Homework Exercise 9

1)

Read about confounding.

Come up with 2 examples of two measurable variables in the context of molecular biology where one is **causally** affecting another (with a well-understood mechanism). Give 2 examples of two measurable variables where you would expect to find an association/correlation between them due to confounding, even though none is causally affecting the other (directly or indirectly). Briefly explain your reasoning.

2)

A)

The threonine amino-acid is coded by 4 different codons. Is any of these codons more likely to code for threonine residues that undergo phosphorylation (resulting phosphothreonine) in human?

B)

We are going to check if the results you arrived at are statistically significant, focusing on the ACG codon.

Watch this introductory video on Fisher's exact test.

Build the following 2x2 contingency table as a Pandas DataFrame:

	Phosphorylated amino-acid	Not phosphorylated
ACG codon	C1	C2
Other codons	C3	C4

Where:

- C1 - The number of threonine residues coded by the **ACG codon** that **undergo phosphorylation**.
- C2- The number of threonine residues coded by the **ACG codon** that **do not undergo phosphorylation**.
- C3 - The number of threonine residues coded by **other codons** that **undergo phosphorylation**.
- C4- The number of threonine residues coded by **other codons** that **do not undergo phosphorylation**.

Use scipy.stats.fisher_exact to obtain a p-value.

Are threonine residues coded by the ACG codon significantly more likely to undergo phosphorylation?

C)

What could explain such significant association? Suggest at least one confounding explanation, and at least one hypothetical biological mechanism that could lead synonymous codons in the RNA level (coding for the same amino acid in the protein level) to affect the phosphorylation outcome of the residue.

Bonus Questions

1)

Play with some Jupyter widgets and demonstrate their capabilities. Show some scenarios in which they can be useful. Make sure to mention progress bars (e.g. FloatProgress).

2)

Explain some of Python's special functions, such as *__len__*, *__str__*, *__repr__*, *__eq__* and *__hash__*. In what scenarios can it be useful to implement these functions in your class? Demonstrate by complementing the DNASeq and RNASeq classes we defined in class with those functions.