

# EXAM II

Samuel Young

Thursday ,April 26th, 2018

- **Part I**

## **1.A Understand the question that is being asked.**

I started by reading "Global mapping of MtrA-binding sites links MtrA to regulation of its targets in *Mycobacterium tuberculosis*" for better insight. In the reading of the meat of the article states that rpfA, rpoB, relF, and rpfB, rpfC promoter regions. It seems that have of found the promoters regions for the MtrA. They found a MtrA-binding site of the sequence upstream. Rv2524c and Rv3246c (chatterjee et. al,102). In the article, they empolyed using MEME software to find the promoters in both regions of rpfA and rpfC(chatterjee et. al, 102). I did similar something similar I used memes to find upwards of 5 sequences that had some likelihood of PSSM similarity to most of the groups. I also removed a majority of the unneed DNA locations and only kept in the DNA binding promoters to our target gene an the hypotheicial genes because I was unsure of what there functionality was.I then download the logos from memes.

## Part II

### 1.B back to the salts mines

In line thriteen, there seems to be a large amount a majority of them that where G and it likely in line 13 there appears to be 2 A that came up in my assumption that I made this actaul frequency was 1 to 21 G for this line. In line six and seventeen they appear t have some what of a similar sizes of C and A's being used In this one I forgot to estimate the size for these one but I know the right size for 6 is C eighteen to four A and the size fo 17 was C ninteen and A of three. So I used information in py to generate the graphs

## Part III

### 1.C eating all the salt

At first I was trying to use the find function on firefox and begin with a base sequeence but once that end up not panning out I was only able to obtain get 3 or 6 letters in to it before thing kept on hitting the bells. I could always consider trying to run against a phoylogneic tree to find the possible sites of similarity.

## • Part IV

### 2.A

There are just about 72 possible ALL and AML data point within the data set. There are a total 47 ALL in the dataset and 25 of the AML in the same data set. So the genes are the rows and the columns are the patients. So first I did a wordcount on the file to see how many rows there are and then minus the 7219-1 to get a close approximation

to the numbers there are 72\* 7218 different genes to look at for the data set.

## Part V

### 2.B

I originally thought that there were at least 500 of these genes had a greater than or equal to 2.000 p-value with a degree of freedom of 60 but apparently there is closer to 2072 genes within this p-value ranging from. What is 20 percent a close approximate to 28 percent. A p-value is a value that is a statistical test that is shown with  $1 - \int_{-\infty}^{cv} f(t)$ .

## Part VI

(

2.C In this part where we were told to make a any algorithm using permutation.py. In which, I removed the first time you called the absolute value and got half that exact same amount. -1.51 as the value for the original. 3.02 when you keep the absolute value. but because you are looking at both tails of the graph in this one you will obtain a more fluid like graph in the case that now you are trying to for values that are outside of the graph in which you will now obtain an estimated 968 out of all the genes.

## Part VII

### 2.D

I was able to obtain a graph but it appears to be right skewed based on the fact that it is only a one sided graphical representation of the data.

With using only genes any number of list to generate this graph as any input.

## **Part VIII**

### **2.E**

So with with the permuatation data that I randomly shuffled around I was able to get much more closer to the .05 precent with a few tweaks to the intial values. Unfortunately, my numbers do randomly shift around alot so there is always a chance of obtaining a prefect number,1.962, because the value will variey 1.565 p-value closer and then there is also a chance of obtain a 3.765 p-value father away form the table. Another weird thing that i realized about my program is that It more closer to 10000 times rather then a thousand fold.

## **• Part IX**

### **3.A**

This question make no sense because you already gave us the final format of the answer.

## **Part X**

### **3.B**

I had a problem with getting the actual graphing to work but I'm a place at which the graph is stored. In which correspond with a memory address.

## • Part XI

### 4.A

”The genomic comparison of *BRAC2* and *BCR/ABL* oncogenes with machines learning algorithms and the ability to find similarities and dissimilarities with the regulation of these two genes.”

## Part XII

### 4.B

- 1 Secondary mutations as a mechanism of cisplatin resistance in BRCA2-mutated cancers It shows secondary characteristics of BRCA-2 Mutation
- 2 BCR/ABL: from molecular mechanisms of leukemia induction to treatment of chronic myelogenous leukemia this talks about possible treatments that used to cure the infection.
- 3 Down-regulation of BCRA1 in BCR-ABL-expressing hematopoietic cells  
This article, so that there are breakdown in repairing the protein with a DNA-dependent protein kinase catalytic subunit.
- 4 Accurate prediction of BRCA1 and BRCA2 heterozygous genotype using expression profiling after induced DNA damage. This article is the one in which I can start by compare the result of other studies in cancer and try and figure out what happens when you use svm.

**Part XIII**

**4.C**

**• Part XIV**

**5**