

## General Concepts

### 1. What is TCGA and why is it important?

The Cancer Genome Atlas (TCGA) is a publicly available database containing comprehensive genomic and clinical data on over 30 types of cancer. One of the TCGA database's most significant features is its multi-omic data inclusion. This allows the dataset to contain information on multiple levels of biological organization, including DNA sequencing, gene expression profiling, and clinical data. By analyzing these various types of data together, researchers can better understand the molecular processes underlying cancer development and progression.

### 2. What are some strengths and weaknesses of TCGA?

**Strengths:** TCGA can generate large amounts of genomic and clinical data from many types of cancer, providing researchers with an invaluable resource for studying the molecular basis of cancer. This open-access database facilitates multidisciplinary collaboration.

**Weakness:** most of the data in the TCGA database are collected during a period of therapy and need long-term data. Besides, some TCGA data may need to be included, complete, or of poor quality, which may limit the usefulness of the data set for specific analyses. It's also hard for it to equally cover all the types of cancer.

## Coding Skills

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

```
cd Users/kojiro/desktop/qbio490/qbio_490_KainingFeng
```

```
git status
```

```
git add file_name
```

```
git commit -m "[informative message about file]"
```

```
git push
```

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

```
if(!require(package_name)){
  install.packages("package_name ")
}
```

```
Library (package_name)
```

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

```
if(!require(BioManager)){
  install.packages("BioManager")
}
if (!require("Bioconductor", quietly = TRUE))
BioManager::install("Bioconductor")
Library (Bioconductor)
```

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

Boolean indexing is a technique used in programming to select elements or subsets of an array or data frame based on Boolean conditions. A Boolean condition is a logical expression that returns TRUE or FALSE for each element in an array or data frame.

In R, boolean indexing can be done by passing a logical vector as an index to the [ ] operator. The operator selects the elements in the array or data frame corresponding to the logical vector's TRUE values. One application is to apply a Boolean mask to a dataframe.

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

```
treatments_mask <- ifelse(colnames(rna_clinical) == "treatments", F,T)
```

- b. boolean indexing

```
rna_clinical <- rna_clinical[,treatments_mask]
```