pgRNAFinder: a web-based tool to design distance independent paired-gRNA

Yuanyan Xiong\textsuperscript{1†}, Xiaowei Xie\textsuperscript{1†}, Yanzhi Wang\textsuperscript{1}, Wenbing Ma\textsuperscript{1}, Puping Liang\textsuperscript{1}, Zhou Songyang\textsuperscript{1,4,5,6,7*}, Zhiming Dai\textsuperscript{2,3,*}

1 Key Laboratory of Gene Engineering of the Ministry of Education and State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-Sen University, Guangzhou 510006, China, 2 School of Data and Computer Science, Sun Yat-Sen University, Guangzhou 510006, China; 3 Guangdong Province Key Laboratory of Big Data Analysis and Processing, Sun Yat-Sen University, Guangzhou 510006, China; 4 Cell-Based Assay Screening Core, 5 Dan L. Duncan Cancer Center, 6 Verna and Marrs Mclean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA, and 7 Key Laboratory of Reproductive Medicine of Guangdong Province, Guangzhou, China

These authors contributed equally to this work.
*Corresponding author: Zhiming Dai. Email: daizhim@mail.sysu.edu.cn Correspondence may also be addressed to Zhou Songyang. Email: songyang@bcm.edu

Abstract

Summary: The CRISPR/Cas System has been shown to be an efficient and accurate genome-editing technique. There exist a number of tools to design the guide RNA sequences and predict potential off-target sites. However, most of the existing computational tools on gRNA design are restricted to small deletions. To address this issue, we present pgRNAFinder, with an easy-to-use web interface, which enables researchers to design single or distance-free paired-gRNA sequences. The web interface of pgRNAFinder contains both gRNA search and scoring system. After users input query sequences, it searches gRNA by 3’ protospacer-adjacent motif (PAM), and possible off-targets, and scores the conservation of the deleted sequences rapidly. Filters can be applied to identify high-quality CRISPR sites. PgRNAFinder offers gRNA design functionality for 8 vertebrate genomes. Furthermore, to keep pgRNAFinder open, extensible to any organism, we provide the source package for local use.

Availability and implementation:

The pgRNAFinder is freely available at http://songyanglab.sysu.edu.cn/wangwebs/pgRNAFinder/, and the source code and user manual can be obtained from https://github.com/xiexiaowei/pgRNAFinder.

Keywords: CRISPR, Cas9, genome editing, large deletions.

1 Introduction

The RNA-guided CRISPR (clustered regularly interspaced short palindrome repeats) with its associated nuclease Cas9 was recently demonstrated to be versatile for genome engineering (Cong, et al., 2013; Jinek, et al., 2012; Mali, et al., 2013). Until now, studies have mainly focused on protein-coding genes, as shown in the genome-wide loss-of-function screening in human (Shalem, et al., 2014). Previous studies have also used CRISPR/Cas system to delete large sequence fragments and knock out non-coding genes in mammalian cells (Canver, et al., 2014; Fulco, et al., 2016; Ho, et al., 2015; Liu, et al., 2017). The CRISPR-Cas system delivers the Cas9 nuclease complexed with a synthetic guide RNA (gRNA) into a cell. The gRNA guides the Cas9 to cleave a desired DNA site in the cell's genome, allowing existing genes to be removed and/or new ones added. To design a “guide” sequence, the gRNA, which is typically 20 nts in length followed by a PAM sequence such as NGG, it is important to consider gRNA sequence itself and potential off-target sites (Fu, et al., 2013; Hsu, et al., 2014; Mali, et al., 2013; Pattanayak, et al., 2013). Many web-delivered and/or standalone software solutions are available for designing highly specific single and paired-gRNA for protein coding genes in vertebrate genomes (Hodgkins, et al., 2015; Perez, et al., 2017). Still, it is paramount to develop a new tool which can design distance-independent paired-gRNA for non-coding genes. Here we present pgRNAFinder, a web-based platform for gRNA design, which integrates three different tools for a comprehensive study of gRNA design based on different algorithms for the first time. Our web interface includes 8 vertebrate genomes for gRNA design. Furthermore, we provide a python-based package that allows users to design paired- and single-gRNA for any organism.
parameters and analyzing the potential off-target cleavage sites (Xie et al., 2014). Moreover, sgRNAas9 searches gRNA without limiting the PAM sequence or the number of mismatching bases. These ideas were applied in gRNA searching in the Step1 of pgRNAFinder. Step1 is performed by the script 01_get_sequence_for_search.py, 02_search_sgRNA.py and 03_get_sgRNA_candidate.py. SSC is a software package developed for predicting sgRNA efficiency from genomic sequences, and outperforms existing methods. Unlike other off-target evaluated methods that are based on the alignment of spacer sequences to the genome, SSC is based on modeling sequence features (Xu et al., 2015). Step2 includes SSC to predict sgRNA efficiency, which is performed by 04_sgRNA_efficiency.py. Off-Spotter is a tool that can quickly and exhaustively identify all genomic sites that satisfy the PAM constraint and are identical or nearly-identical to the provided gRNA (Platsiaka and Rigoutsos, 2015). Therefore, we included Off-Spotter in Step3, as a genome-wide off-target predictor. Step3 is performed by the script 05_offtarget_for_sgRNA.py.

3 Results and Discussion
The web interface of pgRNAFinder is divided into three modules: (i) Single gRNA design module and (ii) short distance paired-gRNA design module. (iii) long distance paired-gRNA design module. Users select or upload a target sequence or genome. There are three options implemented for user input or upload, entering a gene symbol, genome region (by gene structure query), or sequence directly. They will then be prompted to select a species and a gRNA design module. Following sgRNAas9, users can set the length of gRNA as an optional argument in the website, and the default offset of paired-gRNA ranges from -2 to 32 bp (base pairs) (Run, Hsu et al. 2013, Xie, Shen et al. 2014). The default offset of distance-free paired-gRNA ranges from 200-Skb, and the max distance is set as 1Mb, as 1Mb was tested to have very low probabilistic of deletion (Canver, et al., 2014). Max distance set determines the effective off-target pair number count and deletion frequency is inversely related to the deletion size, thus, those parameters need to be set carefully. The strand of gRNA is reported in the result, sense/antisense or both strand gRNA can be selected from the website, and the default offset of distance-free paired-gRNA is different, due to the Cas9nickase system (Run, et al., 2013).

The running time of pgRNAFinder is inversely correlated with gRNA number, and off-target evaluation is the time-consuming step. Off-Spotter is very fast as reported which can complete searching hundreds of thousands of potential off-targets in a few seconds (Platsiaka and Rigoutsos, 2015). For example, we selected human genomic loci chr1:69091-70008 as input and used the long distance paired-gRNA design mode to test the performance of pgRNAFinder. The result was reported in seconds. As shown in Figure S1, the results list the paired-gRNA sites of target loci chr1:69091-70008, information of genomic loci, GC content, strand, sequence, SSC's score, whether overlap with exon, pair's distance, off-target site number and coverage to the gene. Furthermore, the following information was also included: phastCons score, which can be used to check whether the target loci is evolutionary conserved. Moreover, evaluation of the off-target pair score is reported for paired-gRNA. The score is the sum of all intra-chromosomal off-target pairs that are smaller than the preset maximum distance (10kb by default). All information would help researchers choose a more functional and unique gRNA pair. The help page contains detailed information about using the web interface and the illustration of output.

In conclusion, our pgRNAFinder tool has a user-friendly interface that speeds up designing single- and paired-gRNA. Meanwhile, PgRNAFinder provides potential off-target paired number and conservation information of target sequence to improve the gRNA selection.

Funding
The research has been supported by National Natural Science Foundation of China (NSFC) (Grant 31171397, 31271533, and 31301085), Pearl River Nova Program of Guangzhou (2012TS0739 and 201710010044) and National Natural Science Funds of Guangdong Province for Distinguished Young Scholar (2014A030360644).

References