EENdb: a database and knowledge base of ZFNs and TALENs for endonuclease engineering

An Xiao[1], Yingdan Wu[1], Zhipeng Yang[1], Yingying Hu[1], Weiye Wang[1], Yutian Zhang[1], Lei Kong[2], Ge Gao[2], Zuoyan Zhu[1], Shuo Lin[3,*] and Bo Zhang[1,*]

[1]Key Laboratory of Cell Proliferation and Differentiation of the Ministry of Education, [2]State Key Laboratory of Protein and Plant Gene Research, Center for Bioinformatics, College of Life Sciences, Peking University, Beijing 100871, China and [3]Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, CA, USA

ABSTRACT

We report here the construction of engineered endonuclease database (EENdb) (http://eendb.zfgenetics.org/), a searchable database and knowledge base for customizable engineered endonucleases (EENs), including zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs). EENs are artificial nucleases designed to target and cleave specific DNA sequences. EENs have been shown to be a very useful genetic tool for targeted genome modification and have shown great potentials in the applications in basic research, clinical therapies and agricultural utilities, and they are specifically essential for reverse genetics research in species where no other gene targeting techniques are available. EENdb contains over 700 records of all the reported ZFNs and TALENs and related information, such as their target sequences, the peptide components [zinc finger protein-/transcription activator-like effector (TALE)-binding domains, FokI variants and linker peptide/framework], the efficiency and specificity of their activities. The database also lists EEN engineering tools and resources as well as information about forms and types of EENs, EEN screening and construction methods, detection methods for targeting efficiency and many other utilities. The aim of EENdb is to represent a central hub for EEN information and an integrated solution for EEN engineering. These studies may help to extract in-depth properties and common rules regarding ZFN or TALEN efficiency through comparison of the known ZFNs or TALENs.

INTRODUCTION

Engineered endonucleases (EENs) are designed to bind and cleave specific DNA sequences *in vitro* or *in vivo*. Zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) are the two most widely used customizable types of EENs that exhibit many features in common (1). They recognize target sequences via engineerable DNA-binding domains and cut the DNA by non-specific nuclease domains (2,3). With these artificial endonucleases, site-specific DNA double-stranded breaks (DSBs) can be theoretically induced in any given genome, by which targeted genome modification could be easily achieved. Various genetic manipulations, such as gene disruption, gene correction and gene addition, have been designed and achieved in many organisms as well as in cultured cells based on EEN techniques. It is specifically essential for gene targeting in those species where no other reverse genetics approaches are available (4–8). Other types of EENs are also reported, such as triplex-forming oligonucleotides (9–12), engineered homing endonucleases (HEs) (13–15) and fusion proteins of the DNA-binding domains of zinc fingers or HE with full-length restriction endonucleases (16,17); however, they are much more difficult to customize than ZFNs and TALENs, which are based on a modular-repeat structure. Furthermore, a database called LAHEDES for HE-engineering has been reported (18).

Most ZFNs and TALENs function as dimers. Each monomer of these EENs consists of an artificially constructed hybrid protein containing a specific DNA-binding domain, which is derived from zinc finger proteins (ZFPs) for ZFNs (19), or transcription activator-like (TAL) effectors (TALEs) for TALENs (20) and a non-specific cleavage domain, which usually comes from the FokI restriction endonuclease (2) as well as a linker peptide between these two domains and other

*To whom correspondence should be addressed. Tel: +86 10 62759072; Fax: +86 10 62754293; Email: bzhang@pku.edu.cn
Correspondence may also be addressed to Shuo Lin. Tel: +1 310 267 4970; Fax: +1 310 267 4971; Email: shuolin@ucla.edu

framework sequences. The DNA-binding domain of ZFPs or TALEs consists of an array of tandem repeat units, each of which recognizes and binds to one or more nucleotide targets in one strand of the DNA. As for ZFP, each zinc finger repeat approximately targets 3 nt (called a triplet), but internal and external context-dependent effects of neighboring fingers can influence the efficiency and specificity of the corresponding target-protein relations (21). No accurate relationship has been discovered and several strategies have been developed to screen for and construct efficient and specific ZFPs and ZFNs. In contrast, TALE domains show a more predictable one-to-one correspondence between repeat units and their single-nucleotide targets, and no screening is required for TALEN engineering (22,23). Although hundreds of ZFNs or TALENs have been constructed and tested, the targeting efficiency of these EENs varies significantly. In fact, the efficiencies of TALENs in the same work or even targeting the same locus may vary from low to high, or even showing no targeting activity (24–27). Unfortunately, now it is still difficult to predict the efficiency of a given ZFN of TALEN, largely due to the lacking of sufficient knowledge about the properties of these EENs. Through analyzing the rapidly expanding information of these EENs from various resources, it may help to extract common rules regarding the efficiencies of TALENs or ZFNs through comparison of all the evaluated TALENs or ZFNs, respectively. On the other hand, the specificity and potential toxicity of TALENs still need further and careful investigation. In certain applications, the technology of ZFN is more mature, e.g. ZFNs are being tested in gene therapy trials (21,28–30), and are reported recently to be able to be delivered into cultured cells in the form of purified proteins directly (31). The properties of these two types of EENs have been compared with each other in some of the recent works (26,27,32,33).

Other factors in the customizable EEN engineering for TALENs and ZFNs, such as variants of FokI cleavage domains (34–40), as well as the relationship between the length of linker peptides and the length of spacer between the two EEN monomer binding sites (41–46), can also affect the efficiency and specificity of these EENs. Methodologies for the customizable EEN construction and genome modification have been optimized for best performance. The efficiency of EENs and the detection methods as well as the specificity evaluated by off-target tests have been reported previously (47,48). Collection and comparison of the corresponding information may be useful for reviewing/reusing existing EENs and engineering new TALENs or ZFNs.

Some databases such as ZifBASE (49) and ZiFDB (50) have been developed to collect the information of some of the ZFP domains but not the nucleases themselves. So far, hundreds of efficient ZFNs and TALENs have been reported and many methods for their engineering and construction including high-throughput ones (24,51) have been developed. However, a database systematically collecting and classifying the information of all the currently known customized EENs and their targets is unfortunately lacking.

Some tools are available to predict/search for candidate EEN target sites in given DNA sequences. TALE-NT (52), ZiFiT (53) and idTALE (54) are predominant tools for TALEN/TALE target design. ZiFiT also supports ZFN target design and selection covering various strategies, whereas ZFN Sequence Tag and other web-tools (55–57) only support particular types of ZFN design. On the other hand, several strategies have been developed to screen for and construct customized EENs. Polymerase chain reaction (PCR)-based Golden Gate cloning (58,59), vector-based Golden Gate cloning (60–64), Unit Assembly (65), REAL/REAL-Fast (33), dTALE-assembly (54), FLASH (24) and ICA (51) are designed to construct the long and highly repetitive TALE DNA-binding domains, which are difficult to synthesize by ordinary PCR or modular cloning. Modular Assembly (55,66,67), OPEN (68), CoDA (69) and other public (70,71) or commercial strategies (72) are developed to screen for and obtain ZFP domains showing high target binding activity. Apart from these methods, several platforms such as Addgene (http://www.addgene.com) offer plasmids or kits and protocols for EEN engineering with several selections of different methods or strategies, which might be difficult to find and compare all the existing resources and make a proper decision if one does not work on and is not familiar with the field of EEN engineering. A knowledge base providing a route with simple descriptions for and links to these tools and resources will allow users to take it as a starting point and find the most appropriate one for themselves.

Here, we report the construction of EENdb, the first database and knowledge base dedicated to a complete and detailed collection of all the currently available and evaluated ZFN and TALEN data extracted and curated from publications and other sources. It also integrates different EEN (ZFN and TALEN) engineering resources, to provide a relatively complete solution for searching for known EENs as well as for design and construction of new customized ZFNs and TALENs. In current version of EENdb and here below, EENs refer to ZFNs and TALENs.

MATERIALS AND METHODS

Data sources

We manually extracted the information of all ZFNs and TALENs reported to be effective, either targeting endogenous loci or designed for artificial sequences, directly from published research articles, or reconstructed the necessary information if it was not given directly. Non-functional TALENs and ZFNs were also included for comparison and analysis, but numerous inactive ZFNs from screening attempts were omitted. The data of some TALENs showing no detectable activity were also collected via direct submission by the authors.

In total, more than 400 records of ZFNs, more than 300 records of published TALENs and 24 records of directly submitted TALENs have been collected and curated so far. Redundant EENs from the same or different sources were merged. The core information in the database

consists of targeting site sequences, types of ZFN and TALEN forms, critical peptide sequence of the DNA-binding domains, linker peptides and frameworks, FokI variants and other alternate cleavage domains, ZFP or TALE construction strategies and genome modification methods. EEN efficiency with the detection methods and EEN specificity with off-target sites tested were also included if available. In the whole target site for a pair of EENs, the ZFP- or TALE-binding sequences (i.e. half-sites) and spacer sequences between the two half-sites were carefully differentiated.

An NIH-sponsored project aiming to target endogenous genes in zebrafish (*Danio rerio*) regularly constructs arrays of ZFNs and TALENs with only ZFP- or TALE-binding activities tested and releases them in Addgene (http://www.addgene.com/zfc/arrays/ and http://www.addgene.com/talengineering/TALENzebrafish/). EENdb has not collected the information of these EENs unless the cleavage activities were tested and reported. However, a prompt message appears and indicates to the project web pages when users list EENs of all species or of zebrafish in EENdb.

To complement with the EEN tables, we also collected the information of natural and artificial ZFPs (only the C_2H_2 type, which is used as the framework for engineered ZFPs), including those with only binding capability but not cleavage activity reported and artificial TALEs. The 7-aa variable region of each finger of ZFPs (including ZFNs) and the repeat-variable di-residue (RVD) of each repeat unit of TALEs, which are considered to be the most

critical elements determining DNA-binding specificity, were extracted and displayed in association with their target nucleotides.

Design and implementation

EENdb is implemented with open source technologies. The data are stored in a MySQL relational database. The web site is written in PHP scripts and the service is provided by an Apache/PHP server.

RESULTS

EENdb consists of five interrelated sections (Figure 1). The main section is the collection and summary of the information about EENs (including TALENs and ZFNs) called 'TALEN/ZFN'. 'ZFP Domain' and 'TAL Effector' are two sections built to complement with the main section, whereas 'ZFP Domain' provides the collection of the zinc finger DNA-binding domains, especially the 7-aa regions from ZFNs and other ZFPs; 'TAL Effector' collects the information of the TALE DNA-binding domains from TALE proteins other than TALENs. Details about the components and information of EENs, such as variants of the FokI cleavage domains, the scaffolds/frameworks and linker peptides between the DNA-binding domains and the cleavage domains, as well as the information about the construction and application of EENs, are summarized in the section of 'Utilities' and linked to other EEN datasets in each corresponding field. The last section named 'Engineering Resource' provides
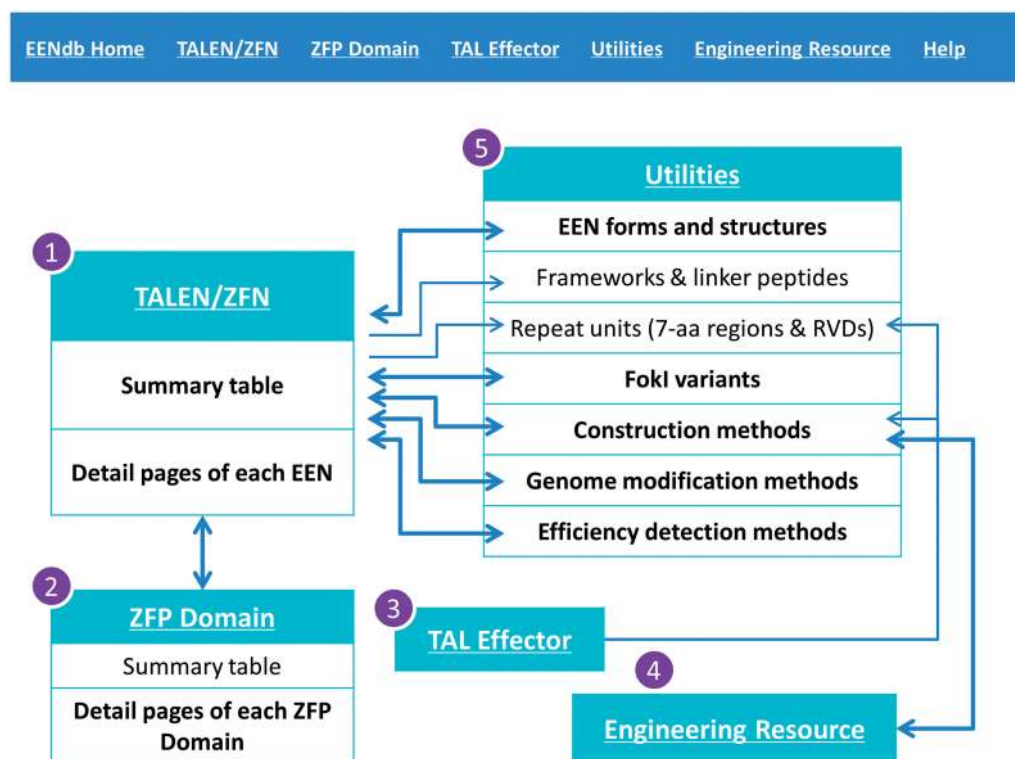
Figure 1. The menu banner of EENdb and the relationship between the five main sections of the database. Contents and organization of these sections are described in the text. The arrows represent links between the sections and web pages.
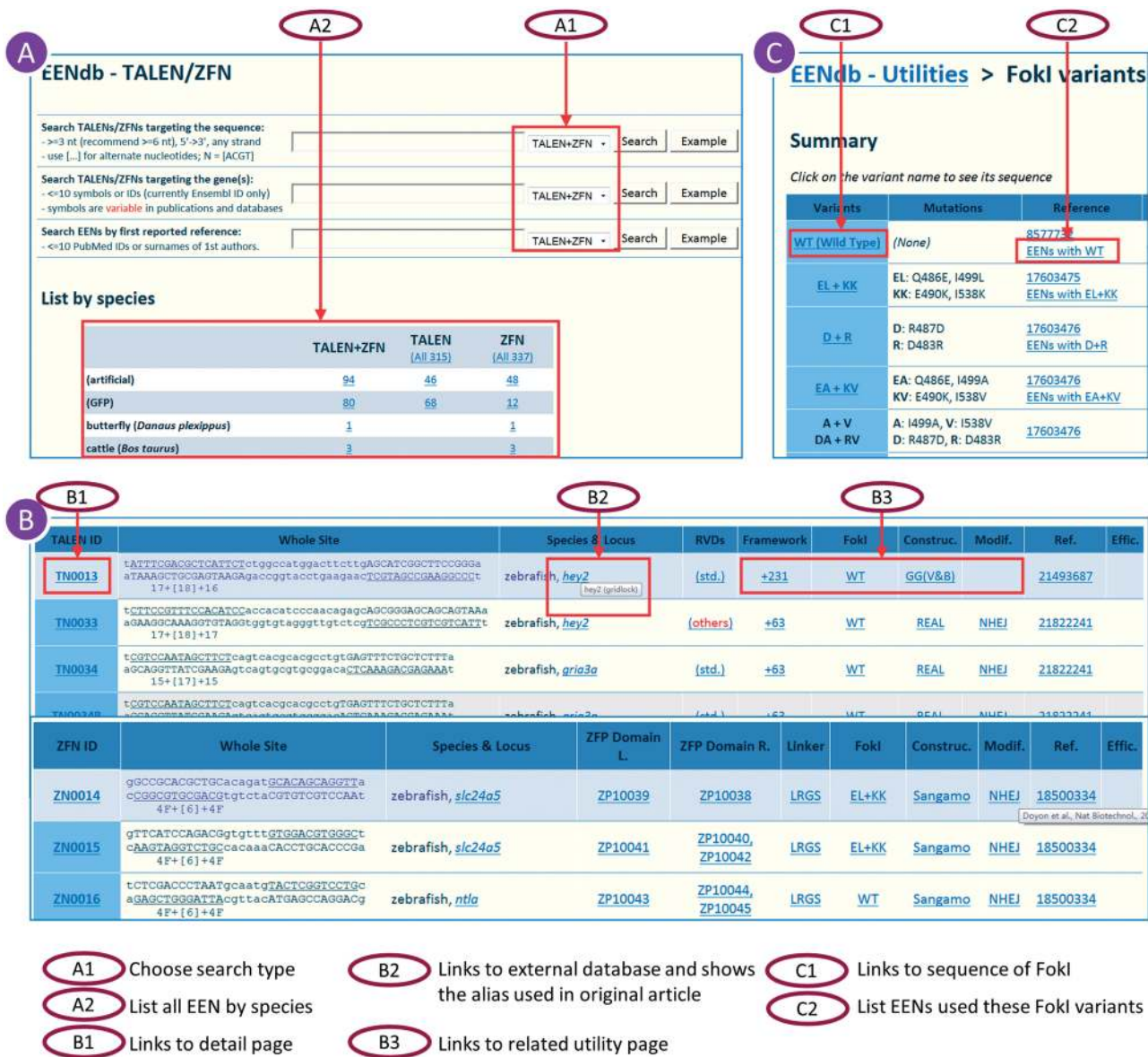
Figure 2. Screenshots of EENdb. (**A**) The section of TALEN/ZFN. (**B**) Part of the summary table of listing all EENs in zebrafish. (**C**) Part of the web page of FokI variants in the 'Utilities' section.

short descriptions of and links to all the available external resources for EEN engineering and is also linked to the EEN construction methods summarized in the 'Utilities'.

The dataset of EENs

Collection of the essential information about EENs is the major component of EENdb. Users can list all TALENs and/or ZFNs designed to target a special species, or easily search by gene symbols, gene IDs, whole or partial target sequences (degenerate nucleotides are also supported), or reference PMIDs (PubMed IDs) or surnames of the first authors from the 'TALEN/ZFN' page (Figure 2A). Results of corresponding ZFNs, TALENs or integration of the two types of EENs are returned, according to the search option set by users. In the search result table or the

table listed by species, summaries of EEN information are listed (Figure 2B). More details can be accessed through the detail page by clicking on the EENdb IDs in the table.

For each EEN, the following information is displayed in the detail page; some of the information is omitted or shown in abbreviation in the summary table.

EENdb ID

EEN data are collected and classified according to their target sites. ID of a ZFN record usually has the format 'ZNxxxx' or 'ZNAxxx' and ID of a TALEN has the format 'TNxxxx' or 'TNAxxx', in which an 'x' refers to a digit. 'ZNxxxx' and 'TNxxxx' are assigned to EENs targeting natural sites, whereas 'ZNAxxx' and 'TNAxxx' are for EENs targeting artificially synthesized DNA sequences. Groups of different EENs targeting similar sequences from the same or different literature are given

serial IDs with one-letter suffixes. For example, TN0002, TN0002B, TN0002C and TN0002D all target the same or overlapped site of human *CCR5* gene. An '-Txxx' suffix indicates that this record contains a previously reported EEN (i.e. the same EEN) but targeting another DNA sequence. In most cases, it represents an off-target site. For example, TN0031-T002 targets *CCR2*, an off-target gene of TN0031, which is originally designed to target *CCR5* gene. In summary tables of the list of a particular species or search results, the EENs are sorted as a default by publication PMID (approximately in the order of publication date) so that EENs from the same publication are arranged together. Alternatively, it can be changed to group by EENdb ID, which is also used in the EEN detail pages, thereby related records with same target sequence and/or off-targets of same EENs are centralized for comparison.

EEN type or form

Whether an EEN is a ZFN or a TALEN can be easily distinguished from the first letter of the ID. In addition, most of the EENs are functional in dimers and the FokI domains are usually fused in the C-terminus of the monomers, but exceptions do exist (73–78). A description of rarely used EEN form and a link to a page with detailed explanation in the 'Utilities' section is provided under the corresponding EENdb ID.

Target site sequence

The sequence of the whole target site recognized by EEN pairs is carefully extracted from or constructed based on publications. The half-sites (i.e. the binding site for a single EEN) of the whole target site are shown in uppercase; the strands bound by the ZFP- or TALE-binding domains are underlined. The spacers are revealed as lowercase letters. Additional one base pair outside the half-sites is also provided for it is important for ZFNs and TALENs in some cases, e.g. to distinguish whether the additional nucleotide is the most commonly used nucleotide T for TALENs (3,24), or to consider the context-dependent effect of ZFN fingers (79,80). The target sequences can be searched by either the forward or the reverse strands.

Other information of the target site

For natural targets, the English and Latin names of species and the gene or genomic locus represented by Ensembl IDs, Ensembl Genomes IDs or RefSeq Accessions are given. Length of the spacer, numbers of fingers in a ZFN and lengths of half-sites of a TALEN are calculated and displayed.

DNA-binding domains of EEN monomers

The key amino acids of the DNA-binding domains are shown here. For ZFNs, the 7-aa variable regions of each finger and a link to the 'ZFP Domain' section are provided. For TALENs, four most commonly used RVDs, each recognizing its own corresponding single-nucleotide target (i.e. NI for nucleotide A, HD for C, NG for T and NN for G), are considered as 'standard code' of RVDs (22,23); other non-standard 'alternative code' of RVDs (e.g. NK or NH for G, NG for ^mC or 5-methylcytosine) (20,81–83) and off-targeted RVDs are marked with colors different from the 'standard' ones for identification.

Other components of the EEN protein

The types of linker peptides between the DNA-binding domains and FokI domains, the variants of FokI cleavage domains as well as the screening and construction methods or strategies for the DNA-binding domains are included if they are known. These items are linked to the related pages in the section of 'Utilities'.

Effectiveness of EEN

Each EEN record also contains the modification method of the target locus [e.g. non-homologous end joining (NHEJ) or homologous recombination (HR)], efficiency of the EENs and the detection methods and *in vitro* cultured cell lines tested if available, heritability of EEN-induced mutations at organism level and specificity of the EEN revealed by the result of off-target evaluation. The modification and efficiency detection methods are also linked to the corresponding 'Utilities' pages for detailed explanation. In some cases, comparisons or collaborations with other EENs are indicated as comments.

References

PMID with a link to the first publication reporting certain pair of EENs is provided, or other information is listed for directly submitted records.

User commenting system

In each detail page of the EEN records, a simple commenting system is available; any correction, complementarity and communication can be contributed by users either with this system or through contacting the administrators of EENdb by email.

The dataset of ZFP domain and TALE

Natural or artificial ZFP domains and artificial TALEs compose two other sections of EENdb in parallel to the section of TALEN/ZFN. The dataset of ZFP domains consists of more than 1000 records and provides much more information about ZFP-binding domains for data mining, e.g. discovering the relationship between the key peptide sequences and the target nucleotide sequences of the nucleases based on statistical analysis of known ZFPs. The dataset of TALEs provides a link to a list of natural TALEs and collects TALEs beyond TALENs, which are both rich in 'alternative code' or off-targeted RVDs and may be helpful for researchers who are interested in the analysis of TALE-binding activities and discovering more information of the DNA-binding properties of TALEs.

The utilities of EENdb

One of the aims of EENdb is to build a knowledge base about EENs. We invested a great deal of effort to organize and refine the information and utilities of EENdb, which offers a global view of the emerging and fast developing

ZFN and TALEN technology and resources. In the 'Utilities' section, detailed information of ZFN and TALEN forms and structures, frameworks and linker peptides, FokI variants and other alternate cleavage domains, repeat units of DNA-binding domains, ZFP- or TALE-construction strategies, genome modification methods and efficiency detection methods are provided (Figure 1). These data can help users to retrieve details about EENs, compare and choose optimal parameters, etc. Most of the utilities also provide links to the 'TALEN/ZFN' section to filter all EENs matching defined conditions.

For example, the page of FokI variants (Figure 2C) lists all known FokI pairs that can only work in hetero-dimers to reduce non-specificity, or that can enhance catalytic activity, or that participate in building a nickase (only cleave one strand of the target DNA) rather than a nuclease (37–39). Users can list the records of all reported EENs containing specific pairs of FokI domains from the EEN dataset via the links under each FokI variants. The sequences of variants with the mutations highlighted can also be found in this page.

Other examples for the usage of the 'Utilities' section are the pages of repeat units of DNA-binding domains, genome modification methods and efficiency detection methods. Similar to the page for FokI variants, users can list the records of all EENs reported to induce HR in different organisms, or all TALENs with particular non-standard 'alternative' RVD via corresponding links. This information may help researchers to choose appropriate method for their own experiments. Only as an example of this attempt, one can easily conclude from the list of all the EENs tested in zebrafish that NHEJ is the only type of genome modification induced by EENs in this species except for one very recent report (84), and the most frequently used detection methods in this species are restriction enzyme-resistance assay (after PCR) and direct sequencing.

EEN engineering tools and resources

As described above, many EEN target prediction tools and construction resources are available for the public. The 'Engineering Resource' section of EENdb generates links and provides short descriptions and comments of these tools and resources, including candidate TALEN/TALE target site prediction tools, candidate ZFN/ZFP target site prediction tools, target- or off-target-site finder of given EENs, resources of EEN construction protocols and materials and other links to news sites and newsgroups. Applicable construction methods supported by any available resources are provided with links to the corresponding pages in the section of 'Utilities'.

Access

EENdb can be freely accessed via http://eendb.zfgenetics.org/. The data of EENdb can be downloaded as tab-separated version (TSV) plain text format in the 'Help' section of the web site.

EENdb welcomes researchers to submit the information of their newly constructed EENs and related experimental data by email, especially the negative data not intended for publication. The feedback form and email address can be found in the 'Help' section of the web site.

DISCUSSION

Continuous updates of the EENdb database will be offered. As more and more EENs are expected to be reported, EENdb will expand to offer links to show custom tracking of known EEN sites in Ensembl and UCSC Genome Browser when necessary.

EENdb has collected the references of all publications first reporting a particular ZFN or TALEN. Other references related to EENs, such as new applications of a known EEN, will also be included in future releases of EENdb.

REFERENCES

1. Baker,M. (2012) Gene-editing nucleases. *Nat. Methods*, **9**, 23–26.
2. 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.
3. 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.
4. Xiao,A., Hu,Y.Y., Wang,W.Y., Yang,Z.P., Wang,Z.X., Huang,P., Tong,X.J., Zhang,B. and Lin,S. (2011) Progress in zinc finger nuclease engineering for targeted genome modification. *Hereditas*, **33**, 665–683.
5. Huang,P., Zhu,Z., Lin,S. and Zhang,B. (2012) Reverse genetic approaches in zebrafish. *J. Genet. Genomics*, **39**, 421–433.
6. Carroll,D. (2011) Genome engineering with zinc-finger nucleases. *Genetics*, **188**, 773–782.
7. DeFrancesco,L. (2011) Move over ZFNs. *Nat. Biotechnol.*, **29**, 681–684.

8. McMahon,M.A., Rahdar,M. and Porteus,M. (2012) Gene editing: not just for translation anymore. *Nat. Methods*, **9**, 28–31.
9. Silanskas,A., Zaremba,M., Sasnauskas,G. and Siksnys,V. (2012) Catalytic activity control of restriction endonuclease–triplex forming oligonucleotide conjugates. *Bioconjug. Chem.*, **23**, 203–211.
10. Dahmen,V. and Kriehuber,R. (2012) Cytotoxic effects and specific gene expression alterations induced by I-125-labeled triplex-forming oligonucleotides. *Int. J. Radiat. Biol.*, **88**, 972–979.
11. Kolganova,N.A., Shchyolkina,A.K., Chudinov,A.V., Zasedatelev,A.S., Florentiev,V.L. and Timofeev,E.N. (2012) Targeting duplex DNA with chimeric alpha, beta-triplex-forming oligonucleotides. *Nucleic Acids Res.*, **40**, 8175–8185.
12. Govan,J.M., Uprety,R., Hemphill,J., Lively,M.O. and Deiters,A. (2012) Regulation of transcription through light-activation and light-deactivation of triplex-forming oligonucleotides in mammalian cells. *ACS Chem. Biol.*, **7**, 1247–1256.
13. Kleinstiver,B.P., Wolfs,J.M., Kolaczyk,T., Roberts,A.K., Hu,S.X. and Edgell,D.R. (2012) Monomeric site-specific nucleases for genome editing. *Proc. Natl Acad. Sci. USA*, **109**, 8061–8066.
14. Baxter,S., Lambert,A.R., Kuhar,R., Jarjour,J., Kulshina,N., Parmeggiani,F., Danaher,P., Gano,J., Baker,D., Stoddard,B.L. *et al.* (2012) Engineering domain fusion chimeras from I-OnuI family LAGLIDADG homing endonucleases. *Nucleic Acids Res.*, **40**, 7985–8000.
15. Ulge,U.Y., Baker,D.A. and Monnat,R.J. Jr (2011) Comprehensive computational design of mCreI homing endonuclease cleavage specificity for genome engineering. *Nucleic Acids Res.*, **39**, 4330–4339.
16. Schierling,B., Dannemann,N., Gabsalilow,L., Wende,W., Cathomen,T. and Pingoud,A. (2012) A novel zinc-finger nuclease platform with a sequence-specific cleavage module. *Nucleic Acids Res.*, **40**, 2623–2638.
17. Fonfara,I., Curth,U., Pingoud,A. and Wende,W. (2012) Creating highly specific nucleases by fusion of active restriction endonucleases and catalytically inactive homing endonucleases. *Nucleic Acids Res.*, **40**, 847–860.
18. Taylor,G.K., Petrucci,L.H., Lambert,A.R., Baxter,S.K., Jarjour,J. and Stoddard,B.L. (2012) LAHEDES: the LAGLIDADG homing endonuclease database and engineering server. *Nucleic Acids Res.*, **40**, W110–W116.
19. Berg,J.M. (1990) Zinc finger domains: hypotheses and current knowledge. *Annu. Rev. Biophys. Biophys. Chem.*, **19**, 405–421.
20. Scholze,H. and Boch,J. (2010) TAL effector-DNA specificity. *Virulence*, **1**, 428–432.
21. Klug,A. (2010) The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Q. Rev. Biophy.*, **43**, 1–21.
22. Moscou,M.J. and Bogdanove,A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 1501.
23. 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.
24. 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.
25. Cade,L., Reyon,D., Hwang,W.Y., Tsai,S.Q., Patel,S., Khayter,C., Joung,J.K., Sander,J.D., Peterson,R.T. and Yeh,J.R. (2012) Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. *Nucleic Acids Res.*, **40**, 8001–8010.
26. Moore,F.E., Reyon,D., Sander,J.D., Martinez,S.A., Blackburn,J.S., Khayter,C., Ramirez,C.L., Joung,J.K. and Langenau,D.M. (2012) Improved somatic mutagenesis in zebrafish using transcription activator-like effector nucleases (TALENs). *PLoS One*, **7**, e37877.
27. 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.
28. Phillips,M.I. (2012) Gene, stem cell, and future therapies for orphan diseases. *Clin. Pharmacol. Ther.*, **92**, 182–192.
29. Olender,S., Wilkin,T.J., Taylor,B.S. and Hammer,S.M. (2012) Advances in antiretroviral therapy. *Top. Antivir. Med.*, **20**, 61–86.
30. Overlack,N., Goldmann,T., Wolfrum,U. and Nagel-Wolfrum,K. (2012) Gene repair of an usher syndrome causing mutation by zinc-finger nuclease mediated homologous recombination. *Invest. Ophthalmol. Vis. Sci.*, **53**, 4140–4146.
31. Gaj,T., Guo,J., Kato,Y., Sirk,S.J. and Barbas,C.F. 3rd (2012) Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. *Nat. Methods*, **9**, 805–807.
32. Kim,H., Um,E., Cho,S.R., Jung,C. and Kim,J.S. (2011) Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat. Methods*, **8**, 941–943.
33. Sander,J.D., Cade,L., Khayter,C., Reyon,D., Peterson,R.T., Joung,J.K. and Yeh,J.R. (2011) Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat. Biotechnol.*, **29**, 697–698.
34. 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.
35. 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.
36. Guo,J., Gaj,T. and Barbas,C.F. 3rd (2010) Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. *J. Mol. Biol.*, **400**, 96–107.
37. Doyon,Y., Vo,T.D., Mendel,M.C., Greenberg,S.G., Wang,J., Xia,D.F., Miller,J.C., Urnov,F.D., Gregory,P.D. and Holmes,M.C. (2011) Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat. Methods*, **8**, 74–79.
38. Kim,E., Kim,S., Kim,D.H., Choi,B.S., Choi,I.Y. and Kim,J.S. (2012) Precision genome engineering with programmable DNA-nicking enzymes. *Genome Res.*, **22**, 1327–1333.
39. Ramirez,C.L., Certo,M.T., Mussolino,C., Goodwin,M.J., Cradick,T.J., McCaffrey,A.P., Cathomen,T., Scharenberg,A.M. and Joung,J.K. (2012) Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Res.*, **40**, 5560–5568.
40. Wang,J., Friedman,G., Doyon,Y., Wang,N.S., Li,C.J., Miller,J.C., Hua,K.L., Yan,J.J., Babiarz,J.E., Gregory,P.D. *et al.* (2012) Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. *Genome Res.*, **22**, 1316–1326.
41. Shimizu,Y., Bhakta,M.S. and Segal,D.J. (2009) Restricted spacer tolerance of a zinc finger nuclease with a six amino acid linker. *Bioorg. Med. Chem. Lett.*, **19**, 3970–3972.
42. Smith,J., Bibikova,M., Whitby,F.G., Reddy,A.R., Chandrasegaran,S. and Carroll,D. (2000) Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res.*, **28**, 3361–3369.
43. Bibikova,M., Carroll,D., Segal,D.J., Trautman,J.K., Smith,J., Kim,Y.G. and Chandrasegaran,S. (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol. Cell. Biol.*, **21**, 289–297.
44. Handel,E.M., Alwin,S. and Cathomen,T. (2009) Expanding or restricting the target site repertoire of zinc-finger nucleases: the inter-domain linker as a major determinant of target site selectivity. *Mol. Ther.*, **17**, 104–111.
45. Kandavelou,K., Ramalingam,S., London,V., Mani,M., Wu,J., Alexeev,V., Civin,C.I. and Chandrasegaran,S. (2009) Targeted manipulation of mammalian genomes using designed zinc finger nucleases. *Biochem. Biophys. Res. Commun.*, **388**, 56–61.
46. Isalan,M. (2012) Zinc-finger nucleases: how to play two good hands. *Nat. Methods*, **9**, 32–34.
47. Pattanayak,V., Ramirez,C.L., Joung,J.K. and Liu,D.R. (2011) Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat. Methods*, **8**, 765–770.
48. Gabriel,R., Lombardo,A., Arens,A., Miller,J.C., Genovese,P., Kaeppel,C., Nowrouzi,A., Bartholomae,C.C., Wang,J.,

Friedman,G. *et al.* (2011) An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat. Biotechnol.*, **29**, 816–823.

49. Jayakanthan,M., Muthukumaran,J., Chandrasekar,S., Chawla,K., Punetha,A. and Sundar,D. (2009) ZifBASE: a database of zinc finger proteins and associated resources. *BMC Genomics*, **10**, 421.

50. Fu,F., Sander,J.D., Maeder,M., Thibodeau-Beganny,S., Joung,J.K., Dobbs,D., Miller,L. and Voytas,D.F. (2009) Zinc Finger Database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays. *Nucleic Acids Res.*, **37**, D279–D283.

51. Briggs,A.W., Rios,X., Chari,R., Yang,L., Zhang,F., Mali,P. and Church,G.M. (2012) Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Res.*, **40**, e117.

52. Doyle,E.L., Booher,N.J., Standage,D.S., Voytas,D.F., Brendel,V.P., Vandyk,J.K. and Bogdanove,A.J. (2012) TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res.*, **40**, W117–W122.

53. Sander,J.D., Maeder,M.L., Reyon,D., Voytas,D.F., Joung,J.K. and Dobbs,D. (2010) ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Res.*, **38**, W462–W468.

54. Li,L., Piatek,M.J., Atef,A., Piatek,A., Wibowo,A., Fang,X., Sabir,J.S., Zhu,J.K. and Mahfouz,M.M. (2012) Rapid and highly efficient construction of TALE-based transcriptional regulators and nucleases for genome modification. *Plant Mol. Biol.*, **78**, 407–416.

55. Gupta,A., Christensen,R.G., Rayla,A.L., Lakshmanan,A., Stormo,G.D. and Wolfe,S.A. (2012) An optimized two-finger archive for ZFN-mediated gene targeting. *Nat. Methods*, **9**, 588–590.

56. Zhu,C., Smith,T., McNulty,J., Rayla,A.L., Lakshmanan,A., Siekmann,A.F., Buffardi,M., Meng,X., Shin,J., Padmanabhan,A. *et al.* (2011) Evaluation and application of modularly assembled zinc-finger nucleases in zebrafish. *Development*, **138**, 4555–4564.

57. Gupta,A., Meng,X., Zhu,L.J., Lawson,N.D. and Wolfe,S.A. (2011) Zinc finger protein-dependent and -independent contributions to the in vivo off-target activity of zinc finger nucleases. *Nucleic Acids Res.*, **39**, 381–392.

58. Sanjana,N.E., Cong,L., Zhou,Y., Cunniff,M.M., Feng,G. and Zhang,F. (2012) A transcription activator-like effector toolbox for genome engineering. *Nat. Protoc.*, **7**, 171–192.

59. 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.

60. 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.

61. 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.

62. 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.

63. 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.

64. 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.

65. 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.

66. Beerli,R.R., Segal,D.J., Dreier,B. and Barbas,C.F. 3rd (1998) Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc. Natl Acad. Sci. USA*, **95**, 14628–14633.

67. Wright,D.A., Thibodeau-Beganny,S., Sander,J.D., Winfrey,R.J., Hirsh,A.S., Eichtinger,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.

68. Maeder,M.L., Thibodeau-Beganny,S., Osiak,A., Wright,D.A., Anthony,R.M., Eichtinger,M., Jiang,T., Foley,J.E., Winfrey,R.J., Townsend,J.A. *et al.* (2008) Rapid "Open-Source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell*, **31**, 294–301.

69. Sander,J.D., Dahlborg,E.J., Goodwin,M.J., Cade,L., Zhang,F., Cifuentes,D., Curtin,S.J., Blackburn,J.S., Thibodeau-Beganny,S., Qi,Y. *et al.* (2010) Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat. Methods*, **8**, 67–69.

70. Meng,X., Noyes,M.B., Zhu,L.J., Lawson,N.D. and Wolfe,S.A. (2008) Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat. Biotechnol.*, **26**, 695–701.

71. Herrmann,F., Garriga-Canut,M., Baumstark,R., Fajardo-Sanchez,E., Cotterell,J., Minoche,A., Himmelbauer,H. and Isalan,M. (2011) p53 Gene repair with zinc finger nucleases optimised by yeast 1-hybrid and validated by Solexa sequencing. *PLoS One*, **6**, e20913.

72. Hansen,K., Coussens,M.J., Sago,J., Subramanian,S., Gjoka,M. and Briner,D. (2012) Genome editing with CompoZr custom zinc finger nucleases (ZFNs). *J. Vis. Exp.*, e3304.

73. Minczuk,M., Papworth,M.A., Miller,J.C., Murphy,M.P. and Klug,A. (2008) Development of a single-chain, quasi-dimeric zinc-finger nuclease for the selective degradation of mutated human mitochondrial DNA. *Nucleic Acids Res.*, **36**, 3926–3938.

74. Mineta,Y., Okamoto,T., Takenaka,K., Doi,N., Aoyama,Y. and Sera,T. (2008) Enhanced cleavage of double-stranded DNA by artificial zinc-finger nuclease sandwiched between two zinc-finger proteins. *Biochemistry*, **47**, 12257–12259.

75. Mino,T., Aoyama,Y. and Sera,T. (2009) Efficient double-stranded DNA cleavage by artificial zinc-finger nucleases composed of one zinc-finger protein and a single-chain FokI dimer. *J. Biotechnol.*, **140**, 156–161.

76. Mori,T., Kagatsume,I., Shinomiya,K., Aoyama,Y. and Sera,T. (2009) Sandwiched zinc-finger nucleases harboring a single-chain FokI dimer as a DNA-cleavage domain. *Biochem. Biophys. Res. Commun.*, **390**, 694–697.

77. Fan,B., Huang,P., Zheng,S., Sun,Y., Fang,C. and Sun,Z. (2011) Assembly and in vitro functional analysis of zinc finger nuclease specific to the 3′ untranslated region of chicken ovalbumin gene. *Anim. Biotechnol.*, **22**, 211–222.

78. 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.

79. Isalan,M., Choo,Y. and Klug,A. (1997) Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proc. Natl Acad. Sci. USA*, **94**, 5617–5621.

80. Isalan,M., Klug,A. and Choo,Y. (1998) Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers. *Biochemistry*, **37**, 12026–12033.

81. Cong,L., Zhou,R., Kuo,Y.C., Cunniff,M. and Zhang,F. (2012) Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat. Commun.*, **3**, 968.

82. Streubel,J., Blucher,C., Landgraf,A. and Boch,J. (2012) TAL effector RVD specificities and efficiencies. *Nat. Biotechnol.*, **30**, 593–595.

83. Deng,D., Yin,P., Yan,C., Pan,X., Gong,X., Qi,S., Xie,T., Mahfouz,M., Zhu,J.K., Yan,N. *et al.* (2012) Recognition of methylated DNA by TAL effectors. *Cell Res.*, **22**, 1502–1504.

84. Bedell,V.M., Wang,Y., Campbell,J.M., Poshusta,T.L., Starker,C.G., Krug,R.G. II, Tan,W., Penheiter,S.G., Ma,A.C., Leung,A.Y.H. *et al.* (2012) *In vivo* genome editing using a high-efficiency TALEN system. *Nature*, **491**, 114–118.