lab7

Theo-60985751, Hannah Yeung-88788427, Jingxuan Chen-85741635

2023-10-23

Reading the data

```
data_clinical = read.table("data_clinical_patient.txt", header = TRUE, sep = "\t")

data_mutation = read.table("data_mutations.txt", header = TRUE, sep = "\t")

data_rna = read.csv("RNAseq_LIHC.csv", header = TRUE)
```

Finding the unique patients

```
unique_patient_clinical = unique(data_clinical$PATIENT_ID)
unique_patient_rna = unique(colnames(data_rna))
unique_patient_mutation = unique(data_mutation$Tumor_Sample_Barcode)

cat("Number of unique patient in RNAseq data:", length(unique_patient_rna),"\n", sep = " ")
```

```
## Number of unique patient in RNAseq data: 425
```

```
cat("Number of unique patient in clinical data:", length(unique_patient_clinical),"\n", sep = " ")
```

```
## Number of unique patient in clinical data: 372
```

```
cat("Number of unique patient in mutation data:", length(unique_patient_mutation),"\n", sep = " ")
```

```
## Number of unique patient in mutation data: 357
```

Changing the "." to "-" in patient ID of data_rna

```
uni_rna<-gsub("\\.", "-", colnames(data_rna[2:425]))

shortened_rna = substr(uni_rna, start = 1, stop = 12)
shortened_mutation = substr(unique_patient_mutation, start = 1, stop = 12)
```

Finding the common patients

```
common_names1 <- intersect(unique_patient_clinical, shortened_mutation)
common_names2 <- intersect(common_names1, shortened_rna)

# unique_patient_clinical
cat("Number of unique patients that we have ALL the clinical,
    mutations, and RNAseq data:", length(common_names2), sep = " ")
```

```
## Number of unique patients that we have ALL the clinical,
##     mutations, and RNAseq data: 352
```