

2012

# TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction

Erin L. Doyle

*Iowa State University*, erinlynddoyle@gmail.com

Nicholas J. Booher

*Iowa State University*

Daniel S. Standage

*Iowa State University*, standage@iastate.edu

Daniel F. Voytas

*University of Minnesota*

Volker P. Brendel

*Indiana University*

Follow this and additional works at: [http://lib.dr.iastate.edu/ent\\_pubs](http://lib.dr.iastate.edu/ent_pubs)



Part of the [Bacteriology Commons](#), [Bioinformatics Commons](#), [Computational Biology Commons](#), [Entomology Commons](#), [Genetics Commons](#), and the [Plant Pathology Commons](#)

The complete bibliographic information for this item can be found at [http://lib.dr.iastate.edu/ent\\_pubs/30](http://lib.dr.iastate.edu/ent_pubs/30). For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

---

# TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction

## Abstract

Transcription activator-like (TAL) effectors are repeat-containing proteins used by plant pathogenic bacteria to manipulate host gene expression. Repeats are polymorphic and individually specify single nucleotides in the DNA target, with some degeneracy. A TAL effector-nucleotide binding code that links repeat type to specified nucleotide enables prediction of genomic binding sites for TAL effectors and customization of TAL effectors for use in DNA targeting, in particular as custom transcription factors for engineered gene regulation and as site-specific nucleases for genome editing. We have developed a suite of web-based tools called TAL Effector-Nucleotide Targeter 2.0 (TALE-NT 2.0; <https://boglab.plp.iastate.edu/>) that enables design of custom TAL effector repeat arrays for desired targets and prediction of TAL effector binding sites, ranked by likelihood, in a genome, promoterome or other sequence of interest. Search parameters can be set by the user to work with any TAL effector or TAL effector nuclease architecture. Applications range from designing highly specific DNA targeting tools and identifying potential off-target sites to predicting effector targets important in plant disease.

## Keywords

Plant Pathology and Microbiology, Genetics Development and Cell Biology

## Disciplines

Bacteriology | Bioinformatics | Computational Biology | Entomology | Genetics | Plant Pathology

## Comments

This article is from *Nucleic Acids Research* 40 (2012): W117–W122, doi:[10.1093/nar/gks608](https://doi.org/10.1093/nar/gks608).

## Rights

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Authors

Erin L. Doyle, Nicholas J. Booher, Daniel S. Standage, Daniel F. Voytas, Volker P. Brendel, John K. VanDyk, and Adam J. Bogdanove

# TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction

Erin L. Doyle<sup>1</sup>, Nicholas J. Booher<sup>2</sup>, Daniel S. Standage<sup>3</sup>, Daniel F. Voytas<sup>4</sup>,  
Volker P. Brendel<sup>5,6</sup>, John K. VanDyk<sup>2</sup> and Adam J. Bogdanove<sup>1,\*</sup>

<sup>1</sup>Department of Plant Pathology and Microbiology, Iowa State University, <sup>2</sup>Department of Entomology, Iowa State University, <sup>3</sup>Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA 50011, <sup>4</sup>Department of Genetics, Cell Biology and Development and Center for Genome Engineering, University of Minnesota, Minneapolis, MN 55455, <sup>5</sup>Department of Biology, Indiana University, Bloomington and <sup>6</sup>School of Informatics and Computing, Indiana University, Bloomington, IN 47405, USA

Received February 26, 2012; Revised April 27, 2012; Accepted May 30, 2012

## ABSTRACT

Transcription activator-like (TAL) effectors are repeat-containing proteins used by plant pathogenic bacteria to manipulate host gene expression. Repeats are polymorphic and individually specify single nucleotides in the DNA target, with some degeneracy. A TAL effector-nucleotide binding code that links repeat type to specified nucleotide enables prediction of genomic binding sites for TAL effectors and customization of TAL effectors for use in DNA targeting, in particular as custom transcription factors for engineered gene regulation and as site-specific nucleases for genome editing. We have developed a suite of web-based tools called TAL Effector-Nucleotide Targeter 2.0 (TALE-NT 2.0; <https://boglab.plp.iastate.edu/>) that enables design of custom TAL effector repeat arrays for desired targets and prediction of TAL effector binding sites, ranked by likelihood, in a genome, promoterome or other sequence of interest. Search parameters can be set by the user to work with any TAL effector or TAL effector nuclease architecture. Applications range from designing highly specific DNA targeting tools and identifying potential off-target sites to predicting effector targets important in plant disease.

## INTRODUCTION

Transcription activator-like (TAL) effectors from the plant pathogenic bacterial genus *Xanthomonas* represent a class of DNA binding proteins that can be readily

engineered to target novel DNA sequences. During infection, TAL effectors are deployed by the bacteria to modulate host gene expression, with each effector directly binding an effector-specific DNA target (1,2). A central region in the protein, composed of a variable number of tandem, near-identical, 33–35 amino acid repeats, determines the target(s) of each TAL effector. Repeat-to-repeat variation occurs primarily at residues 12 and 13 (termed the repeat variable diresidue or RVD). The RVD sequence has been shown both computationally and experimentally to correspond directly to the DNA target site sequence; repeats with different RVDs recognize different nucleotides, in a code-like fashion, with some degeneracy (3,4). Custom TAL effectors can be targeted to novel DNA sequences by assembling an array of repeats that corresponds to the intended target site (5). Designing custom TAL effectors for DNA targeting has proved to be a much simpler and less labor-intensive process than the design of other customizable DNA binding proteins such as zinc fingers (6), and a variety of rapid construction methods for custom TAL effectors and TAL effector fusion proteins have recently been developed (7–12). Increasingly, therefore, TAL effectors and TAL effector-based fusion proteins have been adopted as tools for DNA targeting applications (6).

Site-specific DNA modification has been achieved using TAL effector-endonuclease fusion proteins (TAL effector nucleases or TALENs), which create targeted double-stranded breaks (DSBs) in DNA. TALEN architectures to date combine the central repeat region and some portion of the flanking parts of the TAL effector with the catalytic domain of FokI, which functions as a dimer. Thus, TALENs work in heteromeric pairs, with the monomers binding to opposing target sites oriented 5'–3' on opposite strands across a spacer, allowing the

\*To whom correspondence should be addressed. Tel: +1 515 294 3421; Fax: +1 515 294 3421; Email: [ajbog@iastate.edu](mailto:ajbog@iastate.edu)

C-terminal FokI domains to dimerize and create a DSB in that spacer (5,13–15). In eukaryotes, such breaks are repaired by either non-homologous end joining (NHEJ), which is prone to errors and useful for creating gene knockouts, or by homologous recombination (HR), which replaces the sequences surrounding the break with a supplied template. TALEN-mediated NHEJ and HR have been demonstrated in a variety of cell types and organisms (7,8,13,14,16–20). TAL effectors have also been used as custom transcriptional activators, most effectively in plants in their native form and in mammalian cells with the native acidic activation domain replaced by the VP16 activation domain (or tetrameric derivative VP64) from herpes simplex virus (10,21,22). The TAL effector targeting domain has also been fused to a repressor protein and used successfully for specific gene repression in plants (23).

Custom TAL effector and TALEN architectures using different proportions of the flanking regions on either side of the repeat region have been reported. In the case of TALENs, the optimum spacer length between the TALEN monomers depends on the architecture used (5,13,15,24). In addition, different methods assemble different numbers of repeats in an array, with some allowing a wide range (7–12). Therefore, tools for designing custom TAL effectors and TALENs should allow for a range or a prescribed number of repeats and for TALENs, various spacer lengths. Most TAL effector targets in nature are preceded by a T at the 5'-end (3), but at least one example of a TAL effector target preceded by a C has been identified (25), and some custom TAL effectors have been reported to be active on sites preceded by a C (13). The preference for T at this position is structurally specified by the region of the protein immediately N-terminal to the central repeat region (26,27). Thus, accounting for the '-1 position' nucleotide is important for target prediction and custom TAL effector design as well.

Because of the degeneracy in the TAL effector–DNA recognition code, off-targeting is a concern in the use of TAL effectors and TALENs. RVDs associate preferentially, but not exclusively, with specific nucleotides. TAL effector–target pairs in nature are observed to have up to 50% mismatches (positions at which the RVD is not aligned with its most frequently associated nucleotide) (3), and TAL effector-based custom transcription factors have been shown to activate transcription of off-target genes whose promoters contain sequences similar but not identical to the intended targets (21). If custom TAL effectors and TALENs are to be a widely adopted biotechnology, accurate prediction of potential off-target activity and design to minimize such activity will be crucial. Prediction of potential TAL effector binding sites that takes the degeneracy into account is also important for the identification of disease targets in plant hosts of *Xanthomonas* spp.

To aid researchers in adopting TALENs as gene editing tools, we converted a computational script we had created for TALEN design, called 'TALEN Targeter', to a web-based tool and posted it on a site we called TAL Effector-Nucleotide Targeter (TALE-NT 1.0). The tool designed TALEN pairs according to five design guidelines

based on positional and composition biases observed in known TAL effector–target pairs (7). Herein, we describe a new version of the web site, TALE-NT 2.0, that offers a suite of tools for TALEN and TAL effector design as well as target prediction. Since the publication of TALE-NT 1.0, it was shown that the design guidelines have little effect on TALEN efficiency (12). Therefore, for TALE-NT 2.0, we removed the guidelines from TALEN Targeter and instead allow users to search for TALENs targeting a specific base. We also updated TALEN Targeter to provide users with additional options for entering sequences, customizing their queries to accommodate different TALEN architectures and other preferences, searching for sites preceded by T or C and viewing output. We added two new tools: 'TAL Effector Targeter', for designing custom TAL effectors for single sites, and 'Target Finder', for predicting candidate TAL effector targets, ranked by likelihood, in the genomes or gene promoteromes of several model organisms or in a sequence supplied by the user. The new web site (<https://boglab.plp.iastate.edu/>) makes this suite of tools freely available to the research community. Applications range from designing highly specific DNA targeting reagents and identifying potential off-target sites to predicting effector targets important in plant disease.

## SOFTWARE AND ALGORITHMS

### Programming

TALENT 2.0 tools other than Target Finder are written in the Python programming language and use Biopython libraries for parsing input DNA sequences and other sequence operations (28). Target Finder is written in C and uses Kseq.h (<http://lh3lh3.users.sourceforge.net/parsefastq.shtml>) for sequence parsing.

### Tools for designing custom TALENs and TAL effectors

TALEN Targeter and TAL Effector Targeter design paired and single custom TAL effector repeat arrays, respectively, for targeting DNA sequences of interest. Both tools require one or more FASTA-formatted DNA sequences as input. TALEN Targeter identifies paired monomer (typically heteromeric) binding sites oriented 5'–3' on opposite strands of the DNA and separated by a spacer. TAL Effector Targeter identifies single TAL effector binding sites; users have the option of searching the reverse complement in addition to searching the DNA sequence as entered. In either tool the number of TAL effector repeats can vary across a user-specified range. TALEN Targeter allows users also to specify a range for the spacer length; all possible combinations of repeat numbers and spacer lengths are considered. Identified binding sites are converted into RVD sequences using the four most common RVD–nucleotide pairs (NI-A, HD-C, NN-G and NG-T). Target sites, RVD sequences and other information are returned to the user (the output for all tools is described more fully in the 'Web Interface' section). Arrays designed with TAL Effector Targeter are by default designed to meet the five guidelines developed based on TAL effectors observed in nature, as the effect of

these guidelines on TAL effector transcription factors has not been determined. Users may choose not to enforce one or more guidelines.

### A tool for identifying candidate TAL effector targets

Target Finder allows users to enter an RVD sequence and search for candidate targets in the genome or gene promoter of any of several model organisms or in one or more user-provided, FASTA-formatted DNA sequences. The tool may be used to identify and rank candidate plant genomic targets of TAL effectors important for disease or potential off-target binding sites of custom TAL effector proteins.

To predict and rank sites, the tool uses a simple scoring function developed by Moscou and Bogdanove (3) based on RVD–nucleotide association frequencies found in a set of known TAL effector–target pairs. We have used this scoring function to predict 21 previously unknown and subsequently experimentally verified plant targets for 14 TAL effectors (A. Cernadas, E. Doyle and A. Bogdanove, unpublished results). Briefly, for scoring, for each RVD in the set of TAL effector–target pairs, the frequency with which it pairs with each nucleotide was calculated and then converted to a weighted RVD–nucleotide association probability. For RVDs that are not observed in the set of pairs, each nucleotide association was given an equal probability. The score for a DNA/RVD sequence alignment is found by summing the negative logs of the appropriate RVD–nucleotide association probabilities, such that better alignments have lower scores. A detailed description of the scoring function, including weighted RVD–nucleotide association probabilities, is provided in the Supplementary Material.

Target Finder returns a list of the lowest scoring (best) sites in the queried DNA sequence below a cutoff (discussed in the Web Interface section). Candidate binding sites are not required to conform to any of the design guidelines except that they must be preceded by a T, or C if that option is selected. The tool by default searches for binding sites on both strands of the DNA sequence(s), but users may opt to search only the forward strand.

## WEB INTERFACE

### Design and general features

The TALE-NT 2.0 web site is powered by Drupal 7 on Red Hat Enterprise Linux 6. Job queuing is handled by Celery using Redis as a message broker. All features of the web site have been tested on common web browsers.

TALE-NT 2.0 makes the three tools for design of TAL effectors and prediction of TAL effector targets freely available to all users, with no log-in requirement. Upon submitting a job, users are taken to a bookmarkable page updating them on the status of their query or supplying a link that will take them to their results when the query has finished processing. Processing times are not excessive: a search of the entire rice genome for possible targets of an average length TAL effector (18 RVDs) took <3 min. Nevertheless, users have the option to enter an address

to receive email notification when their job finishes. Results for each tool are displayed in a sortable table by default. Users may also or instead download results as a tab-delimited text file. For Target Finder users can download either or both of two formats, standard or GFF3, described further below.

All tools allow users to design or search for TAL effector binding sites preceded by a 5' T only, C only, or T or C. In our hands, however, with either TALENS or TAL effectors, sites preceded by a C are significantly less active than those preceded by a T, so we suggest using T, which is the default selection.

### TALEN Targeter

The TALEN Targeter web interface allows users to design custom TALENs to target one or more DNA sequences of interest. Users enter their sequence(s) in a text box or upload a file containing the sequence(s). Allowable file size is up to 2 MB. Users may select from four common TALEN architectures (7,13,15,24) with pre-selected ranges for spacer size and numbers of repeats or enter their own ranges for these parameters. Users may choose to allow binding sites to be preceded by a T only, C only, or by T or C. In addition, users have the option to output just TALEN pairs targeting a specific base, a filtered list including up to one TALEN pair for each base or the complete list of all TALEN pairs targeting anywhere in the sequence. The targeted base is defined as the base in the center (or immediately to the left of the center) of the spacer. If users choose to return up to one TALEN pair per base, the pair with the smallest average number of RVDs and shortest spacer targeting a given base will be returned.

Each line of the output describes a pair of monomers that will function together as a TALEN. Information provided includes the name of the input sequence (in case more than one sequence was entered), the starting position and number of repeats for each monomer, the target sequence including the –1 nucleotide and RVD sequences corresponding to each monomer.

The output is ordered by sequence name, with TALENs for each sequence grouped by the start position of the first monomer. The order of TALENs in the output does not relate to how well they might function. No scoring or other prioritization method is used. Users should select TALENs closest to the site of desired cleavage that have the fewest or poorest predicted off-target sites determined using Target Finder and further analysis. This analysis should take into account not only the heterodimer but both of the possible homodimers as well. Users may also wish to choose TALENs that straddle a spacer with a diagnostic restriction endonuclease site to facilitate detection of NHEJ-mediated mutation of the spacer sequence (such restriction endonuclease sites are indicated in the output table) or in the case of HR, a TALEN that uniquely binds the sequence to be replaced, so that following replacement it does not cut again. Users might opt to introduce silent mismatches in the HR template to facilitate this.

### TAL Effector Targeter

TAL Effector Targeter allows users to design a single custom TAL effector repeat array to target a DNA sequence of interest. The array may be used for the design of a TAL effector or any TAL effector-based fusion protein that functions as a monomer. As with TALEN Targeter, users enter sequence(s) in a text box or upload a FASTA-formatted file. Additional text boxes allow users to specify a range for the number of repeats in each TAL effector (default is 15–30). Checkbox options allow users to turn off individual design guidelines. By default, the tool searches only the sequence(s) as entered; a checkbox option includes also the reverse complement sequence(s). Users may also choose to search only for sites preceded by T at the 5'-end, or to allow T and/or C.

Each line of the output describes a single custom TAL effector repeat array. Information returned includes the name of the input sequence (in case multiple sequences were entered), the strand and coordinates of the target, the target sequence including the –1 nucleotide and corresponding RVD sequences.

The output is ordered first by sequence name, with target sites for each sequence sorted by their position in the sequence. As with TALEN Targeter, arrays are not ranked or prioritized. Users should choose according to their own targeting criteria. If there are no target sites in a sequence or no sites in the desired region, users may be able to increase the number of sites by relaxing some of the design guidelines and/or changing the range for number of repeats. We recommend that users leave the percent composition rule enforced to better assure good overall affinity. Although data regarding relative affinities of different RVDs for their partner nucleotide(s) is lacking, based on the published structures (26,27), we predict that the highest affinity interactions are those that occur most frequently in nature. So, if a user chooses not to enforce the percent composition guideline, at a minimum, arrays should be selected that have overall higher numbers of HDs and NGs than NIs and NNs. Also, in most contexts tested, NN pairs with G or A, so we anticipate that minimizing NN content will maximize specificity.

### Target Finder

Target Finder uses the scoring function of Moscou and Bogdanove (3) to identify the best-scoring sites in a DNA sequence for a user-specified string of RVDs. Users input a TAL effector RVD sequence containing from 12 to 35 RVDs, each separated by a space, using single-letter amino acid abbreviations. All possible RVDs constructed using the standard 20 amino acids and "\*" to indicate a missing 13th amino acid are allowed. Users may select a genome or gene promoterome from a drop-down list or provide their own sequence of interest in a text box or as a file upload. Promoterome is defined as the collection of sequences 1000 bp upstream of annotated translational start sites in a genome. Available genomes and promoteromes include those of rice (*Oryza sativa*), *Arabidopsis thaliana*, human (*Homo sapiens*), fruit fly

(*Drosophila melanogaster*), mouse (*Mus musculus*), nematode (*Caenorhabditis elegans*) and zebra fish (*Danio rerio*). Genome sequences were obtained from Ensembl ([www.ensembl.org](http://www.ensembl.org)), except for the rice genome, which was downloaded from the MSU Rice Genome Annotation Project (<http://rice.plantbiology.msu.edu>, Version 6.1). Promoteromes were downloaded from the UCSC Genome Bioinformatics Site (<http://genome.ucsc.edu/index.html>), except for the rice promoterome, which was downloaded from the Rice Genome Annotation Project and the Arabidopsis promoterome, which was downloaded from The Arabidopsis Information Resource (TAIR; <http://www.arabidopsis.org/>). By default, the tool searches both strands, but users may choose to search only the forward strand. Users may choose to search for sites preceded by a 5' T only, C only or T and C.

Target Finder returns all targets scoring under a threshold cutoff defined as a ratio of observed score to best possible score of 3.0 or less. The best possible score is the score for the array aligned to its code-specified, perfect match DNA sequence. Users may relax the cutoff to 4.0. The default threshold of 3.0 was selected because the naturally occurring TAL effector–target site pairs we analyzed to decipher the TAL effector DNA binding code and develop the scoring matrix (3) typically had scores <3.0 times the best possible score for the TAL effector. In addition, for binding sites of naturally occurring TAL effectors predicted using the scoring function, experimentally verified sites also had scores <3.0 times the best possible score (A. Cernadas, E. Doyle and A. Bogdanove, unpublished results). Users wishing to identify more sites should choose the less stringent cutoff. The best possible score for an array is included at the top of the output table or file. Information provided for each target in the output includes the coordinates, the DNA strand being targeted, the score for the target and the target sequence. Targets in each DNA sequence are ordered from best (lowest scoring) to worst (highest scoring). For searches of genomes or promoteromes from the drop-down list, the output in the default displayed table includes a link for each target that configures a custom track to show the target location in a corresponding genome browser when available. The number of targets displayed in the output table is set by default to a maximum of 10, though users can specify a different number. In the downloadable output, all targets below the threshold are included. As noted above, the downloadable output is available in two formats. The standard file includes the same information for each target as is displayed in the browser output table. The GFF3 file contains gene feature coordinates and can be used in many genome browser and related applications.

It is important to point out that the contributions of individual RVD–nucleotide associations to overall binding affinity are not yet worked out. Scoring based on RVD–nucleotide association frequencies provides only an estimate of relative affinities. Thus, not all sites returned may be efficiently bound by the TAL effector, irrespective of their score, and some biologically relevant targets may be missed if their scores rise above the

arbitrary cutoff. Target Finder output should be considered a best estimate of the most probable binding sites for a given TAL effector based on available information. It represents a good starting point for further study to identify true targets or off-targets through experimentation, and a tool for initial assessment of probable relative specificities to choose among multiple arrays that might be available to target a sequence of interest.

A final note about the web interface for Target Finder: TALE-NT 2.0 provides a convenient way to use the tool, but working through the web interface restricts users to the model organism sequences available on the site, or relatively short, user-provided sequences. To study large datasets not included on the site, users can download from TALE-NT 2.0 a C-coded version of Target Finder and run it locally under an open source license.

## CONCLUSION

TALEN Targeter and TAL Effector Targeter are versatile tools that allow design of custom RVD arrays for gene editing, engineered gene regulation and other applications. Although other TALEN design tools exist (TALEN Hit, <http://talen-hit.collectis-bioresearch.com/search>; ZiFit Targeter 4.0 (16), <http://zifit.partners.org/ZiFiTBeta/Introduction.aspx> and idTALE (29), <http://idtale.kaust.edu.sa/index.html>), TALEN Targeter is the only tool that works with any architecture by allowing users to specify ranges for both spacer size and number of repeats. TAL Effector Targeter is the only tool available that targets single custom TAL effector arrays. Although idTALE allows users to search a genome for paired TALEN sites, its search function identifies exact matches only. Target Finder uses a scoring function that allows a biologically relevant number of mismatches in the TAL effector–target alignment, useful for identifying candidate targets of naturally occurring TAL effectors, as well as potential off-targets for custom TAL effector-based proteins.

In addition to these design and targeting tools, TALE-NT 2.0 also includes help pages and tutorials, with guides to interpretation of results. A ‘Protocols and Reagents’ page provides useful links to other TAL effector and TALEN resources. With its wide range of capabilities and content, TALE-NT 2.0 should be a valuable resource for anyone studying TAL effector function or using custom TAL effectors, TALENs or other TAL effector-based proteins for DNA targeting applications.

A new tool, Paired Target Finder, that automates the identification of potential off-target sites for TALENs, as heterodimer or either homodimer, was recently added to the web site.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1.

## ACKNOWLEDGEMENTS

The authors thank D. Mistry, W. Kokulapalan and the Iowa State University Bioinformatics and Computational Biology Laboratory for assistance creating the web site and members of the Bogdanove and Voytas laboratories for beta-testing and suggestions.

## FUNDING

The National Science Foundation [IOS\#0820831] (to A.B. in part and ISO\#1221984 to V.B. in part) and the National Institutes of Health [R01GM098861] (to A.B. and D.V.). Funding for open access charge: NIH grant [R01GM098861].

*Conflict of interest statement.* None declared.

## REFERENCES

- 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.
- Scholze, H. and Boch, J. (2011) TAL effectors are remote controls for gene activation. *Curr. Opin. Microbiol.*, **14**, 47–53.
- Moscou, M.J. and Bogdanove, A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 1501.
- 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.
- 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.
- Bogdanove, A.J. and Voytas, D.F. (2011) TAL effectors: customizable proteins for DNA targeting. *Science*, **333**, 1843–1846.
- Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J. and Voytas, D.F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.*, **39**, e82.
- Li, T., Huang, S., Zhao, X., Wright, D.A., Carpenter, S., Spalding, M.H., Weeks, D.P. and Yang, B. (2011) Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res.*, **39**, 6315–6325.
- Morbitzer, R., Elsaesser, J., Hausner, J. and Lahaye, T. (2011) Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res.*, **39**, 5790–5799.
- 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.
- Weber, E., Gruetzner, R., Werner, S., Engler, C. and Marillonnet, S. (2011) Assembly of designer TAL effectors by golden gate cloning. *PLoS One*, **6**, e19722.
- Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D. and Joung, J.K. (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotechnol.*, **30**, 460–465.
- 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.
- 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.

15. Li,T., Huang,S., Jiang,W.Z., Wright,D., Spalding,M.H., Weeks,D.P. and Yang,B. (2011) TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.*, **39**, 359–372.
16. Sander,J.D., Cade,L., Khayter,C., Reyon,D., Peterson,R.T., Joung,J.K. and Yeh,J.-R.J. (2011) Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat. Biotechnol.*, **29**, 697–698.
17. Hockemeyer,D., Wang,H., Kiani,S., Lai,C.S., Gao,Q., Cassady,J.P., Cost,G.J., Zhang,L., Santiago,Y., Miller,J.C. *et al.* (2011) Genetic engineering of human pluripotent cells using TALE nucleases. *Nat. Biotechnol.*, **29**, 731–734.
18. Wood,A.J., Lo,T.-W., Zeitler,B., Pickle,C.S., Ralston,E.J., Lee,A.H., Amora,R., Miller,J.C., Leung,E., Meng,X. *et al.* (2011) Targeted genome editing across species using ZFNs and TALENs. *Science*, **333**, 307.
19. Huang,P., Xiao,A., Zhou,M., Zhu,Z., Lin,S. and Zhang,B. (2011) Heritable gene targeting in zebrafish using customized TALENs. *Nat. Biotechnol.*, **29**, 699–700.
20. Tesson,L., Usal,C., Menoret,S., Leung,E., Niles,B.J., Remy,S., Santiago,Y., Vincent,A.I., Meng,X., Zhang,L. *et al.* (2011) Knockout rats generated by embryo microinjection of TALENs. *Nat. Biotechnol.*, **29**, 695–696.
21. Morbitzer,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.
22. Geissler,R., Scholze,H., Hahn,S., Streubel,J., Bonas,U., Behrens,S.-E. and Boch,J. (2011) Transcriptional activators of human genes with programmable DNA-specificity. *PLoS One*, **6**, e19509.
23. Mahfouz,M., Li,L., Piatek,M., Fang,X., Mansour,H., Bangarusamy,D. and Zhu,J.-K. (2012) Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein. *Plant Mol. Biol.*, **78**, 311–321.
24. Mussolino,C., Morbitzer,R., Lutge,F., Dannemann,N., Lahaye,T. and Cathomen,T. (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res.*, **39**, 9283–9293.
25. Yu,Y., Streubel,J., Balzergue,S., Champion,A., Boch,J., Koebnik,R., Feng,J.-X., Verdier,V. and Szurek,B. (2011) Colonization of rice leaf blades by an African strain of *Xanthomonas oryzae* pv. *oryzae* depends on a new TAL effector which induces the rice nodulin-3 *Os11N3* gene. *Mol. Plant Microbe Interact.*, **24**, 1102–1113.
26. Mak,A.N.-S., Bradley,P., Cernadas,R.A., Bogdanove,A.J. and Stoddard,B.L. (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science*, **335**, 716–719.
27. Deng,D., Yan,C., Pan,X., Mahfouz,M., Wang,J., Zhu,J.-K., Shi,Y. and Yan,N. (2012) Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science*, **335**, 720–723.
28. Cock,P.J.A., Antao,T., Chang,J.T., Chapman,B.A., Cox,C.J., Dalke,A., Friedberg,I., Hamelryck,T., Kauff,F., Wilczynski,B. *et al.* (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25**, 1422–1423.
29. Li,L., Piatek,M., Atef,A., Piatek,A., Wibowo,A., Fang,X., Sabir,J., Zhu,J.-K. and Mahfouz,M. (2012) Rapid and highly efficient construction of TALE-based transcriptional regulators and nucleases for genome modification. *Plant Mol. Biol.*, **78**, 407–416.