OXFORD

Genome analysis

CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites

Yuki Naito[1,2,*,†], **Kimihiro Hino**[3,†], **Hidemasa Bono**[1,2] **and Kumiko Ui-Tei**[3]

[1]Database Center for Life Science (DBCLS), [2]National Institute of Genetics, Research Organization of Information and Systems, 1111 Yata, Mishima, Shizuoka 411-8540, Japan and [3]Department of Biological Sciences, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
Associate Editor: Alfonso Valencia

Abstract

Summary: CRISPRdirect is a simple and functional web server for selecting rational CRISPR/Cas targets from an input sequence. The CRISPR/Cas system is a promising technique for genome engineering which allows target-specific cleavage of genomic DNA guided by Cas9 nuclease in complex with a guide RNA (gRNA), that complementarily binds to a ~20 nt targeted sequence. The target sequence requirements are twofold. First, the 5′-NGG protospacer adjacent motif (PAM) sequence must be located adjacent to the target sequence. Second, the target sequence should be specific within the entire genome in order to avoid off-target editing. CRISPRdirect enables users to easily select rational target sequences with minimized off-target sites by performing exhaustive searches against genomic sequences. The server currently incorporates the genomic sequences of human, mouse, rat, marmoset, pig, chicken, frog, zebrafish, *Ciona*, fruit fly, silkworm, *Caenorhabditis elegans*, *Arabidopsis*, rice, *Sorghum* and budding yeast.
Availability: Freely available at http://crispr.dbcls.jp/.
Contact: y-naito@dbcls.rois.ac.jp
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Genome engineering is a promising technique to manipulate endogenous chromosomal DNA in a site-specific manner. A novel system that employs the prokaryotic immune defense system based on the clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) protein has been reported as a prominent genome engineering approach (Cho *et al.*, 2013; Cong *et al.*, 2013; Jinek *et al.*, 2013; Mali *et al.*, 2013). Recent studies utilize the RNA-guided endonuclease Cas9 from *Streptococcus pyogenes* and a guide RNA (gRNA), which acts as a guide to define the target site to introduce DNA double-stranded break. A remarkable advantage of the CRISPR/Cas system is that the target DNA sequence is recognized by simple base-pairing complementarity by the gRNA. Thus, the CRISPR/Cas system can be programmed only by changing the gRNA sequence, and the synthesis of the gRNA for

targeting a specific gene is easy at low cost. However, it should be a critical issue to avoid the cleavage of the unintended off-target genes, since double-stranded break results in stable and heritable modification of the genome.

In this study, we present CRISPRdirect (http://crispr.dbcls.jp/), which provides efficient selection of CRISPR/Cas target sites with reduced numbers of potential off-target candidates. CRISPRdirect investigates the entire genome for perfect matches with each candidate target sequence (20 mer) and their seed sequence (12 or 8 mer) flanking the PAM. Users can also browse the detailed list of potential off-target sites that have partial complementarity with the selected sequence. The server incorporates genomic sequences of human, mouse, rat, marmoset, pig, chicken, frog, zebrafish, *Ciona*, fruit fly, silkworm, *Caenorhabditis elegans*, *Arabidopsis*, rice,

Fig. 1. Screenshot from the CRISPRdirect web server. (**A**) Top page. The server accepts either an accession number or a nucleotide sequence as input. (**B**) Typical output of CRISPRdirect. A list of CRISPR/Cas target candidates is displayed. (**C**) A graphical view of target sites demonstrates the position and orientation of each site. (**D**) The results can be exported as tab-delimited text or in JSON format. (**E**) Detailed list of potential off-target sites which visualizes the positions of mismatches and gaps

Sorghum and budding yeast. Currently, several web servers are available for designing CRISPR/Cas gRNAs (Supplementary Table S1). CRISPR Design (Hsu *et al.*, 2013; Ran *et al.*, 2013) performs gRNA selection from an input sequence up to 250 bp, and gRNAs are scored based on predicted off-target interactions. E-CRISP (Heigwer *et al.*, 2014) ranks gRNAs according to on-target specificity and number of off-targets. E-CRISP, ZiFiT (Sander *et al.*, 2010), Cas9 Design (Ma *et al.*, 2013) and CHOPCHOP (Montague *et al.*, 2014) utilize Bowtie (Langmead and Salzberg, 2012) to perform off-target searches allowing mismatches. On the other hand, DNA2.0 gRNA Design Tool (https://www.dna20.com/eCommerce/startCas9) searches for perfect matches with 12 nt seed to identify off-target sites. These servers except CRISPR Design and ZiFiT can process at least 10 kbp of input sequence. Web servers for checking off-target sites for given 20 nt sequences are also available, such as Cas-OFFinder (Bae *et al.*, 2014) and GGGenome (http://GGGenome.dbcls.jp/). These web servers are useful for designing gRNAs for a few input sequences, but processing large number of input sequences requires a laborious process. Even in such cases, CRISPRdirect returns the results quickly and provides a convenient interface for automated gRNA design as described in the Data export and API section, making it a powerful tool for using CRISPR/Cas system on a genome-wide scale.

2 Web Server Implementation

2.1 Overview

The web server accepts an accession number, a genome coordinate or an arbitrary nucleotide sequence up to 10 kbp as input (Fig. 1A) and returns a list of CRISPR/Cas target candidates. Target sequences of 20 nt adjacent to the PAM sequence (e.g. NGG, NRG) are searched from both strands of the input sequence and listed as shown in Figure 1B. The list contains target position, target sequence, additional information on the sequence and the number of target sites in the genome. The additional information on the sequences such as GC content and calculated melting temperature (Tm) are provided, since previous report suggested that sgRNA sequences with very high or low GC content were less effective against their targets (Wang *et al.*, 2014). The presence or absence of TTTT

(four consecutive T's that cause pol III termination) in the target sequence is also indicated in order to avoid TTTT in gRNA vectors with pol III promoter. A detailed description of the web server is provided in Supplementary Methods.

2.2 Off-target evaluation

The number of target sites in the genome (Fig. 1B) is counted using Jellyfish (Marçais and Kingsford, 2011). The column '20 mer+PAM' shows the number of hits with perfect matches for each target sequence (20 mer) adjacent to the PAM. Although the exact length of the completely complementary region necessary for cleavage by CRISPR nucleases is unknown, the mutations within the 'seed' sequence at 8–12 nt immediately adjacent to the PAM are known to impair cleavage, suggesting that this region is the most critical determinant of target specificity (Cong *et al.*, 2013; Fu *et al.*, 2013; Hsu *et al.*, 2013; Pattanayak *et al.*, 2013). Therefore, we built up the columns '12 mer+PAM' and '8 mer+PAM' in order to show the number of hits with perfect matches for their seed sequence (12 or 8 mer, respectively) adjacent to the PAM. Note that the numbers of hits displayed here include both on-target and off-target sites. For instance, one ('1') in these columns indicates that the sequence has only one perfect match with the intended target site. Any number greater than one indicates that there are some potential off-target sites. Thus, in terms of avoiding off-target editing, the smaller the number (but not zero) is, the better. Zero ('0') in these columns means that the sequence has no match in the genomic sequence; such sequences may possibly span over exon–exon junctions, so their use should be avoided. CRISPRdirect highlights the CRISPR/Cas targets that have relatively fewer off-target sites (Fig. 1B and C). A detailed list of off-target candidates can be investigated by clicking the 'detail' link (Fig. 1E). The searches allowing mismatches and gaps (insertions and/or deletions) are performed using GGGenome (http://GGGenome.dbcls.jp/) REST API developed by the authors' group instead of widely used BLAST (Altschul *et al.*, 1990), because BLAST may overlook some potential off-targets as mentioned in our previous work describing siDirect (Naito *et al.*, 2004), a web server for designing functional siRNA with reduced off-target effects. GGGenome quickly searches short nucleotide sequences utilizing suffix arrays and inverse suffix links indexed on solid state drive (SSD). As shown in Supplementary Table S1, off-target searches allowing gaps are not yet available in other existing web tools. However, the most recent report shows that CRISPR/Cas9 system has off-target activity with insertions or deletions between target DNA and gRNA sequences (Lin *et al.*, 2014). Therefore, we consider that off-target searches allowing mismatches and gaps would be a more suitable procedure to list off-target candidates exhaustively. The positions of the mismatches and gaps are visualized in the list (Fig. 1E), which may help predict the potency of off-target editing.

CRISPRdirect incorporates genomic sequences of various organisms to perform off-target searches. Although *Xenopus laevis* has long been used as a preferred model organism among developmental biologists, we incorporated *X.tropicalis* genome instead of *X.laevis* genome, because *X.tropicalis* is diploid while *X.laevis* is allotetraploid which makes it difficult to select specific targets.

There are some loci that are difficult to select specific targets. Typical examples are the histone clusters (NM_021059, etc.) and ribosomal proteins (NM_022551, etc.), which are known to form multigene families. When designing CRISPR targets for such genes, users should manually investigate a detailed list of potential off-target sites (Fig. 1E) and select the sequence that has fewer off-target hits on unrelated loci. Alternatively, if site-specific gRNA could not be designed within intended region, multiple gRNA approaches would be considerable (Guilinger *et al.*, 2014; Ran *et al.*, 2013; Tsai *et al.*, 2014). For such strategy, graphical view of CRISPRdirect results which visualizes the position and orientation of target sites (Fig. 1C) would be helpful for selecting paired gRNAs.

2.3 Data export and API

The results can be exported as tab-delimited text or in JSON format from the bottom of the result page (Fig. 1D). Users can copy–paste the text results into a spreadsheet or text editor for downstream analysis. The results can also be downloaded as a separate file by clicking the 'download' link. Alternatively, tab-delimited text or JSON output can be obtained via API, which is convenient for users to design a number of CRISPR/Cas targets in an automated manner.

References

Altschul,S.F. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

Bae,S. *et al.* (2014) Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*, **30**, 1473–1475.

Cho,S.W. *et al.* (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.*, **31**, 230–232.

Cong,L. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.

Fu,Y. *et al.* (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.*, **31**, 822–826.

Guilinger,J.P. *et al.* (2014) Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.*, **32**, 577–582.

Heigwer,F. *et al.* (2014) E-CRISP: fast CRISPR target site identification. *Nat. Methods*, **11**, 122–123.

Hsu,P.D. *et al.* (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.*, **31**, 827–832.

Jinek,M. *et al.* (2013) RNA-programmed genome editing in human cells. *eLife*, **2**, e00471.

Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.

Lin,Y. *et al.* (2014) CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res.*, **42**, 7473–7485.

Ma,M. *et al.* (2013) A guide RNA sequence design platform for the CRISPR/Cas9 system for model organism genomes. *Biomed. Res. Int.*, **2013**, 270805.

Mali,P. *et al.* (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823–826.

Marçais,G. and Kingsford,C. (2011) A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics*, **27**, 764–770.

Montague,T.G. *et al.* (2014) CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.*, **42**, W401–W407.

Naito,Y. *et al.* (2004) siDirect: highly effective, target-specific siRNA design software for mammalian RNA interference. *Nucleic Acids Res.*, **32**, W124–W129.

Pattanayak,V. *et al.* (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.*, **31**, 839–843.

Ran,F.A. *et al.* (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, **154**, 1380–1389.

Sander,J.D. *et al.* (2010) ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Res.*, **38**, W462–W468.

Tsai,S.Q. *et al.* (2014) Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.*, **32**, 569–576.

Wang,T. *et al.* (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, **343**, 80–84.