**Post Genomics – Fall 2025**

**Homework 3**

**Due: Thursday, September 18, 2025**

**Part 1 (Programming)**

Using the CSV outputs from Homework 2, fill in the python file, Homework3.py, so that it would:

1.1 Merge the 5 normal CSV files together and do the same to the 5 tumor CSV files. The result should be 2 new CSVs, one with Normal variants and another with Tumor variants.

* 1. Run the provided addALT\_Seq().
  2. Remove/Drop duplicate rows (or variants), based on the columns: [“chrom”, “left”, “ref\_seq”, “alt\_seq”, “Patient\_ID”].
  3. After completing the steps above, print the columns and rows of the two Dataframes (code is given in Homework3.py). Provide **whole** screenshot of print statements output with run line. See example below:

A screen shot of a computer

Description automatically generated

**Part 2 (RNAseq analysis)**

This section will walk you through analyzing differences in gene expression of gastric corpus biopsies, both normal and atrophic. Download the transcriptomic zipped folder and extract it. Go step by step through the R commands to determine the significantly up and down regulated genes. (If you use a non-biolinux machine, in R there are three packages that will need to be installed: tximport, DESeq2, and tidyverse if not already done so.) The commands also assume your working directing is in the transcriptomics folder you extracted. Answer the following questions using the last CSV file created and the correlation plot:

1. How many genes were significantly different (adj-p < 0.05)?
2. What does it mean when the log2 expression difference is less than 0?
3. How many genes were upregulated? Downregulated?
4. How many genes are left if you look at differential expression greater than log2 differences greater than 2? How about less than -2?
5. Using the correlation plot from the R code, is there a single sample that may influence the overall reliability of the analysis? Explain how you decided on your answer.

**What to Submit:**

1. **A single PDF with screen shot of output from Part 1.**
2. **Python file with code in Part 1.**
3. **Word document with the answers to part 2.**

**\* Name the files (Your Last Name)\_HW3.pdf and (Your Last Name)\_HW3.py \***