

# IFN646 - Portfolio item 4

---

## Key points

- Individual assignment
  - 6 marks
  - Release date: 22 September 2023
  - Due date: 15 October 2023
- 

## Overview

For this portfolio item, you will focus on differential microbial abundance.

You will need to download some datasets, develop some code, and upload your analysis to Canvas. We strongly recommend using a Jupyter Notebook and submitting an HTML export.

---

## 1. Data for the portfolio item

You are provided with three files:

- A BIOM table that contains abundance counts for various OTUs across a number of samples from the throat and nose of smokers and non-smokers.
- A metadata file that provides additional information about each sample.
- A taxonomy file that allows you to link the ID of each OTU to an actual name.

All the data can be downloaded from Canvas.

---

## 2. Tasks

### Overall goal

The overall goal of the analysis is to understand how smoking impacts the nose and throat microbiomes.

You will perform your analysis using Calour, as we have done in class. Note that for some questions, you may need to write some additional code to process the results.

## Task 1 [1.75 marks]

Identify which bacteria have a differential abundance between the nose and throat (irrespective of the smoking status). [1.5 marks]

Try to report names rather than simply the ID of the OTUs. [0.25 mark]

Note that the taxonomy data is hierarchical. For each line, you get:

- the ID
- the kingdom, with prefix `k__`
- the phylum, with prefix `p__`
- the class, with prefix `c__`
- the order, with prefix `o__`
- the family, with prefix `f__`
- the genus, with prefix `g__`
- the species, with prefix `s__`

If genus and species are available, you should give both. For instance, for ID `1097955`, you should report `Staphylococcus aureus`. Note that we remove the prefixes.

If the species is not available, give the genus only. For instance, for ID `242061`, you should report `Catellatospora`.

If both the genus and species are unavailable, report the most informative level. For instance, `Gaiellaceae` for `104172`, or `Chloroplast` for `47632`.

## Task 2 [1.75 marks]

Identify which bacteria have a differential abundance between smokers and non-smokers (irrespective of the sampling site). [1.5 marks]

Try to report names rather than simply the ID of the OTUs. [0.25 mark]

## Task 3 [2.5 marks]

Identify which bacteria have a differential abundance in the nose between smokers and non-smokers. [1 mark]

Hint: look at the `filter_samples()` method in Calour.

Identify which bacteria have a differential abundance in the throat between smokers and non-smokers. [0.5 mark]

Are these results the same as those obtained in Task 2? Explain why. [0.75 mark]

Try to report names rather than simply the ID of the OTUs. [0.25 mark]

### **3. Submission**

You will submit your analysis through Canvas through the relevant link.

We strongly recommend using a Jupyter Notebook, export your work as an HTML file, and upload this as your submission.

Alternatively, if you prefer not to use Jupyter, you can submit a PDF report that contains both your answers for each task and your Python code as appendix.

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.