

Bioinformatics Project

NOTE: In all Jupyter or R Notebooks below, you must include comment explanations for each code block so that I know you understand what the code is doing.

Select **any 2** of the following:

- 1) RNA Seq: Use the `seurat` package in R to analyze a new vignette. I suggest the cell cycle regression or differential expression testing as possibilities. **Turn in an R Notebook that will run in R studio.**
<https://satijalab.org/seurat/vignettes.html>
- 2) QIIME2: Analyze the Parkinson's Mouse model data set. Produce 3 analyses - create and **turn in 3 different .qzv files** which can be visualized at <https://view.qiime2.org/>
<https://docs.qiime2.org/2023.2/tutorials/pd-mice/#>
- 3) Machine Learning: Complete and **turn in a Jupyter Notebook** analyzing the `gum_pocket_data.txt`:
`MachineLearningProject.ipynb`
- 4) Metagenomics: Use `fast-p` to process the metagenomic file from lab:
insub732_2_R2_fastp.fastq.gz.
 - a. Run through `kaiju` using the nr database. This uses a lot of RAM, so if you do not have a machine with this much RAM you might be able to run it on the `kaiju` webserver: <http://kaiju.binf.ku.dk/server>
 - b. Alternatively, you can use another metagenomics taxonomy program.
 - c. Once you have the output data, make a Jupyter notebook that uses either Python (`pandas`, `ggplot2`) or R to produce graphs of the numbers of bacteria, archaea, eukarya or viruses in the dataset. Also plot the 5 most abundant genera.

Turn in the Jupyter Notebook or the R Notebook along with the output data from Kaiju.