

# BIOI 3500, Advanced Bioinformatics Programming Assignment 8



Assignment 8 :: 50 points :: **Due: 11:59 p.m. Friday, April 3<sup>rd</sup> 2021**

## Objectives

This assignment will give you the opportunity to practice using PostgreSQL.

## Exercises

Answer the following questions using the *gene2pubmed* and *geneinfo* tables as created in the video module “Data-bases (Part 2)”. Note that the taxonomy ID for humans is 9606 and that the taxonomy ID for mouse is 10090. As mentioned in the video lecture, the database I created is a much smaller version of what is in NCBI, to allow you to run it on the Virtual Machine. Therefore, the answers to some questions (e.g., question 4) are not going to be biologically correct.

For each question, record the complete SQL statement(s)—Questions 1 through 8 must be answered using only a single SQL statement; two statements can be used to answer Questions 9 and 10—that you used along with your answer in a text file. Name this file **netID\_Assign8** where **netID** is your UNO NetID. Be sure to include your name in the file.

1. What is the **gene\_id** of the gene with **symbol** CD19?
2. What is the **symbol** and **full name** of the gene with **gene\_id** 7157?
3. How many genes are on **chromosome** 1 (in mouse and human, total)?
4. How many genes are on human **chromosome** 1?
5. What synonyms exist for the gene with the **symbol** MYC?
6. How many PubMed entries are there for **gene\_id** 4609?
7. How many PubMed entries are there for genes with **gene\_ids** 1787 and 7157?
8. How many genes in the *geneinfo* table are human (tax\_id = 9606)?
9. How many PubMed entries are there for the human gene with full name dystrophin?
10. Given that PubMed ID's are issued consecutively, what is the **pubmed\_id** of the most recent paper in PubMed about the human gene TP53?

## How to Submit

Upload your text file to Canvas.