Solutions of Homework Module 10

PROBLEM 1:

Install the "ArrayExpress" package from Bioconductor. Load the yeast microarray data using R commands:

```
library(ArrayExpress)
yeast.raw = ArrayExpress('E-MEXP-1551')
```

(a) Preprocess the raw data set into an expression data set using: the "mas" background correction method, the "quantiles" normalization method, "pmonly" pm correction method and "medianpolish" summary method. Give the R command here for doing this task.

(b) Print out the mean expression values for the first five genes across all samples.

(c) How many genes and how many samples are in the preprocessed expression data set?

Solutions:

1a) the R command is:

```
yeast.raw <- ArrayExpress('E-MEXP-1551')
eset<- expresso(yeast.raw,bgcorrect.method="mas",
                normalize.method="quantiles",
                pmcorrect.method="pmonly",
                summary.method="medianpolish")
exprs.yeast <- exprs(eset)
```

1b)
The mean expression values for the first five genes across all samples are:

```
> apply(exprs.yeast[1:5,], 1, mean)
1769308_at, 1769309_at, 1769310_at, 1769311_at, 1769312_at
 8.936128,    5.666040,      5.650467,      11.380948,      9.752480
```

Solutions of Homework Module 10

1c)

```
> str(exprs.yeast)
 num [1:10928, 1:30] 9.05 5.58 5.7 11.43 9.87 ...
 - attr(*, "dimnames")=List of 2
  ..$ : chr [1:10928] "1769308_at" "1769309_at" "1769310_at" "1769311
_at" ...
  ..$ : chr [1:30] "Gre_MCA_2822" "Gre_MCA_5014" "Gre_MCA_3174" "
Gre_MCA_4108" ...
```

There are 10928 genes and 30 samples

PROBLEM 2:

(a) What is the annotation package for the yeast data set in question 1? Install the annotation package from Bioconductor.

(b) Search the 1769308_at gene GO numbers related to Molecular Function (MF). How many GO numbers do you get?

(c) Find the GO parents of the GO IDs in part (b). How many GO parents are there?

(d) Find the GO children of the GO IDs in part (b). How many GO children are there?

Solutions:

2a)

```
> annotation(yeast.raw)
[1] "yeast2"
```

The annotation package is Yeast2

2b)

```
> length(mf.go1769308_at)
[1] 7
```

No. of numbers related to “Molecular function” is 7

2c)

```
> parents
GO:0016491.is_a GO:0003824.is_a GO:0016616.is_a GO:0016829.is_a G
O:0016853.is_a
  "GO:0003824"   "GO:0003674"   "GO:0016614"   "GO:0003824"   "GO
:0003824"
GO:0004300.is_a GO:0003857.is_a
  "GO:0016836"   "GO:0016616"
> length(parents)
[1] 7
```

There are 7 GO parents.

2d)

```
> length(unlist(ch))
[1] 423
```

The no. of GO children 423

PROBLEM 3:

We work with the patients in stages "B2","B3".

(a) We look for genes expressed differently in stages B2 and B3. Use genefilter to program the Wilcoxon test and the Welch t-test separately for each gene. For each test, we select the genes with p-value<0.001. To save computational time, we set exact=F in the Wilcoxon test function.

(b) Compute a Venn diagram for the Wilcoxon test and the t-test, and plot it.

(c) How many pass the Wilcoxon filter? How many passes both filters?

(d) What is the annotation package for the ALL data set? Find the GO numbers for
“oncogene”.

(e) How many genes passing the filters in (a) are oncogenes?

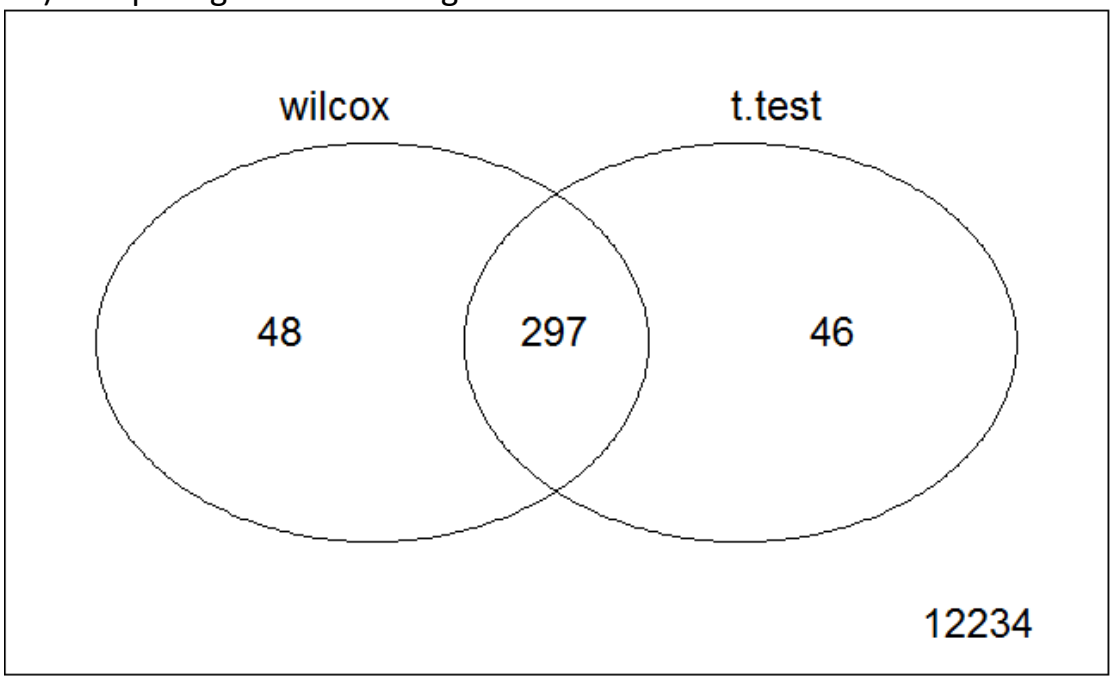
Solutions:

Solutions of Homework Module 10

3a)

```
> patient.B <- exprs(ALL)[,(ALL$BT %in% c("B2","B3"))]
> factor <- droplevels(ALL$BT[ALL$BT %in% c("B2","B3")])
> f1 <- function(x) (wilcox.test(x ~ factor, exact = F)$p.value < 0.001)
> f2 <- function(x) (t.test(x ~ factor)$p.value < 0.001)
> wilcox <- genefilter(patient.B, filterfun(f1))
> t.test <- genefilter(patient.B, filterfun(f2))
```

3b) Comparing with venn diagram :



3c)

Using the venn diagram we got:

Wilcoxon filter : 297+48 = 345

Both the filters : 297

3d)

```
> annotation(ALL)
[1] "hgu95av2"
> print("the oncogene id is ")
[1] "the oncogene id is "
> oncogene.id
[1] "GO:0090402"
```

Solutions of Homework Module 10

The annotation package is " hgu95av2.dg "
the oncogene id is GO:0090402

3e)
genes passing the filters in(a) oncogenes are 0

Problem 4 :

Stages of B-cell ALL in the ALL data. Use the limma package to answer the questions below.

(a) Select the persons with B-cell leukemia which are in stage B1, B2, and B3.

(b) Use the linear model to test the hypothesis of all zero group means. Use "topTable()" to report the top five genes with nonzero means in B3 group.

(c) Use two contrasts to perform analysis of variance to test the null hypothesis of equal group means. Do this with a false discovery rate of 0.01. How many differentially expressed genes are found? Use "topTable()" to report the top five genes that express differently among the three groups.

Solutions:

4a)
```
> all.B <- ALL[,which(ALL$BT %in% c("B1","B2","B3"))]
```

4b)
The top 5 genes with nonzero means in B3 group are :
```
> print(topTable(fit, number=5,adjust.method="fdr"), digits=4)
                B1    B2    B3 AveExpr     F    P.Value  adj.P.Val
AFFX-hum_alu_at 13.42 13.54 13.61   13.53 141230 1.322e-145 1.669e-141
32466_at        12.68 12.72 12.71   12.71 113412 6.882e-142 4.344e-138
31962_at        13.17 13.07 13.05   13.09 107260 6.061e-141 2.551e-137
32748_at        12.08 12.14 12.15   12.12 103287 2.643e-140 8.340e-137
35278_at        12.44 12.47 12.52   12.48 102374 3.736e-140 9.435e-137
```

Solutions of Homework Module 10

The top 5 genes with nonzero means in B3 group are :
AFFX-hum_alu_at
32466_at
31962_at
32748_at
35278_at

The p-values is less than 0.05 , hence we reject the null hypothesis. Thus we can conclude that they are expressed differently.

4c)

```
> sum(fdr.p.data<0.01)
[1] 314
```

The no of genes expressed differently at FDR 0.01 are 314

```
> print(topTable(fit1, number=5,adjust.method="fdr"), digits=4)
         B1...B2  B2...B3 AveExpr     F   P.Value adj.P.Val
1389_at  -1.7852 -0.74038   9.678 49.15 1.532e-14 1.934e-10
1914_at   2.0976  0.35648   4.693 42.20 3.785e-13 2.389e-09
33358_at  1.4890 -0.20733   5.214 29.52 2.837e-10 1.194e-06
38555_at  0.8058  0.62321   6.124 25.93 2.322e-09 7.329e-06
40763_at  1.5921 -0.01192   3.220 23.08 1.337e-08 2.758e-05
```

the top five genes that express differently among the three groups:
1389_at
1914_at
33358_at
38555_at
40763_at