

Hello there!

We hope you are doing well and we are looking forward to meeting you in the next few days. In order to oriente the discussions of that meeting, we ask that you take a look at the following pre-interview exercises.

This is not a an exam, and we don't plan to evaluate you in a pass/fail contexte.

We recommand spending no more than 1 to 2 hours on these exercices. **Please deposit any data, notes and/or code used to complete the exercises in a repo on Github that you can share with us.**

Exercise 1

Retrieve the mRNA sequence of the **NPM1 gene** and download it in the FASTA format. Consider the transcript variant 1.

1. What is the length of the sequence in nucleotides? In codons?
2. What is the GC% of the sequence?
3. Is the CTTAGTAGCTGTGGAGGAA motif present in the sequence? If so, how many time? At which position.s?

Exercise 2

Consider the content of `exercise_2_sequence.fasta` and the sequence downloaded from **Exercise 1**. How similar (or disimilar) are the two sequences? How can you quantify their similarity (or disimilarity)?