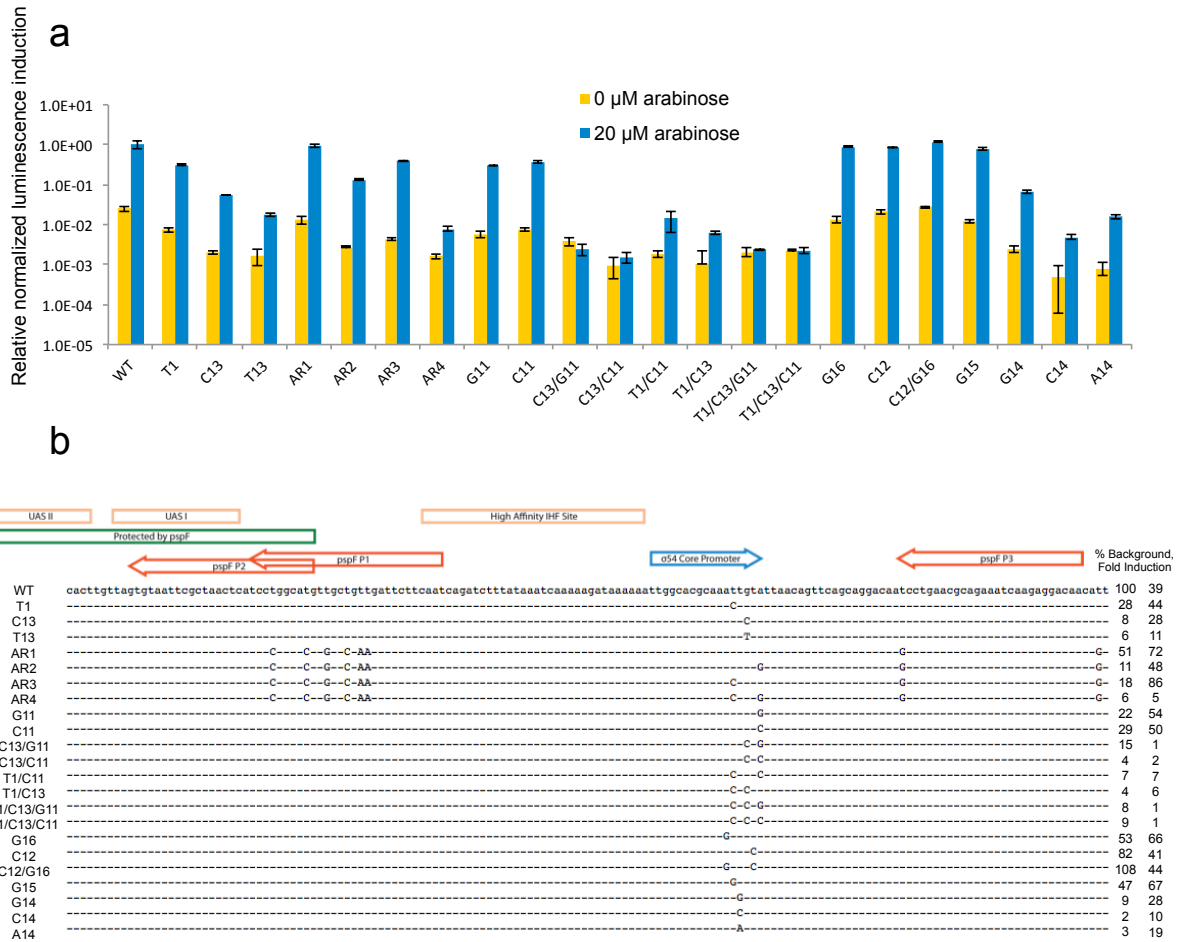


Supplementary Figure 1. Diagram of PACE and optimization of a one-hybrid architecture for PACE. (a) Diagram of PACE. (b) Comparison of pIII-luciferase fold induction (ATc-induced Zif268 expression / non-induced luminescence) resulting from

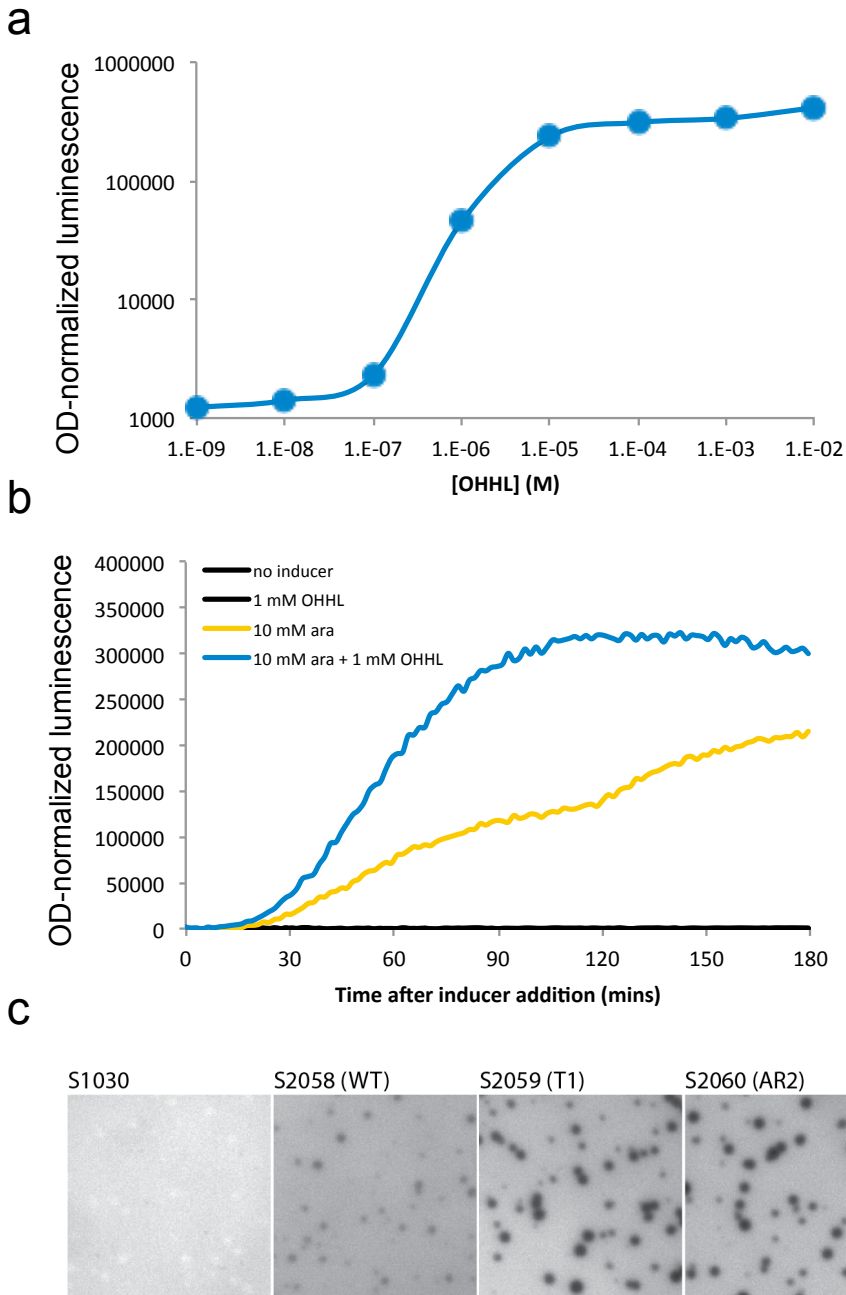
binding of a Zif268 fusion with either the α or ω subunit of RNAP to a Zif268 operator sequence (5'-GCGTGGGCG-3') centered at either -55 or -62. M refers to a medium-length linker between Zif268 and the RNAP subunit (AAATSGGGGAA), and L refers to a longer linker (AAGGGGSGGGGSGGGGSTAAA). (c) Luciferase activity resulting from ATc-induced Zif268 protein binding to either its on-target sequence (5'-GCGTGGGCG-3') or an off-target sequence (5'-GCGTTAGCG-3'). Data represent mean + s.d. ($n = 3$).



Supplementary Figure 2. Chromosomal *pspBC* deletion enables small-molecule control of the phage shock promoter response. (a) Comparison of phage-shock promoter response between S1030 and S1632 cells (see **Supplementary Note 2**). Upon phage infection, activation of a phage shock promoter (PSP) induces bacterial luciferase expression, and can be measured as an increase in luminescence. The phage shock response sensors *pspBC* were deleted from S1632 cells, resulting in no transcriptional activation in the absence or presence of infecting phage. (b) Over-expression of *pspBC* from an arabinose-controlled promoter (P_{BAD}) results in activation of the PSP in a manner independent of phage infection, eliminating variability in transcriptional activation of the promoter. Data represent mean \pm s.d. ($n = 3$).

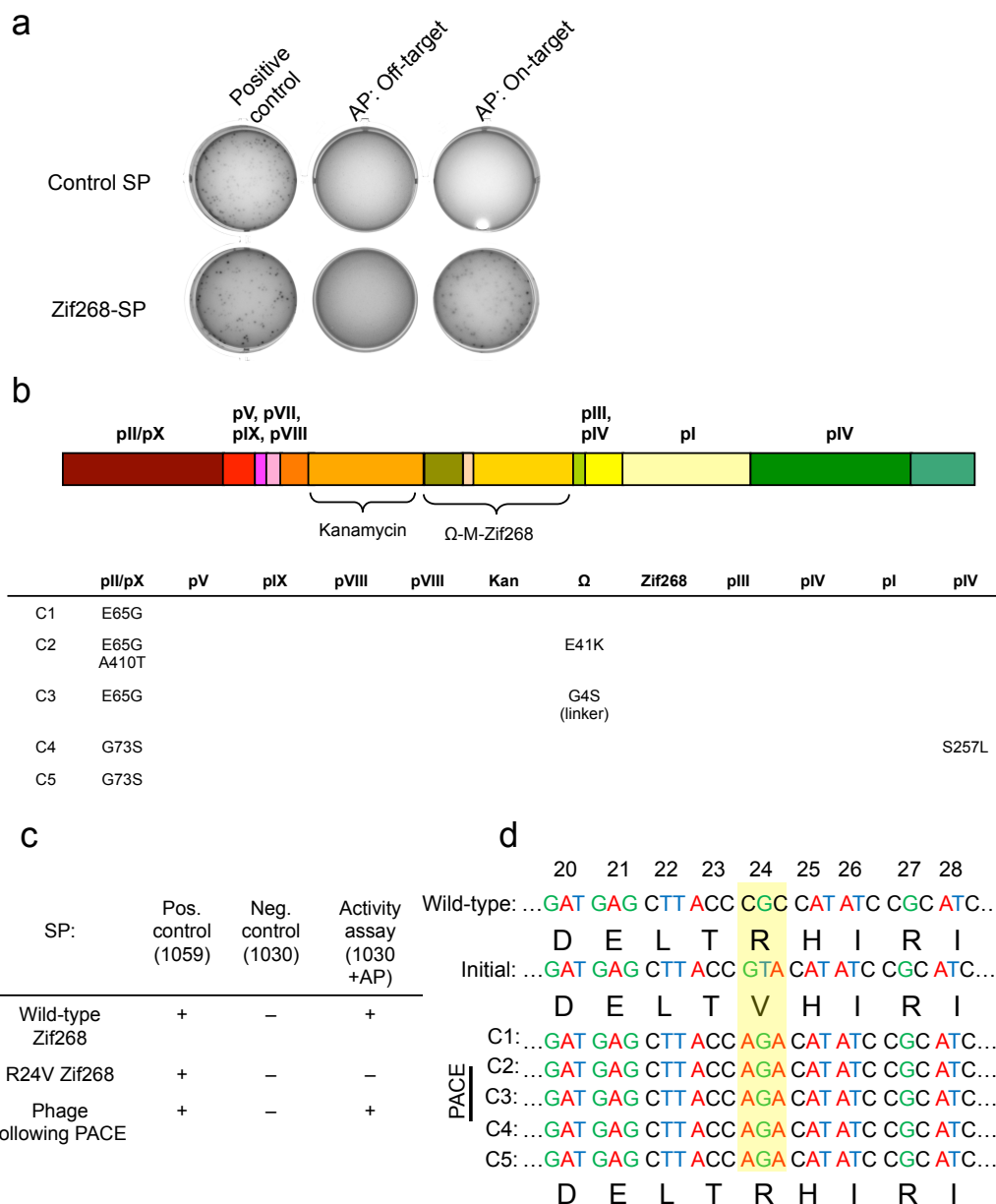


Supplementary Figure 3. Generation of mutant PSP variants with altered dynamic range. Mutants abrogating the efficiency or background transcription of the PSP were constructed and tested through low-level expression of the phage shock sensors *pspBC*, which are master inducers of the phage shock response. Mutations were constructed based on prior knowledge of the PSP architecture¹ and $\sigma 54$ promoter activities². Generally, mutations were focused on the $\sigma 54$ core promoter. The “AR” series carried additional mutations to reduce the strength of $\sigma 70$ cryptic promoters that may influence background transcription levels. (a) Luminescence signal in the presence or absence of 20 μ M arabinose from wild-type and mutant PSP promoters. All readings were normalized to wild-type PSP, which was set to 1. Data represent mean \pm s.d. ($n = 3$). (b) Summary of activity, background levels, and genotypes of mutant promoters assayed in (a). Background levels of all mutant promoters are listed relative to wild-type.



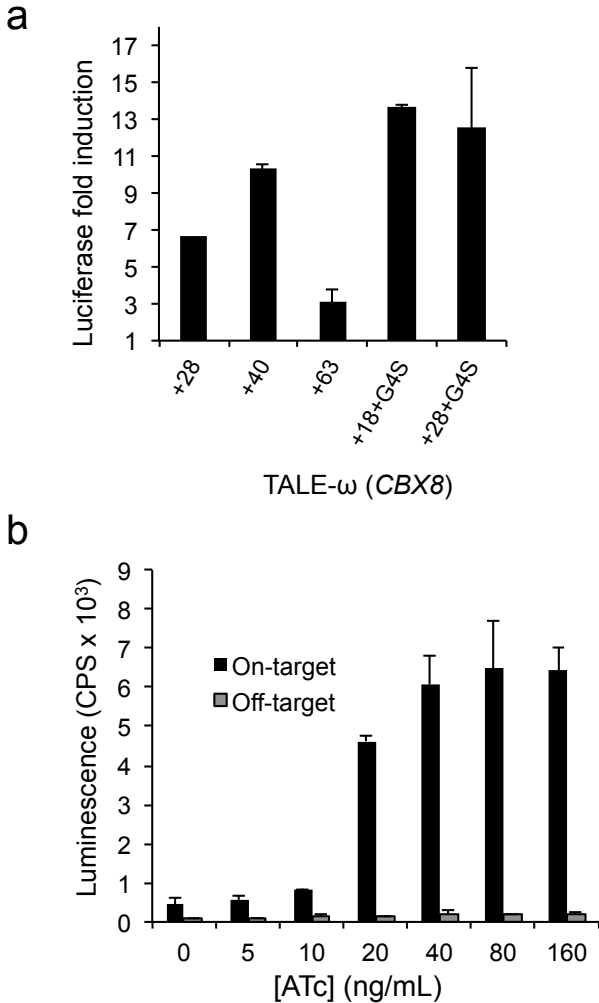
Supplementary Figure 4. Generation of S2060, a bacterial strain for chaperone overexpression and robust visualization of phage plaques. (a) Luminescence resulting from induction of a bacterial luciferase (*luxAB*) cassette driven by the P_{lux} promoter in response to the indicated doses of N-(3-oxohexanoyl)-L-homoserine lactone (OHHL) (the LuxR transcriptional regulator is also controlled by the P_{lux} promoter, only in the opposite direction). Data represent mean \pm s.d. ($n = 3$). (b) Kinetic analysis of OHHL-mediated expression of GroESL (cassette: *luxR-P_{lux}-groESL*) on the folding of LuxAB (cassette: *araC-P_{BAD}-LuxAB*), a known substrate for GroESL. Increased *in vivo*

concentrations of GroESL result in improved folding of LuxAB and rapid saturation of the luminescence response. (c) Comparison of the ability to visualize plaque formation using S1030, S2058, S2059, and S2060 cells. Chromosomally identical strains lacking (S1030) or carrying the *lacZ* and *groESL* cassettes (S2058, S2059, S2060) were infected with WT M13 bacteriophage. The modified strains carry the wild-type (WT) PSP, PSP-T1 or PSP-AR2, respectively. The reduced background and maintained transcriptional activation of the T1 and AR2 variants enables the visualization of phage plaques in top agar supplemented with Blu-Gal, an X-Gal derivative.

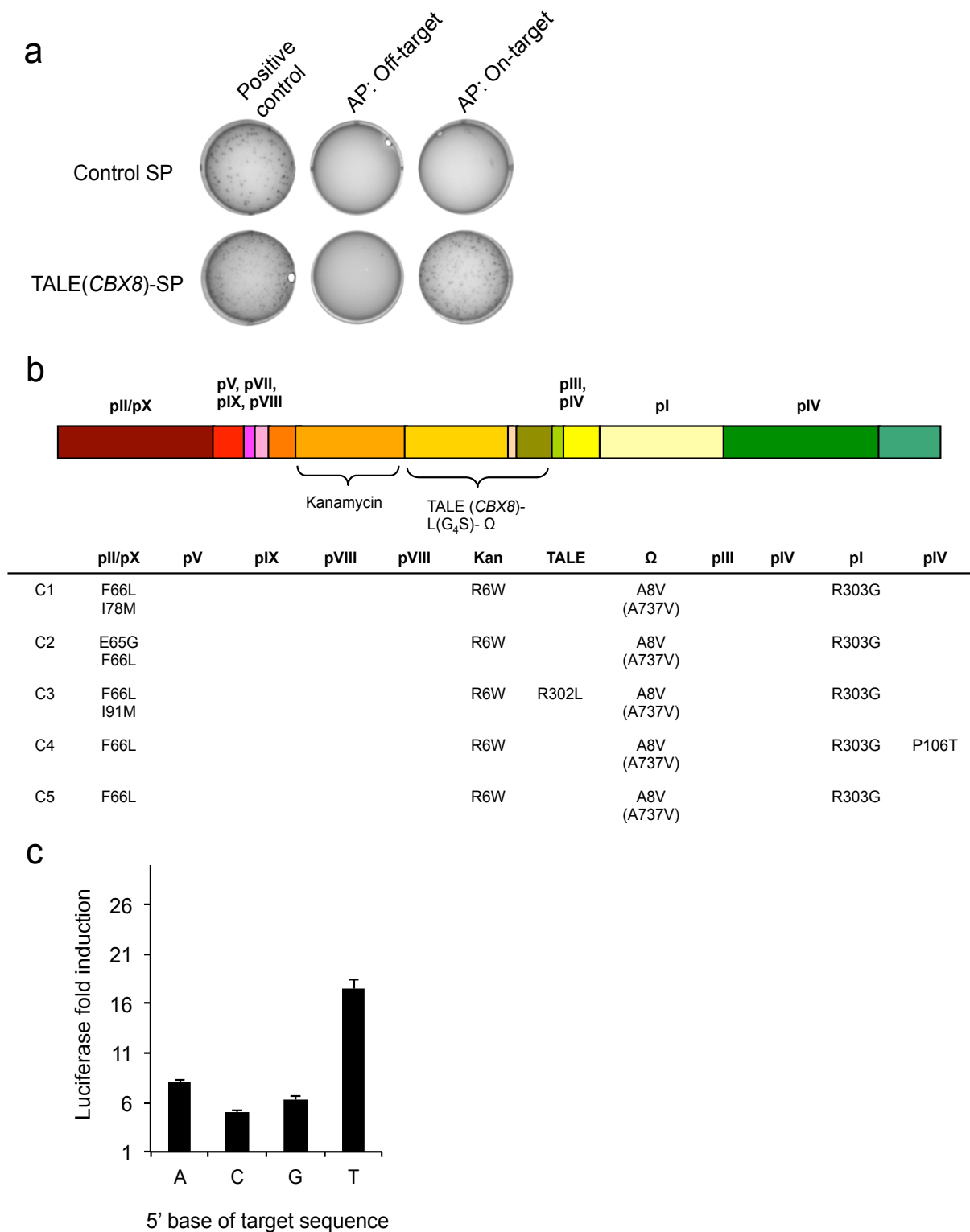


Supplementary Figure 5. *In vitro* and continuous propagation of Zif268 SP, and reversion of an inactive Zif268 mutant to wild-type. (a) Plaque assays of Zif268-SP or a control SP encoding T7 RNAP instead of Zif268 on S2060 cells containing APs encoding either the on- or off-target sequence, or S2208 cells (positive control; see **Supplementary Note 2** for genotype information) (b) Schematic of the relative location of genes in the Zif268-SP, and a summary of mutations arising following 24 h of PACE to optimize the phage backbone and one-hybrid system. (c) Plaque assay results for wild-type Zif268-SP, inactive mutant Zif268-R24V-SP, and evolved SPs derived from a 24 h drift/24 h PACE experiment in the presence of mutagenesis. ‘+’ denotes the

presence of plaques, while ‘–’ denotes the absence of plaques. (d) Genotypes of five phage clones isolated following PACE, all displaying reversion of V24 to R.

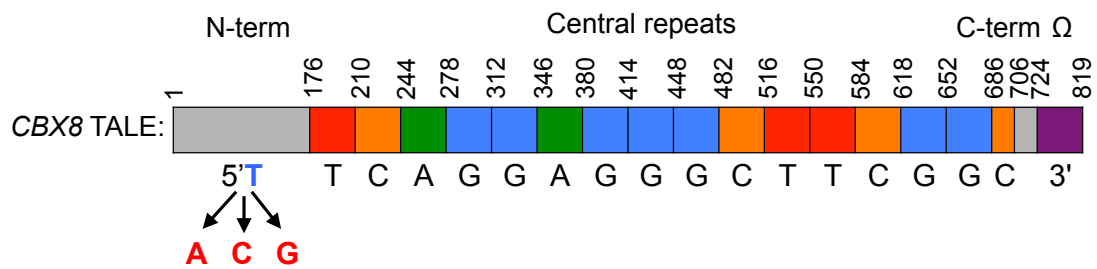


Supplementary Figure 6. Optimization of a TALE one-hybrid architecture for PACE. (a) Comparison of pIII-luciferase fold induction (ATc-induced TALE / non-induced luminescence) resulting from binding of a *CBX8*-targeting TALE- ω fusion construct to the cognate operator sequence (5'-TTCAGGAGGGCTTCGGC-3') centered at -62. The length of the natural TALE C-terminus used as a linker to the ω subunit is indicated. G4S represents the addition of a GGGGS sequence to the end of the C-terminal fragment to increase the flexibility of the linker. (b) Luciferase activity resulting from ATc-induced TALE protein binding to either its on-target sequence (*CBX8*: 5'-TTCAGGAGGGCTTCGGC-3') or an off-target sequence (5' - TTCATAAGGGATTAGGC-3'). Data represent mean + s.d. ($n = 3$).



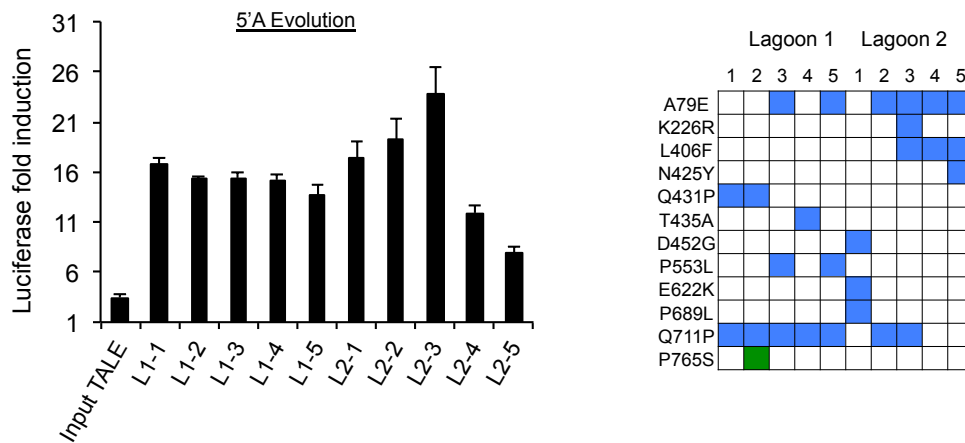
Supplementary Figure 7. *In vitro* and continuous propagation of CBX8-targeting TALE SP, and assay of 5' specificity. (a) Plaque assays of TALE-SP or a control SP encoding T7 RNA polymerase instead of a TALE on S2060 cells containing APs with

either the on- or off-target sequence, or on S2208 cells (positive control). (b) Schematic of the location of genes contained in a *CBX8*-TALE-SP plasmid, and a summary of evolved mutations following 24 h of PACE. (c) pIII-luciferase fold induction (ATc-induced TALE / non-induced luminescence) by binding of a *CBX8*-targeting TALE- ω to *CBX8*-binding sequences beginning with 5' A, C, G, or T. Data represent mean + s.d. ($n = 3$).

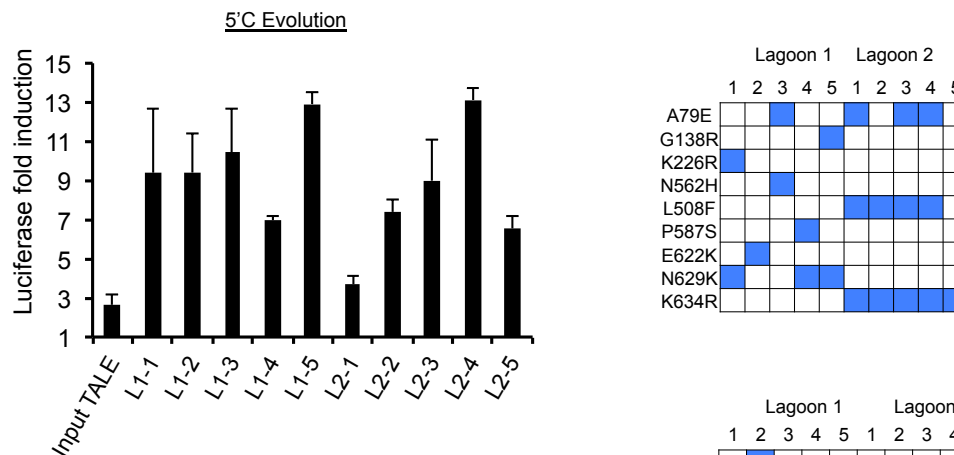


Supplementary Figure 8. Scheme of *CBX8*-targeting TALE and cognate binding sequence.

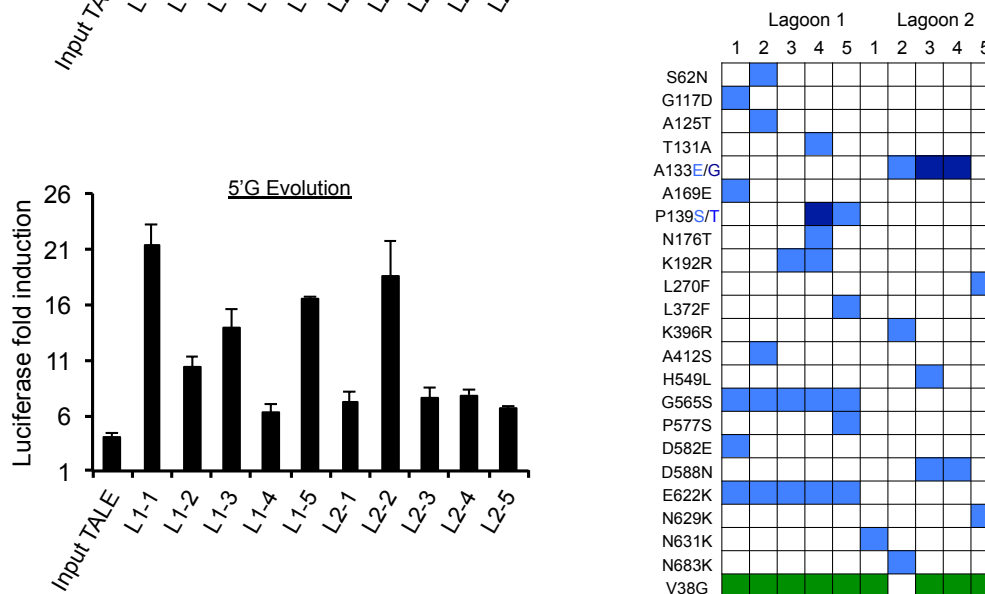
a



b



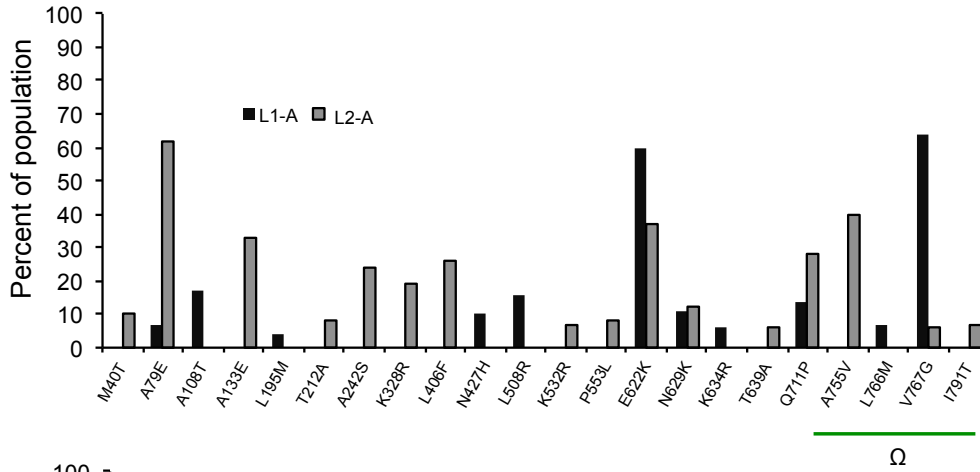
c



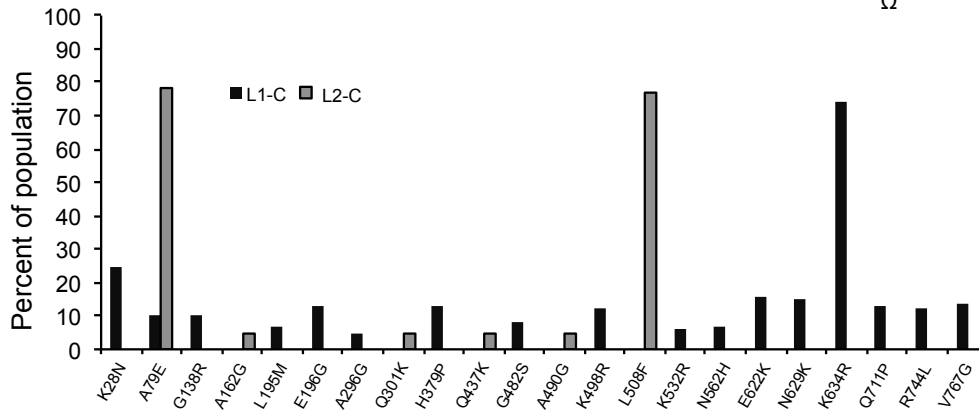
Supplementary Figure 9. Evolution of CBX8-TALE variants with increased activity towards 5' A, C, or G sequences. (a) Left: luciferase activity shown as fold induction (ATc-induced TALE luminescence / non-induced luminescence) for the canonical TALE

(input TALE) and five PACE-evolved clones from either lagoon 1 (L1) or lagoon 2 (L2) evolved to bind a *CBX8*-target sequence beginning with 5' A. Right panel: genotypes for the clones shown in the left panel. (b) Left: same as in (a), but with clones evolved to bind a *CBX8*-target sequence beginning with 5' C. Right: genotypes for the clones shown in the left panel. (c) Left: same as in (a), but with clones evolved to bind a *CBX8*-target sequence beginning with 5' G. Right: genotypes for the clones shown in the left panel. For (a), (b), and (c), blue squares indicate mutations within the TALE domain, and green squares indicate mutations within the ω subunit, and data show mean + s.d. ($n = 3$).

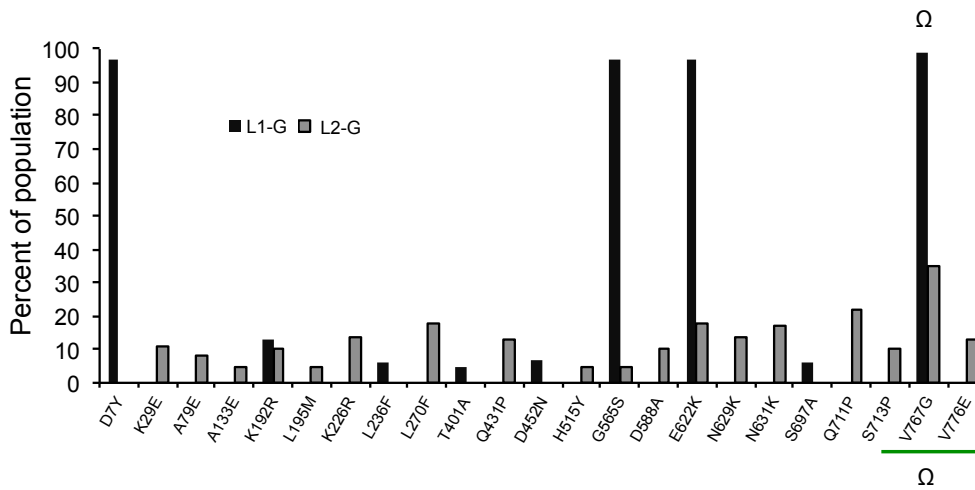
a



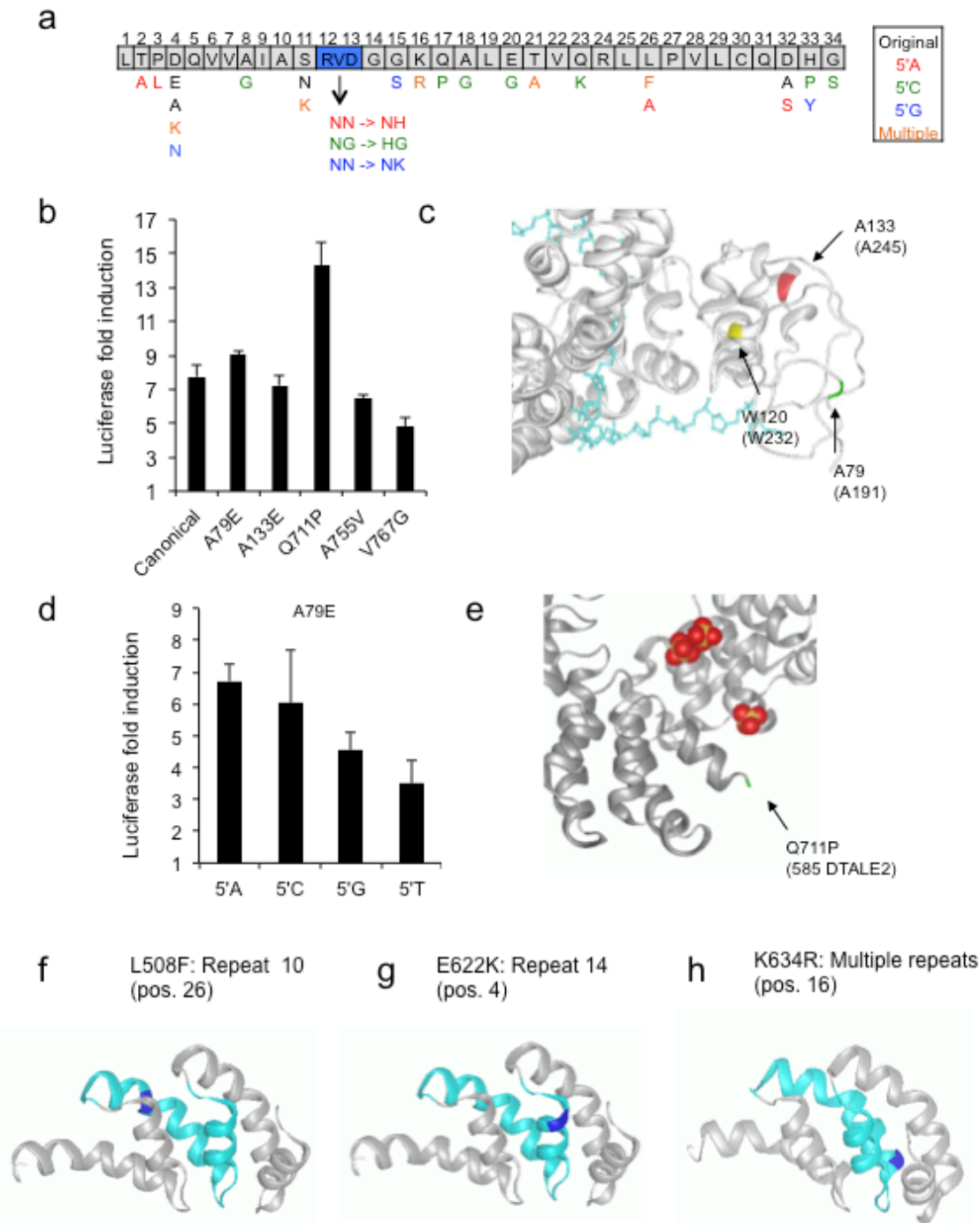
b



c



Supplementary Figure 10. High-throughput sequence analysis of phage populations evolved to bind target sequences beginning with 5' A, C, or G. Frequency of mutations arising in lagoon 1 (L1) or lagoon 2 (L2) following 48 h of PACE in the presence of mutagenesis on *CBX8*-directed target sequences beginning with (a) 5' A, (b) 5' C, or (c) 5' G. Only mutations observed at a frequency $\geq 5\%$ are shown.




Supplementary Figure 11. Characterization of mutations arising from evolution towards 5' A, C, or G target sequence binding. (a) Location of mutations with > 5% prevalence in the population identified in 5' A, C, or G evolutions within the core TALE unit. 'Multiple' refers to equivalent mutations identified in multiple different repeats either in the same experiment or in a separate experiment. (b) Luciferase activity represented as fold induction (ATc-induced TALE luminescence / non-induced luminescence) for the canonical TALE and five mutant constructs using a 5' A *CBX8*-target sequence. (c)

Crystal structure³ showing the location of A133 (red), A79 (green), and W120 (yellow). The corresponding number for each residue in the crystal structure³ is shown in parenthesis. (d) Luciferase activity represented as fold induction (ATc-induced TALE luminescence / non-induced luminescence) for the a *CBX8*-directed TALE with an A79E mutation on *CBX8*-directed target sequences beginning with 5' A, C, G, or T. (e) Crystal structure⁴ indicating the position of the C-terminal residue Q711 (green). Numbering corresponding to the original crystal structure⁴ is shown in parenthesis. Crystal structure⁴ of three TALE repeats showing the positions of the (f) L508, (g) E622, and (h) K634 residues within a core TALE repeat (repeat in light blue, residues in dark blue). Bar graphs in (b) and (d) represent mean + s.d. ($n = 3$).

a


| SP: | Positive control (1059) | Negative control (S1030) | Activity dependent (5'A) | Activity dependent (5'C) | Activity dependent (5'G) | Activity dependent (5'T) |
|-------------------|-------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Initial phage | +++ | – | + | – | – | +++ |
| Evolved 5'A phage | +++ | – | +++ | +++ | +++ | +++ |
| Evolved 5'C phage | +++ | – | +++ | +++ | +++ | +++ |
| Evolved 5'G phage | +++ | – | +++ | +++ | +++ | +++ |

b

mM Theophylline 

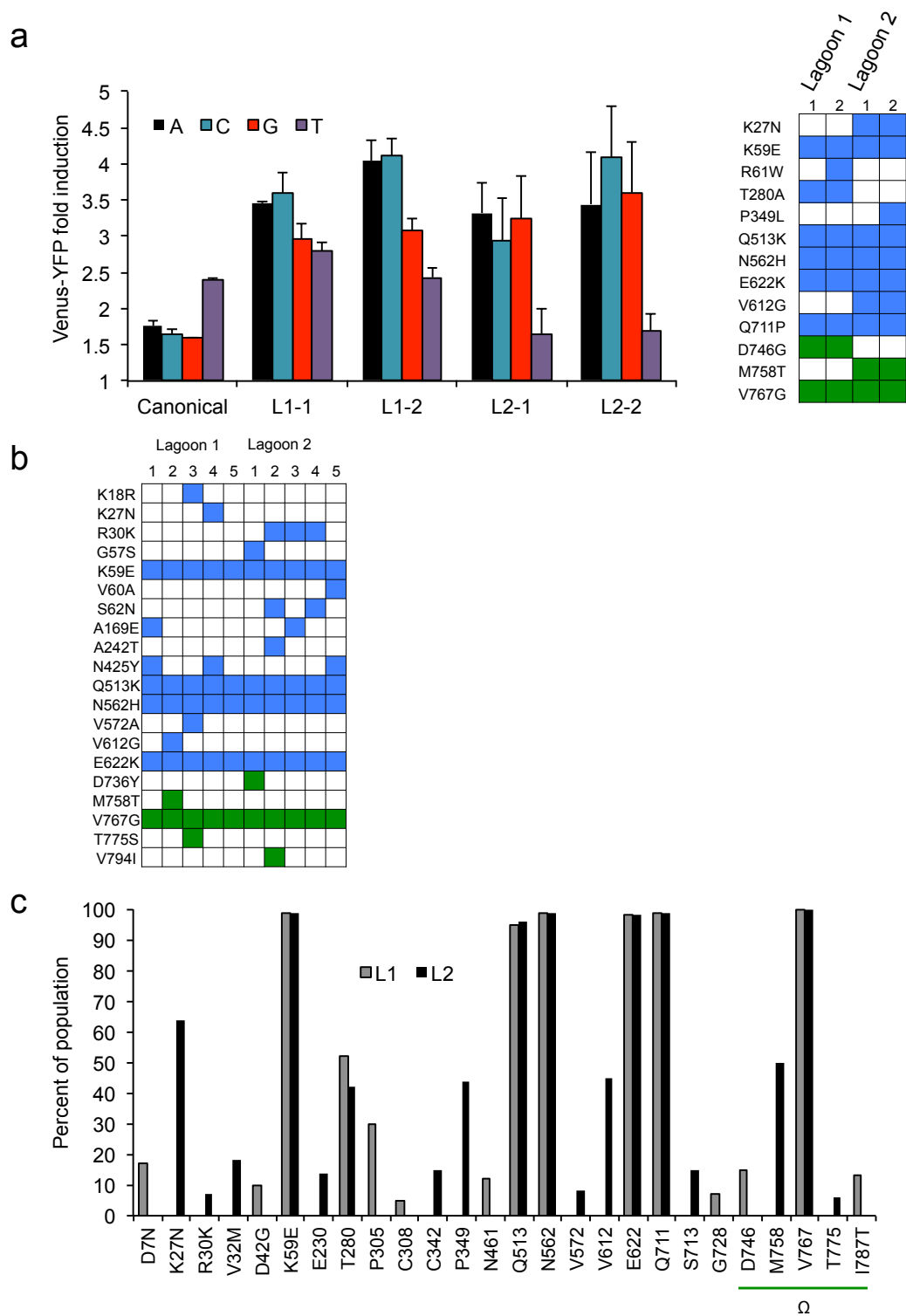
| | 0 | 0.01 | 0.1 | 1 |
|----------------------|-----|------|-----|---|
| Ap5'A/ApNeg5'T | +++ | +++ | + | – |
| Ap5'A/ApNeg5'C | +++ | +++ | ++ | – |
| Ap5'A/ApNeg5'G | +++ | +++ | ++ | – |
| Ap5'A/ApNegScr. | +++ | +++ | +++ | – |
| S1030 (Neg. control) | – | – | – | – |

c

mM Theophylline 

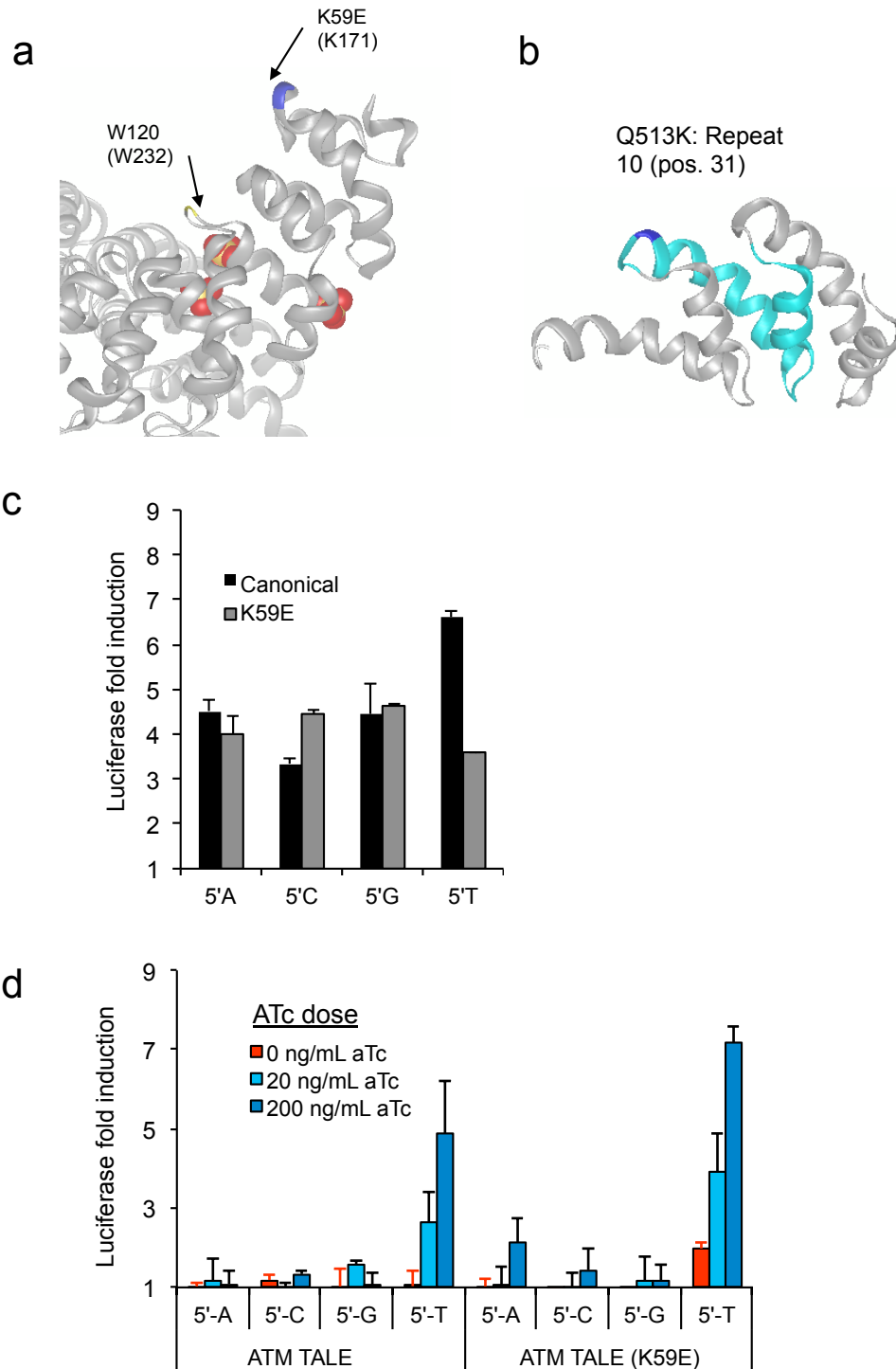
| | 0 | 0.1 | 0.2 | 0.4 | 0.8 |
|----------------------|-----|-----|-----|-----|-----|
| Ap5'A/ApNeg5'T | +++ | + | – | – | – |
| Ap5'A/ApNeg5'C | +++ | ++ | + | – | – |
| Ap5'A/ApNeg5'G | +++ | ++ | + | – | – |
| Ap5'A/ApNegScr. | +++ | +++ | +++ | +++ | + |
| S1030 (Neg. control) | – | – | – | – | – |

Supplementary Figure 12. Specificity of phage evolved to recognize 5' A, C, or G, and negative selection validation. (a) Results of plaque assays of phage pools evolved on *CBX8*-directed target sequences beginning with 5' A, C, or, G on S1059 cells (positive control), or S1030 cells carrying no APs (negative control), or S1030 cells carrying AP containing target sequences beginning with 5' A, C, G, or T. (b) Plaque assays using phage evolved to bind a 5' A target sequence on S1030 cells carrying the indicated combinations of AP/APNeg plasmids in the presence of increasing doses of theophylline. (c) Results of similar experiment to (b) using different doses of theophylline. '–' indicates no plaque formation, '+' indicates weak plaque formation, '++' indicates moderate plaque formation, and '+++ indicates strong phage plaque formation.



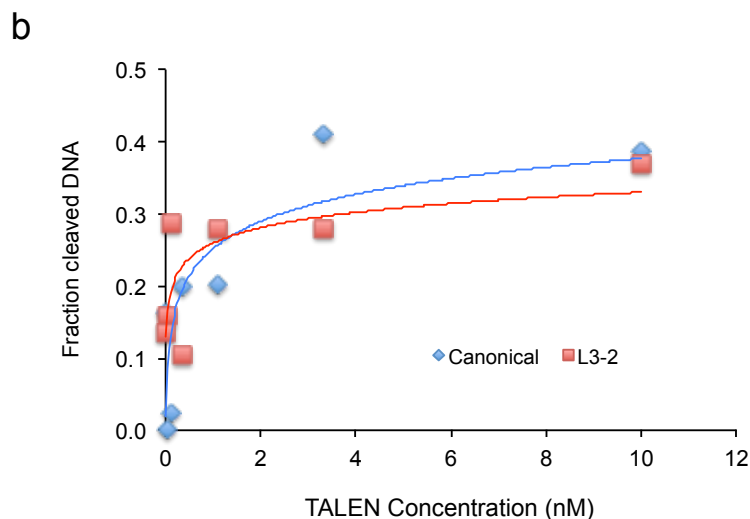
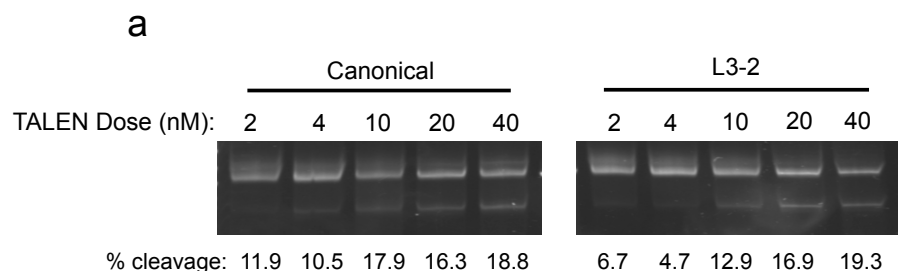
Supplementary Figure 13. Genotypes and phenotypes of evolved clones following negative selection of 5' A-evolved phage against 5' C, G, and T binding. (a) Left:

YFP fluorescence represented as fold induction upon induction of TALE expression, for the canonical TALE and two evolved clones from either L1 or L2 using *CBX8*-target sequences beginning with 5' A, 5' C, 5' G, or 5' T, mean + s.d. ($n = 3$) is shown. Right: mutations in the evolved proteins shown in the left panel are shown. (b) Genotypes of five evolved phage clones from lagoon 1 or lagoon 2 following 144 h of PACE under positive selection for 5' A binding, and negative selection against *CBX8*-target sequences beginning with 5' C, G, or T. (c) Mutations arising in lagoon 1 (L1) or lagoon 2 (L2) following 144 h of dual positive and negative selection PACE. Only mutations arising at a frequency of > 5% are shown. For (a) and (b), blue squares indicate mutations within the TALE domain, and green squares indicate mutations within the ω subunit.

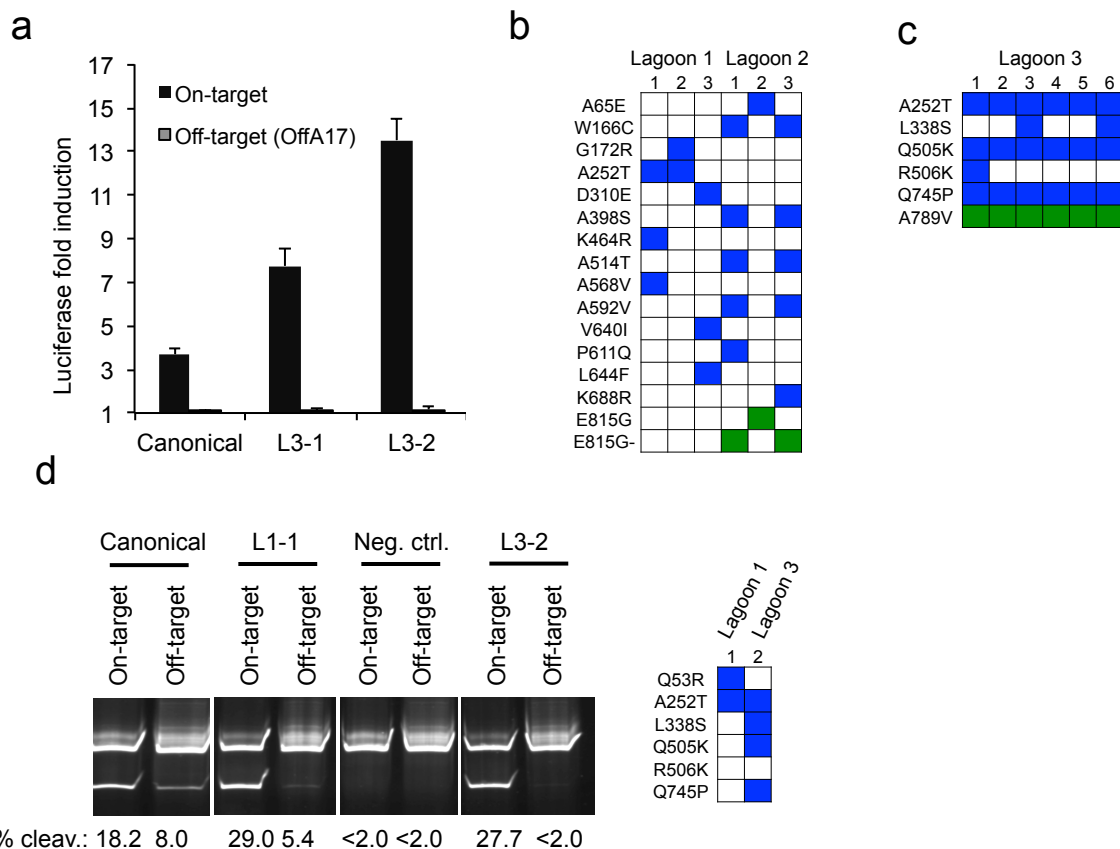


Supplementary Figure 14. Characterization of mutations arising from negative selection PACE against target sequences beginning with 5' C, G, or T. (a) Crystal structure⁴ showing the location of K59 (blue) and W120 (yellow). The corresponding number for each residue in the crystal structure is shown in parenthesis. (b) Crystal structure⁴ of three TALE repeats showing the relative position of the Q513 (repeat in

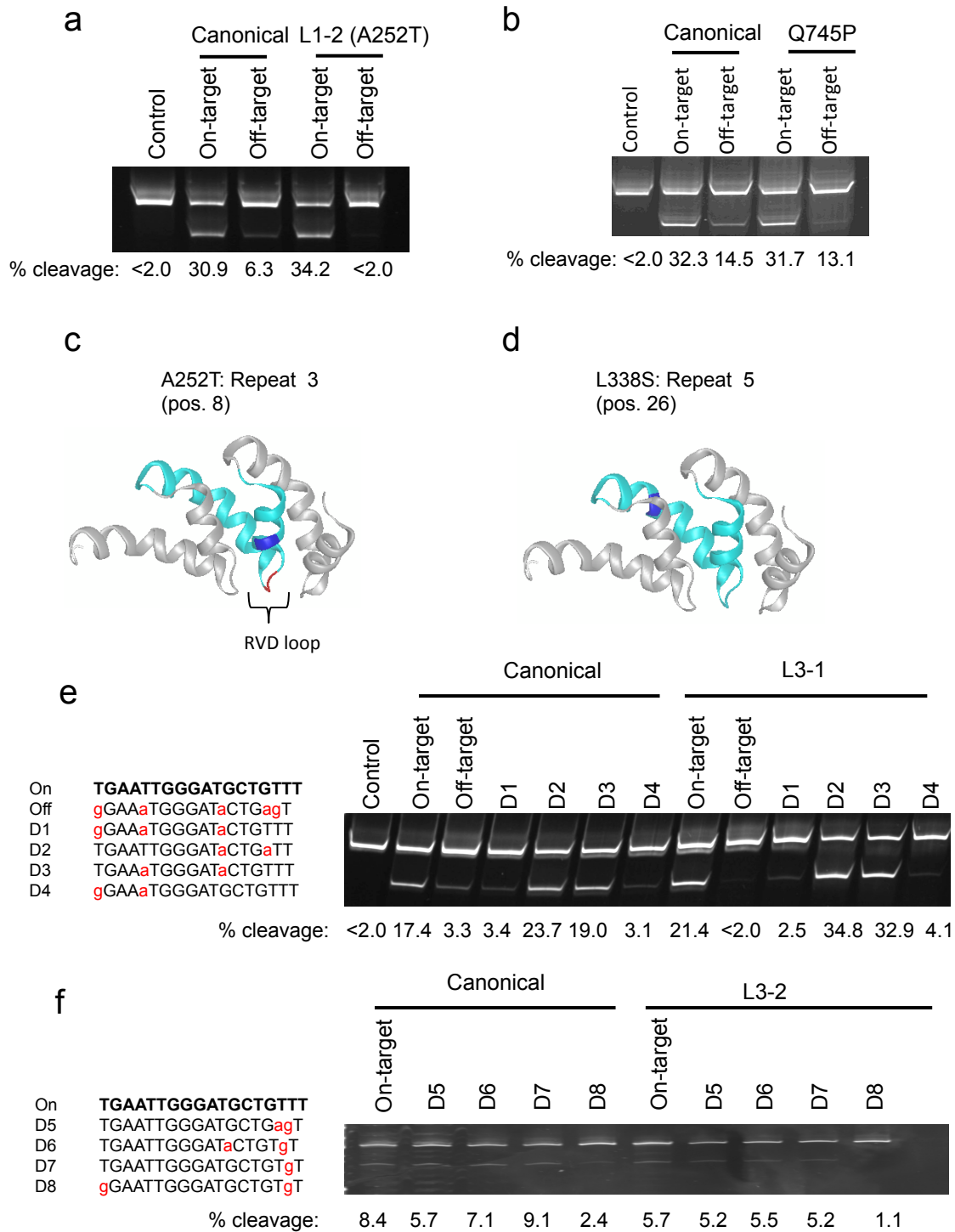
light blue, residue in dark blue). (c) Luciferase activity represented as fold induction (ATc-induced TALE luminescence / non-induced luminescence) for the canonical *CBX8*-directed TALE or a K59E mutant protein on *CBX8*-directed target sequences beginning with 5' A, C, G, or T. (d) Luciferase activity represented as fold induction (ATc-induced TALE luminescence / non-induced luminescence for the indicated doses of ATc) for the canonical *ATM-L*-directed TALE or a K59E mutant protein on *ATM-L*-directed target sequences beginning with 5' A, C, G, or T. Data represent mean + s.d ($n = 3$).



Supplementary Figure 15. Comparison of on-target cleavage efficiency of canonical and evolved L3-2 TALENs. (a) TALEN dose titration showing the relative cleavage efficiencies of the canonical *ATM* TALEN pair or the L3-2 TALEN on 50 ng (~0.75 nM) of a linear 6-kb DNA fragment containing the *ATM* on-target sequence (*ATM*: 5'-TGAATTGGGATGCTGTTT-3'). The top band is non-cleaved DNA, while the bottom band is a cleavage product. Quantified cleavage percentages were determined using densitometry (GelEval), and are shown below each lane. (b) DNA cleavage saturation curves for the canonical *ATM* TALEN pair and the TALEN pair containing the evolved L3-2 TALE. An *in vitro* cleavage assay was performed to measure DNA cleavage of 0.5 ng of DNA containing the *ATM* on-target sequence (~7.5 pM) by either the canonical TALEN pair or the L3-2 TALEN pair at concentrations of 0.01, 0.04, 0.12, 0.37, 1.11, 3.33, or 10 nM. The amount of uncleaved DNA remaining after the reaction was quantified by qPCR. Fraction cleaved DNA was calculated as the amount of cleaved DNA present following completion of each cleavage reaction divided by the total amount of DNA input into each reaction.

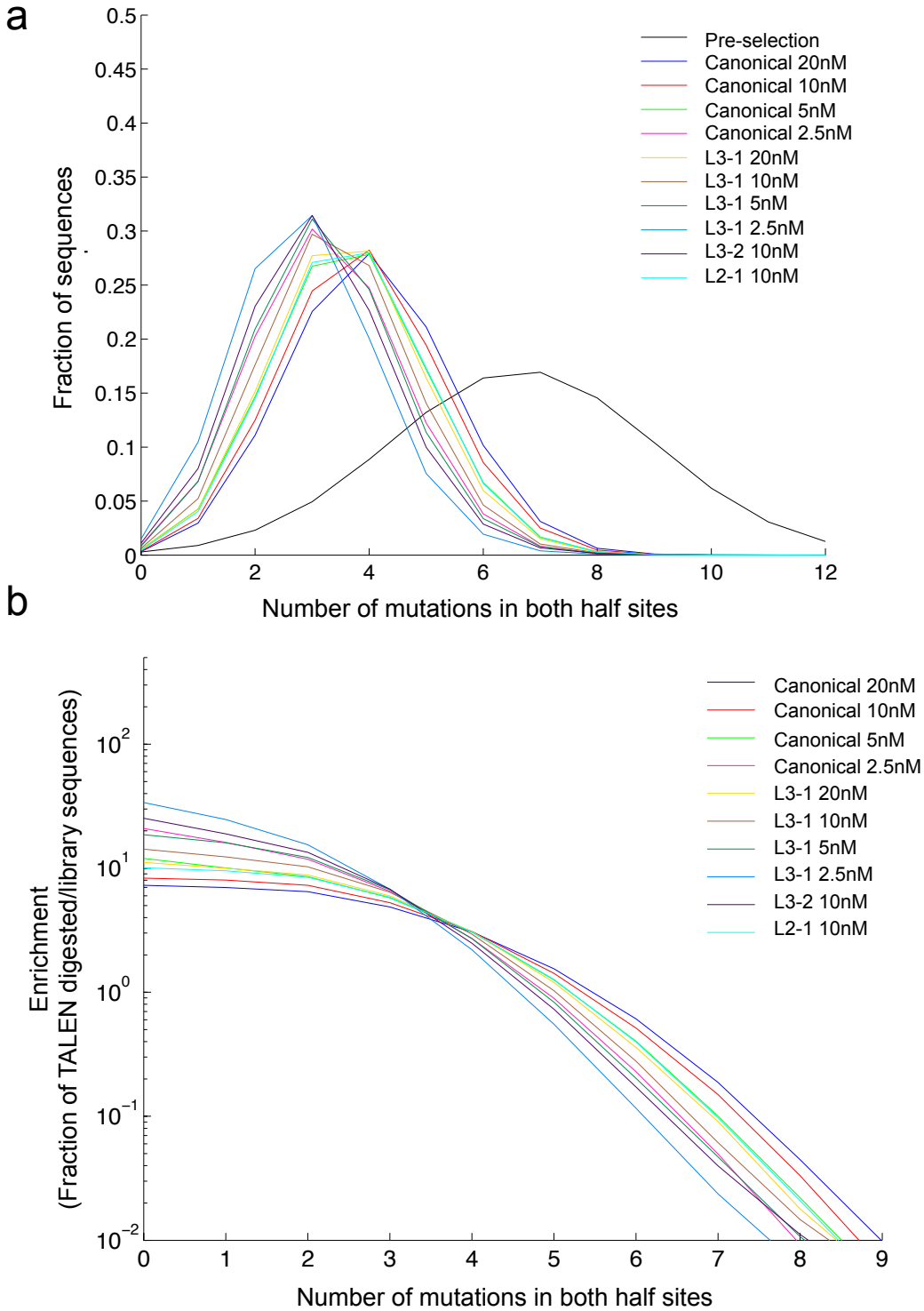


Supplementary Figure 16. Characterization of evolved *ATM-L* TALEs following positive and negative selection PACE. (a) Luciferase activity represented as fold induction (ATc-induced TALE luminescence / non-induced luminescence) for the canonical *ATM-L*-directed TALE or L3-1 and L3-2, on the on-target sequence (*ATM*: 5'-TGAATTGGGATGCTGTTT-3'), or on the off-target sequence OffA17 (OffA17: 5'-GGAAATGGGATACTGAGT-3'). Data represent mean + s.d. ($n = 3$). (b and c) Genotypes of individual evolved phage clones following dual positive and negative selection PACE (against OffA17). Blue squares indicate mutations within the TALE domain, and green squares indicate mutations within the ω subunit. (d) Left: relative cleavage efficiencies of the canonical *ATM* TALEN pair or two TALENs containing an evolved left-half site (L1-1, or L3-2) on a linear 6-kb DNA fragment containing either the *ATM* on-target sequence or the OffA17 off-target sequence. The top band is non-cleaved DNA, while the bottom band is a cleavage product. Quantified cleavage percentages were determined using densitometry (GelEval), and are shown below each lane. Right: mutations in the evolved *ATM*-left half site TALEs used in the left panel.



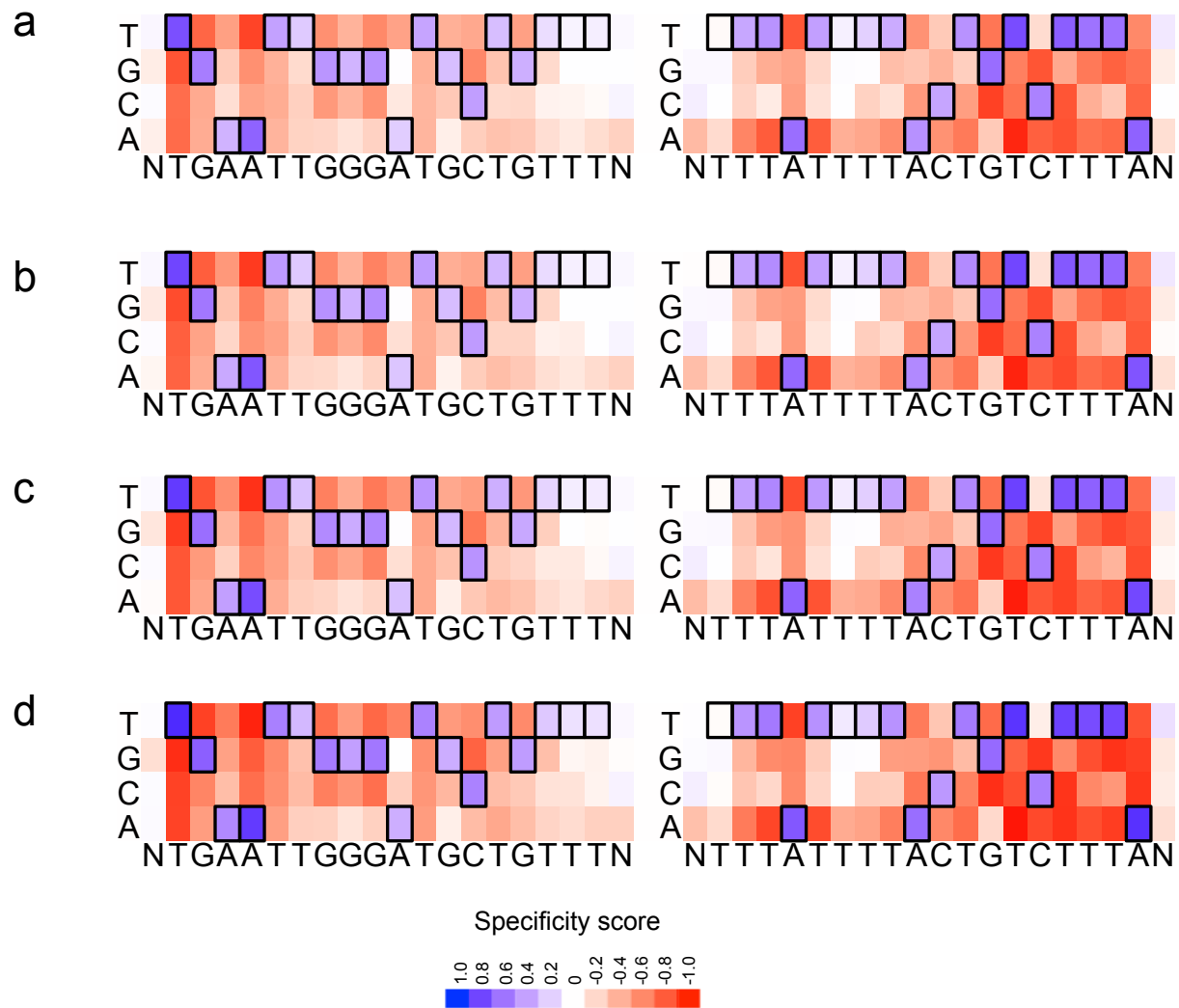
Supplementary Figure 17. Characterization of mutations identified in positive and negative selection *ATM-L* TALE PACE and evolved TALEN specificity. Relative cleavage efficiencies of the canonical *ATM* TALEN pair, or (a) a TALEN (L1-2) containing an evolved left-half site TALE with the A252T mutation and the canonical right half-site TALE on a linear 6-kb DNA fragment containing either the *ATM* on-target

sequence or the OffA17 off-target sequence. The top band is non-cleaved DNA, while the bottom band is a cleavage product. (b) Same as (a), but assaying a TALEN containing a Q745P substitution. Crystal structure⁴ of three TALE repeats showing the relative positions of the (c) A252, and (d) L338 residues within a core TALE repeat (repeat in light blue, residues in dark blue). (e) Relative cleavage efficiencies of the canonical *ATM* TALEN pair, or a TALEN (L3-1) containing an evolved left-half site on a linear 6-kb DNA fragment containing either the *ATM* on-target sequence, the OffA17 sequence, or a derivative of the OffA17 sequence containing a subset of its 5 mutations (D1-D4 listed in the figure). (f) Same as (e) but with derivative sequences containing fewer mutations relative to the on-target sequence (1 or 2 bp). For all cleavage gels, the top band is non-cleaved DNA, while the bottom band is a cleavage product. Quantified cleavage percentages were determined using densitometry (GelEval), and are shown below each lane.

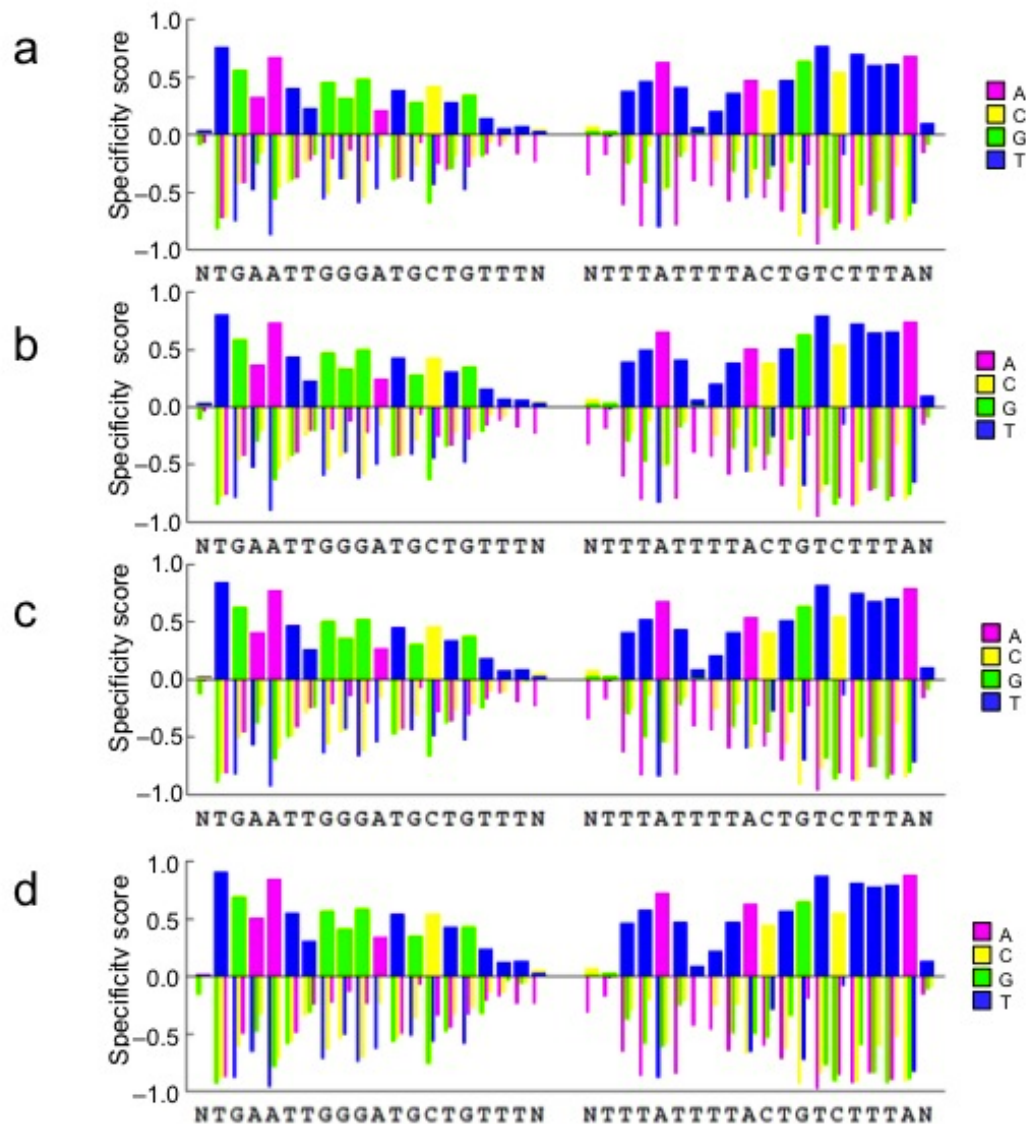


Supplementary Figure 18. Global analysis of *in vitro* TALEN specificity. (a) Sequences surviving selection (TALEN digestion) compared to the pre-selection library as a function of the number of mutations in both half-sites (left and right half-sites combined excluding the spacer) for each of the ten reaction conditions listed. (b)

Enrichment value of on-target (no mutations) and off-target sequences containing one to nine mutations in both half sites (left and right half-sites combined excluding the spacer) for each of the ten reaction conditions listed.

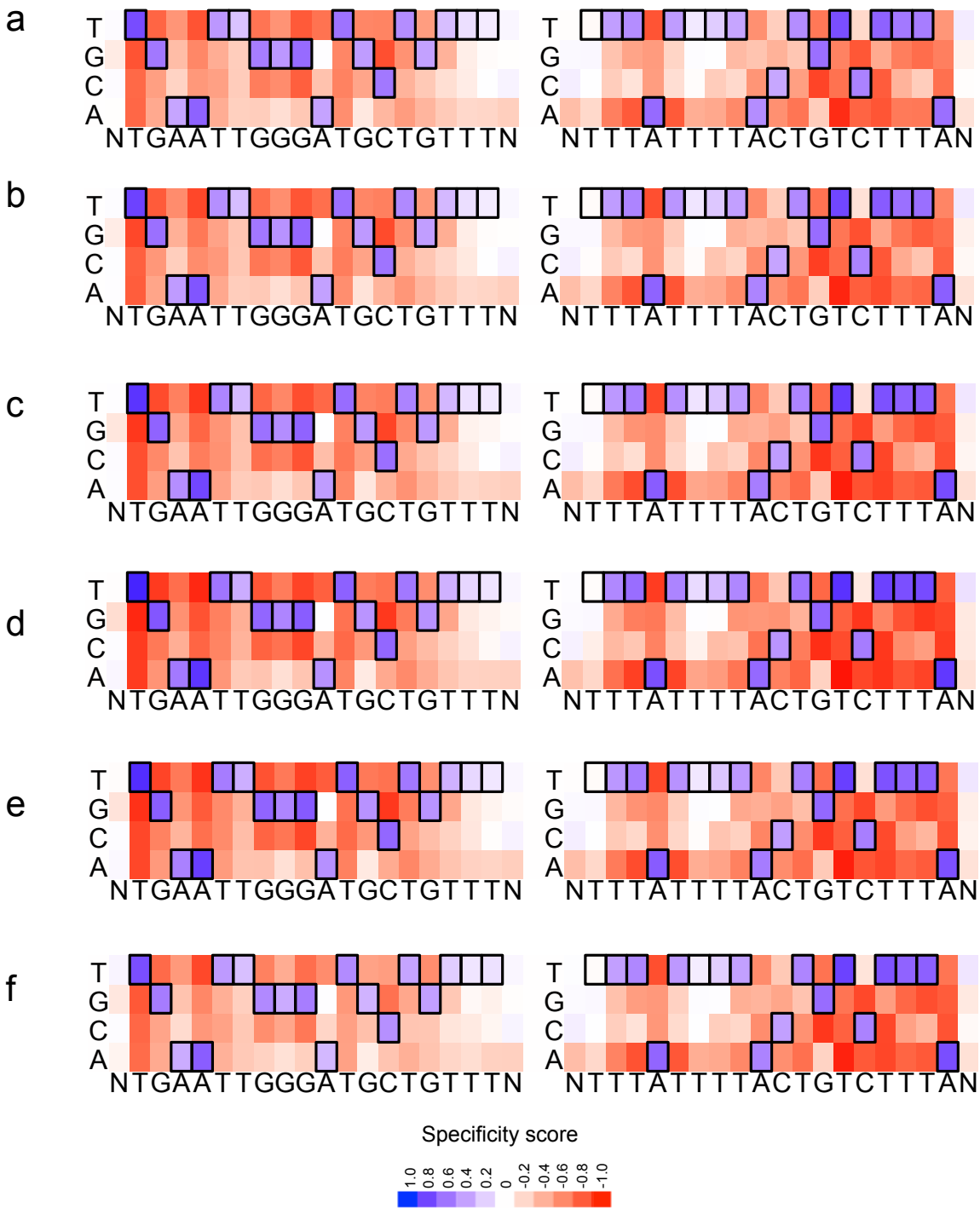


Supplementary Figure 19. Specificity profile heat maps for the canonical *ATM* TALEN pair as a function of concentration. Heat maps showing specificity scores for the canonical TALEN targeting the *ATM* locus used in the cleavage assay at doses of (a) 20 nM, (b) 10 nM, (c) 5 nM, and (d) 2.5 nM. Each position in the left and right half-sites plus a single flanking position (N) are shown. Colors range from dark blue at a score of 1.0 (complete specificity), to white at a score of 0 (no specificity), to dark red at a score of -1.0 (maximum negative score). The cognate base for each position in the target sequence is boxed. For the right half-site, data for the sense strand are displayed.



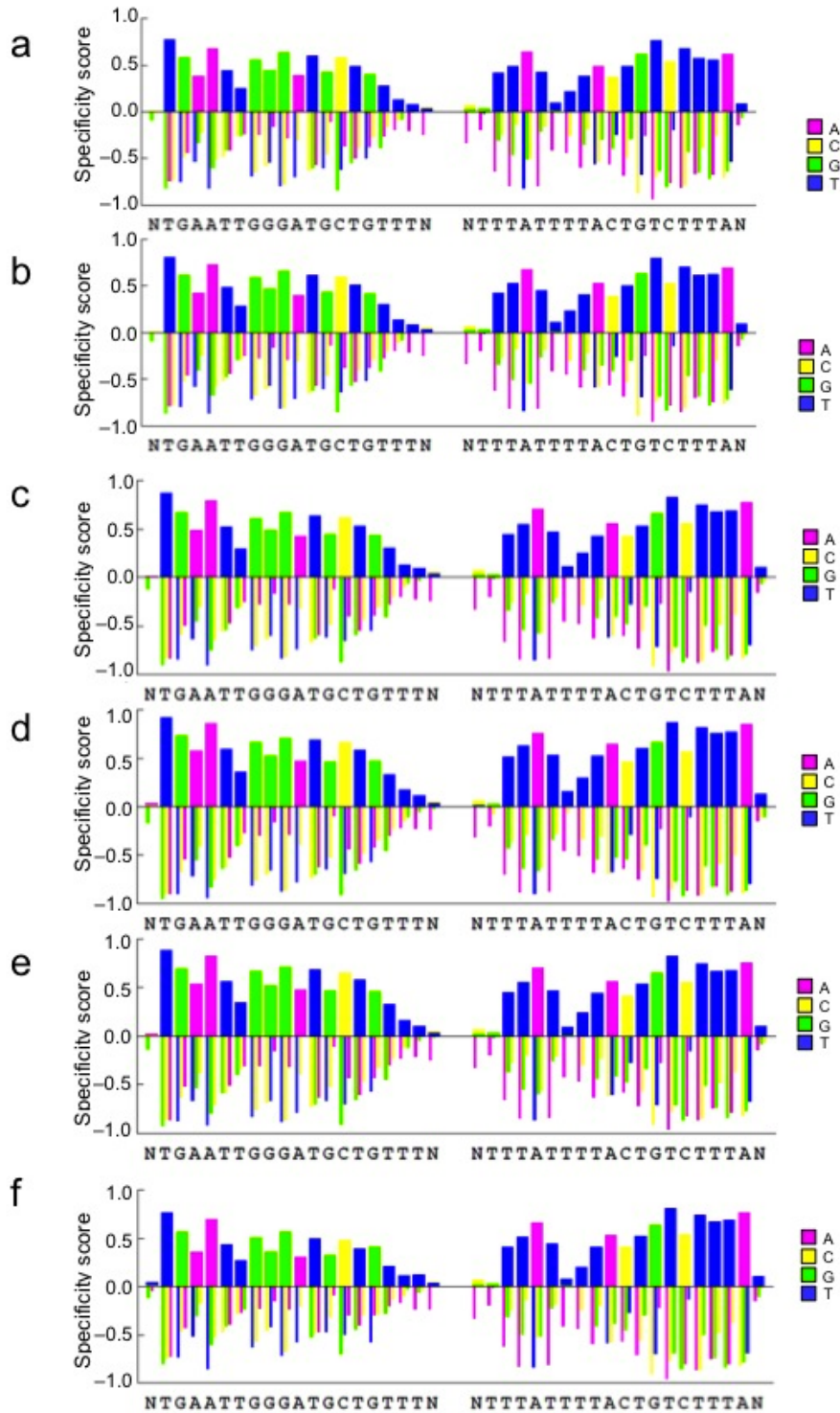
Supplementary Figure 20. Specificity profile bar graphs of the canonical *ATM* TALEN pair as a function of concentration. Bar graphs showing the quantitative specificity score for each nucleotide position for the canonical TALEN targeting the *ATM* locus used in the cleavage assay at doses of (a) 20 nM, (b) 10 nM, (c) 5 nM, and (d) 2.5 nM. Each position in the left and right half-sites plus a single flanking position (N) are shown. A score of zero indicates no specificity, while a score of 1.0 corresponds to perfect specificity. Negative specificity scores range from zero to -1.0, and represent enrichment against that base pair. Specified positions (specificity score >0) were plotted as stacked bars above the axis (multiple specified base pairs at the same position were plotted over each other with the shortest bar in front) while anti-specified base pairs

were plotted as narrow, grouped bars below the axis. For the right half-site, data for the sense strand are displayed.



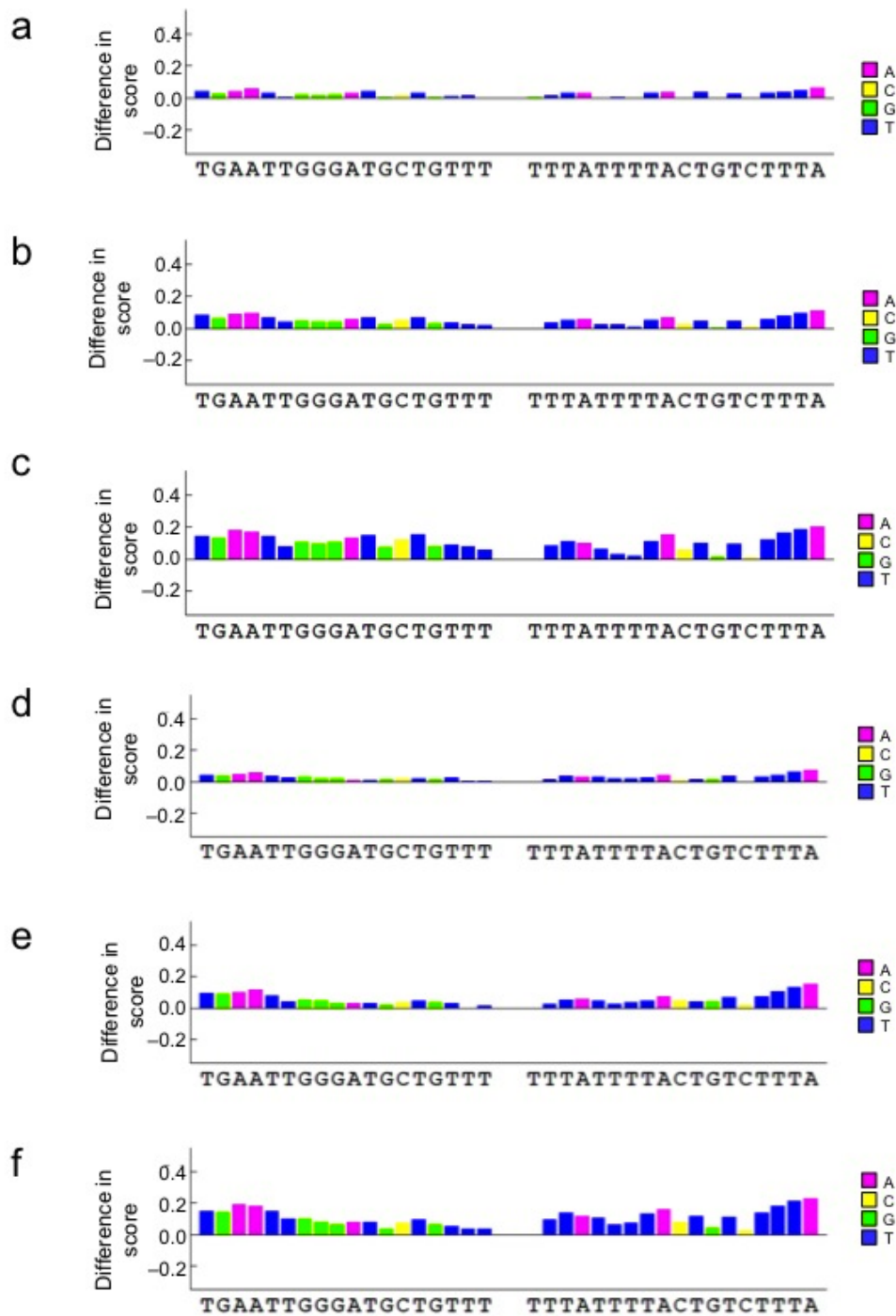
Supplementary Figure 21. Specificity profile heat maps of L2-1, L3-1, and L3-2 *ATM* TALEN pairs. Heat maps showing specificity scores for the L3-1 TALEN pair at doses of (a) 20 nM, (b) 10 nM, (c) 5 nM, and (d) 2.5 nM, or TALEN pairs incorporating L3-2 and L2-1 TALEs at a dose of 10 nM (e and f, respectively). Each position in the left

and right half-sites plus a single flanking position (N) are shown. Colors range from dark blue at a score of 1.0 (complete specificity), to white at a score of 0 (no specificity), to dark red at a score of -1.0 (maximum negative score). The cognate base for each position in the target sequence is boxed. For the right half-site, data for the sense strand are displayed.



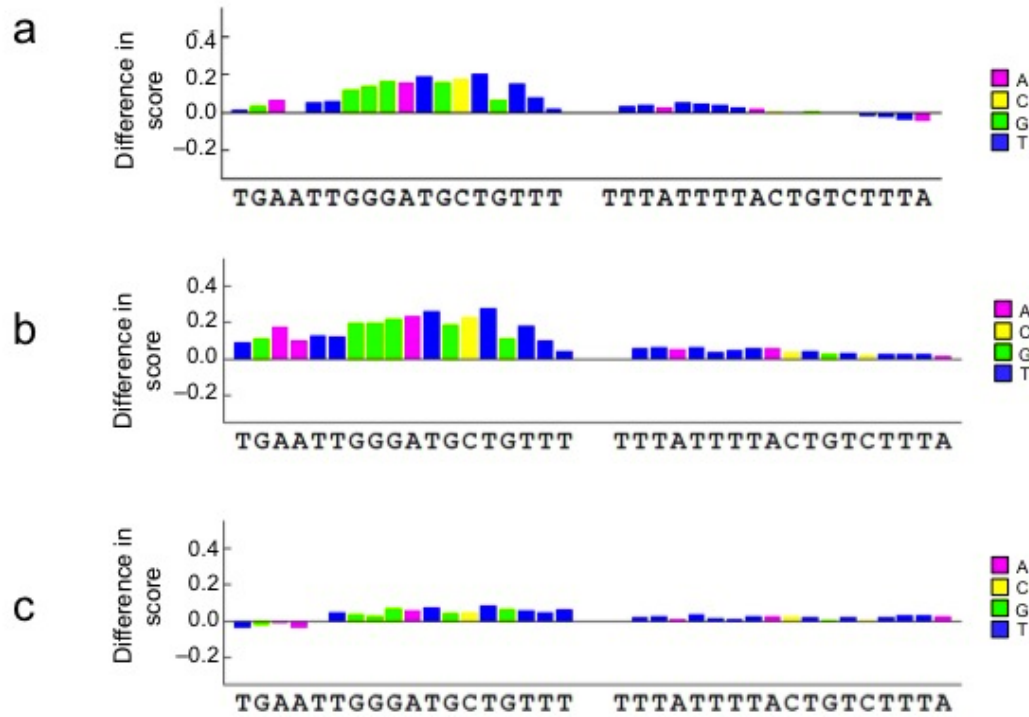
Supplementary Figure 22. Specificity profile bar graphs of L2-1, L3-1, and L3-2 *ATM* TALEN pairs. Bar graphs showing the quantitative specificity score for each nucleotide position for the L3-1 TALEN pair at doses of (a) 20 nM, (b) 10 nM, (c) 5 nM, and (d) 2.5 nM, or TALEN pairs incorporating L3-2 and L2-1 TALEs at a dose of 10 nM (e and f, respectively). Each position in the left and right half-sites plus a single flanking

position (N) are shown. A score of zero indicates no specificity, while a score of 1.0 corresponds to perfect specificity. Negative specificity scores range from zero to -1.0, and represent enrichment against that base pair. Specified positions (specificity score >0) were plotted as stacked bars above the axis (multiple specified base pairs at the same position were plotted over each other with the shortest bar in front) while anti-specified base pairs were plotted as narrow, grouped bars below the axis. For the right half-site, data for the sense strand are displayed.

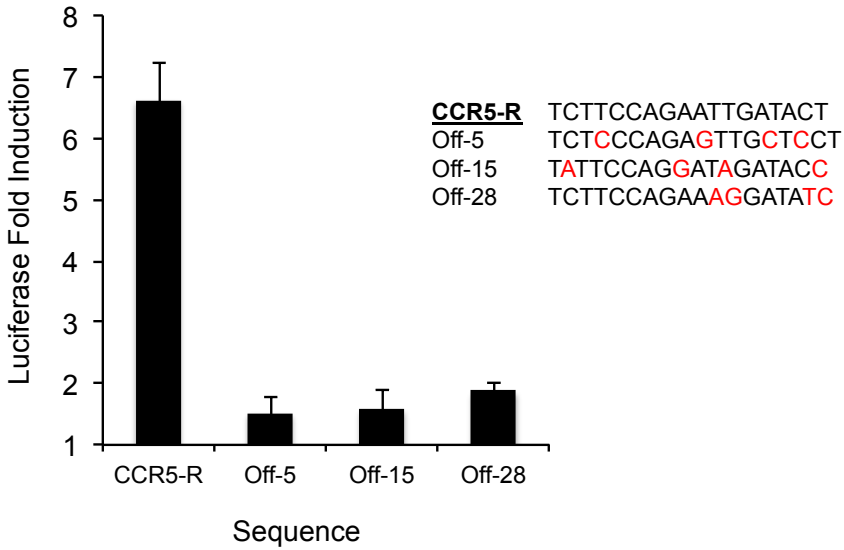


Supplementary Figure 23. Specificity profile difference as a function of TALEN concentration for canonical and L3-1 ATM TALEN pairs. Bar graph indicating the quantitative difference in specificity score at each position between cleavage using the canonical TALEN pair at a dose of 20 nM and (a) 10 nM, (b) 5 nM, and (c) 2.5 nM (calculated as $\text{score}_{\text{lowdose}} - \text{score}_{20\text{nM}}$), or TALEN pairs incorporating an evolved L3-1 ATM TALE at a dose of 20 nM and (d) 10 nM, (e) 5 nM, and (f) 2.5 nM. A score of zero

indicates no change in specificity. For the right half-site, data for the sense strand are displayed.



Supplementary Figure 24. Bar graph showing difference in specificity of the canonical TALEN pair versus the L2-1, L3-1 and L3-2 TALEN pairs. Bar graph indicating the quantitative difference in specificity score at each position between cleavage using the canonical TALEN pair or TALENs incorporating (a) L3-1 or (b) L3-2, or (c) L2-1 TALEs, all at a dose of 10 nM. Difference scores were calculated as $\text{score}_{L2/3} - \text{score}_{WT}$. Bases at each position in the target half-sites are displayed. A score of zero indicates no change in specificity. For the right half-site, data for the sense strand are displayed.



Supplementary Figure 25. Specificity of a CCR5-targeting TALE in the DB-PACE one-hybrid system. Luciferase activity represented as fold induction (ATc-induced TALE luminescence / non-induced luminescence) for a canonical *CCR5-R*-directed TALE on its on-target sequence or one of three previously described off-target sequences⁵ (Off-5, Off-15, Off-28; sequences indicated in the figure). Data represent mean + s.d. ($n = 3$).

Supplementary Table 1. Full target sequences used to study *ATM* TALENs^a.

| Sequence | Left-half sequence | Right-half sequence ^b |
|-----------|--------------------|----------------------------------|
| On-target | TGAATTGGGATGCTGTTT | TTTATTTTACTGTCTTTA |
| OffA1 | TGAATaGGaAataTaTTT | TTTATTTTACTGTtTTTA |
| OffA11 | TGAATTGaGAgaagcaTT | TTTATTTTAtTaTtTTTA |
| OffA17 | gGAAaTGGGATaCTGagT | TTTATgTACTaTtTcTA |
| OffA23 | TagATTGaaATGCTGTTT | TTTtTaTTAtTaTtTTTA |

^aFor *in vitro* cleavage assays, left and right half-site recognition sequences were separated by a constant 18-bp constant spacer sequence (5'-TTAGGTATTCTATTCAA-3'). For high-throughput specificity profiling, a range of spacer lengths was used in the library⁵.

^bFor the right half-site the sense strand is displayed.

Supplementary Table 2. Statistics of sequences selected by TALEN digestion.

| Selection | Seq. count | Mean mut. | Stdev mut. | P-value vs. library^a |
|-----------------------|-------------------|------------------|-------------------|--|
| Canonical 20 nM | 181361 | 3.991 | 1.421 | 4.78×10^{-11} |
| Canonical 10 nM | 180277 | 3.853 | 1.396 | 3.83×10^{-11} |
| Canonical 5 nM | 206958 | 3.662 | 1.367 | 2.40×10^{-11} |
| Canonical 2.5 nM | 343423 | 3.282 | 1.321 | 8.42×10^{-12} |
| L3-1 20 nM | 137886 | 3.617 | 1.343 | 3.69×10^{-11} |
| L3-1 10 nM | 141679 | 3.445 | 1.318 | 2.73×10^{-11} |
| L3-1 5 nM | 190497 | 3.247 | 1.297 | 1.42×10^{-11} |
| L3-1 2.5 nM | 342976 | 2.914 | 1.264 | 5.20×10^{-12} |
| L3-2 10 nM | 187254 | 3.126 | 1.299 | 1.22×10^{-11} |
| L2-1 10 nM | 181692 | 3.67 | 1.35 | 2.83×10^{-11} |
| Pre-selection library | 453246 | 6.811 | 2.311 | NA |

Statistics are shown for the pre-selection library and for DNA surviving each TALEN selection on the *ATM* target sequence. Seq. counts: total counts of high-throughput sequenced and computationally filtered selection sequences. Mean mut.: mean mutations in selected sequences. Stdev. mut.: standard deviation of mutations in selected sequences. Stdev. mut.: standard deviation of mutations in selected sequences.

^aComparisons between the TALEN selection sequence distributions and the corresponding pre-selection library sequence distribution were determined as previously reported⁶ using a one-sided *t*-test.

Supplementary Table 3. Cellular modification rates, sample size and *P* value for high-throughput sequencing of TALEN cleavage in U2OS and 293 cells.

| Cell line | TALEN pair | Site | Indels | Total sequences | Percent modified ^a | <i>P</i> -value ^b |
|-----------|-------------------------|-----------|--------|-----------------|-------------------------------|------------------------------|
| U2OS | Control | On-target | 1 | 10000 | 0.010 | |
| U2OS | Control | OffA1 | 0 | 253702 | <0.001 | |
| U2OS | Control | OffA11 | 5 | 421633 | 0.001 | |
| U2OS | Control | OffA17 | 2 | 438269 | <0.001 | |
| U2OS | Control | OffA23 | 2 | 281288 | 0.001 | |
| U2OS | <i>ATM</i> can. (EL/KK) | On-target | 1100 | 10000 | 11.000 | < 1.0 x 10 ⁻³⁰⁰ |
| U2OS | <i>ATM</i> can. (EL/KK) | OffA1 | 18 | 193251 | 0.009 | 2.8 x 10 ⁻⁷ |
| U2OS | <i>ATM</i> can. (EL/KK) | OffA11 | 144 | 357899 | 0.040 | 5.7 x 10 ⁻⁴² |
| U2OS | <i>ATM</i> can. (EL/KK) | OffA17 | 95 | 569336 | 0.017 | 4.1 x 10 ⁻²¹ |
| U2OS | <i>ATM</i> can. (EL/KK) | OffA23 | 12 | 338944 | 0.004 | 2.8 x 10 ⁻² |
| U2OS | L3-1 (EL/KK) | On-target | 704 | 10000 | 7.040 | 1.3 x 10 ⁻²⁰⁴ |
| U2OS | L3-1 (EL/KK) | OffA1 | 0 | 275541 | <0.001 | |
| U2OS | L3-1 (EL/KK) | OffA11 | 5 | 314087 | 0.002 | |
| U2OS | L3-1 (EL/KK) | OffA17 | 1 | 420626 | <0.001 | |
| U2OS | L3-1 (EL/KK) | OffA23 | 0 | 351725 | <0.001 | |
| U2OS | L3-2 (EL/KK) | On-target | 797 | 10000 | 7.970 | 2.1 x 10 ⁻²³¹ |
| U2OS | L3-2 (EL/KK) | OffA1 | 4 | 235431 | 0.002 | 5.4 x 10 ⁻² |
| U2OS | L3-2 (EL/KK) | OffA11 | 5 | 303362 | 0.002 | |
| U2OS | L3-2 (EL/KK) | OffA17 | 2 | 489318 | <0.001 | |
| U2OS | L3-2 (EL/KK) | OffA23 | 2 | 401692 | <0.001 | |
| 293 | Control | On-target | 74 | 121714 | 0.061 | |
| 293 | Control | OffA11 | 1 | 64667 | <0.001 | |
| 293 | Control | OffA17 | 3 | 191726 | <0.001 | |
| 293 | <i>ATM</i> can. (Homo) | On-target | 23651 | 99604 | 23.745 | < 1.0 x 10 ⁻³⁰⁰ |
| 293 | <i>ATM</i> can. (Homo) | OffA11 | 290 | 38761 | 0.748 | 8.8 x 10 ⁻¹²² |
| 293 | <i>ATM</i> can. (Homo) | OffA17 | 639 | 168318 | 0.380 | 1.2 x 10 ⁻²⁰⁴ |
| 293 | L3-2 (Homo) | On-target | 20944 | 90030 | 23.263 | < 1.0 x 10 ⁻³⁰⁰ |
| 293 | L3-2 (Homo) | OffA11 | 0 | 52317 | <0.001 | |

| | | | | | | |
|-----|----------------|--------|----|--------|-------|-----------------------|
| 293 | L3-2 (Homo) | OffA17 | 37 | 148883 | 0.025 | 9.5×10^{-11} |
|-----|----------------|--------|----|--------|-------|-----------------------|

^aAs previously described⁷, the sensitivity of the high-throughput sequencing method for detecting genomic off-target cleavage is limited by the amount genomic DNA (gDNA) input into the PCR amplification of each genomic target site. Each sample was run with 600 ng of genomic DNA, equivalent to ~198,000 genomes. Thus, the theoretical detection limit of this technique is approximately 1 in 198,000, which we have indicated as <0.001%.

Cellular modification rates are shown as a percentage based on the number of observed sequences containing insertions or deletions (indels) divided by the total number of genomic DNA fragments sequenced.

^b*P* values were calculated as previously reported^{5,6} using a (right) one-sided Fisher's exact test between each TALEN-treated sample and the untreated control sample. *P* values less than the significance threshold, calculated as previously described⁵, are not shown.

Indels are the number of observed sequences containing insertions or deletions consistent with TALEN-induced cleavage, and percent modified corresponds to the number of detected sequences containing indels divided by the total number of genomic DNA fragments sequenced multiplied by one hundred.

Heterodimeric FokI nuclease (EL/KK) TALENs were used for experiments in U2OS cells, while homodimeric FokI nuclease TALENs were used for experiments in HEK 293 cells.

Supplementary Table 4. Primers sequences used for plasmid cloning, and for high-throughput sequencing.

| Primer Name | Primer Sequence |
|---|--|
| Primers used to construct pOHZif268 plasmids | |
| BHP1B | 5'-AGGGTCCCTAAGTCUCCTCAGCAAAACGA-3' |
| BHP1 | 5'-AGACTTAGGGACCCUCATGCGTGGGCGTAGGCACCCCGGGCTTTACT-3' |
| AB0016 | 5'-ACAGTCAGTAGGGCCCUAAAAAAAAACCCCGCCCCTGAC-3' |
| BHP2B | 5'-ACAGTCAGTAGGGCCCUAAAAAAAAACCC-3' |
| BHP2 | 5'-AGGGCCCTACTGACTGUTA CTTCTGTCTTAAATGGATTTTGGTATGCCTCTT-3' |
| BHP3 | 5'-AGCACC GCGGCGGCU GAACGCCCATATGCTTGCCCTGT-3' |
| BHP4 | 5'-AGCCGCCGCGGTG CUGCCGCCGCCGCCGCCGCGCCTCAGGTTTCTTCTTTCACTTCAGGCT-3' |
| JC1062 | 5'-ACCCCTACCCCTTAU...-3' |
| AB0052 | 5'-ATAAGGGGGTAGGGGGUUCTAGATTTTTGTGCGAACTATTCATTTCACTTTTCTC-3' |
| BHP5 | 5'-AGCGGCGGCGGUGGCAGCGGCGGCGGCGGCAGCACCCGCGGCGGCTGAACGCCCATATGCTTGCCCTGT-3' |
| BHP6 | 5'-ACCGCCGCCG CUGCCGCCGCCGCCGCCGCGCCTCAGGTTTCTTCTTTCACTTCAGGCT-3' |
| BHP7 | 5'-AGACTTAGGGACCCUCATGCGTGGGCGGCATGCTTAGGCACCCCGGGCTTTACT-3' |
| BHP8 | 5'-ACCGCCGCCG CUGCCGCCGCCGCCGCCGCGCACGACCTTCAGCAATAGCGGT-3' |
| AB0189 | 5'-ACCCCTACCCCTTAUAAGGAGGAAAAAAAAATGGCACGCGTAACTGTTCAAGAC-3' |
| BHP8B | 5'-ACCCCTACCCCTTAUAAGGAGGAAAAA-3' |
| BHP9 | 5'-AGCCGCCGCGGTG CUGCCGCCGCCGCCGCCGCGCACGACCTTCAGCAATAGCGGT-3' |
| AB0058 | 5'-AGGGCCCTACTGACTGUTAACGACGACCTTCAGCAATAGCG-3' |
| BHP9B | 5'-AGGGCCCTACTGACTGUTAACGACGACCT-3' |
| BHP10 | 5'-AGCACC GCGGCGGCU CGCGTAACTGTTCAAGACGCTGTA-3' |
| BHP11 | 5'-AGCCGCCGCGGTG CUGCCGCCGCCGCCGCCGCGCCTTCTGTCTTAAATGGATTTTGGTATGCCTCTT-3' |
| BHP12 | 5'-ACCCCTACCCCTTAUAAGGAGGAAAAAAAA ATGGAACGCCCATATGCTTGC-3' |
| AB0058 | 5'-AGGGCCCTACTGACTGUTAACGACGACCTTCAGCAATAGCG-3' |
| BHP13 | 5'-AGCGGCGGCGGUGGCAGCGGCGGCGGCGGCAGCACCCGCGGCGGCTCGCGTAACTGTTCAAGACGCTGTA-3' |
| BHP14 | 5'-ACCGCCGCCG CUGCCGCCGCCGCCGCCGCGCCTTCTGTCTTAAATGGATTTTGGTATGCCTCTT-3' |
| Primers used to construct pAPZif268 variants | |
| BHP58 | 5'-ACTACCCATAAUACAAGAAAAGCCCGTCACGG-3' |
| BHP59 | 5'-ATTATGGGTAGUTTCTTGCATGAATCCAT-3' |
| BHP60 | 5'-ACTGCACCTGCAGGUCGGTGCACACA-3' |
| BHP61 | 5'-AGTTTCAAATTTTCATUAAGACTCCTTATTACGCAGTATGTTA-3' |
| BHP62 | 5'-AATGAAATTTGGAAACUTTTTGTACATACCAACCTC-3' |
| BHP64 | 5'-AGGGCCCTACTGACTGUGCGAAAAAACCCCGCCGAAG-3' |

Primers used to construct SPZif268

BHP73SP 5'-ATTACCGTTCATCGAUTCTCTTGTGCTCC-³
BHP74SP 5'-ATCGATGAACGGTAAUCGTAAACTAGCATGT-³
BHP75SP 5'-ATGGCACGCGTAACU GTTCAGGACGC-³
BHP76SP 5'-AGTTACGCGTGCCAU TTTTTTTCCTCCTAAGCTGTTAGAAAACTCAT-³
BHP77SP 5'-ACTTCTGTCTUAAATGGATTTTGGTATGCCTCT-³
BHP78SP 5'-AAGACAGAAGU AATCTAGAAGGAGATTTTCAACATGCTCCCTCAAT-³
BHP75SP 5'-ATGGCACGCGTAACU GTTCAGGACGCTGTAGAGAAAA-³

Primers used to construct pOHCBTAL variants

TAL23-1 5'-ACCCGCCGAAGCCCUCCTGAAATGAGGGTCCCTAAGTCTCCTCAGC-³
TAL23-2 5'-AGGGCTTCGGCGGGUAGGCACCCCGGGCTTTACT-³
TAL23-3 5'-ATCTGGTGGCGGCACGCGTAACTGTUCAGGACGCTGTAGA-³
TAL23-4 5'-AACAGTTACGCGTGCCGCCACCAGAUGGTCATTTCGTTAACGCAG-³
5'-
TAL23-5 ACCCCCTACCCCTTATAAGGAGGAAAAAAAAAUGGACTACAAAGACCATGACGGTGA
TTATAAAGAT-³
TAL23-6 5'-ATTTTTTTTCTCCTTATAAGGGGGTAGGGGGUUCTAGATTTTTGTGCGAACTATT-³
TAL63-1 5'-ACTTCCCATCGAGTCGCGGCACGCGTAACTGTUCAGGACGCTGTAGA-³
TAL63-2 5'-AACAGTTACGCGTGCCGCGACTCGATGGGAAGUTCTCTCGGGAA-³
40FWD 5'-AGTTACGCGTGCCAGACCCUTTTTACTGCATCGAGCGCGGGT-³
40REV 5'-AGGGTCTGGCACGCGTAACUGTTCAGGACGCTGTA-³
28+G4SF
WD 5'-AGTTACGCGTGCGCTGCCGCCGCCACCAAGACAUGCCAACGCCA-³
28+G4SRE
V 5'-ATGTCTTGGTGGCGGCGGCAGCGCACGCGTAACUGTTCAGGACGCT-³
18+G4SF
WD 5'-ACGCGTGCGCTGCCGCCGCCATTTCGTUACGCAGCCAACGCGGGAT-³
18+G4SRE
V 5'-AACGAATGGCGGCGGCAGCGCACGCGUAACTGTTTCAGGACGCTGTA-³
18FWD 5'-ACGCGTGCACTTCGTUACGCAGCCAACGCGGGAT-³
18REV 5'-AACGAATGCACGCGUAACTGTTTCAGGACGCTGTA-³

Primers used to construct SPCBTAL

TALEFW 5'-AGGAAAAAAAAAUGGACTACAAAGACCATGACGGTGATTATAAAGAT-³
SPREV 5'-ATTTTTTTTTCCUCCTAAGCTGTTAGAAAACTCAT-³
TALEREV 5'-ATTAACGACGACCTUCAGCAATAGCGGTAACG-³
SPFWD 5'-AAGTTCGTCGTTAAUCTAGAAGGAGATTTTCAACATGCTCCCT-³
TALEFW 5'-AGGAAAAAAAAAUGGACTACAAAGACCATGACGGTGATTATAAAGAT-³

Primers used to construct pAPNegCBXTAL variants

FWDNEG
SELPLAS
MID
REVCOLE
1 5'-ACCGGATGTCCTCT/IDEOXYU/AAAACGAAAGGCCAGTCTTTTCGACTGAGC-³
FWDCOLE
1 5'-AAATAGGCCG/IDEOXYU/TTGAGATCCTTTTTTCTGCGCG-³
REVNEGS
ELPLASMI
D 5'-ACGGCCTATT/IDEOXYU/GGCCTATTTTCTAAATACATTCAAATATGTATCCGCTC-³
NEGAFW
D 5'-/5PHOS/ATCAGGAGGGCTTCGGCGGGTAGGC-³
NEGCFW
D 5'-/5PHOS/CTCAGGAGGGCTTCGGCGGGTAGGC-³
NEGGFW
D 5'-/5PHOS/GTCAGGAGGGCTTCGGCGGGTAGGC-³

NEGREV 5'-/5PHOS/GTAGGGTCCCTAAGTCTCCTCAGCACC-'3

Primers used to construct pTetCBXTAL plasmids

MINIMIZE
OHFW D 5'-/5PHOS/TGATAATATTAAGAAGTACCACACGGAATATACCTAA-'3

MINIMIZE
OHREV1 5'-/5PHOS/GAGGGTCCCTAAGTCTCCTCAGC-'3

MINIMIZE
OHREV2 5'-/5PHOS/AGTCTCCTCAGCAAACGAAAGGCCAGTCTT-'3

pOH/AP/SP plasmids for the ATM-L TALE were cloned in a similar manner to the CBX8 TALE (refer to primers above for examples)

Primers and gBlocks used for TALEN cleavage assay

PUC19OF
WD 5'-GCGACACGGAAATGTTGAATACTCAT-'3

PUC19OR
EV 5'-CAGCGAGTCAGTGAGCGA-'3

CLONEEV
OATMLFW
D 5'-AAAGACCATGACGG/IDEOXYU/GATTATAAAGATCATGA-'3

CLONEEV
OBACKRE
V 5'-ACCGTCATGGTCTT/IDEOXYU/GTAGTCCATGGTGCTAGCC-'3

CLONEEV
OATMLRE
V 5'-AGCCAACGCGGGA/IDEOXYU/CGGGCCT-'3

CLONEEV
OBACKFW
D 5'-ATCCCGCGTTGGC/IDEOXYU/GCGTTAACGAATG-'3

ON-
TARGET 5'-
ATTTCCGCCTGCGATTTTCAAATCGAAATGGCTCTAGATTTGAATTGGGATGCTGTTTT
TAGGTATTCTATTCAAATTTATTTTACTGTCTTTATTAAGCTTCGTCGGACGTACGTACG
TTTTCCC-'3

OFFA17 5'-
ATTTCCGCCTGCGATTTTCAAATCGAAATGGCTCTAGATTTGAAATGGGATACTGAGTT
TAGGTATTCTATTCAAATTTATTTTACTGTCTTTATTAAGCTTCGTCGGACGTACGTACG
TTTTCCC-'3

OFFA-
17D1 5'-
ATTTCCGCCTGCGATTTTCAAATCGAAATGGCTCTAGATTTGAAATGGGATACTGTTTT
TAGGTATTCTATTCAAATTTATTTTACTGTCTTTATTAAGCTTCGTCGGACGTACGTACG
TTTTCCC-'3

OFFA-
17D2 5'-
ATTTCCGCCTGCGATTTTCAAATCGAAATGGCTCTAGATTTGAATTGGGATACTGATTTT
AGGTATTCTATTCAAATTTATTTTACTGTCTTTATTAAGCTTCGTCGGACGTACGTACGT
TTTTCCC-'3

OFFA-
17D3 5'-
ATTTCCGCCTGCGATTTTCAAATCGAAATGGCTCTAGATTTGAAATGGGATACTGTTTTT
AGGTATTCTATTCAAATTTATTTTACTGTCTTTATTAAGCTTCGTCGGACGTACGTACGT
TTTTCCC-'3

OFFA-
17D4 5'-
ATTTCCGCCTGCGATTTTCAAATCGAAATGGCTCTAGATTTGAAATGGGATGCTGTTTT
TAGGTATTCTATTCAAATTTATTTTACTGTCTTTATTAAGCTTCGTCGGACGTACGTACG
TTTTCCC-'3

OFFAD5 5'-
ATTTCCGCCTGCGATTTTCAAATCGAAATGGCTCTAGATTTGAATTGGGATGCTGAGTT
TAGGTATTCTATTCAAATTTATTTTACTGTCTTTATTAAGCTTCGTCGGACGTACGTACG
TTTTCCC

OFFAD6 ATTTCCGCCTGCGATTTTCAAATCGAAATGGCTCTAGATTTGAATTGGGATACTGTGTT
 TAGGTATTCTATTCAAATTTATTTTACTGTCTTTATTAAGCTTCGTCGGACGTACGTACG
 TTTTCCC-3
 5'-

OFFAD7 ATTTCCGCCTGCGATTTTCAAATCGAAATGGCTCTAGATTTGAATTGGGATGCTGTGTT
 TAGGTATTCTATTCAAATTTATTTTACTGTCTTTATTAAGCTTCGTCGGACGTACGTACG
 TTTTCCC-3
 5'-

OFFAD8 ATTTCCGCCTGCGATTTTCAAATCGAAATGGCTCTAGATTGGAATTGGGATGCTGTGTT
 TAGGTATTCTATTCAAATTTATTTTACTGTCTTTATTAAGCTTCGTCGGACGTACGTACG
 TTTTCCC-3

Primers used to amplify on- and off-target *ATM* genomic sites

ONATMF 5'-GGAGTTCAGACGTGTGCTCTTCCGATCTAGCGCCTGATTTCGAGATCCT-3
 5'-

ONATMR CACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATGCCAAATTCATATGCAAGGC-
 3

OFFA-1F 5'-GGAGTTCAGACGTGTGCTCTTCCGATCTCCTGCCATTGAATTCCAGCCT-3

OFFA-1R 5'-CACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGTCTGCCTTTCTGTCCCC-
 3

OFFA-11F 5'-GGAGTTCAGACGTGTGCTCTTCCGATCTTGCAGCTACGGATGAAAACCAT-3

OFFA-11R 5'-CACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCAGAATACCTCCCCGCCAG-
 3

OFFA-17F 5'-GGAGTTCAGACGTGTGCTCTTCCGATCTGGTGGAAACAATCCACCTGTATTAGC-3

OFFA-17R 5'-CACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAATGTGACACCACCACCGC-
 3

OFFA-23F 5'-GGAGTTCAGACGTGTGCTCTTCCGATCTTGTTTAGTAATTAAGACCCTGGCTTTC-3
 5-

OFFA-23R 'CACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCGACAGGTACAAAGCAGTCCA
 T-3

Supplementary Results

Summary of the PACE platform

Phage-assisted continuous evolution (PACE) allows proteins to undergo directed evolution at a rate ~100-fold faster than conventional methods (**Supplementary Fig. 1a**)⁸. During PACE, host *E. coli* cells continuously dilute an evolving population of filamentous bacteriophages (“selection phage”, SP) in a fixed-volume vessel (a “lagoon”). Dilution occurs faster than cell division but slower than phage replication, ensuring that only the phage can accumulate mutations. Each SP carries an evolving gene instead of gene III, an essential phage gene that is required for infection. Phage encoding active variants trigger host-cell expression of gene III from the “accessory plasmid” (AP) and produce infectious progeny, while phage encoding less active variants produce non-infectious progeny that are diluted out of the lagoon.

Development and validation of a PACE one-hybrid system using Zif268

We developed an assay that transduces cognate DNA-binding of the DBD from Zif268 (residues 333-420)⁹, expressed from a tetracycline-inducible promoter, into activation of pIII-luciferase expression (see **Supplementary Note 1** for plasmid construction). We used this assay to evaluate a variety of DNA operator locations (at -55 and -62 bp with respect to the transcription initiation site)^{10,11} and RNA polymerase fusion architectures using S1030 cells (see **Supplementary Note 2**). Fusing the RNAP ω subunit to the N-terminus of Zif268 with an 11-residue linker (see **Supplementary Note 3** for fusion protein sequence) resulted in ≥ 10 -fold increase in pIII-luciferase production when the consensus Zif268 binding site (5'-GCGTGGGCG-3') was

positioned at -62 (**Supplementary Fig. 1b**). To test the DNA specificity of this system, we created a control construct with an off-target Zif268 binding site in which the middle triplet of the target DNA site was changed to 5'-TTA-3'¹². Only *E. coli* containing the reporter downstream of the on-target sequence, but not those containing the off-target sequence, produced pIII-luciferase (**Supplementary Fig. 1c**), establishing sequence-specific and DNA binding-dependent gene expression.

Propagation of Zif268-based SP on bacteria containing the cognate AP *in vitro*

After developing Zif268-based AP and SP constructs, we developed an *E. coli* strain designated S2060 (see **Supplementary Note 1**) capable of inducing *LacZ* in response to activation of the phage shock promoter, a transcriptional regulatory element that responds to a number of environmental signals including filamentous phage infection¹³ (**Supplementary Figs. 2-4**). This strain can be used in combination with colorimetric *LacZ* substrates such as X-gal to stain bacteria that have been infected with phage. We tested if ω -Zif268-SP could propagate in a DNA-binding activity-dependent manner on S2060 cells containing an AP with the cognate Zif268 binding sequence, or a mutated binding sequence. We observed robust formation of colored plaques, indicative of phage propagation, on cells harboring the on-target AP, but not on cells harboring an AP containing the off-target sequence (**Supplementary Fig. 5a**). These observations demonstrate DNA binding activity-dependent phage propagation.

Continuous propagation of Zif268-SP in DB-PACE

After validating the ability of the Zif268-SP to propagate on the cognate AP in vitro, we next performed an initial PACE experiment to optimize the SP backbone. We continuously propagated SPs encoding Zif268 in PACE over 24 h on host cells carrying the cognate AP plasmid and a mutagenesis plasmid (MP)⁸. After 24 h of PACE, the surviving SPs contained mutations in the phage genes encoding pII/X and pIV, and the fusion protein linker (**Supplementary Fig. 5b**). These results collectively establish a basis for the continuous evolution of DBDs using DNA-binding PACE (DB-PACE).

Reversion of an inactive Zif268 DBD bearing a 3 bp mutation using DB-PACE

Mutation of Arg24 in Zif268 to a small hydrophobic residue is known to abrogate DNA binding¹⁴. We seeded a lagoon with inactive ω -Zif268 SP containing an R24V mutation. After 24 h of neutral drift (mutation in the absence of any selection pressure)¹⁵ followed by 24 h of PACE on host cells containing the cognate AP, the evolved SPs were capable of propagating on the target AP (**Supplementary Fig. 5c**). All of the sequenced phage clones at the end of the 24-h PACE experiment contained the V24R reversion mutation using an Arg codon not present in the wild-type gene (AGA vs. CGC) (**Supplementary Fig. 5d**).

Development and validation of DB-PACE for TALE proteins

To enable DB-PACE to evolve TALE proteins with tailored properties, we first tested a series of C-terminal fusions between a previously reported TALE array targeting *CBX8* (right half-site TALE)¹⁶, and the RNAP ω subunit. Fusions using a linker of 18 or 28 amino acids of the natural TALE C-terminus followed by a GGGGS

sequence resulted in >10-fold gene activation in a luciferase assay (**Supplementary Fig. 6a** and **Supplementary Note 4**). We verified the sequence specificity of the system using an off-target sequence in the luciferase assay (**Supplementary Fig. 6b**), and performed on- and off-target colorimetric plaque assays using TALE- ω SP and host cells harboring the AP containing the target *CBX8* sequence. Similar to our findings with zinc fingers, we observed that both luciferase expression and phage propagation were dependent on the presence of the cognate TALE binding site (**Supplementary Figs. 6b** and **7a**). Next, we performed PACE for 24 h on the cognate AP to optimize the SP backbone. This experiment resulted in several mutations in the phage genome, as well as an A8V substitution in the ω RNAP subunit (**Supplementary Fig. 7b**), collectively suggesting the applicability of DB-PACE to TALE proteins.

Evolution of TALEs with altered 5' specificity using DB-PACE positive selection

We examined the 5' DNA specificity of the *CBX8*-targeting TALE using the gene III-luciferase reporter assay. We observed 2- to 3-fold higher luciferase induction for 5' T over the other bases (**Supplementary Fig. 7c**). Next, we created APs in which the 5' base of the cognate sequence was changed to 5' A, 5' C, or 5' G, and initiated three parallel PACE experiments to evolve TALEs with increased DNA-binding activity for each of these sequences (**Supplementary Fig. 8**). For each experiment, we performed selections in duplicate lagoons (L1 and L2). As shown in **Supplementary Fig. 9**, we isolated phage with up to 6-fold increased activity on 5' A relative to the canonical TALE protein, 5-fold increased activity on 5' C, and 5-fold higher activity on 5' G target sequences.

Analysis of mutations arising following DB-PACE positive selection to alter 5' targeting

Analysis of individual clones from the 5'A, C, and G evolution experiments revealed a variety of mutations occurring throughout the entire TALE protein sequence (**Supplementary Fig. 9**). High-throughput sequencing of $\sim 10^5$ phage revealed that TALE mutations A79E (62% L2), A133E (33% L2), E622K (60% L1, 37% L2), and Q711P (28% L2), were prevalent in 5' A lagoons (**Supplementary Fig. 10a**), while A79E (78% L2), L508F (77% L2), and K634R (74% L1) were dominant in 5' C lagoons (**Supplementary Fig. 10b**), and D7Y (97% L1), G565S (97% L1), and E622K (97% L1) were predominant in the 5' G lagoon (**Supplementary Fig. 10c**). In addition, V767G, corresponding to V38G in the RNAP ω subunit, was common among phage evolved in the 5' A and 5' G lagoons (**Supplementary Fig. 10a,c**). Combined with structure-activity analyses (**Supplementary Fig. 11**), these data reveal substitutions that alter TALE 5' nucleotide specificity and binding activity, and highlight the value of unbiased mutagenesis in DB-PACE.

Validation of DB-PACE negative selection system

To validate our PACE negative selection system, we generated cells carrying an AP requiring recognition of a 5'A-CBX8 sequence, in combination with one of three APNegs bearing 5' C-, G-, or T-CBX8 sequences. Using the TALE- ω SP evolved to bind to the 5' A sequence, we performed plaque assays on each of these strains in the presence of increasing doses of theophylline to modulate pIII-neg production resulting from binding to 5' C, G, or T sequences in the corresponding APNeg. We confirmed that

phage propagation could be suppressed in an activity- and theophylline-dependent manner (**Supplementary Fig. 12b,c**), establishing a negative selection system for DB-PACE.

Dual positive and negative selection DB-PACE to evolve specific 5'A TALE recognition

To perform simultaneous multiplexed negative selection against binding of *CBX8* target sequences beginning with 5' C, G, or T, we mixed three *E. coli* strains in equal proportion, each carrying an APNeg plasmid containing a 5' C, G, or T off-target sequence, together with a positive selection AP harboring the 5' A target site and an MP. The resulting mixed host cell population was used in a 144-h PACE experiment in which phage surviving the previous 5' A PACE experiment were subjected to increasing levels of negative selection stringency (+0.1 mM theophylline every 48 h).

Specificity changes in 5' recognition following positive and negative selection DB-PACE

The fact that clones surviving our dual positive and negative PACE selection displayed a substantial decrease in binding on 5'T sequences while little decrease on 5'C, and 5'G sequences, suggest stronger selection against binding 5' T sequences and weaker selection pressure against binding sequences starting with 5' C or 5' G. This outcome likely resulted from negative selection against the 5'T sequence engaging earlier in the 144 h experiment than negative selection on the 5' C and 5' G sequences. Consistent with this hypothesis, *in vitro* plaque assays showed that a low dose of

theophylline (0.2 mM) is sufficient to suppress evolved 5'-A phage propagation on cells carrying a 5' T-*CBX8* sequence, while a higher dose (0.4 mM) is required to block propagation on 5' C or 5' G sequences (**Supplementary Fig. 12c**). Based on the theophylline titration schedule, phage experienced ~48 h of negative selection against the 5' T sequence, but only 24 h against the 5' C and 5' G sequences. While differences in the genotypes and phenotypes observed in L1 and L2 reflects the stochastic nature of protein evolution, 100% of the evolved TALEs assayed following negative selection exhibited preferences for 5' A over 5' T (**Supplementary Fig. 13a**).

Genotype-phenotype analysis of evolved phage displaying preference for 5'A over 5'T

Sequencing ten individual clones from the end of the experiment revealed an average of nine amino acid substitutions distributed throughout each protein (**Supplementary Fig. 13a,b**), and high-throughput sequence analysis of phage pool genotypes revealed six predominant amino acid substitutions (K59E, Q513K, N562H, E622K, Q711P, V767G) in the 144 h population (**Supplementary Fig. 13c**). Of these, only K59E and Q513K (**Supplementary Fig. 14a,b**) emerged exclusively following negative selection, and only the N-terminal substitution K59E was amenable to study using site-directed mutagenesis due to the highly repetitive nature of TALE repeat arrays. We found that when present in isolation on the *CBX8* TALE K59E decreases affinity for the 5' T target sequence by 2-fold, but has little effect on sequences beginning with 5' A, C, or G (**Supplementary Fig. 14c**). To test if the effect of the K59E mutation is *CBX8*-TALE context-dependent, or if the mutation alters TALE specificity in

a general manner, we introduced this mutation to a different TALE protein targeting the *ATM* locus (see **Supplementary Table 1** for target sequences) and assayed the activity of the resulting mutant on the corresponding 5' A, C, G, or T target sequences. The K59E substitution in the *ATM* TALE increased activity on both 5' A and 5' T sequences by a factor of 2 and 1.5-fold, respectively, indicating that while this position impacts 5' specificity in both TALE proteins, the manner in which the K59E mutation affects DNA binding is context-dependent (**Supplementary Fig. 14d**). These results collectively show that coupled positive and negative selection DB-PACE can rapidly alter TALE 5' DNA specificity in a context-dependent manner by maintaining TALE activity on a site containing a target 5' nucleotide while evolving mutations that decrease binding to other 5' off-target sequences.

Negative selection against a genomic site highly cleaved by the *ATM-L* TALE (OffA17)

To improve the specificity of the *ATM-L* TALE by using DB-PACE to remove binding to the OffA17 off-target site, we first co-transformed the corresponding positive selection AP (on-target) and the negative selection APNeg (bearing the OffA17 binding site) into host cells to enable simultaneous positive and negative selection during PACE. Using these cells, we performed DB-PACE on the *ATM-L* TALE in duplicate lagoons (L1 and L2) at a flow rate of 1.3 vol/h. Increasing quantities of theophylline were added to each lagoon from a starting dose of 0 mM to a final dose of 0.4 mM (+0.1 mM every 24 h) to successively increase negative selection stringency. At 120 h, we pooled the evolved phage populations from L1 and L2, and subjected the mixture to a

subsequent 24 h PACE experiment in a single lagoon (L3) using a fixed concentration of 0.4 mM theophylline and a higher lagoon flow rate of 2.0 vol/h.

Genotype-phenotype analysis of evolved *ATM-L* TALEs

Sequencing of several evolved *ATM-L* TALEs revealed a variety of mutations in L1 and L2 (**Fig. 1c** and **Supplementary Fig. 16b**), but a fairly converged L3 population characterized by A252T, L338S, Q505K, and Q745P (**Fig. 1c** and **Supplementary Fig. 16c**). Analysis of a series of evolved clones revealed that while Q53R in combination with A252T improved specificity substantially, A252T in combination with Q505K, L338S, and Q745P improved specificity by an additional > 2-fold (**Supplementary Fig. 16d**). Although the highly repetitive nature of TALE array genes precludes site-directed mutagenesis studies on residues within repeats, we identified an *ATM-L* TALE variant containing only a single A252T mutation. This mutation in isolation exhibited on-target cleavage activity comparable to that of the canonical TALE, but drastically reduced cleavage of OffA17 (~0.3% compared to 6.3% for the canonical TALE) (**Supplementary Fig. 17a**). We also assayed the effect of the C-terminal Q745P substitution, corresponding to Q711P in the *CBX8*-TALE, in isolation by site-directed mutagenesis and determined that this mutation did not affect on-target or off-target cleavage activity *in vitro* (**Supplementary Fig. 17b**). The dynamic range of the gel-based *in vitro* TALEN cleavage assay was insufficient to distinguish between the specificity enhancement of two clones containing identical genotypes differing only by the presence (L3-2) or absence (L3-1) of L338S (**Fig. 1c**).

Description of specificity profiling technique

To perform high-throughput specificity profiling of our *ATM-L* TALEs we digested a DNA library sufficiently diverse to contain at least ten copies of all DNA sequences with six or fewer mutations from the on-target *ATM* sequence with either the canonical TALEN pair, or with TALEN pairs containing an evolved *ATM-L* TALE (L2-1, L3-1, or L3-2) in combination with the canonical *ATM-R* TALE (see **Supplementary Table 2** for statistics). The specificity profile was generated as previously described⁵. Next, we calculated the enrichment factor for each library member that survived selection by dividing its abundance after selection by its abundance before selection. Mean enrichment values for the on-target sequence ranged from ~8 to 20 across the various samples (**Supplementary Fig. 18a,b**).

Specificity profile of evolved *ATM-L* TALEs

TALEN pairs containing the evolved TALEs L3-1 and L3-2 showed a substantially decreased ability to cleave off-target sequences containing four to nine mutations relative to the canonical TALEN (**Supplementary Fig. 18b**). For example, L3-1 cleaved off-target sequences containing seven mutations ~7-fold less efficiently than the canonical TALEN (both at 2.5 nM), despite cleaving on-target sequences 2-fold more efficiently (**Supplementary Fig. 18b**). These results indicate that the evolved TALEs exhibit general improvements in specificity that are not limited to the OffA17 off-target site used during negative selection PACE, but instead increase the ability of the evolved TALEs to reject other related off-target sequences as well (**Supplementary Fig. 18**).

We calculated specificity scores to directly compare the preference of canonical and evolved *ATM*-L TALEs at each position in the TALEN target site for A, C, G, or T. Scores were calculated by subtracting pre- and post-selection base-pair frequencies, and normalizing values to the maximum possible change of the pre-selection frequency from perfect specificity (1.0) to complete lack of specificity (-1.0). Heat maps and quantitative bar graphs generated for the canonical TALEN pair were in agreement with previously reported observations⁵ (**Fig. 2a** and **Supplementary Figs. 19** and **20**). Cleavage by TALEN pairs incorporating the evolved TALEs L3-1 or L3-2 exhibited substantially increased specificity relative to that of the canonical TALEN at nearly all positions in the left half-site of the *ATM* binding sequence, but no substantial change in specificity in the right half-site that was not used during DB-PACE (**Fig. 2b-d** and **Supplementary Figs. 21-24**). Taken together, these results demonstrate that DB-PACE can be used to greatly reduce TALEN cleavage of a specific genomic off-target sequence, and that the resulting specificity enhancements are not confined to that off-target substrate but instead apply to many other off-target sequences.

***ATM* TALEN DNA cleavage in mammalian cells**

To assay the specificity of our canonical and evolved *ATM*-directed TALENs in cells, we nucleofected U2OS cells with a control plasmid, or plasmids expressing heterodimeric FokI fusions to either the canonical *ATM*-L TALE or evolved L3-1 or L3-2 TALEs, together with a plasmid encoding the canonical *ATM*-R TALE–FokI fusion protein. After 48 h, we harvested genomic DNA and performed high-throughput sequencing analysis to examine cleavage at the on-target site, off-target site OffA17,

and three additional unrelated off-target sites OffA1, OffA11, and OffA23⁵. Cleavage at the on-target *ATM* site was comparable for the canonical and evolved TALENs (**Supplementary Tables 1 and 3**). Importantly, for all four off-target sites, both evolved TALENs exhibited reduced off-target activity relative to the canonical TALEN (**Supplementary Table 3**). These data establish that DB-PACE can be used to improve the specificity of a DNA-binding domain. Moreover, our results demonstrate that the mutations that confer improved specificity during DB-PACE selection can be applied to other TALE effector contexts, such as incorporation into a TALEN pair for genome modification in human cells with improved DNA specificity.

Supplementary Discussion

Structure-activity relationships of novel amino acid substitutions

While two directed evolution studies using combinatorial libraries^{17,18} have supported the original notion that TALE specificity at the 5' position is mediated exclusively by Trp120 (Trp232 in AvrBs3 structure³), our results identify Lys59 and Ala79 as two residues that also determine 5' nucleotide specificity. Mutation of Ala79 or Lys59, which are located within an extended N-terminal DNA-binding region near Trp120⁴, to glutamate resulted in altered 5' specificity (**Supplementary Figs. 11d and 14c,d**). While these residues are not predicted to directly contact DNA (**Supplementary Figs. 11c and 14a**), their effects are likely mediated through their interactions with Trp120, which is predicted to contact the 5' nucleotide⁴. Moreover, using DB-PACE to alter 5' nucleotide targeting identified a large number of additional amino acid

substitutions throughout the entire TALE sequence, and identified context-dependent effects for residues such as K59E that alter 5' specificity in a non-modular fashion (**Supplementary Fig. 14c,d**). Our results therefore support the more recent hypothesis that 5' base specificity is altered in a complex fashion that depends on the context of TALE repeats and their RVD compositions¹⁹⁻²¹. Our findings also suggest that TALE proteins with the most desirable properties, including high activity and high specificity, may contain mutations such as K59E that are not entirely modular but rather specific to the TALE protein of interest. Because such mutations are difficult or impossible to predict using standard TALE design principles, we envision DB-PACE as an ideal method to improve TALE arrays designed by modular assembly.

Our data also support recent observations that TALE activity can be altered in an effector-dependent manner²¹. We observed mutation of Gln711 (or equivalent), present in an unstructured area of the C-terminus (**Supplementary Fig. 11e**), to Pro in all PACE experiments performed. While this substitution doubled the activity of TALE- ω fusions (**Supplementary Fig. 11b**), likely through introduction of a kink in protein backbone that resulted in more effective presentation of the ω RNAP subunit, it had no effect on TALEN activity (**Supplementary Fig. 17b**).

The high efficiency of PACE facilitates the accumulation of many permissive amino acid substitutions in evolving proteins. The results shown in **Supplementary Fig. 11a** reveal novel sequence variability within the normally highly conserved core TALE unit. Of particular note are substitutions D4K/N and S11K within the first helix²², and K16R, T21A, and L26F in the second helix²², all of which arose in multiple evolution experiments, and in some cases, in multiple different TALE array repeats

(**Supplementary Fig. 10**). Position Lys16, as illustrated in the context of a K634R mutation in **Supplementary Fig. 11h**, is adjacent to the RVD loop and makes a non-specific DNA contact²². These results suggest that Arg may be used as a possible alternative at position 16 for this DNA contact. In addition, we observed several RVD substitutions, including replacement of the less specific NN RVD, which targets both G and A²³, with the more specific NK and NH repeats²⁴, as well as substitution of NG with HG, a repeat present in naturally occurring TALEs that also specifies T, but is not typically used in the design of synthetic TALEs (**Supplementary Fig. 11a**)²⁴.

It has been shown that the highly repetitive nature of TALE genes is incompatible with lentiviral delivery vectors due to recombination between repeat units arising from “template-switching” during DNA replication²⁵. The permissive core amino acid substitutions discovered in this work could enable recoding of TALE arrays to decrease sequence homology and thereby improve the manipulability and application scope of TALE proteins.

Remarkably, we observed that a single amino acid substitution, A252T, corresponding to the eighth residue within the third repeat of the *ATM-L* TALE, can greatly diminish binding of the OffA17 off-target site, which contains mutations at target site nucleotides 1, 5, 12, 16, and 17. Structural analysis predicts that Ala252 lies in close proximity to the RVD loop (**Supplementary Fig. 17c**), and it has been suggested that this residue stabilizes the loop^{22,26,27}. It is plausible that mutation of this residue to a larger and more polar Thr residue results in an altered or additional DNA contact, altering specificity of the entire array. The fact that evolved *ATM-L* TALE L3-2 showed even greater specificity than L3-1 (**Supplementary Fig. 24a,b**) and differed only by the

presence of L338S suggests that this position is also a determinant of specificity. Leu338, which corresponds to position 26 in a TALE repeat, is adjacent to Pro339 (**Supplementary Fig. 17d**), a residue that is essential for proper packing of TALE repeats²². We speculate that L338S may adjust repeat packing in a way that decreases excess binding energy and thereby augments specificity.

Supplementary Notes

Supplementary Note 1. Summary of plasmid constructs.

| Name | Class | Antibio. Res. | Origin of Rep. | Promoter | Binding Site | Gene |
|------------------|-------------------------|---------------|----------------|--|--------------------------|------------------------------------|
| pOHZif268-1 | One-hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (Zif268 fusion) | -55 (Zif268) | pIII-luxAB Zif268 DBD-(M)-rpoZ |
| pOHZif268-2 | One-hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (Zif268 fusion) | -55 (Zif268) | pIII-luxAB Zif268 DBD-(L)- rpoZ |
| pOHZif268-3 | One-hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (Zif268 fusion) | -62 (Zif268) | pIII-luxAB Zif268 DBD-(M)- rpoZ |
| pOHZif268-4 | One-hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (Zif268 fusion) | -62 (Zif268) | pIII-luxAB Zif268 DBD-(L)- rpoZ |
| pOHZif268-5 | One-hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (Zif268 fusion) | -55 (Zif268) | pIII-luxAB rpoZ-(M)-Zif268 DBD |
| pOHZif268-6 | One-hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (Zif268 fusion) | -55 (Zif268) | pIII-luxAB rpoZ-(L)-Zif268 DBD |
| pOHZif268-7 | One-hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (Zif268 fusion) | -62 (Zif268) | pIII-luxAB rpoZ-(M)-Zif268 DBD |
| pOHZif268-7: TTA | One-hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (Zif268 fusion) | -62 5'GCGTTA GCG3' | pIII-luxAB rpoZ-(M)-Zif268 DBD |
| pOHZif268-8 | One-hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (Zif268 fusion) | -62 (Zif268) | pIII-luxAB rpoZ-(L)-Zif268 DBD |
| pOHZif268-9 | One-hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (Zif268 fusion) | -55 (Zif268) | pIII-luxAB rpoA-(M)-Zif268 DBD |
| pOHZif268-10 | One-hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (Zif268 fusion) | -62 (Zif268) | pIII-luxAB rpoA-(M)-Zif268 DBD |
| SPZif268 | SP | Kan | F1 | P _{gIII} | - | rpoZ-(M)-Zif268 DBD |
| SPZif268-R24V | SP | Kan | F1 | P _{gIII} | - | rpoZ-(M)-Zif268 DBD-R24V |
| pAPZif268 | AP | Carb | SC101 | P _{lac} | -62 (Zif268) | pIII-luxAB |
| pAPZif268: TTA | AP | Carb | SC101 | P _{lac} | -62 5'GCGTTA GCG3' | pIII-luxAB |
| pOHCBXTAL-1 | One-hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (CBX8 TALE fusion) | -62 (CBX8) | pIII-luxAB TALE(CBX8)-+28-rpoZ |
| pOHCBXTAL-2 | One-hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (CBX8 TALE fusion) | -62 (CBX8) | pIII-luxAB TALE(CBX8)-+40-rpoZ |
| pOHCBXTAL-3 | One-hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (CBX8 TALE fusion) | -62 (CBX8) | pIII-luxAB TALE(CBX8)-+63-rpoZ |

| | | | | | | |
|---|-------------------------|-------|-------|--|---------------------------------------|---|
| pOHCBTAL-4 | One-hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (CBX8) TALE fusion) | -62 (CBX8) | pIII-luxAB TALE(CBX8)- +18G ₄ S-rpoZ |
| pOHCBTAL-4:Offtarget | One-hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (CBX8) TALE fusion) | -62 5' TTCATAA GGGATTA GGC3' | pIII-luxAB TALE(CBX8)- +18G ₄ S-rpoZ |
| pOHCBTAL-4:A79E, A133E, Q711P, A755V V767G | One-hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (CBX8) TALE fusion) | -62 (CBX8) | pIII-luxAB TALE(CBX8)- +18G ₄ S-rpoZ |
| pOHCBTAL-4:5A,L1-1..5, L2-1..5, A79E, K59E | One-hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (CBX8) TALE fusion) | -62 5'ATCAGG AGGGCTT CGGC 3' | pIII-luxAB TALE(CBX8)- +18G ₄ S-rpoZ |
| pOHCBTAL-4:5C,L1-1..5, L2-1..5, A79E, K59E | One-hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (CBX8) TALE fusion) | -62 5'CTCAGG AGGGCTT CGGC 3' | pIII-luxAB TALE(CBX8)- +18G ₄ S-rpoZ |
| pOHCBTAL-4:5G, L1-1..5, L2-1..5, A79E, K59E | One-hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (CBX8) TALE fusion) | -62 5'GTCAGG AGGGCTT CGGC3' | pIII-luxAB TALE(CBX8)- +18G ₄ S-rpoZ |
| pOHCBTAL-5 | One-hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (CBX8) TALE fusion) | -62 (CBX8) | pIII-luxAB TALE(CBX8)- +28G ₄ S-rpoZ |
| SPCBXTAL | SP | Kan | F1 | P _{gIII} | - | TALE(CBX8)- +18G ₄ S-rpoZ |
| pApCBXTAL | AP | Carb | SC101 | P _{lac} | -62 (CBX8) | pIII-luxAB |
| pApCBXTAL:5A | AP | Carb | SC101 | P _{lac} | -62 5'ATCAGG AGGGCTT CGGC 3' | pIII-luxAB |
| pApCBXTAL:5C | AP | Carb | SC101 | P _{lac} | -62 5'CTCAGG AGGGCTT CGGC 3' | pIII-luxAB |
| pApCBXTAL:5G | AP | Carb | SC101 | P _{lac} | -62 5'GTCAGG AGGGCTT CGGC 3' | pIII-luxAB |
| pAPCBXTAL:Offtarget | AP | Carb | SC101 | P _{lac} | -62 5' TTCATAA GGGATTA GGC3' | pIII-luxAB |
| pAPNegCBXTAL:5A | AP-neg | Spect | ColE1 | P _{lac} | -62 5'ATCAGG AGGGCTT CGGC 3' | TheoRibo- 6xHistag-N- C83-Venus |
| pAPNegCBXTAL:5C | AP-neg | Spect | ColE1 | P _{lac} | -62 5'CTCAGG AGGGCTT CGGC 3' | TheoRibo- 6xHistag-N- C83-Venus |
| pAPNegCBXTAL:5G | AP-neg | Spect | ColE1 | P _{lac} | -62 5'GTCAGG AGGGCTT CGGC 3' | TheoRibo- 6xHistag-N- C83-Venus |

| | | | | | | |
|--|--|-------|-------|--|---|---|
| pAPNegCBXT AL:5T | AP-neg | Spect | ColE1 | P _{lac} | -62 5'TTCAGG AGGGCTT CGGC 3' | TheoRibo- 6xHistag-N- C83-Venus |
| pAPNegCBXT AL:Offtarget | AP-neg | Spect | ColE1 | P _{lac} | -62 5' TTCATAA GGGATTA GGC3' | TheoRibo- 6xHistag-N- C83-Venus |
| pTetCBXTAL | Inducible TALE express. | Carb | SC101 | P _{tet} (CBX8 TALE fusion) | - | TALE(CBX8)- +18G ₄ S-rpoZ |
| pTetCBXTAL:L 1-1, L1-2, L2-1, L2-2 | Inducible TALE express. | Carb | SC101 | P _{tet} (CBX8 TALE fusion) | - | TALE(CBX8)- +18G ₄ S-rpoZ |
| SPATMTAL | SP | Kan | F1 | P _{gIII} | - | TALE(ATM- L)+18G ₄ S-rpoZ pIII-luxAB |
| pApATMTAL | AP | Carb | SC101 | P _{lac} | -62 (ATM- L) | |
| pAPNegATMT AL:OffA17 | AP-neg | Spect | ColE1 | P _{lac} | -62 5'- GAAATGG GATACTG AGT3' | TheoRibo- 6xHistag-N- C83-Venus |
| pUC19-On- target | TALEN cleavage | Carb | pMB1 | - | Cleavage site: ATM-L | - |
| pUC19-Off- target, pUC19- OffD1-D4 | TALEN cleavage | Carb | pMB1 | - | Cleavage site: 5'- GAAATGG GATACTG AGT3' or derivative site | - |
| pJG29:L1-1, L1-2, L2-1, L3- 1..L3-4, Q745P | Backbone information previously described ⁵ | | | | | |
| pJG30 | Backbone information previously described ⁵ | | | | | |
| pJG51: L3-2 | Backbone information previously described ⁵ | | | | | |
| pJG52 | Backbone information previously described ⁵ | | | | | |
| pOHATMTAL | One- hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (TALE fusion) | -62 (ATM-L) | TALE(ATM-L)- +18G ₄ S-rpoZ |
| pOHATMTAL:L 3-1, L3-2 | One- hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (TALE fusion) | -62 (ATM-L) | TALE(ATM-L)- +18G ₄ S-rpoZ |
| pOHATMTAL: OffA17 | One- hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (TALE fusion) | -62 5'- GAAATGG GATACTG AGT3' | TALE(ATM-L)- +18G ₄ S-rpoZ |
| pOHATMTAL: OffA17:L3-1, L3-2 | One- hybrid test plasmid | Carb | SC101 | P _{lac} (pIII-luc) P _{tet} (TALE fusion) | -62 5'- GAAATGG GATACTG AGT3' | TALE(ATM-L)- +18G ₄ S-rpoZ |
| pOHATMTAL:5 | One- | Carb | SC101 | P _{lac} (pIII-luc) | -62 | TALE(ATM-L) |

| | | | | | | | |
|--|-------------------------|-------|-------|---|---|---------------------------------------|---------------------------------------|
| 'A,C,G and K59E-5'A,C,G,T | hybrid test plasmid | | | | P_{tet} (TALE fusion) | (ATM-L) or 5'A,C,G sequence variant | or K59E mut)-+18G ₄ S-rpoZ |
| pOHCCR5TAL, pOHCCR5TAL: Off5, Off15, Off28 | One-hybrid test plasmid | Carb | SC101 | P_{lac} (pIII-luc) P_{tet} (TALE fusion) | 5'-TCTTCCA GAATTGA TACT-'3 or off-target site | TALE(CCR5-R)-+18G ₄ S-rpoZ | |
| pAB086a | MP | Chlor | | | RecA- version of pJC184 ¹⁵ | | |

Supplementary Note 2. Genotypes of bacterial strains used.

| Strain | Genotype |
|--------|--|
| S1030 | <i>F'</i> <i>proA+B+</i> Δ (<i>lacIZY</i>) <i>zzf::Tn10 lacI^{Q1} P_{N25}-tetR luxCDE / endA1 recA1 galE15 galK16 nupG rpsL ΔlacIZYA araD139 Δ(ara,leu)7697 mcrA Δ(<i>mrr-hsdRMS-mcrBC</i>) proBA::<i>pir116 araE201 ΔrpoZ Δflu ΔcsgABCDEFG ΔpgaC λ^-</i></i> |
| S1059 | <i>F'</i> <i>proA+B+</i> Δ (<i>lacIZY</i>) <i>zzf::Tn10 lacI^{Q1} P_{N25}-tetR luxCDE / endA1 recA1 galE15 galK16 nupG rpsL ΔlacIZYA araD139 Δ(ara,leu)7697 mcrA Δ(<i>mrr-hsdRMS-mcrBC</i>) proBA::<i>pir116 araE201 ΔrpoZ λ^- pJC175e¹⁵</i></i> |
| S1632 | <i>F'</i> <i>proA+B+</i> Δ (<i>lacIZY</i>) <i>zzf::Tn10 lacI^{Q1} P_{N25}-tetR luxCDE / endA1 recA1 galE15 galK16 nupG rpsL ΔlacIZYA araD139 Δ(ara,leu)7697 mcrA Δ(<i>mrr-hsdRMS-mcrBC</i>) proBA::<i>pir116 araE201 ΔrpoZ Δflu ΔcsgABCDEFG ΔpgaC ΔpspBC λ^-</i></i> |
| S2058 | <i>F'</i> <i>proA+B+</i> Δ (<i>lacIZY</i>) <i>zzf::Tn10 lacI^{Q1} P_{N25}-tetR luxCDE P_{psp} lacZ luxR P_{lux} groESL / endA1 recA1 galE15 galK16 nupG rpsL ΔlacIZYA araD139 Δ(ara,leu)7697 mcrA Δ(<i>mrr-hsdRMS-mcrBC</i>) proBA::<i>pir116 araE201 ΔrpoZ Δflu ΔcsgABCDEFG ΔpgaC λ^-</i></i> |
| S2059 | <i>F'</i> <i>proA+B+</i> Δ (<i>lacIZY</i>) <i>zzf::Tn10 lacI^{Q1} P_{N25}-tetR luxCDE P_{psp}(T1) lacZ luxR P_{lux} groESL / endA1 recA1 galE15 galK16 nupG rpsL ΔlacIZYA araD139 Δ(ara,leu)7697 mcrA Δ(<i>mrr-hsdRMS-mcrBC</i>) proBA::<i>pir116 araE201 ΔrpoZ Δflu ΔcsgABCDEFG ΔpgaC λ^-</i></i> |
| S2060 | <i>F'</i> <i>proA+B+</i> Δ (<i>lacIZY</i>) <i>zzf::Tn10 lacI^{Q1} P_{N25}-tetR luxCDE P_{psp}(AR2) lacZ luxR P_{lux} groESL / endA1 recA1 galE15 galK16 nupG rpsL ΔlacIZYA araD139 Δ(ara,leu)7697 mcrA Δ(<i>mrr-hsdRMS-mcrBC</i>) proBA::<i>pir116 araE201 ΔrpoZ Δflu ΔcsgABCDEFG ΔpgaC λ^-</i></i> |
| S2208 | <i>F'</i> <i>proA+B+</i> Δ (<i>lacIZY</i>) <i>zzf::Tn10 lacI^{Q1} P_{N25}-tetR luxCDE P_{psp}(AR2) lacZ luxR P_{lux} groESL / endA1 recA1 galE15 galK16 nupG rpsL ΔlacIZYA araD139 Δ(ara,leu)7697 mcrA Δ(<i>mrr-hsdRMS-mcrBC</i>) proBA::<i>pir116 araE201 ΔrpoZ Δflu ΔcsgABCDEFG ΔpgaC λ^- pJC175e¹⁵</i></i> |

Supplementary Note 3. DNA and protein coding sequences for the ω -Zif268-DBD fusion protein

Bases 997-1260 of *m. musculus* Zif268, corresponding the zinc finger DNA-binding domain (residues 333-420)⁹, were cloned in downstream of the RNAP ω subunit.

5'-

```
atggcacgcgtaactgttcaggacgctgtagagaaaattggaaccgtttgacctggtactggtcgccgcgctcgcgctc
gtcagatgcaggtaggcggaaaggatccgctggtaccggaagaaaacgataaaaccactgtaatcgcgctgcgcgaa
atcgaagaaggctgatcaacaaccagatcctcgacgttcggaacgccaggaacagcaagagcaggaagccgctg
aattacaagccgttaccgctattgctgaaggctcgtcgtcgggcgggaggcggcagcaccgcgggcgtgaacgcc
catatgcttgccctgtcgagtcctgcatcgccgctttctcgctcggatgagcttaccgcatatccgcatccacacaggc
cagaagcccttcagtgatgaatctgcatgtaacttcagtcgtagtaccacctaccacccacatccgcacccacaca
ggcgagaagcctttgcctgtgacatttggggaggaagttgcccaggagtgatgaacgcaagaggcatacctaaaatcca
ttaagacagaagtaa-3'
```

The protein sequence of the ω -Zif268-DBD fusion protein is shown below. The residues highlighted in blue correspond to the ω subunit, while the residues highlighted in red correspond to the 11-amino acid linker. Residues shown in grey background comprise the Zif268-DBD (residues 333-420)⁹.

```
MARVTVQDAVEKIGNRFDLVLVAARRARQM QVGGKDPLVP
EENDKTTVIALREIEEGLINNQILDVRERQEQQEQEAAELQA
VTAIAEGRRAAGGGGSTAAAPERPYACPVE SCDRRFSRSDEL
TRHIRIHTGQKPFQCRICMRNFSRSDHLTTHIRHTHTGEKPF
A
CDICGRKFARSDERKRHTKIHLRQK Stop
```

Supplementary Note 4. Coding sequences for the *CBX8*- and *ATM-L*- directed TALE- ω fusion proteins.

DNA sequences for the *CBX8*-directed TALE^{16,28} and the *ATM-L* directed TALE^{5,16,28} have previously been reported. The protein sequences of both TALE- ω fusion proteins are included below, indicating the appropriate residue numbering convention used in this manuscript. The non-highlighted residues comprise an N-terminal Flag-tag and NLS sequence, while the residues highlighted in grey correspond to the canonical N-terminal TALE sequence. TALE repeats are highlighted in green, the C-terminal region and linker sequence are highlighted in red, and the ω subunit is highlighted in blue. The DNA and protein sequences for the *CCR5-R* TALE have also been previously reported⁵. The fusion architecture for the *CCR5-R* TALE- ω protein is identical to that of the *CBX8* and *ATM-L*-directed TALEs described below.

***CBX8*-directed TALE- ω fusion protein:**

```
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVG IHRGVPM
VDLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTHAHI
VALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSG
ARALEALLTVAGELRGPPLQLDTGQLLKI AKRGGVTAVEAV
HAWRNALTGAPLN
LTPDQVVVAIASNGGGKQALETVQRLLPVLCQD HG
LTPEQVVVAIASHDGGKQALETVQRLLPVLCQA HG
LTPDQVVVAIASNIGGGKQALETVQRLLPVLCQA HG
LTPAQVVVAIANNNGGKQALETVQRLLPVLCQD HG
LTPDQVVVAIANNNGGKQALETVQRLLPVLCQD HG
LTPEQVVVAIASNIGGGKQALETVQRLLPVLCQA HG
LTPDQVVVAIANNNGGKQALETVQRLLPVLCQA HG
LTPAQVVVAIANNNGGKQALETVQRLLPVLCQD HG
LTPDQVVVAIANNNGGKQALETVQRLLPVLCQD HG
LTPEQVVVAIASHDGGKQALETVQRLLPVLCQA HG
LTPDQVVVAIASNGGGKQALETVQRLLPVLCQA HG
LTPAQVVVAIASNGGGKQALETVQRLLPVLCQD HG
LTPDQVVVAIASHDGGKQALETVQRLLPVLCQD HG
LTPEQVVVAIANNNGGKQALETVQRLLPVLCQA HG
LTPDQVVVAIANNNGGKQALETVQRLLPVLCQA HG
LTPEQVVVAIASHDGGRPALE
```

SIVAQLSRPDPALAAALTNGGGGS ARVTVQDAVEKIGNRFDL
VLVAARRARQMVGKDPVPEENDKTTVIALREIEEGLINN
QILDVRERQEQQEQEAAELQAVTAIAEGRR Stop

ATM-directed TALE-⁰ fusion protein:

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPM
VDLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTHAHI
VALSQHPAALGTAVVKYQDMIAALPEATHEAIVGVGKQWSG
ARALEALLTVAGELRGPPLQLDTGQLLKIAGRGGVTAVEAV
HAWRNALTGAPLN

LTPDQVVVAIANNNGGKQALETVQRLLPVLCQDHG
LTPEQVVVAIASNIGGKQALETVQRLLPVLCQAHG
LTPDQVVVAIASNIGGKQALETVQRLLPVLCQAHG
LTPAQVVVAIASNIGGKQALETVQRLLPVLCQDHG
LTPDQVVVAIASNIGGKQALETVQRLLPVLCQDHG
LTPEQVVVAIANNNGGKQALETVQRLLPVLCQAHG
LTPDQVVVAIANNNGGKQALETVQRLLPVLCQAHG
LTPAQVVVAIANNNGGKQALETVQRLLPVLCQDHG
LTPDQVVVAIASNIGGKQALETVQRLLPVLCQDHG
LTPEQVVVAIASNIGGKQALETVQRLLPVLCQAHG
LTPDQVVVAIANNNGGKQALETVQRLLPVLCQAHG
LTPAQVVVAIASHDGGKQALETVQRLLPVLCQDHG
LTPDQVVVAIASNIGGKQALETVQRLLPVLCQDHG
LTPEQVVVAIANNNGGKQALETVQRLLPVLCQAHG
LTPDQVVVAIASNIGGKQALETVQRLLPVLCQAHG
LTPAQVVVAIASNIGGKQALETVQRLLPVLCQDHG
LTPEQVVVAIASNIGGGRPALE

SIVAQLSRPDPALAAALTNGGGGS ARVTVQDVVEKIGNRFDL
VLVAARRARQMVGKDPVPEENDKTTVIALREIEEGLINN
QILDVRERQEQQEQEAAELQAVTAIAEGRR Stop

Supplementary References

1. Dworkin, J., Jovanovic, G. & Model, P. Role of upstream activation sequences and integration host factor in transcriptional activation by the constitutively active prokaryotic enhancer-binding protein PspF. *J Mol Biol* **273**, 377-88 (1997).
2. Wang, L. & Gralla, J.D. Multiple in vivo roles for the -12-region elements of sigma 54 promoters. *J Bacteriol* **180**, 5626-31 (1998).
3. Stella, S. *et al.* Structure of the AvrBs3-DNA complex provides new insights into the initial thymine-recognition mechanism. *Acta Crystallogr D Biol Crystallogr* **69**, 1707-16 (2013).
4. Gao, H., Wu, X., Chai, J. & Han, Z. Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell Res* **22**, 1716-20 (2012).
5. Guilinger, J.P. *et al.* Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nat Methods* **11**, 429-35 (2014).
6. Pattanayak, V., Ramirez, C.L., Joung, J.K. & Liu, D.R. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Methods* **8**, 765-70 (2011).
7. Zuris, J.A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat Biotechnol* (2014).
8. Esvelt, K.M., Carlson, J.C. & Liu, D.R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499-503 (2011).
9. Choo, Y. & Klug, A. Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. *Proc Natl Acad Sci U S A* **91**, 11163-7 (1994).
10. Hu, J.C., Kornacker, M.G. & Hochschild, A. Escherichia coli one- and two-hybrid systems for the analysis and identification of protein-protein interactions. *Methods* **20**, 80-94 (2000).
11. Durai, S., Bosley, A., Abulencia, A.B., Chandrasegaran, S. & Ostermeier, M. A bacterial one-hybrid selection system for interrogating zinc finger-DNA interactions. *Comb Chem High Throughput Screen* **9**, 301-11 (2006).
12. Maeder, M.L. *et al.* Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell* **31**, 294-301 (2008).
13. Beekwilder, J., Rakonjac, J., Jongsma, M. & Bosch, D. A phagemid vector using the E. coli phage shock promoter facilitates phage display of toxic proteins. *Gene* **228**, 23-31 (1999).
14. Elrod-Erickson, M. & Pabo, C.O. Binding studies with mutants of Zif268. Contribution of individual side chains to binding affinity and specificity in the Zif268 zinc finger-DNA complex. *J Biol Chem* **274**, 19281-5 (1999).
15. Carlson, J.C., Badran, A.H., Guggiana-Nilo, D.A. & Liu, D.R. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nat Chem Biol* **10**, 216-22 (2014).
16. Reyon, D. *et al.* Engineering customized TALE nucleases (TALENs) and TALE transcription factors by fast ligation-based automatable solid-phase high-throughput (FLASH) assembly. *Curr Protoc Mol Biol* **Chapter 12**, Unit 12 16 (2013).

17. Lamb, B.M., Mercer, A.C. & Barbas, C.F., 3rd. Directed evolution of the TALE N-terminal domain for recognition of all 5' bases. *Nucleic Acids Res* **41**, 9779-85 (2013).
18. Tsuji, S., Futaki, S. & Imanishi, M. Creating a TALE protein with unbiased 5'-T binding. *Biochem Biophys Res Commun* **441**, 262-5 (2013).
19. Schreiber, T. & Bonas, U. Repeat 1 of TAL effectors affects target specificity for the base at position zero. *Nucleic Acids Res* **42**, 7160-9 (2014).
20. Meckler, J.F. *et al.* Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Res* **41**, 4118-28 (2013).
21. Doyle, E.L. *et al.* TAL effector specificity for base 0 of the DNA target is altered in a complex, effector- and assay-dependent manner by substitutions for the tryptophan in cryptic repeat -1. *PLoS One* **8**, e82120 (2013).
22. Mak, A.N., Bradley, P., Cernadas, R.A., Bogdanove, A.J. & Stoddard, B.L. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* **335**, 716-9 (2012).
23. Moscou, M.J. & Bogdanove, A.J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501 (2009).
24. Cong, L., Zhou, R., Kuo, Y.C., Cunniff, M. & Zhang, F. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun* **3**, 968 (2012).
25. Holkers, M. *et al.* Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Res* **41**, e63 (2013).
26. Deng, D. *et al.* Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* **335**, 720-3 (2012).
27. Wicky, B.I., Stenta, M. & Dal Peraro, M. TAL effectors specificity stems from negative discrimination. *PLoS One* **8**, e80261 (2013).
28. Miller, J.C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29**, 143-8 (2011).