MetasDB

BIOI 4870 Final project

Zackary Wright

2021

**Abstract\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

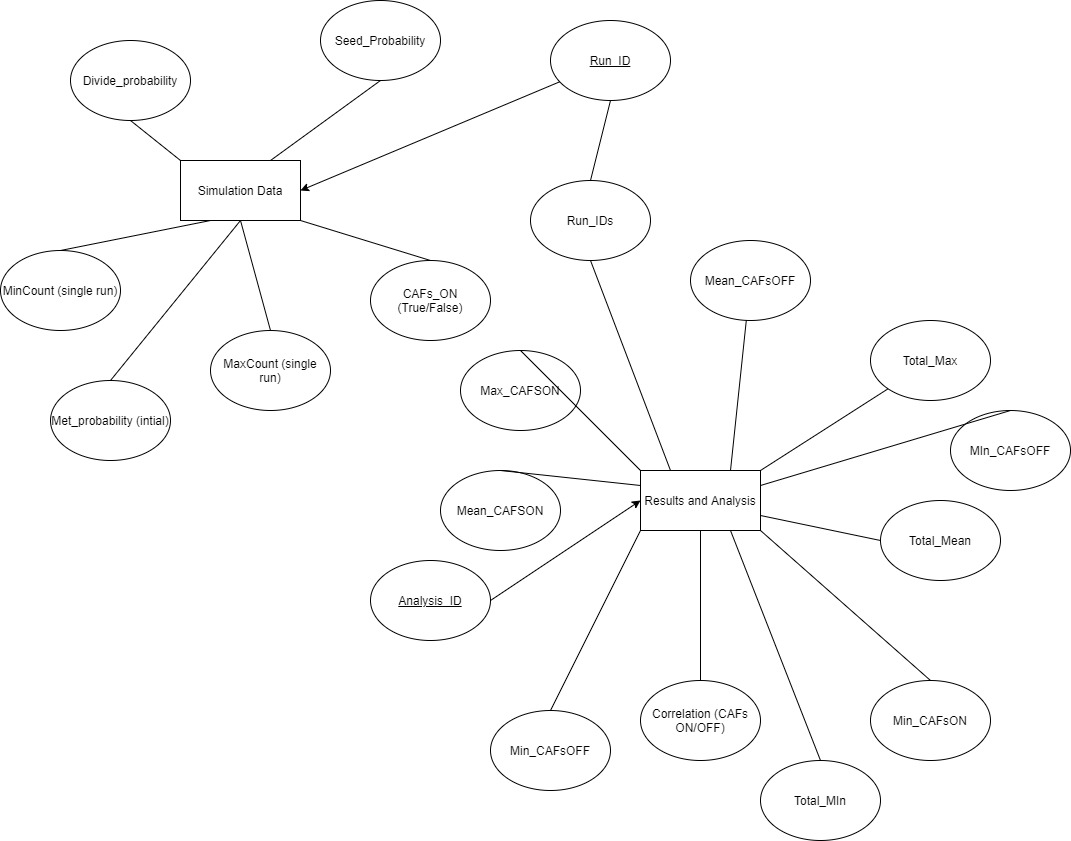
**Cancer associated fibroblasts (CAFs) are fibroblasts that have been found to play a role an inhibitory (or sometimes disinhibitory) effect in cancer metastasis. By researching these fibroblasts, we can gain better insight into the mechanisms of cancer and possibly form drugs that make use of CAFs to help treat cancer patients.[1] The goal of this project is to use data from a CAF simulation built by me last year under the supervision of Dr.Ghersi in which the simulation allows for the measurement of time to metastasis both with and without CAFs, and store that data in a database in which the data can then be analyzed by a user via a webpage. Using the results of the analysis, one can then compare the effects of CAFs on the tumor in the simulation to test whether CAFs do have an effect and what kind of effect the CAFs have.**

**Background\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Modeling is an age-old scientific tool, used for centuries now. From Models of the solar system to the Bohr model, Modeling has helped us understand the universe around us in ways that could not have been done otherwise. With the invention of computers, modeling has taken a whole different form, and is as useful now as it ever has been. Computational modeling software like NetLogo[2] gives us an entire virtual environment to play around in, where researchers can easily add or subtract as many environmental factors as needed, with as little as a few blocks of code. Computational modeling can assist in studying population movements, viral transmissions or even both simultaneously. One of the things computational modeling is helping us do currently is unlocking some of the hidden mechanics of cancer.**

**Cancer tends to be a multi-faceted issue within your cells, not only are the main regulatory cell functions such as apoptosis turned off, but tumor cells also refuse to stop multiplying. To test ways to effectively kill off these cells and keep them from multiplying we must understand the underlying mechanisms that keep these issues from being resolved naturally, in which the equipment, research and labor necessary can be costly. According to cancer.gov, Congress passed a bill in 2019 providing the National Cancer Institute 5.74 billion dollars in 2019[3], and it is not the first year or the first cancer research institute that the government has spent lump sums on. This is where Computational Biology comes in. By instead simulating interactions of cancer cells with drugs/molecules etc. the cost of cancer research materials goes down.**

**Diagram\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**



**Methods\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**The data taken from the simulation is manually edited (mainly for the python script purposes as some of the variable names in the simulation do not translate well into sql), and then fed into a python script I created called processInput.py. The python script has command line arguments so I may specify and change which files are input and output, complete with exception handling for those command lines. The .csv files are then turned into .sql files which are written to create the tables in my database. The php website then handles querying the data for the user and outputs the analyzed data using the system()[4] method.**

**Code and web link**

**All code, example ddl and dml (MetasDB.sql) as well as the Data Dictionary, License and webpage link can be found here:** <https://github.com/human-1914/MetasDB>**. No sample queries are needed to test this database. Simply follow the instructions on the page after you have read the FAQ (linked in the webpage).**

**Data provenance**

**The database draws from three separate .csv files appropriately named Simulation1.csv, Simulation2.csv and Simulation3.csv that hold the simulation data from 3 separate experiments done within the simulation. The data is retrieved from the files via an API of my own writing called processInput.py, which then translates the data to be readable by the mysql. To preserve space on the system as I only have 1 GB of storage space total on the server and each simulation file is about 1 MB, the nature of the webpage is such that the API reproduces and populates the tables each time a new simulation is analyzed, simply by following the on-screen instructions.**

**Results\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**The results of the data analysis seem to be that there is an inhibitory factor to CAFs on tumor metastasis. The average results for runs while CAFs are set to on tend to be higher than when the results of when the CAFs are off. For example, the average steps until metastasis while CAFs are toggled on in the file Simulation 3 is about 1187 steps whereas the average steps until metastasis with CAFs toggled off is approximately 12. This seems to suggest that CAFs have a negative effect on tumor metastasis and while future studies will be needed to confirm this, if true, this finding could mean a new era of CAF based treatment in fighting the spread of cancer in a patient’s body. This could even allow a terminally ill patient to live for a few more years depending on the type of cancer.**

**Discussion/Conclusions\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**This database allows for users to look at the results of a tumor/CAF interaction simulation built by me and allows for one to see the analysis of those results. The result of this work seems to be a promising look into the inner workings of cancer and the mystery that CAFs enters into the problem. There is much work to be done in the field of cancer research, but I truly feel that computational biology can be used as a driving force not only to reduce costs of cancer research but also to speed up the process of developing new cancer treatments overall. The biggest challenge I faced in this project was working with so many languages at once and having them all interact in some way or another. I really went all out to challenge my ability as a programmer, and I even impressed myself to some extent at how far I have come since I entered college. I can say for certain that in this project I used almost every programming skill I learned in the last few years. I do not work with php often so that was the language that challenged me the most. This will be one of the last programming assignments of my undergraduate college experience and I truly consider this to be my Magnum Opus of my college career. I am incredibly proud of this project as simple and sometimes buggy as it may seem.**

**References\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**• O. E. Franco, A. K. Shaw, D. W. Strand, and S. W. Hayward, “Cancer associated fibroblasts in cancer pathogenesis,” Seminars in Cell &amp; Developmental Biology, 05-Nov-2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1084952109002134?casa\_token=n1N\_3M4HEycAAAAA%3AbBDD6nlXZusKKLXhXz8gPZ2r1tLHf7aVbu3xtYBKqg9fFt34FeIuBVr1lN1eCrbj0vJ7m\_e16A. [Accessed: 24-Apr-2021].**

**• NetLogo User Manual (version 6.2). [Online]. Available: https://ccl.northwestern.edu/netlogo/faq.html. [Accessed: 24-Apr-2021].**

* **N. Sharpless, “Plans for NCI's Fiscal Year 2019 Budget,” National Cancer Institute, 18-Dec-2018. [Online]. Available: https://www.cancer.gov/news-events/cancer-currents-blog/2018/sharpless-nci-fy2019-budget#:~:text=For%20FY%202019%2C%20which%20began,million%20for%20the%20Cancer%20Moonshot. [Accessed: 28-Apr-2021].**
* **N/A, “PHP:system,” php, 2001. [Online]. Available: https://www.php.net/manual/en/function.system.php. [Accessed: 28-Apr-2021].**