

Name: Alan Xu

QBIO 490: Directed Research

Midterm Project

3/13/2023

Part I: Review Questions

General Concepts

1. What is TCGA and why is it important?

The Cancer Genome Atlas is a publicly available atlas that includes genomic, epigenomic, transcriptomic and proteomic data from 20,000 different samples that span 33 different cancer types. It is important because it allows any researchers in the world to access its vast sample size and multiomic data to study cancer types that may have poor prognosis or have high public health impact.

2. What are some strengths and weaknesses of TCGA?

TCGA has a relatively large sample size available for analysis, allowing more accurate and precise statistical testing. The fact that TCGA is publicly available makes cancer research more accessible to researchers around the world. However, the cancer types that TCGA host are still limited due to funding limitations. Furthermore, much of the information on patients or tumor samples may have missing values due to various different reasons, making analysis difficult.

Coding Skills

1. What commands are used to save a file to your GitHub repository?

```
git add /*file*/
```

```
git commit -m "/*Your Message*/"
```

```
git push
```

2. What command(s) must be run in order to use a package in R?

```
install.packages("/*Your Package Name*/")
```

```
library(/*Your Package Name*/)
```

3. What command(s) must be run in order to use a Bioconductor package in R?

```
BiocManager::install("/*Your Package Name*/")
```

```
library(*Your Package Name*/)
```

4. What is boolean indexing? What are some applications of it?

Boolean indexing allows the application of a Boolean vector to a row/column in a dataframe to perform some kind of filtering. Some examples include removing NAs from a dataframe or dividing the clinical dataframe into two depending on the initial diagnosis age of the patient.

5. Draw a mock up (just a few rows and columns) of a sample dataframe. Show an example of the following and explain what each line of code does.

- a. an `ifelse()` statement

- b. boolean indexing

SEE R FILE