Nikki Tirrell

AS.410.712.81

April 2023

Final Project Proposal

The Need:

I am in a research group that is taking a computational biology approach to personalized medicine for people with rare diseases. We have access to extensive genomic data for our patients, but the information is spread across multiple pdf documents and spreadsheets. The genomic variants are used to guide our modeling, so it is important to have it readily available. We are racing against time to draw conclusions about their health but using non-optimized methods to store and access their data. I have met other computational biologists and systems biologists that are doing this same type of modeling, so I know that many people would be interested in a better solution.

My Proposed Tool:

I will make a tool that will allow us to store genomic information and easily parse it. It will consist of several components:

* A MySQL database schema that stores the data. This will include basic patient information of patient name and demographics to let the database handle patient data from multiple patients and take into account other determinants of health. There will also be tables for the genomic variants that a patient has, information about those genes, and the frequency of that variant. Finally, there will be tables to store the name of the models created, the processes included in the model, and the genes involved in each process.
* An HTML form with search fields of patient name, patient date of birth and gene name to search for. As a triplet, I know that patient DOB is not a unique enough identifier, and names are also not unique. So, both fields are important to use in the search.
* An HTML page that displays the search results and has been formatted with CSS and JavaScript. This includes the patient information, gene name, gene product, and gene variant information.
* A Python CGI script to take the input from the HTML form, search the MySQL database, and output the search results to the results page

To test this tool, I will create a MySQL database according to the above schema and populate it with sample patient and genomic data.

Significance:

This tool will allow for extensive genomic variant information to be saved in an accessible database. This will cut down on the time that researchers spend looking for relevant information, so that we can focus on our modeling. While the database will take time to populate, this cost will pay off since the information will be simpler to access for the entirety of the model development. My tool will also allow for data from multiple patients to be stored so that we can compare genomic information of health and ill individuals. Finally, the tool will keep track of what models the genomic variants are implemented in, which may aid in drawing connections between biological processes that have not already been made. In time, this tool could also be expanded to include health biomarker information, which is also relevant data in the modeling process.