nature biotechnology

A TALE nuclease architecture for efficient genome editing

Jeffrey C Miller[1,2], Siyuan Tan[1,2], Guijuan Qiao[1], Kyle A Barlow[1], Jianbin Wang[1], Danny F Xia[1], Xiangdong Meng[1], David E Paschon[1], Elo Leung[1], Sarah J Hinkley[1], Gladys P Dulay[1], Kevin L Hua[1], Irina Ankoudinova[1], Gregory J Cost[1], Fyodor D Urnov[1], H Steve Zhang[1], Michael C Holmes[1], Lei Zhang[1], Philip D Gregory[1] & Edward J Rebar[1]

Nucleases that cleave unique genomic sequences in living cells can be used for targeted gene editing and mutagenesis. Here we develop a strategy for generating such reagents based on transcription activator–like effector (TALE) proteins from *Xanthomonas*. We identify TALE truncation variants that efficiently cleave DNA when linked to the catalytic domain of FokI and use these nucleases to generate discrete edits or small deletions within endogenous human *NTF3* and *CCR5* genes at efficiencies of up to 25%. We further show that designed TALEs can regulate endogenous mammalian genes. These studies demonstrate the effective application of designed TALE transcription factors and nucleases for the targeted regulation and modification of endogenous genes.

Methods for cleaving a chromosome at a specific site enable precise engineering of the genomes of higher eukaryotes. By provoking repair of the targeted locus, such methods can increase rates of gene disruption, gene editing or gene addition to levels that enable ready isolation of cells or organisms bearing a desired genetic change[1–6]. Moreover, cleavage-based genome engineering methods are portable to diverse cell types and species as the requisite repair pathways are highly conserved[7]. To date, such methods have been demonstrated in at least nine species that previously lacked effective genome engineering strategies[1,8–10]. These methods have also been used to engineer various mammalian cell types—including human stem cells[11–15]—with retention of full potency and normal growth characteristics. The therapeutic implications of such capabilities have been noted[11,16], and at least three clinical trials are under way that use human cells precisely modified by designed nucleases[1].

A key requirement of such methods is the ability to generate highly specific nucleases for unique genome targets. To this end, we[17] and others[18–21] have generated nucleases for genome engineering by linking the cleavage domain of the FokI restriction enzyme to a designed zinc finger protein (ZFP). The targeting of the resulting zinc finger nuclease (ZFN), which cleaves as a dimer, is mediated by its linked ZFP domain[22]. By varying the number and types of fingers in each ZFP-FokI chimera, a cleavage-competent dimer may be targeted to a user-chosen site of sufficient length (18–36 bp) for unique occurrence in a complex genome. Owing to their combination of specificity, target length and DNA-binding versatility, ZFPs have provided the DNA targeting domain of choice for almost all nucleases used for genome engineering[1–5].

Recently, two groups described an approach for engineering DNA-binding specificities based upon TALEs from *Xanthomonas* plant pathogens[23,24]. TALEs are transcriptional activators that specifically bind and regulate plant genes during pathogenesis[25–29]. Within the TALE structure, a central repeat domain mediates DNA recognition (**Fig. 1a**), with each repeat unit of 33–35 amino acids specifying one target base. The base preference of each unit is determined by two critical, adjacent amino acids referred to as the "repeat variable di-residue" (RVD)[24] (**Fig. 1a**, boxed letters). The relationship between the preferred binding site of a TALE and its successive RVDs appears to constitute a simple code, with each repeat independently specifying its targeted base[23,24] (**Fig. 1b**). TALEs generated with new repeat combinations have been shown to recognize target sequences predicted by this code[23].

These results have spurred interest in the potential use of TALE-nuclease chimeras (TALENs)[30,31] as site-specific endonucleases for selective genome cleavage. Here we report the development of TALENs capable of mediating efficient endogenous gene modification. First, we demonstrate TALE activity in a mammalian cell environment through targeted regulation of episomal reporters and an endogenous gene. Next, we constrain the minimal TALE region required for high-affinity DNA binding and identify truncation variants that enable efficient cleavage *in vitro* when linked to the catalytic domain of FokI. Finally, we use the resultant TALEN architecture to efficiently modify the endogenous human genes *NTF3* and *CCR5*. We thus show that designed TALE proteins can be used as reagents for genome engineering and gene regulation in higher eukaryotes.

RESULTS

Retargeting a natural TALE to an endogenous mammalian sequence

A critical first step for these studies was to choose a TALE that would serve as our initial design framework. We sought to identify a canonical, natural TALE that both exhibited a high degree of specificity and showed evidence of target sequence binding in mammalian cells. To accomplish this, we PCR-amplified and sequenced the coding regions for several TALEs from *Xanthomonas axonopodis* pathovar *citri* genomic DNA (**Supplementary Fig. 1**). By design, the recovered clones lacked codons for their first 152 residues, as this region specifies transport into plant cells and is dispensable for other functions[32]. The proteins were

[1]Sangamo BioSciences, Inc., Richmond, California, USA. [2]These authors contributed equally to this work. Correspondence should be addressed to E.J.R. (erebar@sangamo.com).

Received 15 November; accepted 14 December; published online 22 December 2011; doi:10.1038/nbt.1755

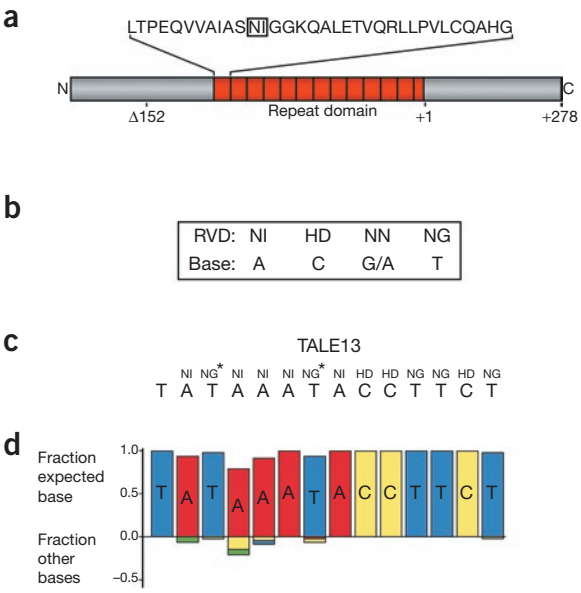

Figure 1 Structure and DNA-binding specificity of TALE proteins. (**a**) Sketch of a TALE from *Xanthomonas*. Red rectangles indicate the central array of tandem repeats that mediate DNA recognition. A typical repeat sequence is provided above, with a box highlighting the RVD (positions 12 and 13) that determines base preference. Gray regions indicate flanking protein segments, which often contain 288 and 278 residues (left and right segments, respectively). Δ152 indicates a truncation point that disrupts TALE transport into plant cells but preserves other functions[32], and which was used as the N terminus for all constructs in these studies. N and C denote N and C termini. (**b**) Base sequence preferences of four common RVDs[23,24], which have been used in recent studies to make TALEs with new specificities[23,30]. (**c**) RVDs (top row of letters) and predicted target bases (second row of letters) for the natural protein TALE13. RVDs are listed in repeat order (1 through 13), whereas the predicted target site is provided with the 5′ on the left. * denotes repeats that contain 33 amino acids, instead of the more typical 34. For exact repeat sequences see **Supplementary Figure 1**. (**d**) Graphical depiction of a SELEX-derived base frequency matrix for a fragment of TALE13 containing the repeat region. (See Online Methods for more details.) At each matrix position, the frequency of the predicted target base is shown above the *x* axis, whereas the remaining base frequencies are shown below the *x* axis. Note that the N-terminal flanking segment specifies the 5′ thymine base[23,24].

characterized by a SELEX assay, and one—TALE13—was identified as being highly selective for its target (**Fig. 1c,d**). Moreover, when linked to a VP16 activation domain, this protein drove a 70-fold induction of reporter gene expression in HEK293 cells (**Supplementary Fig. 2**; compare "+278" versus mock). Based on these results, TALE13 was chosen as the initial design framework for these studies.

We next sought to demonstrate that a designed TALE could regulate an endogenous mammalian gene. Our target for these studies was the human *NTF3* gene, which encodes a secreted nerve growth factor that has therapeutic potential for neurodegenerative diseases[33]. Natural TALEs often contain 18 repeats (**Supplementary Fig. 3**) and activate gene expression by binding to proximal promoter

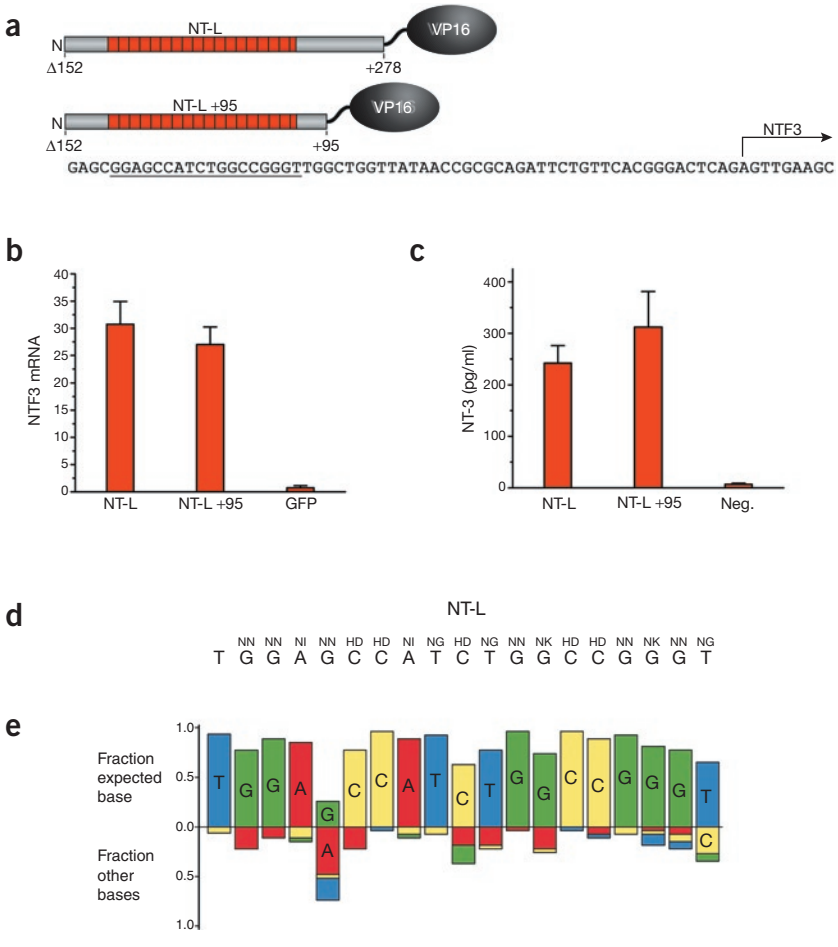

Figure 2 Activation of the endogenous human *NTF3* gene by engineered TALE transcription factors. (**a**) Sketch of the proteins used in this study and their target in the *NTF3* promoter. The designed TALE NT-L or a truncated variant that retained 95 residues of the C-terminal flanking region (NT-L+95) were linked to the VP16 activation domain and expressed in HEK293 cells. (For construct and protein sequences see **Supplementary Methods**.) The sequence at bottom shows the promoter-proximal region of human *NTF3*. Underlined bases indicate the target site for the NT-L TALE repeat domain. The hooked arrow shows the start site of native *NTF3* transcription. N denotes the N terminus of each protein. (**b**) Relative *NTF3* mRNA levels in HEK293 cells expressing either NT-L or NT-L+95 fused to the VP16 activation domain. GFP indicates cells transfected with a control plasmid that expresses enhanced GFP. Measurements were performed in quadruplicate and error bars indicate s.d. (**c**) Levels of NT-3 protein (encoded by *NTF3*) secreted from HEK293 cells expressing either NT-L or NT-L+95 fused to the VP16 activation domain. Measurements were performed in duplicate and error bars indicate s.d. Neg. indicates cells transfected with an empty vector control. (**d**) RVDs (top row of letters) and expected binding site (second row of letters) for NT-L and NT-L+95. For the design of NT-L and all other TALEs in this study, we used the four common and well-characterized RVDs[23,24] listed in **Figure 1b**, plus one additional design—NK—as an alternative for recognition of guanine[24]. (**e**) Graphical depiction of a SELEX-derived base frequency matrix for NT-L+95. For additional detail see legend to **Figure 1d**.

elements[34]. Therefore, to perform these studies we designed a protein in which the repeat domain of TALE13 was replaced with a new repeat array that specified an 18-bp site within the proximal promoter of *NTF3*. This yielded the protein "NT-L" (for *NTF3* Left) (**Fig. 2a**). Next, NT-L or a truncation variant (NT-L+95) was fused to the VP16 activation domain, expressed in human HEK293 cells, and tested for activation of the endogenous *NTF3* locus. We observed strong (>20-fold) induction of both *NTF3* transcript (**Fig. 2b**) and protein product (**Fig. 2c**). Consistent with these results, SELEX analysis revealed a high degree of specificity for the targeted locus (**Fig. 2d,e**). These results indicate that a designed TALE transcription factor can regulate an endogenous mammalian gene.

Development of a TALE-nuclease architecture

Having shown that NT-L can activate endogenous *NTF3* when linked to VP16, we next sought to convert this protein into a nuclease that could cleave its native chromosomal target. These activities proceeded in three stages. First, we used gene regulation studies to constrain the minimal region required for high-affinity binding of TALE proteins to DNA. We reasoned that trimming extraneous peptide would enable a more constrained attachment of the FokI cleavage domain, which could improve the catalytic activity of TALE-FokI fusions. In an initial study, we found that trimming the C-terminal flanking segment of NT-L to 95 residues preserved full activation of endogenous *NTF3* (**Fig. 2b,c**). A reporter study of TALE13 supported this observation and further showed that trimming the C-terminal segment to 23 residues impaired activity (**Supplementary Fig. 2**). Next, we generated a C-terminal truncation series of TALE13, linked these proteins to the catalytic domain of wild-type FokI and screened these candidate TALENs as homodimers for cleavage of targets bearing 2–24 bp between monomer binding sites (**Supplementary Fig. 4**). This study yielded several useful observations. First, TALENs bearing the 95-residue, C-terminal segment lacked detectable nuclease activity. This result may reflect distinct structural requirements for DNA cleavage versus gene regulation (e.g., the +95 variant may display the FokI domain in a way that prevents interaction with DNA). Second, shorter C-terminal segments allowed efficient DNA cleavage, which identified candidate TALEN architectures for cellular studies. Third, for all TALENs tested, we observed reduced cleavage on targets bearing fewer than 10–12 bp between TALEN binding sites. This provided an important guide to our design of TALENs for *in vivo* studies, as it indicated that heterodimers targeted to endogenous loci would cleave most effectively if individual TALEN binding sites were separated by at least this distance.

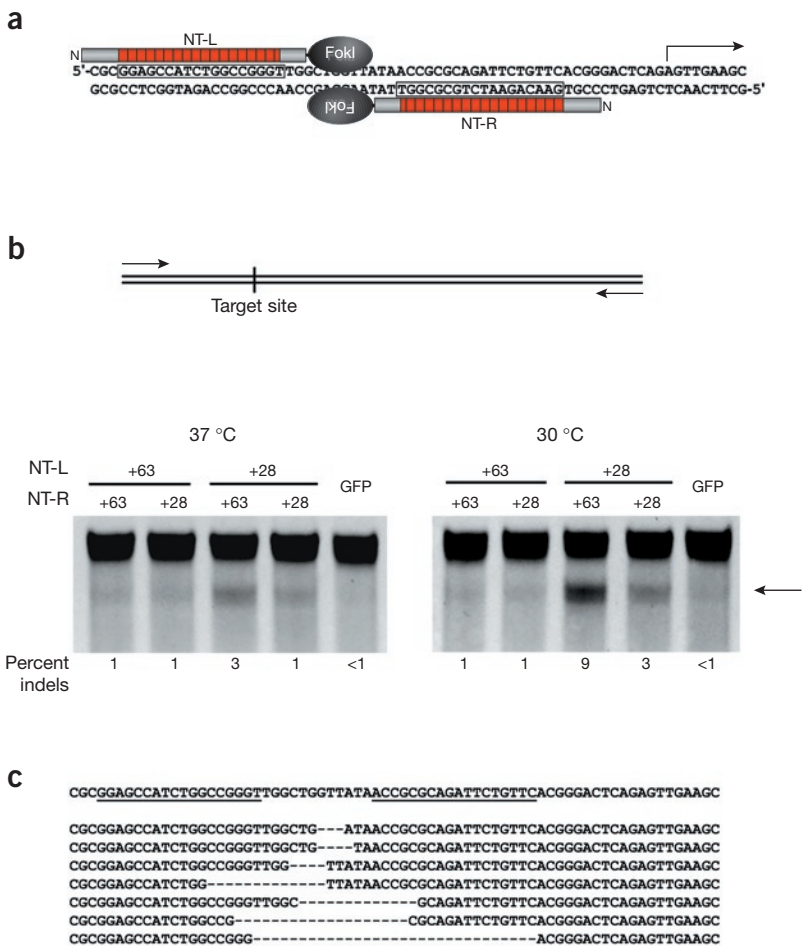

Figure 3 Modification of the endogenous *NTF3* locus in human cells. (**a**) Sketch of the designed TALENs NT-L and NT-R bound to their target sites (boxed bases) in the *NTF3* promoter. The hooked arrow shows the start site of *NTF3* transcription. Ovals indicate the cleavage domain from wild-type FokI, which was the catalytic domain used for all TALENs in these studies. N denotes the N terminus of each protein. (**b**) Surveyor nuclease assay for minor insertions and deletions. Top, sketch of the amplicon used for this assay. The NT-L/NT-R cleavage target is centered at bp 46 of the 272-bp amplicon. Arrows indicate PCR primers. Bottom, annotated assay gels. Lane headings indicate the TALEN dimer used for modification of each sample, where +28 and +63 refer to the C-terminal truncation variant of NT-L or NT-R. GFP indicates negative control samples from cells transfected with a plasmid expressing enhanced GFP. The arrow indicates the larger fragment produced by a Surveyor nuclease digest of amplicons bearing a mismatch at the site of NT-L/NT-R cleavage (the smaller fragment was not visible on this gel). Numbers beneath each lane indicate the percentage of modified alleles. Example lane traces are provided in **Supplementary Figure 6**. (**c**) Mutated alleles identified from sequence analysis of 84 cloned amplicons from cells that expressed NT-L+28 and NT-R+63 at 30 °C. Underlines highlight binding sites for NT-L and NT-R. Dashes indicate deleted bases. Uncropped gels are shown in **Supplementary Figure 12**.

In the final step in our development of an endogenously active TALEN architecture, we designed a heterodimer partner for NT-L (called NT-R; **Fig. 3a** and **Supplementary Fig. 5**), expressed both proteins in human K562 cells as fusions with the cleavage domain of wild-type FokI, and assayed the targeted locus for gene modification. We tested each designed TALEN using two different C-terminal subregions (+28 and +63) that had shown substantial activity in our *in vitro* cleavage study. When TALENs were expressed at 37 °C (our standard conditions) we observed up to 3% modification as measured by the Surveyor assay[35,36] (**Fig. 3b**, left gel panel, NT-L+28/NT-R +63 and **Supplementary Fig. 6**). This increased to 9% when the TALEN-transfected cells were subjected to transient hypothermia[37] (**Fig. 3b**, right gel panel). Moreover, Sanger sequencing of the resultant amplicons identified 7 mutated alleles out of 84 analyzed and also revealed a mutation spectrum (minor deletions)

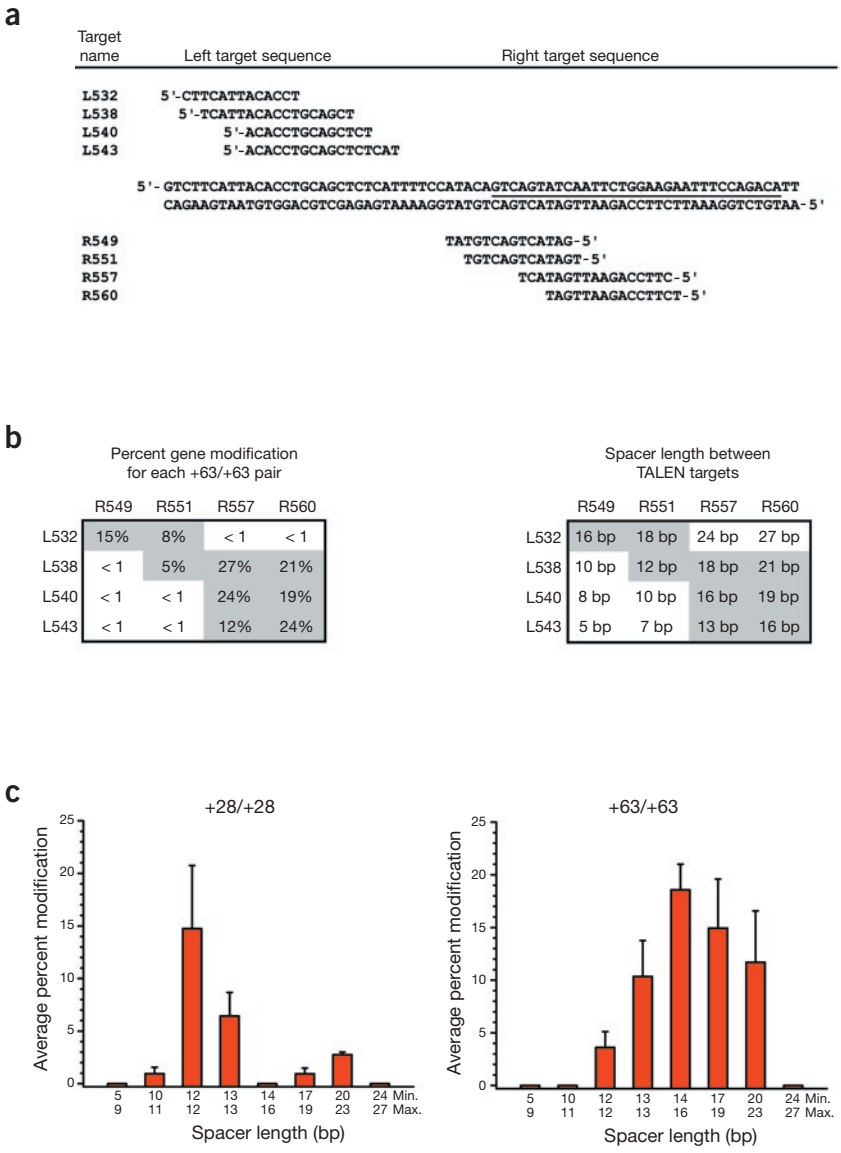

Figure 4 Modification of endogenous human *CCR5*. (**a**) Target names and sequences of the TALENs tested in this study. Targets are aligned with the corresponding region of CCR5 (double-stranded sequence at center). Each target is named according to whether it is a left or right target (L or R) and the position of its 3′ base in the CCR5 open reading frame (e.g., L532 is a target for a left TALEN and the 3′ T of this target is bp 532 of the CCR5 open reading frame). The underlined segment of CCR5 sequence indicates bases deleted in the Δ32 variant. For more detail regarding our choice of target sites see **Supplementary Methods.** (**b**) Gene modification levels induced by the +63/+63 versions of all 16 nuclease combinations as gauged by the Surveyor nuclease assay. The left grid provides modification levels whereas the right grid indicates the separation distance (or 'spacer') between nuclease target sites. An entry of "<1" indicates that modification levels were below the detection limit of 1%. In both grids, pairs showing detectable activity are shaded gray. (**c**) Average gene modification activities for +28/+28 and +63/+63 TALEN pairs from this study and one additional study (**Supplementary Fig. 9**), binned according to the spacer length between TALEN targets. Max. and Min. indicate the longest and shortest target separation distance (in bp) in each bin. The number of data points in each bin was for 5–9 bp: $N = 4$; 10–11 bp: $N = 4$; 12 bp: $N = 3$; 13 bp: $N = 3$; 14–16 bp: $N = 5$; 17–19 bp: $N = 4$; 20–23 bp: $N = 3$; and 24–27 bp: $N = 2$. Error bars indicate s.e.m..

consistent with error-prone break repair by means of nonhomologous end joining[38] (NHEJ) (**Fig. 3c**). TALENs were also shown to induce NHEJ-mediated capture of an oligonucleotide duplex at the endogenous *NTF3* locus in K562 cells[39] (**Supplementary Fig. 7**). These results show that our TALEN architecture can drive NHEJ-mediated genome modification at an endogenous locus in a mammalian cell.

Efficient modification by NHEJ and homology directed repair (HDR)

To gauge the generality of this result, we sought to demonstrate TALEN-mediated gene modification at an additional locus: the site of the Δ32 deletion within the human *CCR5* gene[40]. For this study, we designed and tested a panel of TALENs with a range of target lengths (13–17 bp) and of separation distances (or 'spacers'; 5–27 bp) to identify values compatible with endogenous activity. To accomplish this, we first designated a cluster of four 'left' and four 'right' binding sites at the location of the Δ32 deletion, which defined a matrix of 16 heterodimer targets (**Fig. 4a**). Next, two alternative proteins were generated for each binding site, bearing a C-terminal segment of either 28 or 63 residues. Finally, all pairwise combinations of left and right proteins ($8 \times 8 = 64$ total) were expressed in K562 cells under our standard conditions (37 °C) and assayed for modification of the endogenous locus. **Supplementary Figure 8** summarizes all data from this study, and **Figure 4b** provides data for all +63/+63 pairings, the most active subset.

These results provide insights into the function of our TALENs. First, they show that TALENs can drive efficient gene modification, with four of the +63/+63 pairings yielding >20% mutated alleles (**Fig. 4b**). Moreover, TALENs appear to be generally well-tolerated, as most TALEN-treated cell pools retain a stable modification level after 10 d of growth (<25% signal loss; see **Supplementary Fig. 8b**). Our results also reveal clear patterns of activity versus target spacing. Within the +63/+63 TALEN panel, all nuclease pairs separated by 12–21 bp yielded at least 5% modification, whereas no activity (<1% modification) was detected for any pair of targets with a separation distance lying outside this range (**Fig. 4b**). In contrast, the +28/+28 panel yielded maximal activity for a narrower separation range of 12–13 bp (**Supplementary Fig. 8**). This difference in behavior was also observed in a gene modification study using a second TALEN panel targeted to a different locus in *CCR5* (**Supplementary Fig. 9**). This difference in behavior is also apparent in **Figure 4c**, which bins and averages gene modification efficiencies according to the number of bases separating TALEN monomer target sites.

Having demonstrated efficient NHEJ-mediated modification, we next sought to show that our TALEN architecture could induce gene editing through HDR, the other major cellular DNA repair pathway. For this study, we targeted a second locus within *CCR5* previously used as a safe harbor for transgene integration[15]. We screened 24 TALEN pairs for their ability to induce modification through NHEJ and identified the two most active combinations, which yielded mutation levels of 21% and 18% (**Supplementary Fig. 9**, pairs highlighted in gray). Next, these TALENs were introduced into K562 cells with a

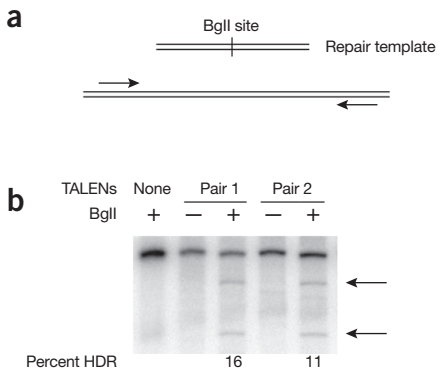

Figure 5 TALEN-mediated gene editing by HDR. (**a**) Experimental overview. HDR using the indicated repair template transfers a 46-bp insert bearing a BglI restriction site into the targeted locus. PCR amplification, followed by BglI digestion, allows detection and quantification of the gene editing event. Arrows indicate PCR primers. (**b**) Annotated assay gel. Arrows indicate the fragments generated by BglI digestion. Lane headings indicate which TALENs were used (pair 1, pair 2; see **Supplementary Fig. 9** for details), as well as whether the resultant amplicons were treated with BglI. Uncropped gels are shown in **Supplementary Figure 12**.

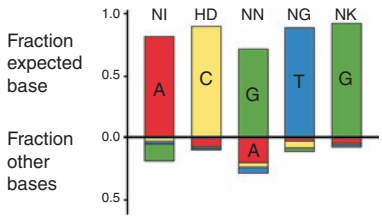

Figure 6 Base preferences of the five RVDs used in this study, as determined by averaging SELEX-derived base preferences for NT-L (**Fig. 2e**) and NT-R (**Supplementary Fig. 5**), as well as two additional designed TALEs "VEGF-1" and "CCR5-1" (**Supplementary Fig. 10**). For each RVD (listed at top), the average frequency of its preferred target base is shown above the *x* axis, whereas frequencies of the remaining bases are shown below the *x* axis.

donor DNA fragment designed to transfer a 46-bp insert encoding a BglI restriction site into the targeted locus. Subsequent PCR and BglI digestion revealed efficient editing, with up to 16% of alleles possessing the inserted sequence (**Fig. 5**). This result shows that the TALEN architecture described here can induce precise gene editing by HDR.

DISCUSSION

In these studies we have developed a site-specific nuclease architecture by linking TALE truncation variants to the catalytic domain of FokI. We have shown that the resulting TALENs can induce gene modification in human cells by means of the two major eukaryotic DNA repair pathways, NHEJ and HDR[7]. We have also demonstrated three distinct applications for a TALEN-induced double-stranded break: NHEJ-mediated gene disruption, NHEJ-mediated capture of a DNA fragment and gene editing by HDR. This work shows that TALENs can be applied as genome engineering agents.

Perhaps the most obvious approach for developing a TALEN architecture would have been to link the minimal central TALE repeat domain to the catalytic domain of FokI. However, observations from other groups[30] as well as our initial studies (data not shown) revealed that such constructs lack activity. An alternative approach would be to use a full-length TALE for DNA recognition. However, this would interpose a large span of excess peptide between the TALE repeats and FokI cleavage domain, which could impair activity (e.g., by serving as a very long, entropically unfavorable linker). Thus we used a third approach, which examined TALE truncation variants. This approach yielded two TALEN architectures that retain just 28 or 63 of the 278 original C-terminal residues, and that mediate efficient genome modification.

Our TALEN architectures contrast with two others that were described while this work was in progress and which retain most (231 residues)[30] or all of the C terminus[31]. These proteins were not validated in studies targeting bona fide endogenous native loci but were instead characterized by episome cleavage in yeast[30,31], a setting that does not necessarily predict function on endogenous genes in more complex genomes[41,42]. Moreover, consistent with their larger C-terminal segments, these proteins also appear to cleave across spacer sequences with a somewhat wider range of lengths and larger mean size than the TALENs developed in these studies.

The properties of our TALENs provide an interesting comparison with ZFNs, which are the most widely used designed nucleases for genome engineering. Perhaps the most notable property of our TALENs is their high cleavage activity, which enables gene modification efficiencies that approach the levels induced by many ZFNs (see studies listed in Table 1 of ref. 1). However, the activities fall short of several ZFN pairs[16,35,43,44], and it will be interesting to see whether TALENs can be engineered to increase their activity further. A second point of comparison is relative size. Per recognized base, TALE repeats are 3–4 times larger than ZFPs, and it is possible that this will impose limitations for certain delivery methods. The very high levels of TALE repeat homology may also complicate the assembly and maintenance of certain constructs. Regarding other potential comparisons—for example, specificity, ease of design, portability to more sensitive cell types—we anticipate that future experiments will provide a fuller picture of the relative strengths of each platform.

In addition to showing gene modification, these studies also demonstrate targeted endogenous gene regulation by designed TALE proteins. This capability may find use in a range of research[45] and biotechnology[46,47] applications. Designed activators, for example, are useful for increasing the dose of a gene product that requires natural splice variant ratios for proper function[48], or that is toxic when overexpressed[49]. Transcriptional repression has provided a strategy for reducing the activity of classically nondruggable gene targets[50]. Although TALEs function naturally as transcription factors[25,26], prior studies had not examined whether these proteins could regulate gene expression outside of their native plant context. The experiments presented here show that designed TALEs can be used as targeted gene regulators.

In addition, to our knowledge no previous study has provided unbiased assessment of the DNA binding preferences of natural and designed TALE proteins. Using a SELEX assay, we identified binding sites for TALE13 (**Fig. 1d**) as well as four engineered TALEs (**Fig. 2** and **Supplementary Figs. 5** and **10**). We observed that 73 of the 76 TALE repeats in these five proteins yielded the base preference specified by its RVD. This generally predictable targeting behavior provides further support for the use of TALEs as designable DNA-binding proteins. We interpret the rare exceptions (e.g., the adenine preference of the fourth repeat in NT-L; the guanine preference of the seventh repeat of NT-R) as evidence for as yet poorly characterized context effects—such as contacts from neighboring repeats—that can change the specificity of an RVD. These studies also yielded sufficient data to gauge the average base preference for five different RVDs: four common RVDs used previously[23,30] (**Fig. 1b**) and an alternative—NK—that we tested for recognition of guanine[24,51]. Consistent with earlier experimental studies[23], the common RVDs—NI, HD, NN and NG—selectively bound adenine, cytosine, guanine and thymine, respectively (**Fig. 6**). Moreover, NN exhibited substantial binding to a second base, adenine (**Fig. 6**). In contrast to this behavior, NK showed a much stronger preference for guanine, representing a potential improvement for recognition of this base.

The description of rules for TALE-DNA recognition[23,24] raised the prospect of new classes of TALE fusion proteins for targeted gene modification and regulation in higher eukaryotes. Our studies have both examined and furthered these possibilities. Using SELEX analyses, we have assessed the predictive power of TALE-DNA recognition rules and have validated a new RVD for potentially improved recognition of guanine. In our cellular studies, we have demonstrated TALE function in higher eukaryotic cells outside of their native (plant) context. Finally, by developing a TALEN architecture that efficiently cleaves DNA, we have shown that designed TALE proteins can enable targeted and precise modification of endogenous genes.

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 S. Abrahamson for critically reviewing this manuscript, and also E. Wolffe and D. Guschin for helpful comments and suggestions. We also thank S. Orlando, Y. Santiago, S. Lussier, A. Vincent and S. Lam for technical support.

AUTHOR CONTRIBUTIONS

J.C.M. and S.T. designed studies, performed experiments and analyzed data. G.Q., K.A.B., J.W., D.F.X., X.M., D.E.P., K.L.H., and G.J.C. performed studies. S.T, G.Q., E.L., S.J.H., G.P.D., L.Z., and I.A. developed new procedures and assembled constructs. F.D.U., H.S.Z, M.C.H., L.Z., P.D.G. and E.J.R. supervised studies and designed experiments. J.C.M., P.D.G. and E.J.R. 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://npg.nature.com/reprintsandpermissions/.

1. Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. & Gregory, P.D. Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* **11**, 636–646 (2010).
2. Davis, D. & Stokoe, D. Zinc finger nucleases as tools to understand and treat human diseases. *BMC Med.* **8**, 42 (2010).
3. Camenisch, T.D., Brilliant, M.H. & Segal, D.J. Critical parameters for genome editing using zinc finger nucleases. *Mini Rev. Med. Chem.* **8**, 669–676 (2008).
4. Carroll, D. Progress and prospects: zinc-finger nucleases as gene therapy agents. *Gene Ther.* **15**, 1463–1468 (2008).
5. Cathomen, T. & Joung, J.K. Zinc-finger nucleases: the next generation emerges. *Mol. Ther.* **16**, 1200–1207 (2008).
6. Galetto, R., Duchateau, P. & Paques, F. Targeted approaches for gene therapy and the emergence of engineered meganucleases. *Expert Opin. Biol. Ther.* **9**, 1289–1303 (2009).
7. Hoeijmakers, J.H. Genome maintenance mechanisms for preventing cancer. *Nature* **411**, 366–374 (2001).
8. Ochiai, H. *et al.* Targeted mutagenesis in the sea urchin embryo using zinc-finger nucleases. *Genes Cells* **15**, 875–885 (2010).
9. Takasu, Y. *et al.* Targeted mutagenesis in the silkworm Bombyx mori using zinc finger nuclease mRNA injection. *Insect Biochem. Mol. Biol.* **40**, 759–765 (2010).
10. Morton, J., Davis, M.W., Jorgensen, E.M. & Carroll, D. Induction and repair of zinc-finger nuclease-targeted double-strand breaks in *Caenorhabditis elegans* somatic cells. *Proc. Natl. Acad. Sci. USA* **103**, 16370–16375 (2006).
11. Holt, N. *et al.* Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. *Nat. Biotechnol.* **28**, 839–847 (2010).
12. Benabdallah, B.F. *et al.* Targeted gene addition to human mesenchymal stromal cells as a cell-based plasma-soluble protein delivery platform. *Cytotherapy* **12**, 394–399 (2010).
13. Zou, J. *et al.* Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell* **5**, 97–110 (2009).
14. Hockemeyer, D. *et al.* Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat. Biotechnol.* **27**, 851–857 (2009).
15. Lombardo, A. *et al.* Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat. Biotechnol.* **25**, 1298–1306 (2007).
16. Perez, E.E. *et al.* Establishment of HIV-1 resistance in CD^{4+} T cells by genome editing using zinc-finger nucleases. *Nat. Biotechnol.* **26**, 808–816 (2008).
17. Urnov, F.D. *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646–651 (2005).
18. Bibikova, M., Golic, M., Golic, K.G. & Carroll, D. Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics* **161**, 1169–1175 (2002).
19. Porteus, M.H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**, 763 (2003).
20. 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).
21. 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).
22. Kim, Y.G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. USA* **93**, 1156–1160 (1996).
23. Boch, J. *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509–1512 (2009).
24. Moscou, M.J. & Bogdanove, A.J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501 (2009).
25. Römer, P. *et al.* Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. *Science* **318**, 645–648 (2007).
26. Kay, S., Hahn, S., Marois, E., Hause, G. & Bonas, U. A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science* **318**, 648–651 (2007).
27. Kay, S. & Bonas, U. How Xanthomonas type III effectors manipulate the host plant. *Curr. Opin. Microbiol.* **12**, 37–43 (2009).
28. Bogdanove, A.J., Schornack, S. & Lahaye, T. TAL effectors: finding plant genes for disease and defense. *Curr. Opin. Plant Biol.* **13**, 394–401 (2010).
29. Boch, J. & Bonas, U. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu. Rev. Phytopathol.* **48**, 419–436 (2010).
30. Christian, M. *et al.* Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186, 757–761 (2010).
31. Li, T. *et al.* TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.* published online, doi:10.1093/nar/gkq704 (10 August 2010).
32. Szurek, B., Rossier, O., Hause, G. & Bonas, U. Type III-dependent translocation of the Xanthomonas AvrBs3 protein into the plant cell. *Mol. Microbiol.* **46**, 13–23 (2002).
33. Chao, M.V., Rajagopal, R. & Lee, F.S. Neurotrophin signalling in health and disease. *Clin. Sci. (Lond.)* **110**, 167–173 (2006).
34. Kay, S., Hahn, S., Marois, E., Wieduwild, R. & Bonas, U. Detailed analysis of the DNA recognition motifs of the Xanthomonas type III effectors AvrBs3 and AvrBs3Deltarep16. *Plant J.* **59**, 859–871 (2009).
35. Miller, J.C. *et al.* An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.* **25**, 778–785 (2007).
36. Guschin, D.Y. *et al.* A rapid and general assay for monitoring endogenous gene modification. *Methods Mol. Biol.* **649**, 247–256 (2010).
37. Doyon, Y. *et al.* Transient cold shock enhances zinc-finger nuclease-mediated gene disruption. *Nat. Methods* **7**, 459–460 (2010).
38. Jeggo, P.A. DNA breakage and repair. *Adv. Genet.* **38**, 185–218 (1998).
39. Orlando, S.J. *et al.* Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. *Nucleic Acids Res.* **38**, e152 (2010).
40. Liu, R. *et al.* Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* **86**, 367–377 (1996).
41. Townsend, J.A. *et al.* High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature* **459**, 442–445 (2009).
42. Redondo, P. *et al.* Molecular basis of xeroderma pigmentosum group C DNA recognition by engineered meganucleases. *Nature* **456**, 107–111 (2008).
43. DeKelver, R.C. *et al.* Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. *Genome Res.* **20**, 1133–1142 (2010).
44. Doyon, Y. *et al.* Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat. Methods* published online, doi: 10.1038/nmeth.1539 (5 December 2010).
45. Cuthbert, G.L. *et al.* Histone deimination antagonizes arginine methylation. *Cell* **118**, 545–553 (2004).
46. Reik, A. *et al.* Enhanced protein production by engineered zinc finger proteins. *Biotechnol. Bioeng.* **97**, 1180–1189 (2007).
47. Liu, P.Q. *et al.* Isogenic human cell lines for drug discovery: regulation of target gene expression by engineered zinc-finger protein transcription factors. *J. Biomol. Screen.* **10**, 304–313 (2005).
48. Whitlock, P.R., Hackett, N.R., Leopold, P.L., Rosengart, T.K. & Crystal, R.G. Adenovirus-mediated transfer of a minigene expressing multiple isoforms of VEGF is more effective at inducing angiogenesis than comparable vectors expressing individual VEGF cDNAs. *Mol. Ther.* **9**, 67–75 (2004).
49. Ozawa, C.R. *et al.* Microenvironmental VEGF concentration, not total dose, determines a threshold between normal and aberrant angiogenesis. *J. Clin. Invest.* **113**, 516–527 (2004).
50. Tan, S. *et al.* Zinc-finger protein-targeted gene regulation: genomewide single-gene specificity. *Proc. Natl. Acad. Sci. USA* **100**, 11997–12002 (2003).
51. Morbitzer, R., Römer, 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* (2010).

ONLINE METHODS

TALE constructs. DNA fragments encoding TALE13 and the three other TALE proteins described in **Supplementary Figure 1** were isolated from *Xanthomonas axonopodis* Starr and Garces pathovar *citri* genomic DNA (American Type Culture Collection (ATCC) no. 49118) by PCR amplification with the following PCR primers: 5′-ACGTGGATTCATGGTGGATCTACGCACGCTC-3′ and 5′-TACGTCCGCGGTCCTGAGGCAATAGCTCCATCA-3′. These primers were designed to amplify a region of natural TALE genes corresponding to residues 153–278 of AvrBS3. PCR products were cloned into pCR2.1 (Invitrogen) and sequenced. TALE13-VP16 fusion constructs were generated by linking DNA encoding TALE13 to the VP16 activation domain, and cloning into the pcDNA3 mammalian expression vector (Invitrogen). Full sequences of all TALE-VP16 constructs are provided in **Supplementary Methods**. The ΔR1-13 variant was generated by PCR and deleted all 13 repeats of TALE13 as well as the first 14 residues of the C-terminal flanking region.

A DNA fragment encoding the TALE repeat domain of NT-L was generated by PCR-based gene synthesis as described[52]. RVDs for individual repeats were chosen using the previously described TALE design code[23] except that NK[24] (instead of NN) was used in the 12th and 16th TALE repeats to recognize guanine. To generate the VEGF-1 and CCR5-1 TALE constructs, we first created overlapping fragments (each encoding 2–5 repeats) by PCR using NT-L as a template and primers that specify desired RVD changes; full-length TALEs were then assembled from such fragments using overlapping PCR. The NT-R+63 construct was generated by complete gene synthesis (DNA2.0). Shorter C-terminal truncations were generated by PCR.

To generate TALENs, we cloned the DNA encoding the TALE repeats and flanking regions into pVAX-based vectors by means of Acc65I and BamHI digestion and ligation, which generates an in-frame fusion gene comprising an N-terminal nuclear localization signal and a C-terminal wild-type FokI cleavage domain. TALENs targeting the two *CCR5* loci were generated by combining portions of synthetic sequence (DNA2.0) and cloning into a TALEN vector derived from the NT-R+63 construct with the appropriate C-terminal flanking region through an ApaI restriction site (encoding the three residues 'GAP' before the first repeat) and an HpaI restriction site (encoding the three residues 'ALT' after the last repeat). These larger scale assemblies used only the NN RVD to target guanine bases as the better average performance of the NK RVD had not been determined when the assemblies were initiated. The complete amino acid and DNA sequences for all TALEN constructs used in this paper are shown in **Supplementary Methods**.

SELEX assay. This assay was carried out essentially as described[16]. Briefly, oligonucleotide target libraries were chemically synthesized using a mixture of all four phosphoramidites at from 18 to 26 positions. These randomized regions were flanked by discrete sequences of 23–24 bp in length. Library sequences were:

N18TA: 5′-CAGGGATCCATGCACTGTACGTTTNNNNNNNNNNNNNN-NNNNNAAACCACTTG ACTGCGGATCCTGG-3′

N22TA: 5′-CAGGGATCCATGCACTGTACGTTTNNNNNNNNNNNNNNNN-NNNNNNNAAACC ACTTGACTGCGGATCCTGG-3′

N23TA: 5′-CAGGGATCCATGCACTGTACGTTTNNNNNNNNNNNNN-NNNNNNNNNNNAAAC CACTTGACTGCGGATCCTGG-3′

N26: 5′-CAGGGATCCATGCACTGTACGTTNNNNNNNNNNNNNNNN-NNNNNNNNNNNAACC ACTTGACTGCGGATCCTGG-3′

Library oligonucleotides were then converted to double-stranded duplex by annealing with the 3′ library primer (5′-CCAGGATCCGCAGTCAAGTGG) and incubation under the following conditions: 2 nmol library 6 nmol primer in 100 µl 1× PCR Master (Roche) supplemented with higher concentrations of each dNTP (1.2 mM final concentration of each dNTP) annealed at 94 °C 5 min, 58 °C 10 min, 72 °C 10 min.

For the first cycle of the assay, fragments encoding TALE proteins were amplified by PCR using a 5′ primer bearing a T7 promoter: 5′-GCTTACTGGCTTATCGAAATTAATACGACTCACTATAGG-GAGACGAATTCACCACCATGGTGGATCTACGCACGCTCG-3′ and a 3′ primer encoding an in-frame hemagglutinin (HA)-epitope tag: 5′-CACGTACTTCAGCTTTTATCAGGCGTAGTCGGGCACGTCGTAGG GGTAGCCGCTCATCCCGAACTGCGTCA-3′ for TALE13+95, NT-L+95, CCR5-1+95 and VEGF-1+95 or 5′-CACGTACTTCAGCTTTTATTAGGCGT AGTCGGGCACGTCGTAGGGGTAGCCCGCGACTCGATGGGAAGTTC-3′ for NT-R+63. Protein was then expressed by adding the PCR product to a TnT-coupled transcription-translation system (Promega) and then mixed with 200 pmol of library duplex and 0.2 µg biotinylated anti-HA antibody (Roche) in a total volume of 100 µl of SELEX buffer (0.5% Tween 20, 10 µM $ZnCl_2$, 0.5 mM $MgCl_2$, 20 µg/ml poly dIdC, 0.01% BSA in PBS without $CaCl_2$). After incubation for 50 min protein-DNA-antibody complexes were captured on streptavidin-coated magnetic beads (Invitrogen) and washed six times with SELEX buffer.

Bound DNA target was then amplified by PCR using the 3′ library primer (above) and the 5′ library primer (5′-CAGGGATCCATGCACTGTACG), and then used as input for additional cycles of enrichment. Protein expression and binding conditions for these subsequent cycles were identical to the conditions used in the first round. After three cycles, recovered DNA fragments were cloned and sequenced. Sequences were then aligned, purged of exact duplicates and condensed to the base frequency plots provided in **Figures 1d** and **2e**, and **Supplementary Figures 5** and **10**. The number of unique sequences used to generate these plots were as follows: **Figure 1d**, 48 sequences; **Fig. 2e**, 27 sequences; **Supplementary Fig. 5**, 45 sequences; **Supplementary Fig. 10**, 25 (VEGF-1) and 17 sequences (CCR5-1). Sequence alignments are provided in **Supplementary Figure 11**. The initial TALE13 construct was also characterized with the full +278 C-terminal flanking region and yielded results similar to those shown in **Figure 1d**; the data for the +95 variant of TALE13 were shown to allow easier comparisons to the other SELEX data sets.

Endogenous gene activation assay for *NTF3*. HEK293 cells were transiently transfected by the indicated TALE constructs in Lipofectamine 2000 (Invitrogen). Forty-eight hours after transfection, the cells were harvested, and total RNA prepared using the high pure RNA isolation kit (Roche Diagnostics) according to the manufacturer's recommendations. *NTF3* mRNA expression was measured by real-time quantitative RT-PCR using Taqman chemistry in a 96-well format on an ABI 7700SDS machine (Perkin-Elmer). As a normalization control, *GAPDH* mRNA expression was also measured in a similar fashion. The Taqman primer and probe set for *NTF3* and *GAPDH* were as follows: NTF3-For (5′-GATAAACACTGGAACTCTCAGTGCAA-3′), NTF3-Rev (5′-GCCAGCCCACGAGTTTATTGT-3′) and NTF3-Pro (5′-FAM-CAAACCTACGTCCGAGCACTGACTTCAGA-BHQ1-3′) for NTF3, and GAPDH-For (5′-CCATGTTCGTCATGGGTGTGA-3′), GAPDH-Rev (5′-CATGGACTGTGGTCATGAGT-3′), GAPDH-Pro (5′-FAM-TCCTGCACCACCAACTGCTTAGCA-TAMRA-3′) for GAPDH, respectively. NT-3 protein levels in culture supernatants were measured using an ELISA kit (Promega G7640).

Surveyor nuclease assay for *NTF3* modification. Human K562 cells (ATCC) were grown in RPMI medium supplied with 5% FBS (Invitrogen) and transfected with 400 ng of each TALEN plasmid using the Amaxa Nucleofector 96-well Shuttle System (Solution SF, Program 96-FF-120) (Amaxa Biosystems/Lonza) according to manufacturer's protocols. Genomic DNA was isolated 72 h after transfection using the QuickExtract DNA purification solution (Epicentre Biotechnologies). PCR was performed for 35 cycles (95 °C, 30 s; 60 °C, 30 s; 68 °C, 40 s) using Accuprime HiFi DNA polymerase (Invitrogen) and the primer pairs (LZNT3-F4: 5′-GAAGGGGTTAAGGCGCTGAG-3′; LZNT3-1077R: 5′-AGGGACGTCGACATGAAGAG-3′) that amplify the intended target regions. The Surveyor nuclease (Surveyor mutation detection kit; Transgenomic) was used to treat the DNA according to manufacturer's instruction. The products were resolved on a 10% polyacrylamide gel and bands were visualized by staining with ethidium bromide and image capture with a FluorChem SP gel imaging station (Alpha Innotech). The percentage NHEJ was calculated as previously described[36]. See the **Supplementary Methods** for an overview of the Surveyor assay.

Gene modification of endogenous *CCR5*. To screen TALEN pairs for NHEJ-mediated gene modification, K562 cells were cultured in RPMI1640 media (Invitrogen) supplemented with 10% (vol/vol) FBS, 2 mM L-glutamine, 100 U/ml penicillin, and 100 mg/ml streptomycin. Cells ($1–2 \times 10^5$) were

doi:10.1038/nbt.1755

nucleofected with TALEN expression plasmids (400 ng each) using the Amaxa 96-well shuttle system (Amaxa Biosystems/Lonza) according to manufacturer's instructions. Cells were collected 3 d and 10 d after transfection and genomic DNA was extracted using the QuickExtract DNA Extraction Solution (Epicentre Biotechnologies) according to suppliers' instructions. Frequency of gene modification by NHEJ was evaluated by the Surveyor Nuclease Assay, as described previously[16,36]. Primers used for monitoring the Δ32 locus were CCR5_409_20_F, 5′-TTAAAAGCCAGGACGGTCAC-3′; CCR5_754_20_R, 5′-TGTAGGGAGCCCAGAAGAGA-3′. Primers used for monitoring the other *CCR5* locus were previously described[16]. See the **Supplementary Methods** for an overview of the Surveyor assay.

To demonstrate *CCR5* gene editing by HDR, we cultured K562 cells in RPMI1640 media (Invitrogen) supplemented with 10% (vol/vol) FBS, 2 mM L-glutamine, 100 U/ml penicillin, and 100 μg/ml streptomycin. Cells ($1–2 \times 10^5$) were nucleofected with TALEN expression plasmids (400 ng each) and 400 ng of HDR donor DNA, which was designed to insert 46 bp in between the TALEN binding sites which contained a BglI target site. Cells were collected 3 d after transfection and genomic DNA was extracted using the DNeasy Tissue kit (Qiagen). Frequency of gene modification by homologous recombination was evaluated by a restriction fragment length polymorphism (RFLP) assay, which was performed based on the insertion of a BglI site between the CCR5 TALEN binding sequences within the *CCR5* gene. Briefly, a pair of *CCR5* primers (5′-CTGCCTCATAAGGTTGCCCTAAG-3′ and 5′-CCAGCAATAGATGATCCAACTCAAATTCC-3′) located outside the CCR5 homologous region of donor molecules was used to PCR amplify a 2.5 kb *CCR5* fragment in the presence of [α-32P]dATP and dCTP. PCR products were passed through a G-50 column (GE Healthcare) and digested with BglI. The products were resolved on a 10% polyacrylamide gel, the gel dried and RFLP knock-in quantified using a phosphorimager (Molecular Dynamics). The ratio of cleaved to total products was calculated as a measurement of frequency of targeted integration.

Donor DNA fragment for this study consisted of two large homology arms bearing the sequence of the *CCR5* locus flanking the TALEN binding sites, and a patch sequence of 51 bp between the arms (patch sequence: TAGATCAGTGAGTATGCC**CTGAT**GGCG-TCTGGACTGGATGCCTCGTCTAGA). The patch contained a BglI restriction site (underlined) along with 5 bp that was homologous to the sequence between TALEN targets (in bold) for a net insertion of 46 bp. The complete DNA sequence of the donor is shown in **Supplementary Methods**.

52. Hoover, D.M. & Lubkowski, J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* **30**, e43 (2002).

doi:10.1038/nbt.1755