**BCH 519 – Week 11**

**Suruchi JaiKumar Ahuja - 50170070**

**Question 1**

cut –f1 count\_file.txt > genename.txt

cut –f2-7 count\_file.txt > newfile.txt

head newfile.txt

module load R/3.1.2

R

> library(edgeR)

Loading required package: limma

> x=read.table("/user/suruchij/newfile.txt")

> head(x)

V1 V2 V3 V4 V5 V6

1 0 0 0 0 23 0

2 0 0 0 0 23 0

3 0 0 0 0 23 0

4 0 0 0 0 23 0

5 105 125 96 169 129 132

6 45 47 48 84 60 69

> groups=c("1","1","1","2","2","2")

> y=DGEList(counts=x,group=groups)

> y=calcNormFactors(y)

> y = estimateCommonDisp(y)

> y = estimateTagwiseDisp(y)

> et=exactTest(y)

> head(et$table)

logFC logCPM PValue

1 5.85003520 5.359977 0.0002971473

2 5.85003520 5.359977 0.0002971473

3 5.85003520 5.359977 0.0002971473

4 5.85003520 5.359977 0.0002971473

5 -0.10633332 9.819815 0.3974901422

6 0.09810559 8.741220 0.6248149748

> de1 = decideTestsDGE(et,adjust.method="BH",p.value=0.05)

> summary(de1)

[,1]

-1 344

0 463

1 392

> ahuja.outputTable.txt = cbind(et$table,de1)

> head(ahuja.outputTable.txt)

logFC logCPM PValue de1

1 5.85003520 5.359977 0.0002971473 1

2 5.85003520 5.359977 0.0002971473 1

3 5.85003520 5.359977 0.0002971473 1

4 5.85003520 5.359977 0.0002971473 1

5 -0.10633332 9.819815 0.3974901422 0

6 0.09810559 8.741220 0.6248149748 0

**Question 2 : Volcano plot**

> x=et$table$logFC

> head(x)

[1] 5.85003520 5.85003520 5.85003520 5.85003520 -0.10633332 0.09810559

> y=-log10(et$table$PValue)

> head(y)

[1] 3.5270283 3.5270283 3.5270283 3.5270283 0.4006736 0.2042486

> pdf("ahuja.volcanoplot.pdf")

> pdf(plot(x,y))

**Question 3 : MA Plot**

> x=et$table$logCPM

Loading required package: edgeR

Loading required package: limma

> y=et$table$logFC

> pdf("ahuja.MAplot.pdf")

> pdf(plot(x,y))

**Question 4 : Heat Maps**

> x=read.table("/user/suruchij/newfile.txt")

> head(x)

V1 V2 V3 V4 V5 V6

1 0 0 0 0 23 0

2 0 0 0 0 23 0

3 0 0 0 0 23 0

4 0 0 0 0 23 0

5 105 125 96 169 129 132

6 45 47 48 84 60 69

> y=DGEList(counts=x,group=groups)

> y=calcNormFactors(y)

> logcpm <- cpm(y,prior.count = 2,log = TRUE)

> pdf("ahuja.heatmap.pdf")

> pdf(heatmap(logcpm))

**Question 5 : GO term analysis**

cut –f1 count\_file.txt > genename.txt

cut –f2-7 count\_file.txt > newfile.txt

head newfile.txt

paste genename.txt ahuja.outputTable.txt > ahuja1.txt

head ahuja1.txt

Python script : nano code.txt

readfile = open('ahuja1.txt','r')

file = readfile.readlines()

for line in file:

    r = line.split("\t")

    if (r[-4:] != '0'):

       print ' '.join(line)

python code.txt > ahuja.xls