

# IFN646 - Portfolio item 3

---

## Key points

- Individual assignment
  - 6 marks
  - Release date: 7 September 2023
  - Due date: 24 September 2023
- 

## Overview

For this portfolio item, you will focus on evaluating CRISPR guide RNA sequence specificity.

You will need to download some datasets, develop some code, and upload it to Canvas. This code will be marked automatically, so it is important that you respect the structure we are giving you as a scaffold.

---

## 1. Data for the portfolio item

- A number of complete genomes are available. You are allocated a single one, following this rule:
  - If your student number ends between 00 and 24, you will work with *E. coli*.
  - If it ends between 25 and 49, you will work with *S. aureus*.
  - If it ends between 50 and 74, you will work with *S. bongori*.
  - And if it ends between 75 and 99, you will work with *S. pyogenes*.
- A set of guides. Again, you are allocated only one of them:
  - If the previous two digits in your student number are between 00 and 49, you will work with the set of reads that corresponds to that number.
  - If they are between 50 and 99, subtract 50, and work with the corresponding set of reads.

For instance, if your student number is n10764294, you will work with *S. pyogenes* (because it ends with 94), and with the *S\_pyogenes.fa-42-guides.txt* set (because the previous two digits are 42).

Similarly, if your student number is n9188414, you will work with *E. coli* (because it ends with 14), and with the *E\_coli.fa-34-guides.txt* set (because the previous two digits are 84, and  $84 - 50 = 34$ ).

All the data is available on Canvas. Make sure to download the genome and the set of guides that you have been assigned in the question shown on Canvas.

If you do not work with the assigned set of guides, your results will be incorrect and you will not be able to get full marks.

---

## 2. Tasks

### Overall goal

As we discussed in the lecture, specificity is a crucial property when designing guides. You want to minimise the risk of off-target modifications.

You will implement some code for the two methods covered in the lecture: identifying the number of sites with 1, 2, 3 or 4 mismatches, and calculating a risk score based on the position of these mismatches.

We have provided some code, to get you started and to help you with an ambiguous aspect of the Zhang score (how  $d$  is calculated).

### Task 1 [2 marks]

Extract all off-target sites for your allocated genome.

Complete the method `TaskA(..)` in the Python boilerplate code. Your method is to return a list of off-target sites. It will look like a longer version of this example:

```
[
    'CGTACAGGAAACACAGAAAA',
    'AGAAAAAAGCCCGCACCTGA',
    'AAAGCCCGCACCTGACAGTG'
]
```

### Task 2 [2 marks]

For your given set of guides, find how many off-target sites exist for each guide.

Complete the method `TaskB(..)` in the Python boilerplate code.

Your method is to return a list of lists, where the sub-lists contain items in the following order:

1. The guide sequence
2. The number of identical off-target sites
3. The number of off-target sites that contained one mismatch to the guide sequence
4. The number of off-target sites that contained two mismatches to the guide sequence
5. The number of off-target sites that contained three mismatches to the guide sequence
6. The number of off-target sites that contained four mismatches to the guide sequence

For example:

```
[
    ['GTCACCACCGTCGCCAGGACCGG', 0, 1, 3, 2, 3],
    ['GGTTGACTCTGGCGTAGGTCCGG', 1, 0, 1, 2, 1],
    ['AGCCAGATAATATAGTGTTTAGG', 1, 0, 0, 0, 1]
]
```

### Task 3 [2 marks]

Implement the off-target score by Zhang. Evaluate the specificity of each candidate guide.

Complete the method `TaskC(...)` and `SingleScore(...)` in the Python boilerplate code.

Your method is to return a list of lists, where the sub-lists contain items in the following order:

1. The guide sequence
2. The Zhang score for this guide

A sample of this is below.

```
[
    ['GTCACCACCGTCGCCAGGACCGG', 70.23],
    ['GGTTGACTCTGGCGTAGGTCCGG', 85.91],
    ['AGCCAGATAATATAGTGTTTAGG', 98.32]
]
```

---

## 3. Submission

You will submit your code on Canvas using the "Portfolio Item 3" link.

Submission will close at 11.59pm on the due date.

---

## 4. Academic honesty

This is an individual assessment, and you need to submit your own work. Standard plagiarism checks will be performed on your code. We also reserve the right to select some submissions and ask students to explain the reasoning behind their answers.