

Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes

Ting Li¹, Sheng Huang¹, Xuefeng Zhao², David A. Wright¹, Susan Carpenter³, Martin H. Spalding¹, Donald P. Weeks⁴ and Bing Yang^{1,*}

¹Department of Genetics, Development and Cell Biology, ²Laurence H. Baker Center for Bioinformatics and Biological Statistics, ³Department of Animal Science, Iowa State University, Ames, IA 50011 and

⁴Department of Biochemistry, University of Nebraska, Lincoln, NE 68588, USA

Received February 20, 2011; Revised and Accepted March 16, 2011

ABSTRACT

Recent studies indicate that the DNA recognition domain of transcription activator-like (TAL) effectors can be combined with the nuclease domain of FokI restriction enzyme to produce TAL effector nucleases (TALENs) that, in pairs, bind adjacent DNA target sites and produce double-strand breaks between the target sequences, stimulating non-homologous end-joining and homologous recombination. Here, we exploit the four prevalent TAL repeats and their DNA recognition cipher to develop a 'modular assembly' method for rapid production of designer TALENs (dTALENs) that recognize unique DNA sequence up to 23 bases in any gene. We have used this approach to engineer 10 dTALENs to target specific loci in native yeast chromosomal genes. All dTALENs produced high rates of site-specific gene disruptions and created strains with expected mutant phenotypes. Moreover, dTALENs stimulated high rates (up to 34%) of gene replacement by homologous recombination. Finally, dTALENs caused no detectable cytotoxicity and minimal levels of undesired genetic mutations in the treated yeast strains. These studies expand the realm of verified TALEN activity from cultured human cells to an intact eukaryotic organism and suggest that low-cost, highly dependable dTALENs can assume a significant role for gene modifications of value in human and animal health, agriculture and industry.

INTRODUCTION

Technologies for precise and efficient gene editing in living cells hold great promise in both basic and applied research, including therapeutic interventions for genetic diseases. These technologies exploit the ability of endonucleases to cause chromosomal double-stranded DNA breaks (DSBs) and stimulate the subsequent breakage repair mechanisms in living cells (1–3). The two widely conserved, major repair pathways in eukaryotes are non-homologous end-joining (NHEJ) and homologous recombination (HR). NHEJ repair often results in mutagenic deletions/insertions and substitutions in the targeted gene. DSBs also stimulate HR between the endogenous target gene locus and an exogenously introduced homologous donor DNA carrying desired genetic alterations (4–6). At the forefront of these methods are custom-designed zinc-finger nucleases (ZFNs), which are hybrid proteins derived from the DNA binding domains of zinc-finger (ZF) proteins and the non-specific cleavage domain of the endonuclease FokI. However, despite their promise, widespread adoption of ZFNs is hampered by a bottleneck in custom-engineering ZFs with high specificity and affinity for the DNA target sites (7,8). Further, ZFN utility is somewhat limited by the number and location of potential target sites within a genome. Alternative strategies that overcome limitations in current technologies for targeted genome editing could greatly accelerate adoption of artificial nucleases for efficient gene disruption and gene replacement in a variety of heretofore recalcitrant eukaryotic organisms.

ZFN efficacy depends almost solely on the DNA binding specificity of their ZF domains (9), which theoretically can be supplanted by any high-fidelity DNA

*To whom correspondence should be addressed. Tel: +1 515 294 2968; Fax: +1 515 294 5256; Email: byang@iastate.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2011. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.5>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

binding domain (10–12). Transcription activator-like effector (TALE) proteins, a large group of bacterial plant pathogen proteins, have emerged as alternatives to ZF proteins. TALE proteins contain a varying number of centrally located tandem 34-amino-acid repeats that mediate binding to a specific DNA target sequence, referred to as the effector binding elements (EBE). Each repeat is nearly identical except for two variable amino acids at positions 12 and 13, known as repeat variable di-residues (RVD). Polymorphism in the number of repeats (a range of 13–33) and in the RVD composition collectively determines the DNA binding specificity of individual TALE proteins. Remarkably, recognition of a specific DNA sequence is based on a fairly simple code wherein one base of the DNA target site is recognized by the RVD of one repeat (i.e. one repeat/one nucleotide). The sequential repeat arrangement in a single TALE protein thus specifies the contiguous DNA sequence that will be bound by that TALE protein (13,14). TALE recognition studies also reveal a preference for certain RVDs over others in recognizing a particular nucleotide. In most cases, the RVDs asparagine and isoleucine (NI), histidine and aspartic acid (HD), asparagine and glycine (NG) and two asparagines (NN) combine to recognize the nucleotides ‘A’, ‘C’, ‘T’ and ‘G’, respectively (13,14). The prevalence of these four RVDs in the native TALEs made it possible to use them exclusively to *de novo* synthesize or assemble TALE repeat arrays of up to 13 repeat units to target DNA sequences in plant and human cells (15–17). The addition of FokI nuclease domains to the C-termini of two paired synthetic TAL effectors (as in ZFNs) should allow for highly gene-specific gene targeting. Because of the modular nature of TALEs and the potential to use long DNA recognition sites, custom-made designer TALE nucleases (dTALENs) may overcome limitations of current ZFN technologies and may significantly advance the use of targeted genome editing for challenging long-term opportunities such as therapeutic repair of genetic diseases. Toward this end, we and others have recently shown that TALEs can be linked with FokI nuclease domain to direct targeted cleavage of DNA containing a specific EBE (17–19). More recently, Miller *et al.* (20) demonstrated TALE-mediated editing of endogenous genes in cultured human cells.

Here, we report development of a modular assembly technology for custom-engineering dTALENs and characterization of 10 such dTALENs for targeted gene modification. The method uses four basic repeats with RVDs of NI, NG, NN and HD to generate 48 ready-to-use modules of single-TAL repeat units that can be used to assemble up to 23 repeat units in any predetermined order. The modular assembly method is simple, fast and inexpensive and can be performed in most academic or industrial molecular biology laboratories. Remarkably, all 10 dTALENs demonstrated efficient gene knockout and/or gene replacement in tests with three different chromosomal genes in yeast (*Saccharomyces cerevisiae*). The high success rate and facile synthesis of potent dTALENs against a variety of chromosomal targets further establishes dTALENs as an emerging and viable technology for precise gene modification in living cells. As previously practiced, all materials

described in the present study will be provided to other laboratories upon request.

MATERIALS AND METHODS

Yeast strains and growth conditions

Yeast strains YPH499, YPH500 and RFY231 as previously described (18,21) were grown in nutrient medium YPD or synthetic complete medium (SC) lacking the appropriate nutrients. 5-fluoroorotic acid (5-FOA) was used at 0.1% in SC medium as described (22). α -amino adipate (α -AA) was used at 0.2% in SC medium lacking the normal nitrogen source, but containing a small amount of lysine (30 mg/l) (23). The adenine limited medium is SC medium containing a limited amount of adenine (10 mg/l) and lacking leucine and histidine (24).

dTALEN constructs

Four repeats, each encoding the RVD of NI, NG, HD or NN from AvrXa7, were used as the ‘core’ repeats. [More recently, NK has been substituted for NN in the recognition of G nucleotides based on the observations of Miller *et al.* (20).] Using a combination of 12 forward and 11 reverse primers (Supplementary Table S1), 12 repeat sets were constructed from the 4 ‘core’ repeats. For construction of the first 8–mer repeat array, combinations of the PCR primers were: TAL-Sph-F and TALCgct-R, TALCgct-F and TALcctt-R, TALcctt-F and TALctct-R, TALctct-F and TALcgtt-R, TALcgtt-F and TALcttt-R, TALcttt-F and TALcatt-R, TALcatt-F and TALccct-R and TALccct-F and TAL/Pst-R for repeat Set 1–8, respectively. For ligation of the second and the third 8–mer repeat arrays, two additional repeat sets each corresponding to Set 1 were constructed by using primers TAL/Pst-F&TALcctt-R and TAL/Bsr-F&TALcctt-R, respectively; another two repeat sets corresponding to Set 8 were also constructed by using primers TALccct-R and TAL/Bsr-R, and Tailcatt-F&TAL-Sph-R, respectively. In total 12 repeat sets were generated and then individually digested with BsmBI. Based on the base entity and order of the preselected target DNA sequence (e.g. any sequence combination of 22 bp long), one repeat from each of eight repeat sets was sequentially selected for one ligation reaction to construct the 8–mer repeat array. Each set of ligated DNA was directly cloned into pGEM-T for sequencing. Once confirmed, the first 8–mer array was digested with SphI and PstI, the second 8–mers with PstI and BsrGI, and the third array with BsrGI and SalI. The three purified DNA fragments were ligated into pSK/AvrXa7 (18) that was digested with SphI and SalI, resulting in pSK/dTALE plasmids. The repeat regions of individual dTALENs were used to replace the TALE repeat domain of AvrXa7 in pCP3M-AvrXa7-FN or pCP4M-AvrXa7-FN (18), resulting in chimeric genes encoding the fusion of the full-length dTALE and the C-terminal FokI homodimeric cleavage domain. The expression level of these nuclease genes should be moderate due to the low copy number (about one copy per cell) of the centromeric plasmids pCP3M and pCP4M and the strong promoter from the translation elongation factor 1 α gene (25).

Yeast SSA assay for dTALEN activity

The individual target sites were constructed into pCP5 (18) in a tail-to-tail (3' to 3') orientation with AvrXa7 EBE. The oligonucleotide sequences for all TALEN EBEs were provided in [Supplementary Table S2](#). AvrXa7 EBE ($\times 7$ -EBE-F&R) was cloned into the PstI and SpeI sites of pCP5. The EcoRI site immediately upstream of SpeI site enabled all other TALEN EBEs to be cloned individually between EcoRI and SpeI sites, resulting in the reporter plasmids for individual TALENs in combination with AvrXa7-FN. The assay for each individual TALEN paired with AvrXa7-FN was performed in a manner similar to that described (18). The assays were performed in triplicate.

Targeted gene disruption of *URA3*, *LYS2* and *ADE2* in yeast

The *URA3* gene (*ura3-52*) that was insertionally inactivated by the transposon, Ty1, in YPH500 was restored to a functional *URA3*, resulting in strain YPH500a. Similarly, the target sequences for ZFNs (Zif268 and BCR-ABL) and for TALENs of AvrXa7 and PthXo1 were individually integrated into the *URA3* gene immediately downstream of the start codon and used to restore the *ura3-52* mutant, resulting in strain YPH500b and YPH500c, respectively. YPH500b was transformed with plasmids pCP3M/Zif268-FN and pCP4M/BCR-ABL-FN. DNA fragment for the ZF protein Zif268 was PCR amplified from pMal-Zif268 (kindly provided by David J. Segal) using primers Zif268-F and Zif268-R and cloned in frame with FokI cleavage domain in pCP3M. Construct pCP4M/BCR-ABL-FN was described previously (18). YPH500c was transformed with pCP3M/AvrXa7-FN and pCP4M/PthXo1-FN, two plasmids previously described (18). YPH500a was transformed with plasmids expressing the paired dTALENs U3a_L and U3a_R and U3b_L and U3b_R. The respective yeast strains were transformed with plasmids pCP3M and pCP4M as a negative control for each paired nucleases. The transformants were grown on SC medium lacking histidine and leucine for 5 days before plating on the SC medium containing 0.1% 5-FOA for selection of resistant colonies and in parallel on SC medium without 5-FOA to test for plating efficiency. Genomic DNA extracted from a number of 5-FOA-resistant colonies for each pair of nucleases was used for PCR amplification of the relevant regions. The PCR products were sequenced using the respective primers. Similarly, gene disruption of *LYS2* and *ADE2* was performed in yeast strain RFY231. See [Supplementary Data](#) for detailed information regarding the creation of these strains and gene disruption.

HR-based *URA3* gene replacement stimulated by dTALENs

Donor DNA constructs were made each with the ORF of *URA3* deleted (Δ ura3) and replaced by the *NPTII* expression cassette (Δ ura3::Kan). YPH500c was transformed with the donor construct and a plasmid expressing the TALEN pair AvrXa7-FN and PthXo1-FN, dTALENs

U3a-L and -R, U3b-L and -R, or with the plasmids lacking a nuclease gene as control. The transformed cells were grown in SC media lacking histidine and leucine for 5 days, then plated on SC medium supplemented with either 5-FOA (for Δ ura3) or YPD medium supplemented with 200 mg/l of G418 (for Δ ura3::Kan), the duplicated cells were in parallel plated on SC medium or YPD medium to determine plating efficiency. See [Supplementary Data](#) for additional details.

Cell growth assay

YPH500 cells were transformed with individual plasmids as indicated in the text. Three single colonies for each plasmid were picked and grown to a concentration of $OD_{600} = 1.0$. The cells were serially diluted and applied to appropriate solid medium for growth, which was observed daily for 5 days.

Solexa sequencing and data analysis

The genomic DNA from each of five yeast strains derived from YPH500, which is congenic with S288C (26), was extracted as described by Philippsen *et al.* (27). The DNA processing and sequencing were performed according to manufacturer's instruction for the Illumina/Solexa Genome Analyzer II at the Iowa State University DNA facility. The Illumina short reads for each strain were aligned to the yeast reference genome S288C to assemble each genome using the BWA software (28). The consensus sequences and polymorphisms among the five sequenced strains and S288C were delineated using SAMtools (29).

RESULTS

Tractability of yeast for testing gene modification by TALENs

We chose yeast, a classic eukaryotic model for studies of HR (30), as a platform to develop and test TALEN-based technologies for targeted gene modification based on the recent breakthroughs in the area of TALE research (13,14,17,18). First, we determined if the yeast chromosome, which has not been subjected to genome modification using any artificial nucleases, is tractable for TALE nuclease-based modification. For this experiment, the EBEs for AvrXa7 and PthXo1 were precisely integrated in-frame between the first and second codons of the yeast *URA3* gene on chromosome 5. In parallel, DNA target sequences for the known ZFNs Zif268 and BCR-ABL were inserted into the identical *URA3* gene site for comparison of activities conferred by these two types of nucleases (Figure 1 and [Supplementary Figure S2](#)). The resulting yeast strains were prototrophic in uracil-free medium and sensitive to 5-FOA, indicating that the chimeric *URA3* genes were functional and intact. Yeast strains bearing the chimeric *URA3* genes were transformed with plasmids expressing the paired TALENs or ZFNs, grown for 5 days, plated on medium containing 5-FOA to select cells with an inactivated *URA3* gene. Approximately 0.9% of cells (out of $\sim 10^6$ cells) expressing paired TALENs produced 5-FOA resistant colonies while

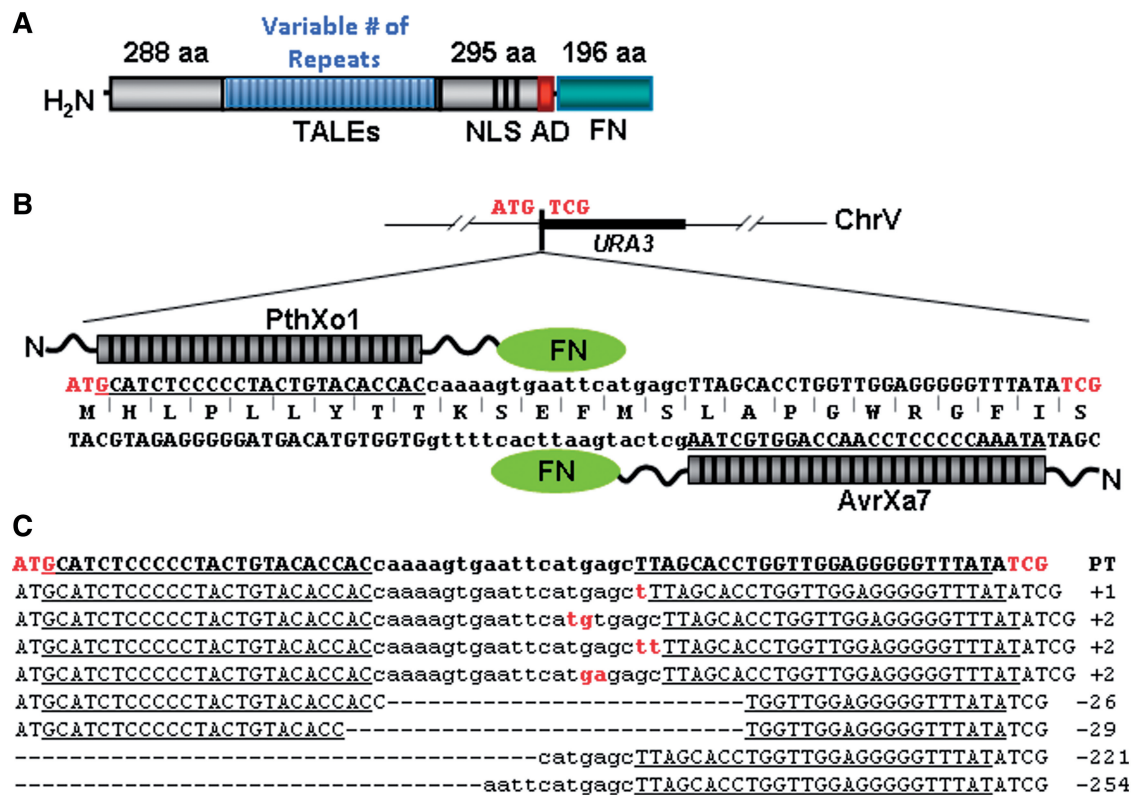


Figure 1. DNA alterations by natural TAL effector-derived TALENs at chromosomally integrated paired EBE target sites in the *URA3* gene. (A) Schematic of TALENs used in the study. Full-length TALEs were fused with the homodimeric cleavage domain of FokI (FN). The number of amino acids in the separate domains is shown above each region (NLS, nuclear localization motifs; AD, transcription activation domain). (B) Target sequences between the first and second codons of the yeast *URA3* gene targeted for cleavage by hybrid TALENs derived from AvrXa7 and PthXo1. (C) Alignment of genomic sequences of mutants and their parental strain at the TALEN target site. The number of nucleotides inserted (lowercase letters in red) or deleted (dashes) compared to parental sequences (PT) from each mutant colony is indicated on the right of each sequence.

~0.3% of cells expressing paired ZFNs yielded 5-FOA resistant colonies. In contrast, no 5-FOA resistant colonies were observed among ~10⁶ cells containing plasmids lacking a functional nuclease gene (a ratio of <0.0001%). Sequenced PCR products from the relevant alleles revealed all of the selected 5-FOA-resistant clones harbor mutations (insertions and/or deletions that caused frame shift in the *URA3* genes) at the nuclease target sites (Figure 1 and Supplementary Figure S2). These results established similar gene disruptions and comparable mutation rates elicited by ZFNs and TALENs targeted against genes in a native yeast chromosomal environment.

Modular dTALEN assembly

To fully realize the potential of TALENs, they must be custom-engineered to target any chromosomal DNA sequence of interest. By exploiting the repeat homology and the unique recognition sequence of the type IIS restriction enzyme BsmBI within each repeat of AvrXa7 or any TALE, we developed a method to assemble repeat domains in an exact predetermined order to recognize a specific DNA sequence in any gene of choice from any organism. Briefly, four AvrXa7 ‘core’ repeats whose coding RVDs each recognize one of four nucleotides (i.e. NI, NG, NN and HD, respectively, for A, T, G and C)

were used to construct independent modules (single repeats) whose 5'- and 3'-ends were designed to form a unique 4-bp overhang with single-base polymorphism after digestion with BsmBI. The BsmBI site is immediately downstream of codons 18 and 19 of each repeat and BsmBI cleaves these two triplets into a 5' overhang of 4-bp at each end. The 18th and 19th codons are GCG CTG and can be modified into eight variant triplets GC(A, T, G or C) (T or C) TG (Figure 2A). Therefore, combinations of eight or fewer such overhangs, one at each end of a single repeat, were created without altering the encoded amino acids. This allowed the ordered ligation of eight or fewer repeat modules in any predetermined sequence for construction of sub-arrays of repeats (Figure 2B). Each sub-array was cloned into the cloning vector pGEM-T and sequenced to confirm the correct number and order of the repeats. Similarly, multiple sub-arrays (two and three in this study) were further assembled to match the order of nucleotides at the preselected genomic site. The AvrXa7-FN nuclease (18) lacking its repeat domain was used as the scaffold for the assembled repeat domains, resulting in a finished dTALENs (Figure 2C).

To test the feasibility of our approach, we selected five distinct dual target sites (two in *URA3*, two in *LYS2* and

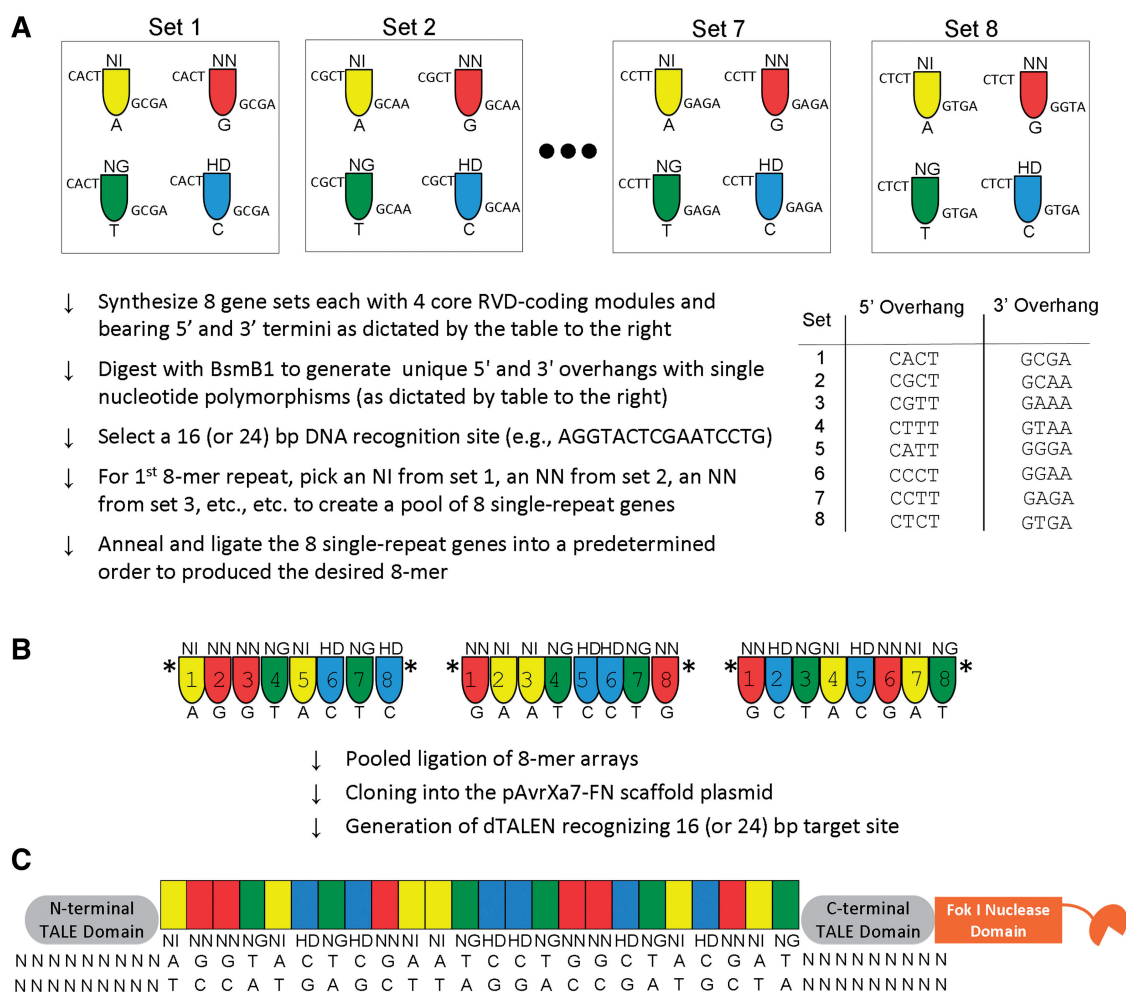


Figure 2. Design and modular construction of dTALENs. (A) TALEN repeat gene sets for modular construction of multi-repeat TALEN genes. Each set contains four single-repeat genes each encoding one of the four 'core' TALEN repeat modules containing NI, NG, NN and HD RVDs with binding specificity for A, T, G and C nucleotides, respectively. The core repeats in each set (boxed) contains a 5'- and 3'-termini unique to that particular set—as listed in the inset table. For construction of a TALEN repeat array recognizing a specific 8 nt DNA recognition sequence (e.g. AGGTACTC), an NI repeat gene is selected from Set 1, an NN repeat gene from Set 2, an NN repeat gene from Set 3, an NG repeat gene from Set 4, etc. The 5'- and 3'-terminal regions are designed such that BsmBI digestion results in generation of 5' and 3' 4-base overhangs. Because of the unique complementarity of 3' overhangs from Set 1 genes with the 5' overhang of genes from Set 2, the complementarity of the 3' overhangs of Set 2 with 5' overhangs from Set 3, etc., the annealing and ligation of overhangs results in one, and only one, ordered alignment of the eight repeat genes that, when translated, will specifically recognize and bind (in the present example) the AGGTACTC DNA recognition sequence. (B) Once two (or three) such blocks of repeats are constructed, they are combined in a similar ordered fashion to create a TALE repeat region that is capable of binding a specific 16 (or 24) bp target site. (C) The assembled TALE repeat domains are cloned into the TALEN repeat-deficient pAvrXa7-FN scaffold to create a candidate dTALEN.

one in *ADE2*) based on the criteria: (i) 'T' preceding each target sequence; (ii) avoiding G-rich blocks; and (iii) 17–20 bp spacer between the two inverted EBE target sites. Accordingly, 10 dTALENs were synthesized for gene targeting based on the preselected DNA coding sequences of the three yeast genes (Figure 3A). All 10 dTALENs were expressed in yeast to levels comparable to those of hybrid nucleases made from PthXo1 or AvrXa7, except U3b-L which had somewhat lower expression (Supplementary Figure S3).

To test the function of newly synthesized dTALENs and to reveal their relative DNA cleavage activity, we modified a transient and plasmid-borne single-strand annealing (SSA) assay (31,32) as a facile analytical tool.

This method uses a yeast plasmid carrying a *lacZ* gene divided into upstream and downstream portions by insertion of two opposing EBEs, one recognized by a proven AvrXa7 TALEN (18) and the other EBE recognized by a candidate dTALEN. The separated *lacZ* fragments share at the 3'-end of the upstream fragment and the 5'-end of the downstream fragment a 125 bp segment of identical *lacZ* sequence. If a functional dTALEN is co-expressed in yeast cells with the AvrXa7 TALEN along with the target *lacZ* gene, it will bind to its target EBE sequence adjacent to the AvrXa7 TALEN and, thereby, create a DSB (Figure 3B). The duplicated *lacZ* sequences are thus available for HR and restoration of an intact *lacZ* gene. The amount of β -galactosidase activity produced by

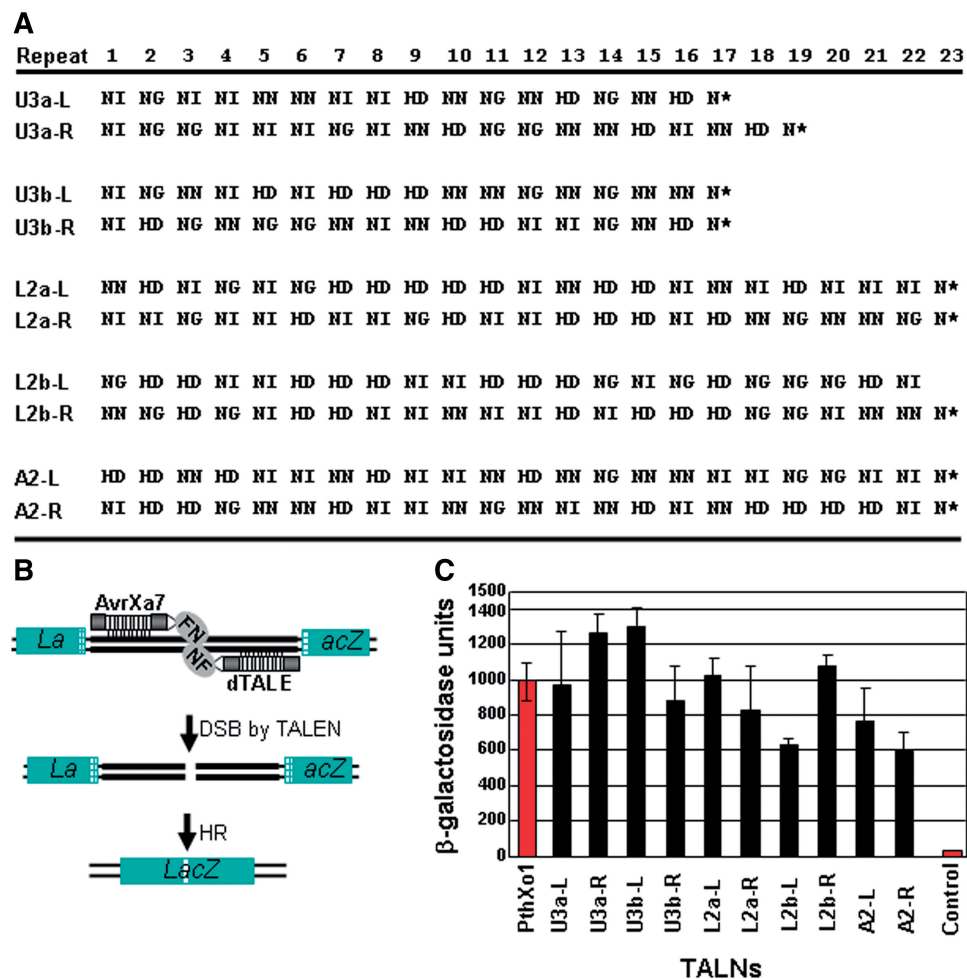


Figure 3. Yeast SSA assay of modularly assembled dTALENs targeting three endogenous yeast genes. (A) RVD sequences within repeat modules of 10 custom-synthesized dTALENs. N* represents dTALEN repeat modules with the 13th amino acid missing. (B) Schematic of the yeast SSA assay for measuring dTALEN activity based on plasmid-borne HR. Individual candidate dTALENs are assayed in combination with AvrXa7-FN for their ability to stimulate the recombination between the duplicated regions of *LacZ* gene (hatched boxes), leading to formation of a functional *lacZ* gene. (C) Activities of individual dTALENs in creating DSBs as detected in a β -galactosidase assay. Control denotes the β -galactosidase activity (<5 U) of yeast cells lacking a functional TALEN gene. Error bars denote SD; $n = 3$.

the transformed yeast cells thus provides a measure of the amount of DNA cleavage supported by the candidate dTALEN. We initially tested this assay by first pairing the proven AvrXa7-FN TALEN with another proven TALEN, PthXo1-FN (18). The activity of this TALEN pair provides a standard against which the activity of any candidate dTALEN can be judged (Figure 3C). The 10 newly produced dTALENs, each designed to recognize a specific 17–23 base DNA sequence in different yeast genes, were all found in the yeast SSA tests to function nearly as well or better than the standard AvrXa7/PthXo1 TALEN pair (Figure 3C).

Efficient gene modification by dTALEN-induced NHEJ and HR

As a final evaluation of the function of the five pairs of dTALENs, we tested their ability to elicit site-specific DNA alterations at the preselected target sites in the *URA3*, *LYS2* and *ADE2* genes, which all have easily

scored knock-out phenotypes. Yeast cells were transformed with individual pairs of dTALEN-expressing plasmids and grown for 5 days on SC medium to allow accumulation and activity of the heteromeric dTALEN pair. Two yeast cultures were transformed separately with one or the other pair of dTALENs targeting the *URA3* gene and plated for 5-FOA selection of cells lacking a functional *URA3* gene. Likewise, two yeast cultures were transformed separately with one or the other pair of dTALENs targeting the *LYS2* gene and plated for α -amino adipate (α -AA) selection of cells with *LYS2* gene mutations. *URA3* and *LYS2* mutants were obtained at a rate of $\sim 10^{-4}$ to 10^{-2} mutants/total cells (Figure 4A). Yeast cells transformed with the dTALENs pair targeting the *ADE2* gene were plated on medium containing limiting adenine concentrations that result in the formation of pink colonies if a functional *ADE2* gene is not present. Pink colonies appeared with a frequency of 0.15% (Figure 4A). In contrast, $\sim 10^6$ yeast cells carrying

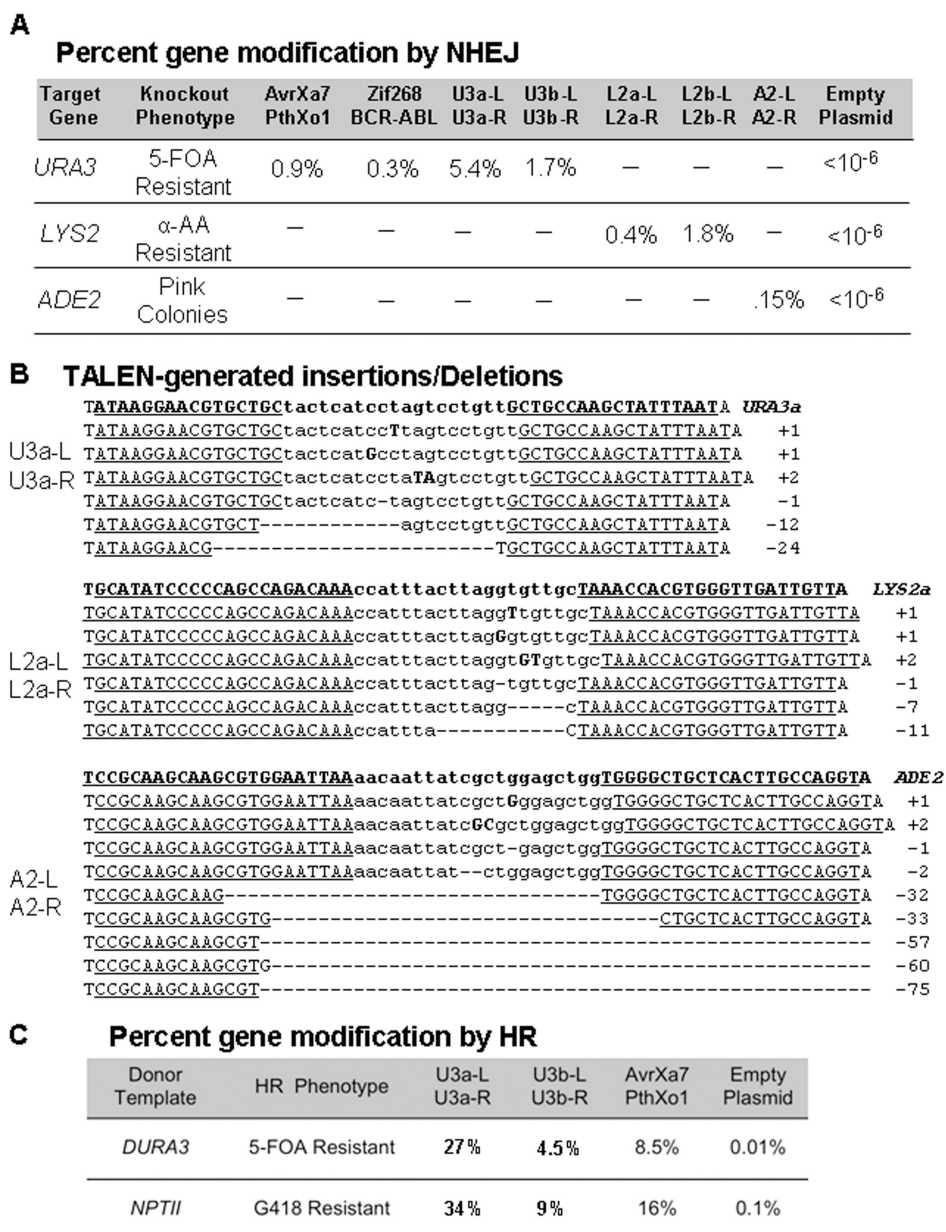


Figure 4. dTALEN-induced gene modifications by NHEJ and HR. (A) The frequency of gene disruption induced by five sets of paired dTALENs at five specific gene target sites as measured by the numbers of colonies with the indicated mutant phenotypes. '—' denotes not applicable. (B) TALEN-induced insertion/deletion mutations at three of five gene loci tested. (Mutations at the other two target sites are provided in Supplementary Figure S4.) Genomic sequences from each mutant clone at the relevant loci are aligned with the respective wild-type sequences. dTALEN target sites are underlined. The number of nucleotides inserted (bold uppercase letters) or deleted (dashes) is indicated to the right of each sequence. (C) TALEN-induced HR as measured by the percentage of yeast colonies displaying the indicated phenotypes.

plasmids lacking a functional dTALEN gene pair yielded no colonies resistant to 5-FOA or α-AA or with a pink color. Sequence analysis of PCR-amplified genomic DNA from the relevant target sites in several putative *URA3*, *LYS2* and *ADE2*-gene knock-out mutants revealed that all alleles harboured mutations at the dTALEN target sites as expected. A high proportion of the mutated loci contained deletions in a range from 1 to 75 bp (Figure 4B and Supplementary Figure S4).

One experimentally and practically important virtue of DSBs caused by agents such as ZFNs (3) is that they

increase rates of recombination between the DNA sequences within a broken gene and homologous endogenous or exogenously-supplied DNA sequences, which enables powerful gene replacement research opportunities. To determine if dTALEN-mediated DSBs stimulate HR, we targeted the *URA3* gene for breakage with artificial dTALEN pairs in the presence of two different exogenously supplied DNA fragments, one containing a *URA3* gene interrupted by a neomycin phosphotransferase II (*NPTII*) expression cassette and the other a DNA fragment with the *URA3* ORF deleted. Both fragments

contained at their 5' ends a 0.5 kb segment homologous to the 5' end of the *URA3* gene and, at their 3' ends, a 0.2 kb segment from the 3' end of the *URA3* gene. Yeast cells of YPH500c, a strain containing the functional *URA3* with the integrated EBEs for AvrXa7 and PthXo1, were transformed with one set of three different paired TALENs (i.e. U3a-L and U3a-R, U3b-L and U3b-R, and AvrXa7-FN & PthXo1-FN) and one of the two modified *URA3* gene constructs. The transformants were plated after 5 days of incubation on either selective medium containing 5-FOA (deleted *URA3* construct) or medium containing the neomycin-like antibiotic, G418 (*NPTII* interrupted *URA3*). Cells transformed with TALENs and the *URA3* ORF deletion construct (*DURA3*) showed frequencies of 5-FOA-resistant colonies in the range of 4.5–27%. The negative control [transformed with donor DNA and plasmids lacking a nuclease gene (Empty Plasmid)] yielded 5-FOA-resistant colonies at a rate of 0.01% (Figure 4C). The frequency of gene replacement for cells transformed with the *NPTII*-interrupted *URA3* gene construct (*NPTII*) was in the range of 9–34% with the negative control (Empty Plasmid) displaying only ~0.1% gene replacement activity (Figure 4C). Overall, the enhancement of TALEN-induced gene replacement was between 100- and 2700-fold higher than the control. The scale and consistency in the stimulation of HR by dTALENs suggests they have the potential to promote HR when used in eukaryotic cells that lack other sufficiently robust mechanisms to facilitate HR.

Genome-wide undesired mutations caused by TALENs

Some ZFNs have been found to be associated with toxicity in living cells (33–39). Whether such effect also exists for TALENs is unknown. To test the possibility, yeast cells were transformed with plasmids encoding six pairs of TALENs (one pair of native TALE-derived nucleases and five pairs of synthetic dTALENs targeting sequences in a size range between 17 and 27 bp) and one pair of known ZFNs [Zif268 and BCR-ABL with target sequences of 9 bp (33,40)]. The transformed cells were grown in SC medium lacking leucine and histidine. During 5 days of growth, TALEN-expressing yeast cells displayed no distinct phenotype in terms of cell viability and proliferation compared to the control, as did yeast strains with ectopic expression of paired ZFNs (Figure 5). Yeast cells expressing individually each of the eight introduced nuclease genes showed similar results (Supplementary Figure S6). The results indicate the lack of any apparently deleterious effects on the viability of yeast cells expressing the tested TALENs and ZFNs under our experimental conditions.

Undesired genetic mutations (genotoxicity) due to promiscuous cleavage also have been reported for ZFNs (33,40,41), but it is unknown whether TALENs also induce such genotoxicity. It is possible that such mutations occurred in our cell survival experiment but did not visibly affect the viability of the yeast cells. The haploid nature and relatively small size (~12 Mb) of the yeast genome in combination with the next-generation sequencing technology enabled us to investigate any

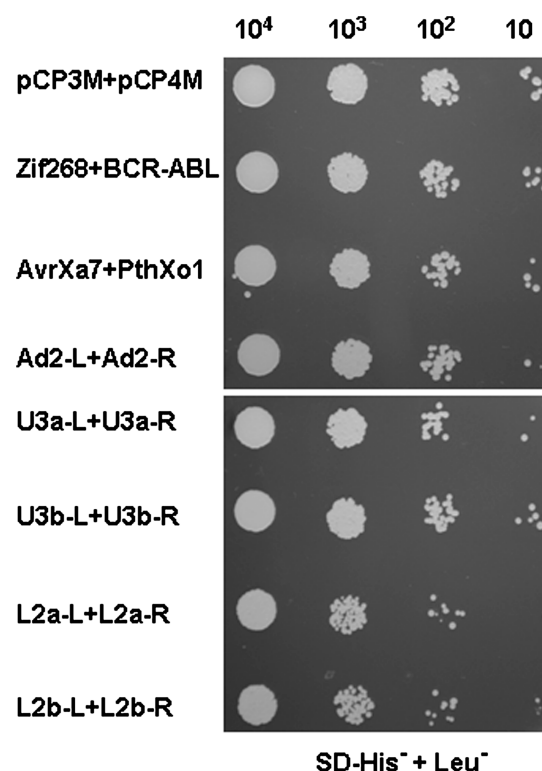


Figure 5. Growth of yeast cells expressing the paired TALEN and paired ZFN genes. Strains containing plasmids encoding the indicated nuclease genes (on the left side of each row) or lack thereof (pCP3M and pCP4M) were serially diluted (from 10^4 to 10 cells), applied as spots to SC medium plates lacking histidine (for pCP3M and its derived plasmids) and leucine (for pCP4M and its derived plasmids) and allowed to grow for 4 days.

potential genome-wide undesired effects of TALENs. Five strains, including the parental strain YPH500 and four mutants, were chosen to investigate the occurrence of unintended mutations in addition to the site-specific mutations at the *URA3* locus mediated by the respective nucleases. The four mutant strains included one that contained a deletion in the chimeric *URA3* gene at the integrated EBE site targeted by the paired nucleases of natural TAL effectors AvrXa7 and PthXo1, one that contained an insertion in the chimeric *URA3* gene with integrated target sequences for the ZFNs Zif268 and BCR-ABL, and the other two that each contained a deletion mutation in the wild-type *URA3* genes induced by the paired dTALENs U3a-L&-R and U3b-L&-R, respectively. The five strains were sequenced using Illumina/Solexa Genome Analyzer II, and their genomes were assembled with coverage depths [(number of reads × average read length) / (size of genome)] in a range from 135 to 170 × (sequences available upon request). The assembled genomic sequence of each mutant strain was first screened for possible mutations at the sites that matched or loosely matched (defined as at least two-thirds of match as the cut-off) either intended sub-dual target sequence for the respective nucleases (Supplementary Figure S5). No mutations were found at those locations other than the

site-specific mutations mediated by the respective pairs of TALENs and ZFNs in each genome (data not shown). However, alignment of the genomes of the individual nuclease treated strains with that of parental strain revealed a few mutations in each strain (in a range from 3 to 5 mutations) (Supplementary Table S4). These mutations were almost all nucleotide substitutions instead of predominant deletions/insertions at the nuclease target sites and they, therefore, were highly likely to be simultaneous mutations which might or might not be associated with the nuclease treatment.

DISCUSSION

We have developed a simple, cost-effective method to assemble functional TALE DNA binding domains, combine them with a nuclease module and confirm their activity against specific gene targets *in vivo*. Using this method, dTALENs were readily produced and rapidly validated using a facile yeast SSA assay. Importantly, these dTALENs were then demonstrated to function as pairs to mediate efficient gene disruption by NHEJ and gene replacement by HR at specific yeast chromosomal loci.

The DNA binding domain encoded by a TALE gene is modular by nature, having 13–33 RDV units that bear a simple code for target site recognition (13,14). However, the high-repeat homology imposes technical difficulty when using PCR-based *de novo* gene synthesis methods (42) for construction of the lengthy repeat arrays required for high-level DNA specificity. To simplify gene synthesis and reduce dTALEN production costs, a modular assembly technique was developed. TALE repeat domains have been assembled from the single-RVD coding repeats to bind DNA targets predicted by the cipher; but they were assembled in a random way, not in a predetermined order and not based on the target DNA sequence (13). Our method involves creation of 48 ready-to-use modules and their assembly into repeat arrays in a controlled order based on any user chosen target DNA sequences (Figure 2). The ready-to-use modules can also be adapted for high-throughput dTALEN synthesis. By manipulating the combination of 5' and 3' unique overhangs of the modular sets, dTALENs with varying number of repeat units (up to 23 bp in this study) can be designed to better meet the criteria in choosing the dual target sites. Our facile modular TALEN assembly method contrasts with another recently reported method that involves several rounds of PCR amplifications and ligations to assemble individual repeats into 12 repeat TAL effectors for gene activation in human cells (16).

A modified plasmid-based yeast SSA *lacZ* assay was employed to determine activity of the candidate dTALENs before use for final gene targeting. This test system is based on assays developed to initially validate engineered meganucleases and ZFNs (31,32) and was used previously in a preliminary form for TALEN testing (17,18). The original yeast SSA assay relied on inverted repeat DNA targets and engineered homodimeric nucleases (32). For our modified assay, one candidate dTALEN was

paired with the proven TALEN AvrXa7-FN (18) to target two adjacent EBEs, one for recognition by the candidate dTALEN and the other for binding of AvrXa7-FN. The paired TALENs derived from the natural TALEs AvrXa7 and PthXo1 were used as a standard. Another advantage of the modified SSA assay is that one EBE and the spacer remain constant so only the second EBE has to be replaced to test a different dTALEN. When testing many dTALENs, this represents a significant ($\sim 5\times$) cost saving due to using much shorter oligonucleotides to create a single (versus double) TALEN site while allowing rapid evaluation of the activity of each individual candidate dTALEN. Moreover, construction of 3' to 3' EBE sites containing non-identical DNA sequences avoids the technical challenges associated with cloning and stably maintaining inverted EBE sites containing identical TALEN recognition sequences. From a broader perspective, this plasmid-based transient assay should be readily adapted as a facile screening tool in animal and plant cell culture systems.

The 10 dTALENs engineered for this study were designed in pairs to target the native sequences of the yeast genes, *URA3*, *LYS2* and *ADE2*. Each gene has a selectable or easily scored phenotype, so any dTALEN activity should be readily apparent. Expression of dTALEN pairs under appropriate selections indicated that all five dTALEN pairs were functional in creating phenotypic mutant strains whose site-specific DNA alterations were all confirmed by genotyping (Figure 4 and Supplementary Figure S4). The activity of the five dTALEN pairs was comparable to the standard TALEN pair, AvrXa7 and PthXo1. Two pairs of dTALENs also were tested for their ability to mediate HR by targeting *URA3* for gene replacements and compared against HR stimulated by the standard TALEN pair. The results indicated that dTALENs and standard TALENs increase HR at a comparable rate (Figure 4). The low rates of gene disruption *via* NHEJ in present study are probably attributable to the cryptic NHEJ repair pathway in yeast. The NHEJ repair of DSBs in yeast is mostly accurate. For example, repair to DSBs with 4-base overhangs led to an error frequency of only $\sim 1\%$ (43,44). Only cells with mutagenic alterations in the intentionally targeted genes might be recovered (for *URA3* and *LYS2*) or detected (for *ADE2*) by using our procedures. Thus, while the number of dTALENs examined was somewhat limited, these experiments establish that dTALENs are active *in vivo* in promoting NHEJ- and HR-mediated gene modifications at endogenous loci in an intact, free-living, eukaryotic organism and, thus, verify the power of TALENs for targeted genome editing in eukaryotes beyond the TALEN-mediated gene disruption previously demonstrated in cultured human cells (20).

Taken together, the results of this study indicate that the modular assembly technique is valid, that the yeast SSA assay is a reliable and facile indicator of TALEN activity and that functional dTALENs can target a diverse range of genomic loci. Our study also establishes that yeast, as an intact eukaryotic organism, is a reliable platform to develop and potentially refine engineered nuclease-based technology for targeted genome editing.

The designer TALEN technology described here may overcome a number of critical hurdles (e.g. high cost, limited plasticity and high rates of failure) faced in designing and producing other types of nucleases with DNA targeting capabilities (7,8). The opportunity to build dTALENs with quite long DNA recognition domains bodes well for developing TALENs with exceptionally high accuracy in targeting any gene in any organism, including eukaryotes with highly complex genomes. Such technology has significant potential in experimental biology and medicine and in the development of products with value for human and animal health, agriculture and in a wide range of life sciences industries.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Dr Dan Voytas for providing the components of the yeast SSA system and Dr Keith Joung and Dr David Segal for providing materials to make constructs of ZFNs BCR-ABL and Zif268, respectively.

FUNDING

The National Science Foundation (Grant 0820831 to B.Y., MCB-0952323 to M.H.S., MCB-0952533 and EPSCoR-1004094 to D.P.W.); Department of Energy (DE-EE0003373 to D.P.W.); Department of Energy Advanced Research Projects Agency-Energy Program (DEAR0000010 to M.H.S.); Center for Integrated Animal Genomics at Iowa State University (to S.C. and B.Y.). Funding for open access charge: The National Science Foundation award (0820831).

Conflict of interest statement. None declared.

REFERENCES

- Cathomen,T. and Joung,J.K. (2008) Zinc-finger nucleases: The next generation emerges. *Mol. Ther.*, **16**, 1200–1207.
- Carroll,D. (2008) Progress and prospects: Zinc-finger nucleases as gene therapy agents. *Gene Ther.*, **15**, 1463–1468.
- Urnov,F.D., Rebar,E.J., Holmes,M.C., Zhang,H.S. and Gregory,P.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.*, **11**, 636–646.
- Jasin,M. (1996) Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet.*, **12**, 224–228.
- Vasquez,K.M., Marburger,K., Intody,Z. and Wilson,J.H. (2001) Manipulating the mammalian genome by homologous recombination. *Proc. Natl Acad. Sci. USA*, **98**, 8403–8410.
- Bibikova,M., Beumer,K., Trautman,J.K. and Carroll,D. (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science*, **300**, 764.
- Ramirez,C.L., Foley,J.E., Wright,D.A., Muller-Lerch,F., Rahman,S.H., Cornu,T.I., Winfrey,R.J., Sander,J.D., Fu,F., Townsend,J.A. *et al.* (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat. Methods*, **5**, 374–375.
- Kim,J.S., Lee,H.J. and Carroll,D. (2010) Genome editing with modularly assembled zinc-finger nucleases. *Nat. Methods*, **7**, 91–92.
- Kim,Y.G., Cha,J. and Chandrasegaran,S. (1996) Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proc. Natl Acad. Sci. USA*, **93**, 1156–1160.
- Boch,J. and Bonas,U. (2010) Xanthomonas AvrBs3 family-type III effectors: Discovery and function. *Annu. Rev. Phytopathol.*, **48**, 419–436.
- White,F.F., Potnis,N., Jones,J.B. and Koebnik,R. (2009) The type III effectors of xanthomonas. *Mol. Plant Pathol.*, **10**, 749–766.
- Bogdanove,A.J., Schornack,S. and Lahaye,T. (2010) TAL effectors: Finding plant genes for disease and defense. *Curr. Opin. Plant Biol.*, **13**, 394–401.
- Boch,J., Scholze,H., Schornack,S., Landgraf,A., Hahn,S., Kay,S., Lahaye,T., Nickstadt,A. and Bonas,U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509–1512.
- Moscou,M.J. and Bogdanove,A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 1501.
- Moritz,R., Romer,P., Boch,J. and Lahaye,T. (2010) 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.
- Zhang,F., Cong,L., Lodato,S., Kosuri,S., Church,G.M. and Arlotta,P. (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat. Biotechnol.*, **29**, 149–153.
- Christian,M., Cermak,T., Doyle,E.L., Schmidt,C., Zhang,F., Hummel,A., Bogdanove,A.J. and Voytas,D.F. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, **186**, 757–761.
- Li,T., Huang,S., Jiang,W.Z., Wright,D., Spalding,M.H., Weeks,D.P. and Yang,B. (2010) TAL nucleases (TALENs): Hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.*, **39**, 359–372.
- Mahfouz,M.M., Li,L., Shamimuzzaman,M., Wibowo,A., Fang,X. and Zhu,J.K. (2011) 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.
- Miller,J.C., Tan,S., Qiao,G., Barlow,K.A., Wang,J., Xia,D.F., Meng,X., Paschon,D.E., Leung,E., Hinkley,S.J. *et al.* (2011) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.*, **29**, 143–148.
- Kolonin,M.G. and Finley,R.L. Jr (1998) Targeting cyclin-dependent kinases in drosophila with peptide aptamers. *Proc. Natl Acad. Sci. USA*, **95**, 14266–14271.
- Boeke,J.D., Trueheart,J., Natsoulis,G. and Fink,G.R. (1987) 5-fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol.*, **154**, 164–175.
- Chattoo,B.B., Sherman,F., Azubalis,D.A., Fjellstedt,T.A., Mehnert,D. and Ogur,M. (1979) Selection of lys2 mutants of the yeast *Saccharomyces cerevisiae* by the utilization of alpha-aminoadipate. *Genetics*, **93**, 51–65.
- Ugolini,S. and Bruschi,C.V. (1996) The red/white colony color assay in the yeast *saccharomyces cerevisiae*: Epistatic growth advantage of white ade8-18, ade2 cells over red ade2 cells. *Curr. Genet.*, **30**, 485–492.
- Mumberg,D., Muller,R. and Funk,M. (1995) Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene*, **156**, 119–122.
- Johnston,M. and Davis,R.W. (1984) Sequences that regulate the divergent GAL1-GAL10 promoter in *saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **4**, 1440–1448.
- Philippsen,P., Stotz,A. and Scherf,C. (1991) DNA of *saccharomyces cerevisiae*. *Methods Enzymol.*, **194**, 169–182.
- Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, **25**, 1754–1760.
- Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and 1000 Genome Project Data Processing Subgroup. (2009) The sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
- Aylon,Y. and Kupiec,M. (2004) DSB repair: The yeast paradigm. *DNA Repair*, **3**, 797–815.

31. Epinat,J.C., Arnould,S., Chames,P., Rochaix,P., Desfontaines,D., Puzin,C., Patin,A., Zanghellini,A., Paques,F. and Lacroix,E. (2003) A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res.*, **31**, 2952–2962.
32. Townsend,J.A., Wright,D.A., Winfrey,R.J., Fu,F., Maeder,M.L., Joung,J.K. and Voytas,D.F. (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature*, **459**, 442–445.
33. Porteus,M.H. and Baltimore,D. (2003) Chimeric nucleases stimulate gene targeting in human cells. *Science*, **300**, 763.
34. Alwin,S., Gere,M.B., Guhl,E., Effertz,K., Barbas,C.F. III, Segal,D.J., Weitzman,M.D. and Cathomen,T. (2005) Custom zinc-finger nucleases for use in human cells. *Mol. Ther.*, **12**, 610–617.
35. Porteus,M.H. (2006) Mammalian gene targeting with designed zinc finger nucleases. *Mol. Ther.*, **13**, 438–446.
36. Miller,J.C., Holmes,M.C., Wang,J., Guschin,D.Y., Lee,Y.L., Rupniewski,I., Beausejour,C.M., Waite,A.J., Wang,N.S., Kim,K.A. *et al.* (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.*, **25**, 778–785.
37. Szczepek,M., Brondani,V., Buchel,J., Serrano,L., Segal,D.J. and Cathomen,T. (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat. Biotechnol.*, **25**, 786–793.
38. Cornu,T.I., Thibodeau-Beganny,S., Guhl,E., Alwin,S., Eichinger,M., Joung,J.K. and Cathomen,T. (2008) DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Mol. Ther.*, **16**, 352–358.
39. Bibikova,M., Golic,M., Golic,K.G. and Carroll,D. (2002) Targeted chromosomal cleavage and mutagenesis in drosophila using zinc-finger nucleases. *Genetics*, **161**, 1169–1175.
40. Wright,D.A., Thibodeau-Beganny,S., Sander,J.D., Winfrey,R.J., Hirsh,A.S., Eichinger,M., Fu,F., Porteus,M.H., Dobbs,D., Voytas,D.F. *et al.* (2006) Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat. Protoc.*, **1**, 1637–1652.
41. Beumer,K., Bhattacharyya,G., Bibikova,M., Trautman,J.K. and Carroll,D. (2006) Efficient gene targeting in drosophila with zinc-finger nucleases. *Genetics*, **172**, 2391–2403.
42. Hoover,D.M. and Lubkowski,J. (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.*, **30**, e43.
43. Wilson,T.E. and Lieber,M.R. (1999) Efficient processing of DNA ends during yeast nonhomologous end joining. Evidence for a DNA polymerase beta (Pol4)-dependent pathway. *J. Biol. Chem.*, **274**, 23599–23609.
44. Moore,J.K. and Haber,J.E. (1996) Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **16**, 2164–2173.