

FLASH assembly of TALENs for high-throughput genome editing

Deepak Reyon^{1–4}, Shengdar Q Tsai^{1,3,4}, Cyd Khayter¹, Jennifer A Foden¹, Jeffry D Sander^{1,3} & J Keith Joung^{1,3}

Engineered transcription activator–like effector nucleases (TALENs) have shown promise as facile and broadly applicable genome editing tools. However, no publicly available high-throughput method for constructing TALENs has been published, and large-scale assessments of the success rate and targeting range of the technology remain lacking. Here we describe the **fast ligation-based automatable solid-phase high-throughput (FLASH)** system, a rapid and cost-effective method for large-scale assembly of TALENs. We tested 48 FLASH-assembled TALEN pairs in a human cell–based EGFP reporter system and found that all 48 possessed efficient gene-modification activities. We also used FLASH to assemble TALENs for 96 endogenous human genes implicated in cancer and/or epigenetic regulation and found that 84 pairs were able to efficiently introduce targeted alterations. Our results establish the robustness of TALEN technology and demonstrate that FLASH facilitates high-throughput genome editing at a scale not currently possible with other genome modification technologies.

Engineered transcription activator–like (TAL) effector repeat domains have generated much interest as a new approach for creating customized DNA-binding proteins^{1–3}. TAL effector repeats are highly conserved, 33- to 35-amino-acid sequences found in naturally occurring TAL effectors encoded by *Xanthomonas* bacteria. One TAL effector repeat binds to a single DNA base pair, and the identities of amino acids at two positions (known as repeat variable di-residues (RVDs)) have been associated with specificities for different nucleotides^{4,5}. TAL effector repeats can be joined together into more extended arrays capable of recognizing novel target DNA sequences. Such engineered arrays have been fused to gene regulatory domains to create customized transcription factors^{4,6–11} and to nonspecific nuclease domains to create targeted TALENs^{6,12–22}. Repair of TALEN-induced double-strand breaks by either non-homologous end-joining (NHEJ) or homology-directed repair can induce efficient alteration of endogenous genes in yeast¹⁶, plants¹⁷, nematodes¹⁸, zebrafish^{21,22}, rats²⁰, and human somatic^{6,15,17} and pluripotent stem cells¹⁹.

Although the ability to design TALENs targeted to nearly any DNA sequence of interest has been highlighted as an important potential advantage of the technology^{1–3}, only a very limited number of endogenous genes (17 in total to our knowledge) have been altered using TALENs in the published literature to date^{6,15–22}. The TALENs used in these studies were constructed on different architectures and composed of variable numbers of TAL effector repeats, making it difficult to ascertain whether these parameters affect the efficiencies of nuclease activity. In addition, a recent report has suggested limits to the targeting range of TALENs based on a computational analysis of naturally occurring TAL effector binding sites¹⁷.

The simplicity of TALEN design raises the exciting prospect that large-scale pathway- and genome-wide gene modification projects might be possible, but currently no publicly available, cost-effective and high-throughput method for constructing these nucleases exists. Many different methods for constructing TALENs have been described^{8–11,16,17,21,22}, most of which use variations on the Golden Gate cloning method. However, none of these methods are readily adaptable for automated high-throughput production due to requirements for PCR, gel isolation of fragments, and/or passage and characterization of intermediate constructs^{8–11,16,17,21,22}. In addition, many of these methods enable production of TAL effector repeat arrays composed of only a certain fixed numbers of repeats^{8,9,11,16}. A commercial high-throughput platform exists—Cellestis Bioresearch has announced the capability to produce 7,200 TALENs per year—but details of this proprietary method are not publicly available.

Here we describe the development and optimization of the FLASH assembly method for rapid construction of large numbers of TAL effector repeat arrays. We used FLASH to construct 48 TALEN pairs targeted to a diverse range of EGFP-reporter gene sequences and found that 100% of these nucleases were active in a human cell–based assay. We also made FLASH TALEN pairs targeted to 96 human genes involved in cancer or epigenetic regulation and used them to rapidly introduce targeted alterations into 84 of these genes. **Our results provide large-scale experimental support for the broad and robust targeting range of TALEN technology and show that FLASH can enable rapid, high-throughput gene editing not currently possible with engineered zinc-finger nucleases (ZFNs) or meganucleases.**

¹Molecular Pathology Unit, Center for Computational and Integrative Biology, and Center for Cancer Research, Massachusetts General Hospital, Charlestown, Massachusetts, USA. ²Department of Genetics, Development and Cell Biology, Interdepartmental Graduate Program in Bioinformatics & Computational Biology, Iowa State University, Ames, Iowa, USA. ³Department of Pathology, Harvard Medical School, Boston, Massachusetts, USA. ⁴These authors contributed equally to this work. Correspondence should be addressed to J.K.J. (jjoung@partners.org) or J.D.S. (jsander@partners.org).

Received 7 December 2011; accepted 26 February 2012; published online 8 April 2012; doi:10.1038/nbt.2170

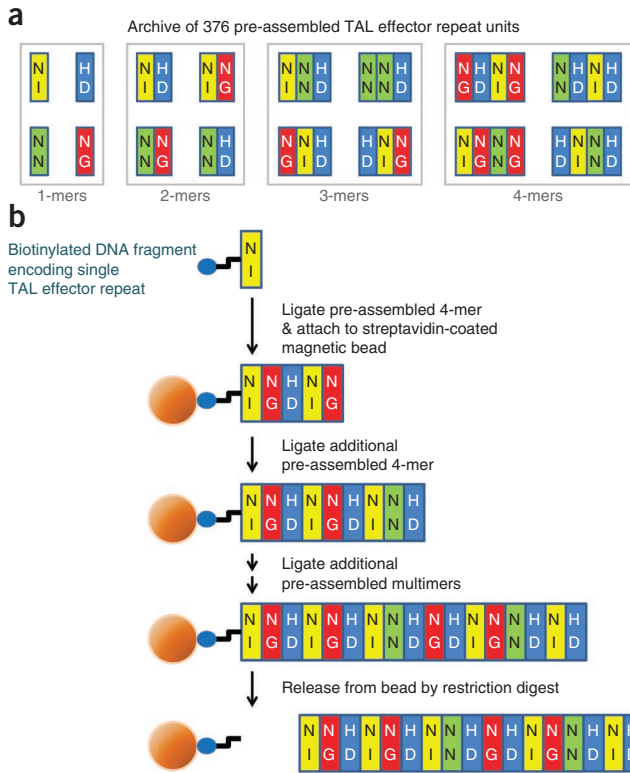


Figure 1 Graphical overview of the FLASH assembly method. **(a)** Archive of 376 TAL effector repeat-encoding plasmids required to practice FLASH. Plasmids encoding one, two, three and four TAL effector repeats (colored rectangles) harboring various RVDs (represented by two upper-case letters within the rectangles) were constructed as described in Online Methods. **(b)** Schematic overview of the FLASH assembly process. A DNA fragment encoding a single TAL effector repeat and labeled on its 5' end with biotin (blue oval) is initially ligated to a second DNA fragment encoding four specific TAL effector repeats and then attached to a streptavidin-coated magnetic bead (orange sphere). Additional DNA fragments encoding pre-assembled TAL effector repeats are ligated until an array of the desired length is assembled. The DNA fragment encoding the full-length TAL effector repeat array is then cleaved from the bead by restriction digestion.

each TALEN pair contained the same number of repeats (ranging from 8.5 to 19.5), and pairs were targeted to sites possessing a 16-bp spacer sequence between the half-sites (**Supplementary Table 3**). We tested each of these 48 TALEN pairs in human cells for its ability to disrupt the coding sequence of a chromosomally integrated *EGFP* reporter gene. In this assay, NHEJ-mediated repair of TALEN-induced breaks leads to loss of EGFP expression, which we quantitatively assessed 2 and 5 d after transfection. (All TALEN pairs we made targeted sites located at or upstream of nucleotide position 503 in *EGFP*, a position we have previously shown will disrupt *EGFP* function when mutated with a ZFN²³.) Notably, all 48 TALEN pairs showed significant *EGFP* gene-disruption activities, similar to those of four *EGFP*-targeted ZFN pairs previously made by the oligomerized pool engineering (OPEN) method (**Fig. 2a**). These results demonstrate that TALENs containing as few as 8.5 TAL effector repeats possess significant nuclease activity and provide large-scale evidence of the robustness of TALENs in human cells.

Requantification of the percentage of EGFP-disrupted cells at day 5 post-transfection revealed that a greater number of shorter-length TALENs (e.g., those composed of 8.5 to 10.5 repeats) showed significant reductions in the percentage of EGFP-disrupted cells than longer-length TALENs (**Fig. 2a**). This trend can also be observed by comparing mean percentages of EGFP-disrupted cells for TALENs of various lengths at day 2 and day 5 post-transfection (**Fig. 2b**) and by plotting mean ratios of day 2/day 5 EGFP-disrupted cells for each TALEN length (**Fig. 2c**). An analysis of variance (ANOVA) of data from **Figure 2c** demonstrated a significant effect of TALEN length on the day 2/day 5 ratio of EGFP-disrupted cells ($P = 8.9 \times 10^{-8}$). One potential explanation for this effect is cellular toxicity associated with expression of shorter-length TALENs. Consistent with this hypothesis, in cells transfected with plasmids encoding shorter-length TALENs, greater reductions occurred in the percentage of tdTomato-positive cells from day 2 to day 5 post-transfection (**Fig. 2d**) (a tdTomato-encoding plasmid was co-transfected with the TALEN expression plasmids on day 0). ANOVA of data from **Figure 2d** also showed a significant effect of TALEN length on the day 2/day 5 ratio of tdTomato-positive cells ($P = 1.5 \times 10^{-8}$). Taken together, our results suggest that although shorter-length TALENs are as active as longer-length TALENs, the former can be more cytotoxic in human cells.

Our EGFP experiments also provided an opportunity to assess five previously described computationally derived design guidelines¹⁷. All 48 of the sequences we targeted in *EGFP* failed to meet one or more of these guidelines (**Supplementary Discussion** and **Supplementary Table 3**), although all of these targets did meet the requirement for a 5' T in both half-sites. Our 100% success rate for these 48 sites demonstrates that TALENs can be readily obtained for target sequences that do not follow four of the guidelines. In addition, for each of these four guidelines, we did not find any statistically significant correlation

RESULTS

Automated assembly of TAL effector repeat arrays

Implementation of FLASH relies on an archive of 376 plasmids that encode one, two, three or four TAL effector repeats consisting of all possible combinations of the NI, NN, HD or NG RVDs (Online Methods, **Fig. 1a**, and **Supplementary Tables 1** and **2**). DNA fragments encoding TAL effector repeats are assembled in an iterative fashion on solid-phase magnetic beads, an innovation that permits automation of serial restriction digest, purification and ligation steps on a liquid-handling instrument (Online Methods, **Supplementary Methods** and **Fig. 1b**). Because the final fragment to be ligated can encode one, two or three TAL effector repeats, arrays consisting of any desired number of TAL effector repeats can be assembled (**Supplementary Fig. 1**). DNA fragments encoding the final full-length TAL effector repeat array are released from the beads by restriction enzyme digestion (Online Methods, **Supplementary Methods** and **Fig. 1b**).

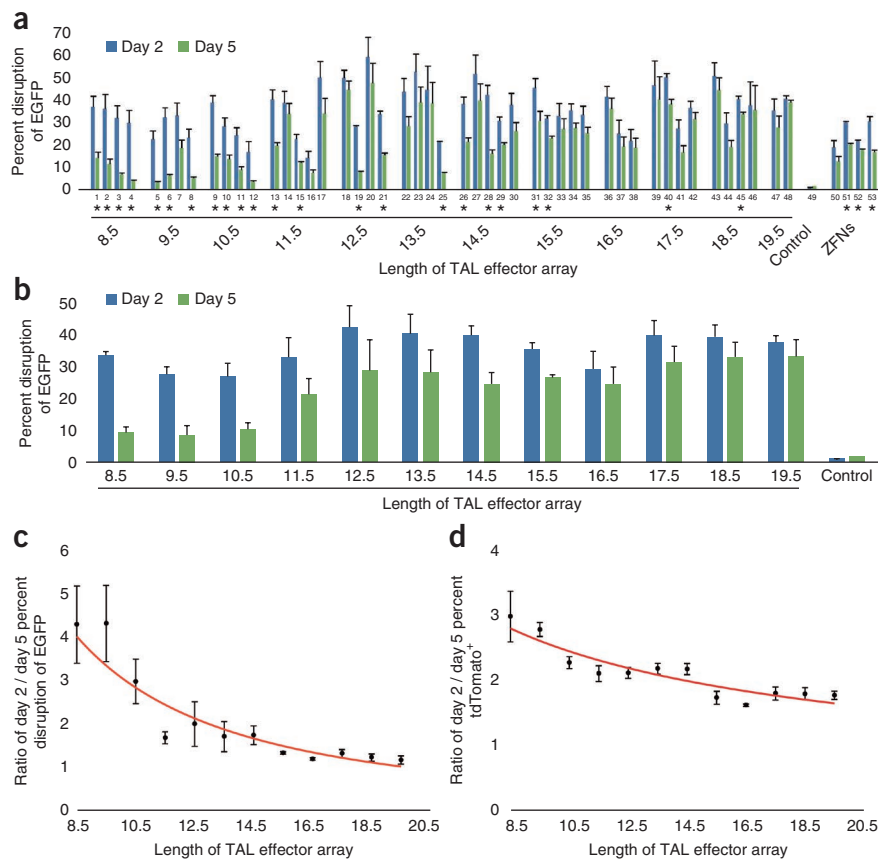
We optimized FLASH so that it can be efficiently practiced in 96-well format using a robotic liquid-handling workstation (Online Methods and **Supplementary Methods**). With automation, we can assemble DNA fragments encoding up to 96 different TAL effector repeat arrays in less than 1 d. We also adapted FLASH so that medium-throughput assembly can be performed manually in 1–2 d using multi-channel pipettes (data not shown). Fragments assembled using either approach can be cloned into expression vectors (e.g., for expression as a TALEN) to generate sequence-verified plasmids in less than 1 week (**Supplementary Methods**). Using automated FLASH, we can make sequence-verified TAL effector expression plasmids for <\$100 each including the cost of labor.

Large-scale testing of FLASH-assembled TALENs

To perform a large-scale test of the robustness of TALENs in human cells, we used FLASH to construct plasmids encoding 48 TALEN pairs targeted to different sites in the *EGFP* reporter gene. Monomers in



Figure 2 Gene-disruption activities of 48 FLASH-assembled TALENs targeted to *EGFP*. (a) Activities of 48 TALEN pairs and four ZFN pairs in the *EGFP* gene-disruption assay. Percentages of EGFP-disrupted cells as measured 2 and 5 d after transfection of U2OS cells bearing a chromosomally integrated *EGFP* reporter gene with nuclease-encoding plasmids are shown. Target sites for the 48 TALEN pairs are shown in **Supplementary Table 3**. The four ZFN pairs were made by OPEN and have been previously described²³. Mean percent disruption of EGFP and s.e.m. from three independent transfections are shown. All 48 TALEN pairs and all four ZFN pairs tested induced statistically significant ($P < 0.05$) increases in EGFP-disrupted cells relative to control transfected cells on both day 2 and day 5. *, TALEN and ZFN pairs for which statistically significant differences exist between values of EGFP-disrupted cells on day 2 and day 5 ($P < 0.05$). (b) Mean EGFP-disruption activities from a grouped by length of the TALENs. (c) Ratio of mean percent EGFP disruption values from day 2 to day 5. Ratios were calculated for groups of each length TALEN using the data from b. Values >1 indicate a decrease in the average percentage of EGFP-disrupted cells at day 5 relative to day 2. (d) Ratio of mean tdTomato-positive cells on day 2 and day 5 grouped by various lengths of TALENs. tdTomato-encoding control plasmids were transfected together with nuclease-encoding plasmids on day 0. Error bars, s.e.m.



between guideline violation and the level of TALEN activities on either day 2 or day 5 post-transfection (Fig. 3). We also did not find a significant correlation between the total number of guideline violations and the level of TALEN activity (Fig. 3). Thus, failure to meet four of the five previously described design guidelines when choosing target sequences does not adversely affect activities of TALENs made for those sites.

High-throughput alteration of endogenous human genes

We next tested the efficiency of TALENs for modifying endogenous genes in human cells by using FLASH to engineer TALEN pairs for targets in 96 different human genes: 78 genes implicated in human cancer²⁴ and 18 genes involved in epigenetic regulation (Supplementary Table 4). For each gene, we designed TALENs that cleave near the N terminus of the protein coding sequence, although in

a small number of cases the presence of repetitive sequences led us to target alternate sites in downstream exons or introns (Supplementary Table 4). Guided by our results with the *EGFP* TALENs and by spacer lengths defined in earlier studies⁶, we constructed TALENs composed of 14.5, 15.5 or 16.5 repeats designed to cleave sites with 16, 17, 18, 19 or 21 bp spacers.

We tested the activities of our 96 TALEN pairs at their intended endogenous gene targets using a modified T7 endonuclease I (T7EI) assay (Supplementary Fig. 2)^{15,25}. In this assay, 84 of the 96 TALEN pairs showed efficient NHEJ-mediated mutagenesis at their intended target sites, an overall success rate of ~88% (Table 1 and Supplementary Fig. 3). Efficiencies of TALEN-induced mutagenesis ranged from 2.5% to 55.8% with a mean of 22.2%. To provide molecular confirmation of the mutations identified by T7EI assay, we sequenced target loci for 11 different TALEN pairs that induced

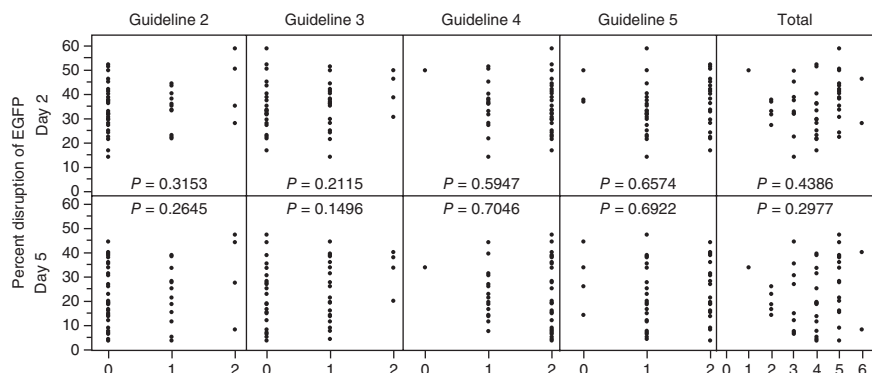


Figure 3 Computationally-derived design guidelines do not show statistically significant correlation with TALEN activity levels. Each graph plots the mean EGFP-disruption activity of each of our 48 *EGFP*-targeted TALEN pairs against the number of half-sites within each target site that fail to meet one of four previously described design guidelines (guidelines 2, 3, 4, or 5; see Supplementary Discussion¹⁷) or against the total number of guideline violations in each target site (total guideline violations). Plots are shown for EGFP-disruption activities measured on day 2 and day 5 post-transfection. Guideline violation data are from **Supplementary Table 3**. Correlation P -values are shown for each plot.



Table 1 TALEN-induced mutation frequencies for 96 endogenous human genes involved in cancer and epigenetic regulation

Gene	Mean indel mutation frequency (%) ± s.e.m.	Gene	Mean indel mutation frequency (%) ± s.e.m.
<i>ABL1</i>	22.5 ± 7.1	<i>HOXC13</i>	10.5 ± 0.3
<i>AKT2</i>	14.1 ± 7.3	<i>HOXD11</i>	None
<i>ALK</i>	12.7 ± 2.9	<i>HOXD13</i>	None
<i>APC</i>	48.8 ± 9.8	<i>JAK2</i>	44.9 ± 16.9
<i>ATM</i>	35.5 ± 15.6	<i>KIT</i>	None
<i>AXIN2</i>	2.5 ± 0.6	<i>KRAS</i>	9.4 ± 0.9
<i>BAX</i>	14.7 ± 11.6	<i>MAP2K4</i>	11.9 ± 7.1
<i>BCL6</i>	14.9 ± 5.9	<i>MDM2</i>	33.0 ± 20.2
<i>BMPR1A</i>	50.4 ± 16.4	<i>MET</i>	40.4 ± 10.7
<i>BRCA1</i>	44.5 ± 15.5	<i>MLH1</i>	44.9 ± 6.3
<i>BRCA2</i>	41.6 ± 10.5	<i>MSH2</i>	27.5 ± 10.4
<i>CBX3</i>	35.2 ± 22.6	<i>MUTYH</i>	24.9 ± 8.4
<i>CBX8</i>	13.5 ± 3.4	<i>MYCL1</i>	17.3 ± 0.6
<i>CCND1</i>	40.5 ± 2.2	<i>MYC</i>	13.4 ± 4.0
<i>CDC73</i>	36.3 ± 7.7	<i>MYCN</i>	16.3 ± 11.6
<i>CDH1</i>	None	<i>NBN</i>	46.3 ± 15.5
<i>CDK4</i>	21.5 ± 17.4	<i>NCOR1</i>	29.6 ± 13.1
<i>CHD4</i>	9.6 ± 0.1	<i>NCOR2</i>	3.3 ± 0.6
<i>CHD7</i>	11.4 ± 2.7	<i>NTRK1</i>	None
<i>CTNNB1</i>	26.0 ± 8.1	<i>PDGFRA</i>	16.0 ± 4.3
<i>CYLD</i>	24.7 ± 2.3	<i>PDGFRB</i>	16.0 ± 3.2
<i>DDB2</i>	15.8 ± 7.2	<i>PHF8</i>	22.2 ± 6.1
<i>ERCC2</i>	55.8 ± 12.7	<i>PMS2</i>	26.9 ± 9.5
<i>ERCC5</i>	None	<i>PTCH1</i>	27.5 ± 15.9
<i>EWSR1</i>	14.3 ± 8.2	<i>PTEN</i>	31.5 ± 11.7
<i>EXT1</i>	9.5 ± 3.0	<i>RARA</i>	13.4 ± 6.1
<i>EXT2</i>	4.0 ± 1.2	<i>RBBP5</i>	15.7 ± 9.5
<i>EZH2</i>	41.3 ± 2.6	<i>RECQL4</i>	22.1 ± 16.2
<i>FANCA</i>	9.7 ± 5.0	<i>REST</i>	None
<i>FANCC</i>	23.7 ± 17.8	<i>RET</i>	5.4 ± 1.8
<i>FANCE</i>	None	<i>RNF2</i>	None
<i>FANCF</i>	46.0 ± 7.7	<i>RUNX1</i>	25.1 ± 6.9
<i>FANCG</i>	26.9 ± 16.2	<i>SDHB</i>	36.4 ± 19.2
<i>FES</i>	12.6 ± 10.6	<i>SDHC</i>	13.7 ± 3.4
<i>FGFR1</i>	17.4 ± 6.2	<i>SDHD</i>	42.0 ± 7.8
<i>FH</i>	20.9 ± 11.8	<i>SETDB1</i>	33.5 ± 6.1
<i>FLCN</i>	11.1 ± 4.4	<i>SIRT6</i>	43.3 ± 3.1
<i>FLT3</i>	None	<i>SMAD2</i>	3.9 ± 1.6
<i>FLT4</i>	9.9 ± 5.0	<i>SS18</i>	31.4 ± 7.9
<i>FOXO1</i>	8.5 ± 1.1	<i>SUZ12</i>	13.1 ± 0.4
<i>FOXO3</i>	7.3 ± 2.3	<i>TFE3</i>	17.3 ± 2.4
<i>GLI1</i>	21.5 ± 12.4	<i>TGFBR2</i>	None
<i>HDAC1</i>	10.8 ± 3.0	<i>TLX3</i>	None
<i>HDAC2</i>	4.2 ± 0.9	<i>TP53</i>	19.9 ± 3.6
<i>HDAC6</i>	21.4 ± 2.1	<i>TSC2</i>	30.7 ± 22.7
<i>HMGA2</i>	3.0 ± 1.5	<i>VHL</i>	19.4 ± 1.1
<i>HOXA13</i>	7.6 ± 3.1	<i>XPA</i>	12.9 ± 2.2
<i>HOXA9</i>	6.4 ± 2.7	<i>XPC</i>	31.4 ± 4.2

varying efficiencies of mutagenesis (Fig. 4). As expected, this experiment revealed characteristic insertion or deletion mutations (indels) at the target gene sites with frequencies similar to those observed by T7EI assay (compare Fig. 4 and Table 1).

DISCUSSION

The FLASH system described here will enable any interested researcher, core facility or institution to produce TAL effector repeat arrays in an inexpensive and high-throughput manner. With FLASH, DNA fragments are assembled on solid-phase magnetic beads rather than in solution, thereby enabling serial enzymatic reactions to be performed without the need for column-based wash or purification steps. FLASH also avoids the need for gel isolation or analysis of intermediate constructs, both of which are labor-intensive and difficult to automate. When done with a liquid-handling robotic instrument, FLASH enables assembly of DNA fragments encoding 96 arrays in less than a day.

Our large-scale testing of 144 FLASH-assembled TALEN pairs provides the most comprehensive test of TALEN technology performed to date. One hundred percent of the 48 EGFP-targeted TALEN pairs and ~88% of the 96 endogenous gene-targeted TALEN pairs we produced cleave their targets in human cells with mutation efficiencies similar to those induced by ZFNs engineered by the selection-based OPEN method. The nucleotide composition of the 96 EGFP TALEN half-sites and the 168 endogenous gene TALEN half-sites for which we successfully made TALENs is quite diverse, reflecting DNA sequences composed of variable numbers and percentages of G, A, T and C bases (Supplementary Figs. 4 and 5). We do not know the precise reason(s) why 12 TALEN pairs targeted to endogenous human genes failed to show activity in our T7EI assay. Possible explanations include inhibitory effects of chromatin structure, DNA modification, or inefficient expression and/or folding of particular TALENs. Nonetheless, the high success rate we observed suggests that prescreening TALENs using other surrogate assays (e.g., yeast-based reporter assays^{12,17}) may be unnecessary and that TALENs might be tested directly at the endogenous gene target in the cell type or organism of interest.

We note that all of the TALENs we assembled using FLASH were made using a particular framework of TAL effector repeats and N- and C-terminal sequences, as described⁶. This framework has been used to construct nucleases that function efficiently in nematodes¹⁸, zebrafish²¹, rats²⁰, and human somatic⁶ and pluripotent stem cells¹⁹, and we therefore predict that FLASH-assembled TALENs will also likely show high activities and high success rates in other cell types and organisms. A related question to test in future experiments will be whether TALENs constructed using other architectures described in the literature^{8–11,15,16,17,22} will be as robust. Our results demonstrate that the targeting range of TALENs is substantially greater than that suggested in a study that described five design guidelines for choosing potential cleavage sites¹⁷ (Supplementary Discussion). These guidelines limit the targeting range of TALENs to approximately one site in every 35 bp of DNA sequence¹⁷ and have been implemented in web-based TALENT software¹⁷ (<https://boglab.plp.iastate.edu/>). We were able to successfully make active TALENs for 131 full target sites that do not meet one or more of four of these design guidelines. Furthermore, we did not find any statistically significant correlation between failure to meet these four guidelines and the activity levels of our 48 EGFP-targeted TALENs. The discrepancy between these computationally derived guidelines and our experimental results may be because the rules were derived from sites bound by monomeric TAL effectors, whereas TALENs function as dimers.

By systematically making and testing TALENs for target sites of various lengths, we uncovered an inverse correlation between the number of TAL effector repeats in a TALEN and the degree of associated cytotoxicity. In our EGFP-reporter experiments, shorter TALENs were as active as longer ones but were generally more cytotoxic, presumably because of their greater potential for binding to off-target sites. Our findings suggest that cytotoxicity might be minimized by constructing longer TALENs (e.g., with 14.5–19.5 TAL effector repeats), a hypothesis that can be tested in future experiments. Even with this restriction and other limitations on the length of the spacer sequence, we estimate that on average more than three TALEN pairs can be targeted per base pair of random DNA sequence (Supplementary Discussion).

Production-scale use of FLASH should enable the construction of thousands of TALEN pairs per year. We have already made >600 TALEN-encoding plasmids using FLASH (Supplementary Tables 3 and 4 and data not shown). In production mode, it should be straightforward for two scientists to construct a set of 96 TAL effector repeat arrays at least three times per week, enabling the generation of >7,200





Figure 4 DNA sequences and frequencies of FLASH TALEN-induced mutations at endogenous human genes. For each endogenous gene target, the wild-type (WT) sequence is shown at the top with the TALEN target half-sites highlighted in yellow and the translation start codon of the gene (ATG) underlined and highlighted in bold type. Deletions are indicated by red dashes against a gray background and insertions by lowercase letters against a light blue background. The sizes of the insertions (+) or deletions (Δ) are indicated to the right of each mutated site. The number of times that each mutant was isolated is shown in parentheses. Mutation frequencies are calculated as the number of mutants identified divided by the total number of sequences analyzed. Note that for several of the genes, we also identified larger deletions that extend beyond the sequences of the TALEN target sites. For each gene, sequencing was done with the same genomic DNA used to carry out the corresponding T7E1 assays shown in **Supplementary Figure 3**.

TALEN pairs per year. Our cost for making a pair of sequence-verified TALEN plasmids using FLASH (including labor) is <\$200. This low per-unit cost will be particularly important for large-scale gene editing projects or for academic core facilities interested in making large numbers of nucleases. Because we have not yet fully optimized the FLASH method, the cost of producing a pair of TALENs could likely be further reduced.

We have also adapted the FLASH method so that it can be performed in medium throughput. In this modified protocol, the overall approach remains unchanged but manipulations are carried out manually using a multi-channel pipette rather than with a liquid-handling robot. We have successfully used this protocol to assemble dozens of TAL effector repeat arrays in 1–2 d (data not shown). This alternative smaller-scale protocol provides access to FLASH for laboratories that do not have automated liquid-handling equipment.

An important issue for future investigation is the extent of undesired off-target alterations introduced by TALENs. Off-target sites for one TALEN pair in the human genome have already been identified using a computational approach¹⁹. Application of improved methods for

identification of nuclease off-target cleavage events^{26,27} to TALENs may reveal additional off-target sites. Whole exome or genome sequencing, as recently done with human induced pluripotent stem cells modified by ZFNs²⁸ and with yeast modified by TALENs¹⁶, might also be informative. All TALENs we constructed by FLASH harbor the wild-type FokI domain and therefore may form unwanted homodimers capable of inducing off-target mutations. As previously demonstrated by others, using obligate heterodimeric FokI domains may reduce formation of undesirable homodimers^{27,29–31}. Until off-target sites can be comprehensively identified, users of TALENs will need to account for these undesired potential effects, as is done for ZFNs.

In summary, FLASH should enable any researcher to rapidly, efficiently and precisely alter virtually any gene or DNA sequence of interest without the need for specialized protein engineering expertise or for extensive screening to identify active nucleases. The capability of FLASH to produce TALENs in high or medium throughput will broaden the scope of gene modification experiments that can be performed by individual laboratories and core facilities, enabling pathway- or genome-wide projects. In this study we have modified more endogenous genes



than any other individual report using ZFNs, meganucleases or TALENs. Although we have focused here on TALENs, FLASH should also inspire innovative applications involving the fusion of engineered TAL effector repeat arrays to other functional domains to create targetable chimeric proteins. All reagents needed to practice FLASH and all TALEN expression plasmids we assembled will be available to academic researchers by request (<http://www.talengeneering.org/>).

METHODS

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

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

ACKNOWLEDGMENTS

We thank T. Cathomen for providing polyclonal U2OS-EGFP reporter cells and a T7EI protocol, Y. Fu for deriving the clonal U2OS-EGFP cell line, D. Dobbs for support and encouragement, and M. Maeder, J. Angstman and C. Ramirez for helpful comments. This work was supported by a US National Institutes of Health (NIH) Director's Pioneer Award DP1 OD006862 (J.K.J.), NIH P50HG005550 (J.K.J.), the Jim and Ann Orr MGH Research Scholar Award (J.K.J.), NIH T32 CA009216 (J.D.S.) and National Science Foundation DBI-0923827 (D.R.).

AUTHOR CONTRIBUTIONS

J.D.S. and J.K.J. conceived of the FLASH method. D.R., S.Q.T., C.K., J.A.F. and J.D.S. performed experiments. D.R., S.Q.T., C.K., J.A.F., J.D.S. and J.K.J. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at <http://www.nature.com/naturebiotechnology/>.

Published online at <http://www.nature.com/naturebiotechnology/>.

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

1. Bogdanove, A.J. & Voytas, D.F. TAL effectors: customizable proteins for DNA targeting. *Science* **333**, 1843–1846 (2011).
2. Bogdanove, A.J., Schornack, S. & Lahaye, T. TAL effectors: finding plant genes for disease and defense. *Curr. Opin. Plant Biol.* **13**, 394–401 (2010).
3. Scholze, H. & Boch, J. TAL effectors are remote controls for gene activation. *Curr. Opin. Microbiol.* 47–53 (2011).
4. Boch, J. *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509–1512 (2009).
5. Moscou, M.J. & Bogdanove, A.J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501 (2009).
6. Miller, J.C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29**, 143–148 (2011).
7. Morbitzer, R., Romer, P., Boch, J. & Lahaye, T. Regulation of selected genome loci using *de novo*-engineered transcription activator-like effector (TALE)-type transcription factors. *Proc. Natl. Acad. Sci. USA* **107**, 21617–21622 (2010).
8. Morbitzer, R., Elsaesser, J., Hausner, J. & Lahaye, T. Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res.* **39**, 5790–5799 (2011).
9. Zhang, F. *et al.* Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat. Biotechnol.* **29**, 149–153 (2011).
10. Geissler, R. *et al.* Transcriptional activators of human genes with programmable DNA-specificity. *PLoS ONE* **6**, e19509 (2011).
11. Weber, E., Gruetzner, R., Werner, S., Engler, C. & Marillonnet, S. Assembly of designer TAL effectors by Golden Gate cloning. *PLoS ONE* **6**, e19722 (2011).
12. Christian, M. *et al.* Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 757–761 (2010).
13. Li, T. *et al.* TAL nucleases (TALENs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.* **39**, 359–372 (2011).
14. Mahfouz, M.M. *et al.* *De novo*-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc. Natl. Acad. Sci. USA* **108**, 2623–2628 (2011).
15. Mussolino, C. *et al.* A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res.* 9283–9293 (2011).
16. Li, T. *et al.* Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res.* **39**, 6315–6325 (2011).
17. Cermak, T. *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* **39**, e82 (2011).
18. Wood, A.J. *et al.* Targeted genome editing across species using ZFNs and TALENs. *Science* **333**, 307 (2011).
19. Hockemeyer, D. *et al.* Genetic engineering of human pluripotent cells using TALE nucleases. *Nat. Biotechnol.* **29**, 731–734 (2011).
20. Tesson, L. *et al.* Knockout rats generated by embryo microinjection of TALENs. *Nat. Biotechnol.* **29**, 695–696 (2011).
21. Sander, J.D. *et al.* Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat. Biotechnol.* **29**, 697–698 (2011).
22. Huang, P. *et al.* Heritable gene targeting in zebrafish using customized TALENs. *Nat. Biotechnol.* **29**, 699–700 (2011).
23. 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).
24. Vogelstein, B. & Kinzler, K.W. Cancer genes and the pathways they control. *Nat. Med.* **10**, 789–799 (2004).
25. Kim, H.J., Lee, H.J., Kim, H., Cho, S.W. & Kim, J.S. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res.* **19**, 1279–1288 (2009).
26. 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**, 765–770 (2011).
27. Gabriel, R. *et al.* An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat. Biotechnol.* **29**, 816–823 (2011).
28. Yusa, K. *et al.* Targeted gene correction of α_1 -antitrypsin deficiency in induced pluripotent stem cells. *Nature* **478**, 391–394 (2011).
29. Miller, J.C. *et al.* An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.* **25**, 778–785 (2007).
30. Doyon, Y. *et al.* Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat. Methods* **8**, 74–79 (2011).
31. Szczepek, M. *et al.* Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat. Biotechnol.* **25**, 786–793 (2007).



ONLINE METHODS

Construction of a plasmid archive encoding pre-assembled TAL effector repeats. We sought to construct TAL effector repeat arrays using the same architecture, as previously described⁶, in which four distinct TAL effector repeat backbones that differ slightly in their amino acid and DNA sequences occur in a repeated pattern. We designated the first, N-terminal TAL effector repeat in an array as the α unit. This is followed by β , γ and δ units and then an ϵ unit that is essentially identical to the α unit except for the different positioning of a Type IIS restriction site on the 5' end (required to enable creation of a unique overhang on the α unit needed for cloning). The ϵ unit is then followed again by repeats of β , γ , δ and ϵ units. Due to constraints related to creation of a 3' end required for cloning, slightly modified DNA sequences were required for TAL effector repeat arrays that end with a C-terminal γ or ϵ unit. We designated these variant units as γ^* and ϵ^* .

For each type of TAL effector repeat unit (that is, α , β , γ , δ , ϵ , γ^* and ϵ^*), we commercially synthesized (Genscript) a series of four plasmids, each harboring one of the four repeat variable di-residues (RVDs) that specifies one of the four DNA bases (NI = A; HD = C; NN = G; NG = T). Full DNA sequences of these plasmids are provided in **Supplementary Table 1** and **Supplementary Fig. 6**. For all 28 of these plasmids, the sequence encoding the TAL effector repeat domain is flanked on the 5' end by unique XbaI and BbsI restriction sites and on the 3' end by unique BsaI and BamHI restriction sites. Additionally, the overhangs generated by digestion of any plasmids encoding units designed to be adjacent to one another (e.g., β and γ , or δ and ϵ) with BsaI and BbsI are complementary. Using these 28 different plasmids and serial ligation through the BsaI and BbsI restriction sites, as previously described²¹, we assembled an archive of all possible combinations of $\beta\gamma\delta\epsilon$, $\beta\gamma\delta$, $\beta\gamma$, $\beta\gamma^*$ and $\delta\epsilon^*$ repeats. In total, this archive consisted of 368 different plasmids encoding 256 $\beta\gamma\delta\epsilon$ combinations, 64 $\beta\gamma\delta$ combinations, 16 $\beta\gamma$ combinations, 16 $\beta\gamma^*$ combinations and 16 $\delta\epsilon^*$ combinations (**Supplementary Table 2**). These 368 plasmids plus eight of the original 28 plasmids encoding single TAL effector repeats (four α and four β plasmids) are required to practice FLASH. With this archive of 376 plasmids listed in **Supplementary Table 2**, FLASH can be used to construct TAL effector repeat arrays of any desired length and composition (**Supplementary Fig. 1**).

Preparation of TAL effector repeat-encoding DNA fragments for FLASH assembly. We provide here an overview of the FLASH assembly method and of the process for subcloning of FLASH-assembled DNA fragments into TALEN expression vectors but more detailed step-by-step protocols can be found in the **Supplementary Methods**. To prepare DNA fragments encoding α units for use in FLASH assembly, we performed 20 rounds of PCR with each α unit plasmid as a template using primers oJS2581 (5'-Biotin-TCTAGAGAAGACAAGAACCCTGACC-3') and oJS2582 (5'-GGATCCGGTCTCTTAAGGCCGTGG-3'). The resulting PCR products are biotinylated on the 5' end. Each α PCR product was then digested with 40 units of BsaI-HF restriction enzyme to generate 4-bp overhangs, purified using the QIAquick PCR purification kit (Qiagen) according to manufacturer's instructions except that the final product was eluted in 50 μ l of 0.1 \times EB.

To prepare DNA fragments encoding β , $\beta\gamma\delta\epsilon$, $\beta\gamma\delta$, $\beta\gamma$, $\beta\gamma^*$ and $\delta\epsilon^*$ repeats, we digested 10 μ g of each of these plasmids with 50 units of BbsI restriction enzyme in NEBuffer 2 for 2 h at 37 °C followed by serial restriction digests performed in NEBuffer 4 at 37 °C using 100 units each of XbaI, BamHI-HF and Sall-HF enzymes that were added at 5-min intervals. The latter set of restriction digestions are designed to cleave the plasmid backbone to ensure that this larger DNA fragment does not interfere with subsequent ligations done during the FLASH assembly process. These restriction digest reactions were then purified using the QIAquick PCR purification kit according to manufacturer's instructions except that the final product was eluted in 180 μ l of 0.1 \times EB.

Automated FLASH assembly. All steps of FLASH assembly were performed using a Sciclone G3 liquid-handling workstation (Caliper) in 96-well plates and using a SPRiplate 96-ring magnet (Beckman Coulter Genomics) and a DynaMag-96 Side magnet (Life Technologies). In the first step of FLASH, a biotinylated α unit fragment is ligated to the first $\beta\gamma\delta\epsilon$ fragment and then the resulting $\alpha\beta\gamma\delta\epsilon$ fragments are bound to Dynabeads MyOne C1 streptavidin-coated magnetic beads (Life Technologies) in 2 \times B&W Buffer. Beads are then

drawn to the side of the well by placing the plate on the magnet and then washed with 100 μ l B&W buffer with 0.005% Tween 20 (Sigma) and again with 100 μ l 0.1 mg/ml bovine serum albumin (BSA) (New England Biolabs). Additional $\beta\gamma\delta\epsilon$ fragments are ligated by removing the plate from the magnet, resuspending the beads in solution in each well, digesting the bead-bound fragment with BsaI-HF restriction enzyme, placing the plate on the magnet, washing with 100 μ l B&W/Tween20 followed by 100 μ l of 0.1 mg/ml BSA, and then ligating the next fragment. This process is repeated multiple times with additional $\beta\gamma\delta\epsilon$ units to extend the bead-bound fragment. The last fragment to be ligated is always a β , $\beta\gamma^*$, $\beta\gamma\delta$ or $\delta\epsilon^*$ unit to enable cloning of the full-length fragment into expression vectors (note that fragments that end with a $\delta\epsilon^*$ unit are always preceded by ligation of a $\beta\gamma$ unit).

The final full-length bead-bound fragment is digested with 40 units of BsaI-HF restriction enzyme followed by 25 units of BbsI restriction enzyme (New England Biolabs). Digestion with BbsI releases the fragment from the beads and generates a unique 5' overhang for cloning of the fragment. Digestion with BsaI-HF results in creation of a unique 3' overhang for cloning.

Subcloning of TAL effector repeat array-encoding DNA fragments into TALEN expression vectors. We subcloned DNA fragments encoding our FLASH assembled TAL effector repeat arrays into one of four TALEN expression vectors. Each of these vectors includes a CMV promoter, a translational start codon optimized for mammalian cell expression, a triple FLAG epitope tag, a nuclear localization signal, amino acids 153 to 288 from the TAL effector 13 protein (as numbered in ref. 6), two unique and closely positioned Type IIS BsmBI restriction sites, a 0.5 TAL effector repeat domain encoding one of four possible RVDs (NI, HD, NN, or NG for recognition of an A, C, G, or T nucleotide, respectively), amino acids 715 to 777 from the TAL effector 13 protein and the wild-type FokI cleavage domain. All DNA fragments assembled by FLASH possess overhangs that enable directional cloning into any of the four TALEN expression vectors that has been digested with BsmBI. All four of the TALEN expression vectors (each possessing a different 0.5 TAL effector repeat) are already available from Addgene and full sequences of these plasmids are freely available on a web page dedicated to these constructs: <http://www.addgene.org/talengineering/expressionvectors/>.

To prepare a TALEN expression vector for subcloning, we digested 5 μ g of plasmid DNA with 50 units of BsmBI restriction enzyme (New England Biolabs) in NEBuffer 3 for 8 h at 55 °C. Digested DNA was purified using 90 μ l of Ampure XP beads (Agencourt) according to manufacturer's instructions and diluted to a final concentration of 5 ng/ μ l in 1 mM TrisHCl. FLASH-assembled TAL effector repeat arrays were ligated into TALEN expression vectors using 400 U of T4 DNA Ligase (New England Biolabs). Ligation products were transformed into chemically competent XL-1 Blue cells. Typically, six colonies were picked for each ligation and plasmid DNA isolated by an alkaline lysis miniprep procedure. Simultaneously, the same colonies were screened by PCR using primers oSQT34 (5'-GACGGTGGC TGTC AAATACCAAGATATG-3') and oSQT35 (5'-TCTCCTCCAGTTCA CTTTTGACTAGTTGGG-3'). PCR products were analyzed on a QIAxcel capillary electrophoresis system (Qiagen). Miniprep DNA from clones that contained correctly sized PCR products were sent for DNA sequence confirmation with primers oSQT1 (5'-AGTAACAGCGGTAGAGGCAG-3'), oSQT3 (5'-ATTGGGCTACGATGGACTCC-3') and oJS2980 (5'-TTAATTC AATA TATTCATGAGGCAC-3'); oSQT1 anneals at the 5' end of the TAL effector repeat array coding sequence and enables sequencing of the N-terminal half of the assembled array, oSQT3 anneals at the 3' end of the TAL effector repeat array coding sequence and enables sequencing of the C-terminal half of the assembled array, and oJS2980 primes within the coding sequence of the FokI domain (downstream of oSQT3) and enables sequencing and verification of the C-terminal 0.5 TAL effector repeat domain. An example of DNA sequence for a FLASH-assembled TALEN is shown in **Supplementary Figure 7**.

We typically screened six colonies for each assembly as described above, followed by six additional colonies if necessary. With this approach, we have routinely identified one or more sequence-verified clones for >90% of assembly reactions. These percentages are derived primarily from experiments designed to construct DNA fragments encoding 16.5 TAL effector repeats. We have observed slightly lower frequencies of success when screening for clones encoding more than 16.5 TAL effector repeats.



EGFP TALEN activity and toxicity assays. EGFP reporter assays were performed in a clonal U2OS human cell line bearing an integrated construct that constitutively expresses an EGFP-PEST fusion protein. This clonal line was derived from a polyclonal U2OS EGFP-PEST reporter line (a gift from Dr. Toni Cathomen, Hannover Medical School). Clonal U2OS EGFP-PEST cells were cultured in Advanced DMEM (Life Technologies) supplemented with 10% FBS, 2 mM GlutaMax (Life Technologies), penicillin/streptomycin and 400 µg/ml G418. Cells were transfected in triplicate with 500 ng of each TALEN plasmid DNA and 50 ng ptdTomato-N1 plasmid DNA using a Lonza 4D-Nucleofector System, Solution SE and program DN-100 according to manufacturer's instructions. 1 µg of ptdTomato-N1 plasmid alone was transfected in triplicate as a negative control. Cells were assayed for EGFP and tdTomato expression at 2 and 5 d post-transfection using a BD FACSAriaII flow cytometer.

A one-tailed independent two-sample *t*-test was used to test for increased percentages of EGFP disruption in cells transfected with nucleases (TALENs or ZFNs) compared with control transfected cells (Fig. 2a). A one-tailed independent two-sample *t*-test was used to test for differences in the percentage of EGFP-disrupted cells transfected with TALENs or ZFNs at day 2 and day 5 post-transfection (Fig. 2a). Analysis of variance (ANOVA) was performed on the ratio of day 2/day 5 EGFP-disrupted cells as a function of TAL effector array length (Fig. 2c). ANOVA was performed on the ratio of day 2/day 5 tdTomato+ cells as a function of TAL effector array length (Fig. 2d).

PCR amplification and sequence verification of endogenous human genes.

PCR reactions to amplify targeted loci were performed using the primers shown in Supplementary Table 5. For most loci, we were able to use standard PCR conditions with Phusion Hot Start II high-fidelity DNA polymerase (Thermo Fisher) performed according to manufacturer's instructions for 35 cycles (98 °C, 10 s denaturation; 68 °C, 15 s annealing; 72 °C, 30 s extension). For loci that did not amplify under standard conditions, we used one of the following modifications: (i) the addition of betaine to a final concentration of 1.8 M, (ii) touchdown PCR ([98 °C, 10 s; 72–62 °C, –1 °C/cycle, 15 s; 72 °C, 30 s]_{10 cycles}, [98 °C, 10 s; 62 °C, –1 °C/cycle, 15 s; 72 °C, 30 s]_{25 cycles}) with 1.8 M betaine and (iii) the addition of 3% or 5% DMSO and an annealing temperature of 65 °C. PCR products were analyzed for correct size on a QIAxcel capillary electrophoresis system. Correctly sized products were treated with

ExoSap-IT (Affymetrix) to remove unincorporated nucleotides or primers and sent for DNA sequencing to confirm the endogenous gene sequence.

T7 Endonuclease I assay for quantifying NHEJ-mediated mutation of endogenous human genes.

U2OS-EGFP cells were cultured and transfected in duplicate as described above. Genomic DNA was isolated from cells transfected with TALEN-encoding or control plasmids using a high-throughput magnetic-bead based purification system (Agencourt DNAdvance) according to the manufacturer's instructions. PCR to amplify endogenous loci was performed for 35 cycles as described above and fragments were purified with Ampure XP (Agencourt) according to manufacturer's instructions. 200 ng of purified PCR product was denatured and reannealed in NEBuffer 2 (New England Biolabs) using a thermocycler with the following protocol: 95 °C, 5 min; 95–85 °C at –2 °C/s; 85–25 °C at –0.1 °C/s; hold at 4 °C³². Hybridized PCR products were treated with 10 U of T7 Endonuclease I at 37 °C for 15 min in a reaction volume of 20 µl. Reactions were stopped by the addition of 2 µl 0.5 M EDTA, purified with Ampure XP, and quantified on a QIAxcel capillary electrophoresis system using method OM500. The sum of the area beneath TALEN-specific cleavage peaks (expressed as a percentage of the parent amplicon peak, denoted fraction cleaved) is used to estimate gene modification levels using the following equation as previously described³²:

$$\% \text{ gene modification} = 100 \times (1 - (1 - \text{fraction cleaved})^{1/2})$$

Sequence confirmation of endogenous gene mutations. 11 genes that showed evidence of mutations in the T7 Endonuclease I assay were chosen for independent confirmation via Sanger sequencing. PCR products corresponding to these 11 sites were cloned into PCR-4-Blunt-Topo using a Topo cloning kit (Life Technologies) and transformed into XL-1 Blue cells. Plasmid DNA was isolated for multiple colonies from each transformation and then sent for DNA sequencing using either the T7 (5'-TAATACGACTCACTATAG GG-3') and T3 (5'-ATTAACCCTCACTAAAGGGA-3') primers or the M13F (5'-GTAAAACGACGGCCAG-3') and T3 primers.

32. Guschin, D.Y. *et al.* A rapid and general assay for monitoring endogenous gene modification. *Methods Mol. Biol.* **649**, 247–256 (2010).