

The standalone version of sgRNA Scorer 2.0 will identify and score guide RNA sequences identified from an input FASTA sequence. There are 4 python scripts in total. The user only needs to run one wrapper script called "identifyAndScore.py" which can be run in your terminal using the python command. An example file containing the eGFP sequence is provided for testing.

**Requirements:**

Python 2.7 or greater; python packages: BioPython and and Sci-kit learn.

**Description of main python script (identifyAndScore.py):**

The wrapper python script which will run the other three python scripts in the directory. The parameters can be listed either as single dash (-i) or double dash (--input) when specified. The PAM sequence parameter single dash is lower case L.

**Required parameters:**

- i / --input: input file in FASTA format
- o / --output: final output file with guide RNA sequences and scores
- p / --pamOrientation: PAM orientation. Value must either be 5 or 3
- s / --spacerLength: length of spacer sequence. For example, SpCas9 uses a 20 bp spacer and SaCas9 uses a 21 bp spacer
- l / --pamSequence: PAM sequence. For SpCas9, the value here is NGG and for SaCas9, the value would be NNGRRT

**Examples of how you would run the script (using eGFP sequence file as an example):**

Identifying and scoring all SpCas9 guide RNA sites:

```
python identifyAndScore.py -i eGFP.fasta -o eGFP.SpCas9.tab -p 3 -s 20 -l NGG
```

Identifying and scoring all SaCas9 guide RNA sites:

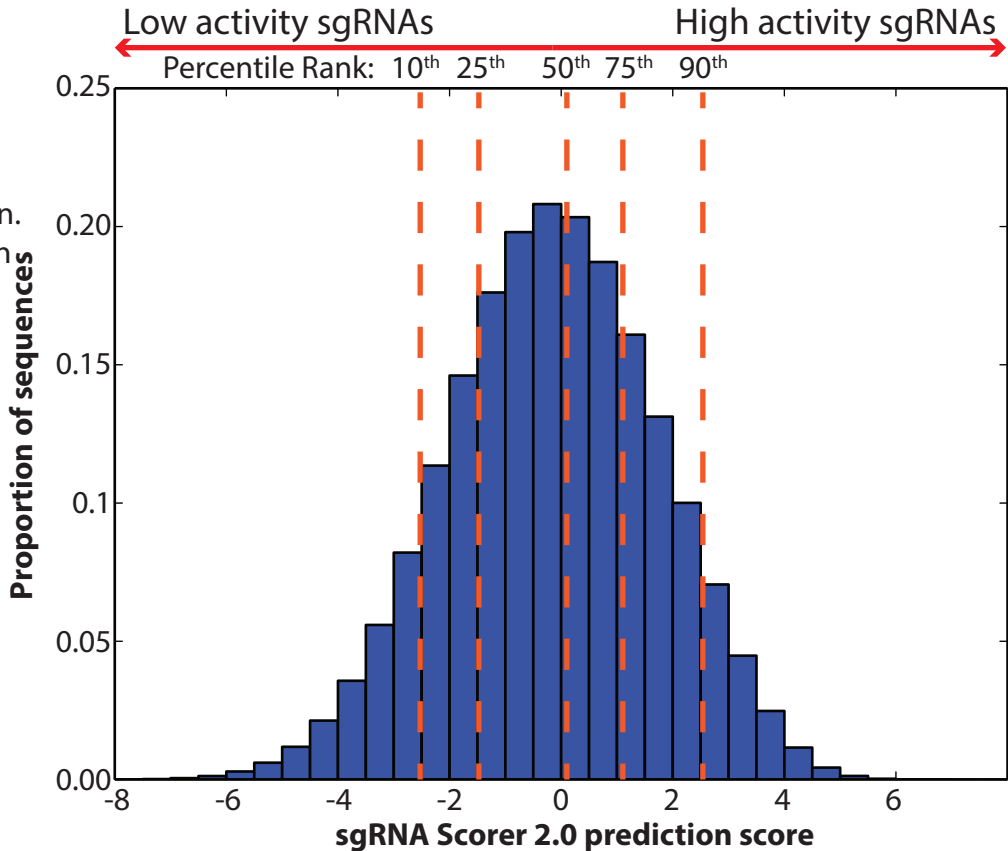
```
python identifyAndScore.py -i eGFP.fasta -o eGFP.SaCas9.tab -p 3 -s 21 -l NNGRRT
```

Identifying and scoring all AsCpf1 guide RNA sites:

```
python identifyAndScore.py -i eGFP.fasta -o eGFP.Cpf1.tab -p 5 -s 20 TTTN
```

**Interpreting your results**

The score value returned for the guide RNA ranges from -8 to 8. In figure 1 (right), we generated over 7,500,000 guide RNA sequences and plotted the score distribution. For example, a score under -2.59 would be in the lowest 10% of predicted activity while a score greater than 2.27 would be in the upper 10%.



**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.