CRISPR Optimal Target Finder

OVERVIEW

CRISPR Optimal Target Finder uses the TagScan[1] genome searching algorithm to identify all sequences with similarity to CRISPR target queries in selected genomes. These sequences are filtered using algorithms based on large-scale analyses of CRISPR/Cas9 specificity in cell lines[2-7] and animals[8,9] to identify potential off-target cleavage sites for a given CRISPR target. These and earlier studies demonstrate that the PAM-proximal region of the CRISPR target sequence (also referred to as the "seed") is more critical for specificity than the distal 8 nucleotides (Fig. 1)[2-7,9-11]. Our algorithms consider both the number and location of mismatches to evaluate all potential off-target cleavage sites.

Fig. 1. Example target sequence.

```
Distal   Proximal     PAM
GATCGATG|TTGAATGCCGAT TGG
```

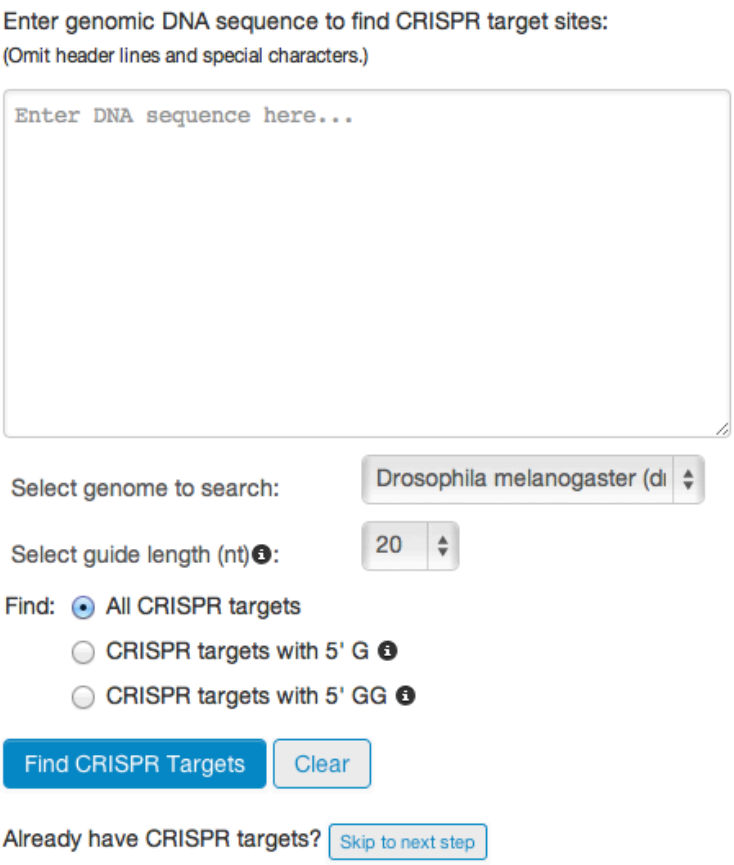
In transformed cell lines, CRISPR sequences adjacent to an NAG PAM sequence can also be cleaved at ~20% efficiency[5]. This has not been observed in animals to date, so the program allows the user to choose whether or not NAG-adjacent sites are considered in the evaluation of CRISPR target sequence specificity.

The rules behind our algorithms are detailed below and will be updated as new data becomes available.

Step 1. Find CRISPR Targets:

Paste the genomic DNA sequence you wish to target in the window (up to 10 kb is allowed). Include only ATGC, no headers or special characters. Select the genome you wish to search. The following options are currently available:

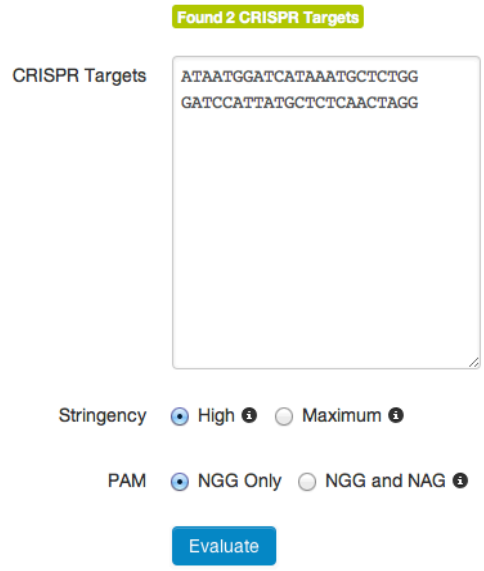
Drosophila melanogaster (dm3)
D. simulans (annotation DsimV2)
D. simulans (DroSim1)
D. yakuba (DroYak2)
D. mauritiana (Dmau_MS17)
D. sechellia (DroSec1)
D. ananassae (dana_r1.3_FB2011_07)
D. erecta (dere_r1.3_FB2011_08)
D. persimilis (dper_r1.3_FB2010_02)
D. pseudoobscura (dpse_r3.1_FB2013_03)
D. virilis (DroVir3)
D. mojavensis (dmoj_r1.3_FB2011_05)
D. willistoni (dwil_r1.3_FB2010_02)
D. grimshawi (dgri_r1.3_FB2010_02)
Anopheles gambiae (AgamM1)
Anopheles gambiae (AgamS1)
Aedes aegypti (AaegL1)
Apis mellifera (apiMel3)
Tribolium castaneum (TriCas2)
T. castaneum (T cas 4.0 draft)
Caenorhabditis elegans (ce10)



Specify whether you would like the program to identify all CRISPR Targets in the input sequence, only those that start with a G for efficient expression from the U6 promoter when supplying gRNA as plasmid DNA, or only those that start with GG for efficient expression from the T7 promoter if preparing gRNA as RNA. Alternatively, we and others have found that replacing the 5' nucleotide with a G or simply adding a G to the 5' end of your gRNA yields sufficient transcription. All gRNAs with strings of 5 or more Ts, which serve as termination sequences for PolIII, are excluded. Using a shorter gRNA can increase Cas9 nuclease specificity[12]. Choose your desired gRNA length with the dropdown box. Click the Find CRISPR Targets button to return all CRISPR targets within your genomic sequence.

If you already have CRISPR target sequences to evaluate, click Skip to next step and insert those CRISPR sequences in the following page (step 2 below). Include the PAM sequence for each target site you enter.

Step 2. Evaluate CRISPR Targets:



The CRISPR Target window lists all identified CRISPR targets and their associated PAM sequences. In this example 2 CRISPR target sites beginning with G were identified.

To evaluate these sequences for potential off-target cleavage sites, select the following parameters:

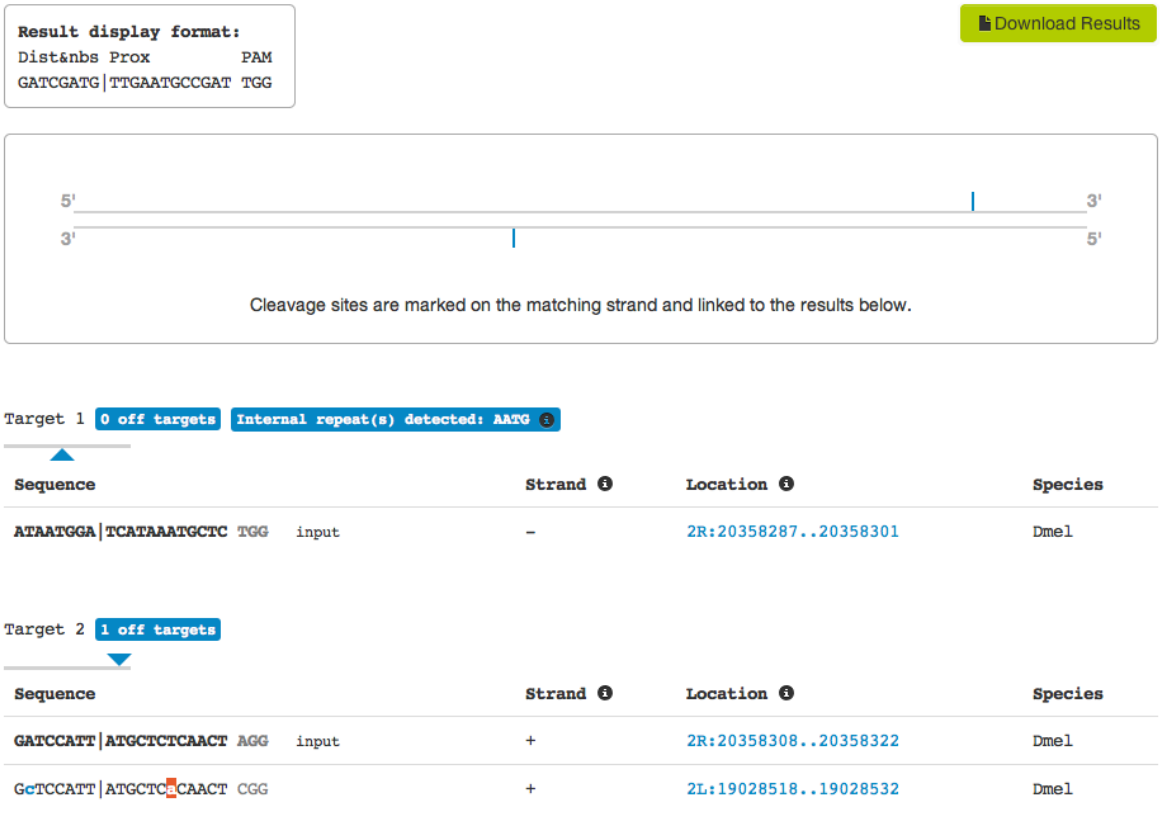
1. Stringency
Maximum stringency criteria are based on the composition of off-target cleavage sites observed in large-scale analyses carried out in cell lines[2-7] and define potential off-target sites elsewhere in the genome as sites with:
i. perfect matches (0 mismatches) to the proximal sequence
ii. 1 mismatch in the proximal sequence AND fewer than 5 mismatches in the distal sequence
iii. 2 mismatches in the proximal sequence AND fewer than 2 mismatches in the distal sequence

High stringency criteria are based on the composition of off-target cleavage sites observed in large-scale analyses conducted in animals[8,9] and define potential off-target sites elsewhere in the genome as sites with:
i. perfect matches (0 mismatches) to the proximal sequence
ii. 1 mismatch in the proximal sequence AND fewer than 2 mismatches in the distal sequence.

2. PAM sequence
Choose whether or not NAG-adjacent sequences are included in the evaluation of potential off-target sites.

Click Evaluate and the program will use the selected parameters to identify all potential off-target sites for each CRISPR target sequence entered.

Step 3: Results:



If a genomic DNA fragment was entered, a schematic illustrating the locations of valid target sites within the fragment is displayed. The target sites are marked on the matching strand with respect to the input

sequence. The target sites are listed below the schematic and returned in order from ‘best’ to ‘worst’ based on the nature and number of potential matches elsewhere in the genome. The ‘best’ CRISPR targets have no potential off-target matches elsewhere in the genome based on the specified parameters. In the example above, the first two target sequences evaluated meet this standard using the ‘high’ and ‘NGG only’ parameters. CRISPR targets for which potential off-target cleavage sites were identified are ranked next. CRISPR targets with off-target sites that all contain mismatches in the PAM-proximal region, which significantly reduce off-target cleavage, are ranked above CRISPR targets with any off-target sites containing perfect matches to the proximal sequence. Within each category, CRISPR targets are ranked by the total number of off-target sites identified. In the example above, the second-best CRISPR target has a single predicted off-target site that contains one mismatch in the proximal region and one mismatch in the distal region.

For each evaluated target sequence, the program returns the full list of the potential off-target sites. Mismatches to the target sequence are highlighted in either red (PAM-proximal) or blue (distal region). The strand, chromosome location and a GBrowse link are provided for each off-target site.

Results data can be downloaded as a CSV (Excel) file by clicking the 

button.

References

1. Iseli, C., Ambrosini, G., Bucher, P. & Jongeneel, C.V. Indexing strategies for rapid searches of short words in genome sequences. *PLoS ONE* **2**, e579 (2007).

2. Cho, S.W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res* **24**, 132-41 (2014).

3. Cradick, T.J., Fine, E.J., Antico, C.J. & Bao, G. CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res* **41**, 9584-92 (2013).

4. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**, 822-6 (2013).

5. Hsu, P.D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827-32 (2013).

6. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-6 (2013).

7. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* **31**, 839-43 (2013).

8. Chiu, H., Schwartz, H.T., Antoshechkin, I. & Sternberg, P.W. Transgene-free genome editing in Caenorhabditis elegans using CRISPR-Cas. *Genetics* **195**, 1167-71 (2013).

9. Yang, H. *et al.* One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* **154**, 1370-9 (2013).

10. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A* **108**, 10098-103 (2011).

11. Wiedenheft, B. *et al.* RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc Natl Acad Sci U S A* **108**, 10092-7 (2011).

12. Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M. & Joung, J.K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* **32**, 279-84 (2014).