

## Calculation of Rule Set 2 score

### Requirements:

Software: Python-2.7

Packages:

1. scikit-learn 0.16.1
2. pickle
3. pandas
4. numpy
5. scipy

### Inputs:

1. `--seq` : 30mer nucleotide sequence, which should be of the form NNNN20merNGGNNN
2. `--aa-cut` : This is an optional argument. Specify the position at which the sgRNA cuts the amino acid sequence. Default value: None
3. `--per-peptide` : This is an optional argument. Specify the percentage of protein cut by the sgRNA. Default value: None

The optional arguments are helpful for prediction but if this information is unknown, a separate model will be used for prediction where these features are turned off.

### Output:

The Rule Set 2 score for the input sequence will be printed on the terminal.

### Running the code:

To run this code, make the “analysis” folder your working directory.

- To run full model prediction, type on the terminal:  
`python rs2_score_calculator.py --seq <30mer sequence> --aa-cut <amino acid cut position> --per-peptide <Percent peptide cut by sgRNA>`
- To use only sequence information, type on the terminal:  
`python rs2_score_calculator.py --seq <30mer sequence>`

### Examples:

```
python rs2_score_calculator.py --seq TGGAGGCTGCTTTACCCGCTGTGGGGGCGC
```

Output: Rule set 2 score: 0.5656

```
python rs2_score_calculator.py --seq TGGAGGCTGCTTTACCCGCTGTGGGGGCGC --aa-cut 254 --per-peptide 87
```

Output: Rule set 2 score: 0.5309