Lab 8 | Jon Lee

Code

Lab 8 | BINF 6310 | Spring 2020 | Jon Lee

1. We return to our RNA seq dataset of E. Coli genes from mice. The URL is here: http://afodor.github.io/classes/stats2015/longitdunalRNASeqData.zip (http://afodor.github.io/classes/stats2015/longitdunalRNASeqData.zip)

Read and normalize the counts table (“nc101_scaff_dataCounts.txt “ into R). For example:

setwd(“C:\somewhere”)

myT<-read.table(“nc101_scaff_dataCounts.txt”,sep=",header=TRUE,row.names=1)

remove rare genes myT <- myT[apply(myT,1, median)> 5,]

myTNorm <- myT for (i in 1:ncol(myT)) { colSum = sum(myT[,i]) myTNorm[,i] = myTNorm[,i]/colSum }

Hide

```
myT <- read.table("nc101_scaff_dataCounts.txt", sep = "\t", header = TRUE, row.names = 1
)

#Remove rare genes
myT <- myT[apply( myT, 1, median)> 5,]

myTNorm <- myT
for( i in 1:ncol(myT))
{
    colSum = sum(myT[,i])
  myTNorm[,i] = myTNorm[,i]/colSum
}
```

(The first 3 columns are “day 2”, the next 3 columns are “week 12” and the last 5 are “week 18”). Remember, that day 2 is before the mice have inflammation symptoms, week 12 is associated with inflammation and week 18 is associated with cancer.

2. For every row in the normalized spreadsheet, run three t-tests (“day 2” vs. “week 12”, “day 2” vs. “week 18” and “week 12” vs. “week 18”). At a p < .05 threshold fill in the following table:

Hide

```
#Table for Problem 2
problem2 <- matrix(nrow = 3, ncol = 3)
colnames(problem2) <- c("# of genes significant at p < 0.05 uncorrected", "# of genes si
gnificant at p < 0.05 BH FDR corrected", "# of genes significant at p < 0.05 Bonferrnoi
 corrected")
rownames(problem2) <- c("day 2 vs. week 12", "day 2 vs. week 18", "week 12 vs. week 18")

problem2
```

```
                    # of genes significant at p < 0.05 uncorrected
day 2 vs. week 12                                              NA
day 2 vs. week 18                                              NA
week 12 vs. week 18                                            NA
                    # of genes significant at p < 0.05 BH FDR corrected
day 2 vs. week 12                                                 NA
day 2 vs. week 18                                                 NA
week 12 vs. week 18                                               NA
                    # of genes significant at p < 0.05 Bonferrnoi corrected
day 2 vs. week 12                                                    NA
day 2 vs. week 18                                                    NA
week 12 vs. week 18                                                  NA
```

Hide

```
#t-test for day 2 vs. week 12
day2week12 <- vector()

for(i in 1:nrow(myTNorm))
{
  vals1 <- as.numeric(myTNorm[i,1:3])
  vals2 <- as.numeric(myTNorm[i,4:6])
  day2week12[i] <- t.test(vals1, vals2)$p.value
}

#t-test for day 2 vs. week 18
day2week18 <- vector()

for(i in 1:nrow(myTNorm))
{
  vals1 <- as.numeric(myTNorm[i,1:3])
  vals2 <- as.numeric(myTNorm[i,7:11])
  day2week18[i] <- t.test(vals1, vals2)$p.value
}

#t-test for week 12 vs. week 18
week12week18 <- vector()

for(i in 1:nrow(myTNorm))
{
  vals1 <- as.numeric(myTNorm[i,4:6])
  vals2 <- as.numeric(myTNorm[i,7:11])
  week12week18[i] <- t.test(vals1, vals2)$p.value
}
```

Hide

```
#num of genes significant at p < 0.05 uncorrected
day2week12PVals <- 0
for(i in 1:length(day2week12))
{
  if(day2week12[i] < 0.05)
  {
    day2week12PVals =  day2week12PVals + 1
  }
}
print(day2week12PVals)
```

```
[1] 398
```

Hide

```
day2week18PVals <- 0
for(i in 1:length(day2week18))
{
  if(day2week18[i] < 0.05)
  {
    day2week18PVals =  day2week18PVals + 1
  }
}
print(day2week18PVals)
```

```
[1] 1243
```

Hide

```
week12week18PVals <- 0
for(i in 1:length(week12week18))
{
  if(week12week18[i] < 0.05)
  {
    week12week18PVals =  week12week18PVals + 1
  }
}
print(week12week18PVals)
```

```
[1] 810
```

Hide

```
#num of genes significant at p < 0.05 BH FDR corrected
day2week12BHP <- p.adjust(day2week12, method = "BH")
day2week12BHPPVals <- 0
for(i in 1:length(day2week12BHP))
{
  if(day2week12BHP[i] < 0.05)
  {
    day2week12BHPPVals =  day2week12BHPPVals + 1
  }
}
print(day2week12BHPPVals)
```

```
[1] 0
```

Hide

```
day2week18BHP <- p.adjust(day2week18, method = "BH")
day2week18BHPPVals <- 0
for(i in 1:length(day2week18BHP))
{
  if(day2week18BHP[i] < 0.05)
  {
    day2week18BHPPVals =  day2week18BHPPVals + 1
  }
}
print(day2week18BHPPVals)
```

```
[1] 279
```

Hide

```
week12week18BHP <- p.adjust(week12week18, method = "BH")
week12week18BHPPVals <- 0
for(i in 1:length(week12week18BHP))
{
  if(week12week18BHP[i] < 0.05)
  {
    week12week18BHPPVals =  week12week18BHPPVals + 1
  }
}
print(week12week18BHPPVals)
```

```
[1] 15
```

Hide

```
#num of genes significant at p < 0.05 Bonferroni corrected
day2week12TotPVals <- length(day2week12)
day2week12AdjPVal <- 0.05/day2week12TotPVals
print(day2week12AdjPVal)
```

```
[1] 1.255335e-05
```

Hide

```
day2week12AdjPVals <- 0
for(i in 1:length(day2week12))
{
  if(day2week12[i] < day2week12AdjPVal)
  {
    day2week12AdjPVals =  day2week12AdjPVals + 1
  }
}
print(day2week12AdjPVals)
```

```
[1] 0
```

Hide

```
day2week18TotPVals <- length(day2week18)
day2week18AdjPVal <- 0.05/day2week18TotPVals
print(day2week18AdjPVal)
```

```
[1] 1.255335e-05
```

Hide

```
day2week18AdjPVals <- 0
for(i in 1:length(day2week18))
{
  if(day2week18[i] < day2week18AdjPVal)
  {
    day2week18AdjPVals =  day2week18AdjPVals + 1
  }
}
print(day2week18AdjPVals)
```

```
[1] 3
```

Hide

```
week12week18TotPVals <- length(week12week18)
week12week18AdjPVal <- 0.05/week12week18TotPVals
print(week12week18AdjPVal)
```

```
[1] 1.255335e-05
```

Hide

```
week12week18AdjPVals <- 0
for(i in 1:length(week12week18))
{
  if(week12week18[i] < week12week18AdjPVal)
  {
    week12week18AdjPVals =  week12week18AdjPVals + 1
  }
}
print(week12week18AdjPVals)
```

```
[1] 1
```

Hide

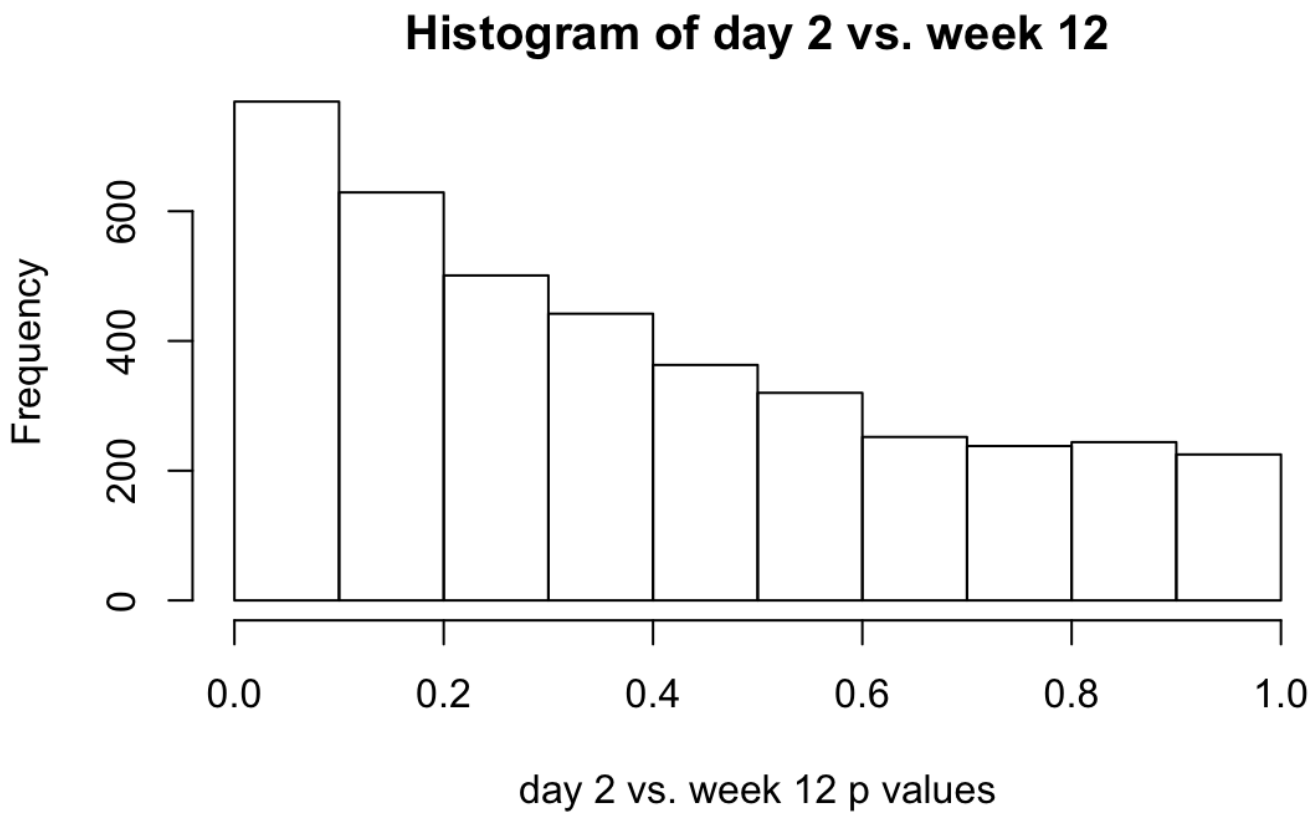
```
#input values back into Problem 2 table
day2week12Data <- c(day2week12PVals, day2week12BHPPVals, day2week12AdjPVals)
day2week18Data <- c(day2week18PVals, day2week18BHPPVals, day2week18AdjPVals)
week12week18Data <- c(week12week18PVals, week12week18BHPPVals, week12week18AdjPVals)

problem2[1,] <- day2week12Data
problem2[2,] <- day2week18Data
problem2[3,] <- week12week18Data

print(problem2)
```

```
                  # of genes significant at p < 0.05 uncorrected
day 2 vs. week 12                                           398
day 2 vs. week 18                                          1243
week 12 vs. week 18                                         810
                  # of genes significant at p < 0.05 BH FDR corrected
day 2 vs. week 12                                               0
day 2 vs. week 18                                             279
week 12 vs. week 18                                            15
                  # of genes significant at p < 0.05 Bonferrnoi corrected
day 2 vs. week 12                                                     0
day 2 vs. week 18                                                     3
week 12 vs. week 18                                                   1
```

3. Make histograms of all the uncorrected p-values for each of the three companions. Are any of the distributions uniform?

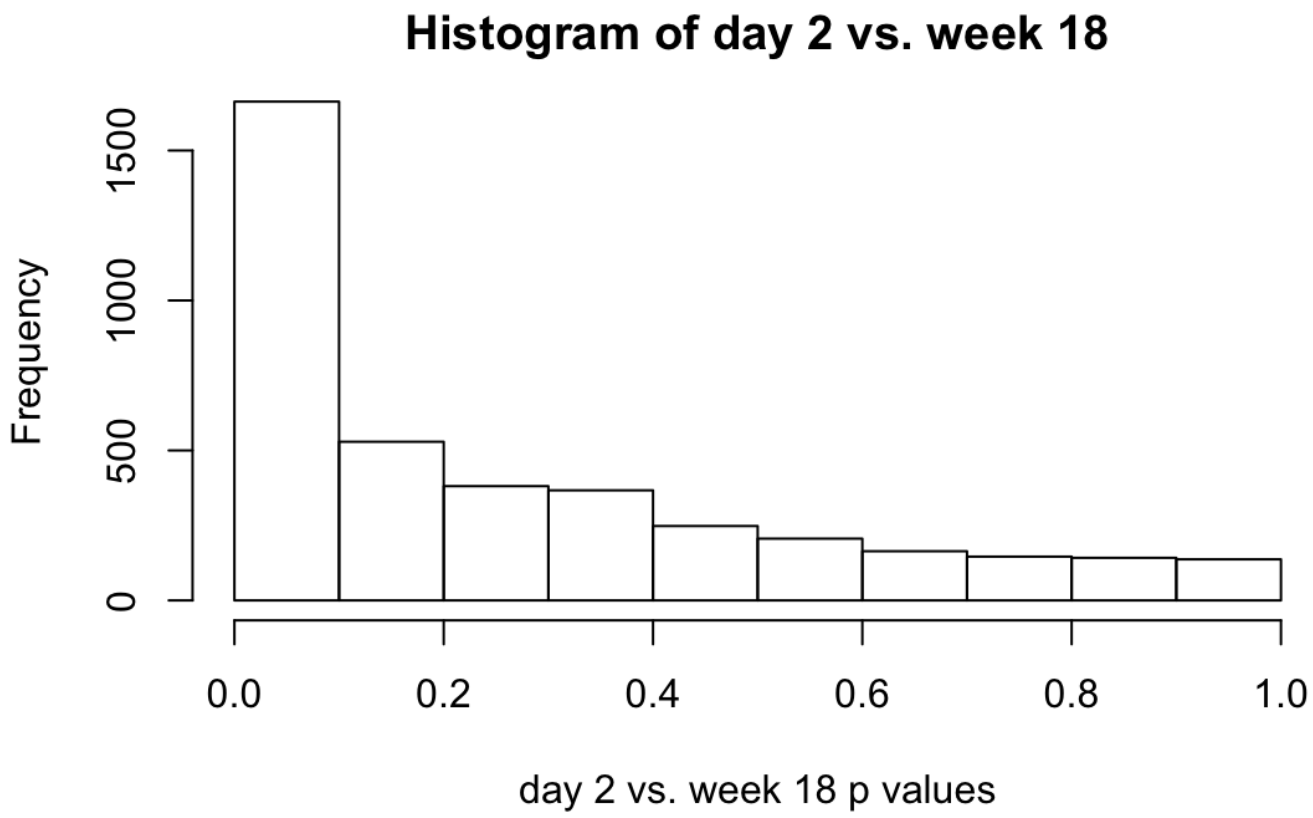
Hide

```
hist(day2week12, xlab = "day 2 vs. week 12 p values", main = "Histogram of day 2 vs. wee
k 12")
```



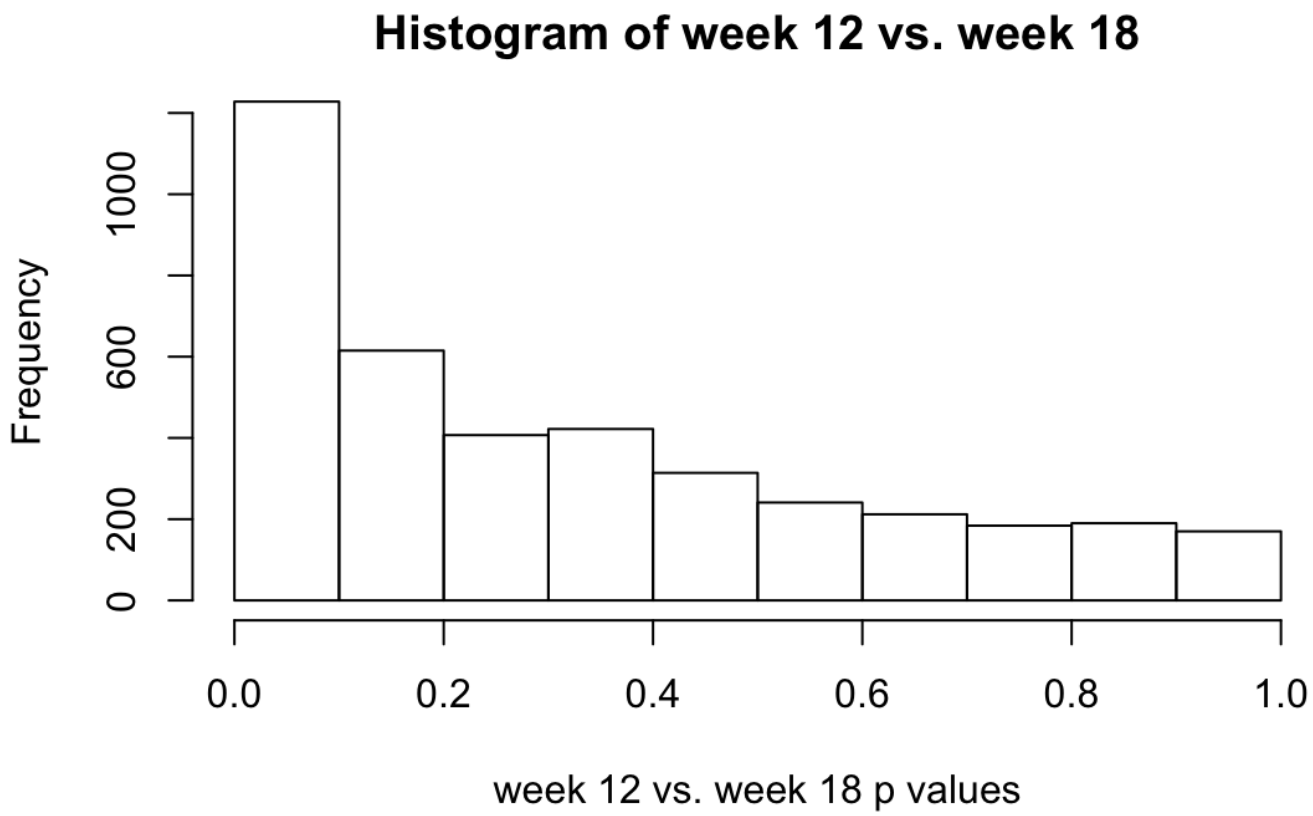
Hide

```
hist(day2week18, xlab = "day 2 vs. week 18 p values", main = "Histogram of day 2 vs. wee
k 18")
```



Hide

```
hist(week12week18, xlab = "week 12 vs. week 18 p values", main = "Histogram of week 12 v
s. week 18")
```



Answer:

None of the distributions of p-values look to be uniform.

4. Based on these data, when is the biggest shift in the transcriptome? Which samples are most different from one another?

Answer:

Based on the data, the biggest shift occurs between day and week 18, becuase it has the largest number of significant p-valeues.