sgRNA Scorer 2.0 - Manual

Getting started

Name:

E-mail address:

Please make sure to put in a valid e-mail address as a link to the results will be emailed to you

CRISPR system

CRISPR system: | S. pyogenes (PAM: NGG) |

| ✓ S. pyogenes (PAM: NGG) |
| S. aureus (PAM: NNGRRT) |
| AsCpf1 (PAM: TTTN) |
| N. meningitidis (PAM: NNNNGATT) |
| S. thermophilus 1 (PAM: NNAGAAW) |
| S. thermophilus 3 (PAM: NGGNG) |
| Other |

CRISPR System: There are six preloaded systems. S. pyogenes is default, which is the most commonly used system. In addition, we have included systems for which we have some level of experimental validation. For both S. aureus Cas9 and S. thermophilus 1 Cas9, our algorithm does very well in terms of predicting activity. There is also an option for "Other" if the CRISPR system does not exist in our system.

Custom settings

Spacer length (14-24): | 20 | PAM sequence (e.g. NGG): | NGG | PAM orientation: | 3' |

Spacer Length

| 14 |
| 15 |
| 16 |
| 17 |
| 18 |
| 19 |
| ✓ 20 |
| 21 |
| 22 |
| 23 |
| 24 |

For Cas9 S. pyogenes, this value is 20. This can be set to anywhere between 14 and 24.

PAM sequence

For Cas9 S. pyogenes, this value is NGG. This can be customized if a new CRISPR system is specified. Valid characters must be IUPAC symbols

PAM orientation

For Cas9 S. pyogenes, this value is 3' while for Cpf1, this value is 5'. If 5' is specified, the reverse weighting scheme will be used

sgRNA Scorer 2.0 - Manual

Sequence input

Select input (paste sequence or upload file, FASTA format, limit to 1 Mb of sequence total)

◉ Paste sequence

◯ Upload file

[Choose File] no file selected

The user has the option of either pasting in sequences in FASTA format or uploading a FASTA file.
Please try to avoid using special characters in the sequence name such as quotation markers or apostrophes
FASTA format is as follows:
>sequence1Name
ATTGGTGGTGTGGGTAACCCT
>sequence2Name
GGCCTGACCGTGCC

....

Off-target analysis

Run CasFinder off-target analysis? ☐ [Human (Hg19)]

The user has the option of performing off-target analyses of the sgRNAs that are identified from the inputted sequence. The algorithm used is CasFinder by Aach et al (2014) BioRxiv. We support off-target analysis in the species listed on the right and in parentheses, is the version of the genome in the UCSC database. If this checkbox is not selected, sgRNA scorer will just identify all possible sgRNAs based on the chosen parameters and score them with our model.

- ✓ Human (Hg19)
- Mouse (Mm10)
- Cat (Fc5)
- Chicken (Gg4)
- Cow (Bt8)
- C. elegans (Ce10)
- Dog (Cf3)
- Drosphila melanogaster (Dm6)
- African Elephant (loxAfr3)
- Pig (susScr3)
- Rat (Rn6)
- Rhesus macaque (rheMac3)
- S. cerevisiae (sacCer3)
- Zebrafish (Dr10)

sgRNA Scorer 2.0 - Manual
Results zip file

In each zip file containing the results, the following files are present (shown by file extensions):

*.fasta - the file containing the original sequence(s) that were uploaded or pasted in
*.FinalOutput.txt - result file containing sgRNA sequences identified and scored

If CasFinder was run, the following additional files are generated. Please see the CasFinder documentation (http://biorxiv.org/content/biorxiv/suppl/2014/05/12/005074.DC1/005074-1.pdf) as to what each of these files are:
*.CasValue_v2.pl.info.txt, *.CasFinderOutput.data.txt, *.CasFinderOutput.info.txt,
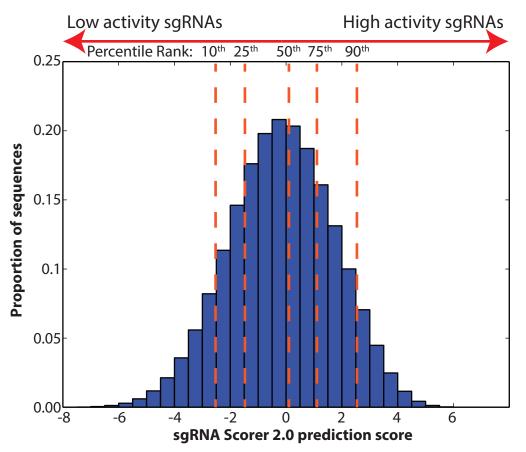***.CasFinderOutput.stat.txt**

Description of result file

SeqID - unique identifier for the sgRNA sequence

Sequence - nucleotide sequence of the sgRNA sequence (including PAM)

Score - score of sgRNA sequence based on our classifier. The higher the score, the better the predicted activity. Guide scores should typically be used in a *relative* manner as opposed to *absolute*. For example, if one guide is scored 2.5 and another guide is scored 1.3, the first guide would be considered better than the second guide. The score is purely based on on-target activity and **does not** incorporate off-target activity. Figure 1 below shows the distribution scores of ~7,500,000 guide RNA sequences when classified by sgRNA Scorer 2.0

# of hits - the total number of target sites (including unintended off-targets) of the guide in the genome analyzed.

Figure 1. Score distribution of ~7,500,000 sgRNA sequences classified with sgRNA Scorer 2.0. The orange lines (with scores in parentheses) indicate the 10th (-2.59), 25th (-1.43), 50th (-0.14), 75th (1.15), and 90th (2.27) percentiles.