Hw 4

1. Load Data

```
data <- read.csv('GSE37704.csv')
rownames(data) <- data[,1]
data <- data[,-1]
```

2. Preprocess Data

2.1 Normalize Data by count per million

```
normData <- lapply(data, function(x){return(x/sum(x) * 1000000)})
df <- as.data.frame(normData, row.names = rownames(data), col.names = names(normData))
summary.data.frame(df)
```

```
##    control_1           control_2           control_3
##  Min.   :    0.000   Min.   :    0.000   Min.   :    0.000
##  1st Qu.:    0.070   1st Qu.:    0.123   1st Qu.:    0.098
##  Median :    6.603   Median :    6.471   Median :    6.390
##  Mean   :   50.485   Mean   :   50.485   Mean   :   50.485
##  3rd Qu.:   36.315   3rd Qu.:   36.545   3rd Qu.:   36.194
##  Max.   :28286.585   Max.   :27042.737   Max.   :27266.460
##    Hoxa1KN_1           Hoxa1KN_2           Hoxa1KN_3
##  Min.   :    0.000   Min.   :    0.000   Min.   :    0.000
##  1st Qu.:    0.119   1st Qu.:    0.106   1st Qu.:    0.092
##  Median :    6.084   Median :    6.119   Median :    6.075
##  Mean   :   50.485   Mean   :   50.485   Mean   :   50.485
##  3rd Qu.:   35.967   3rd Qu.:   35.702   3rd Qu.:   36.268
##  Max.   :25486.423   Max.   :26528.477   Max.   :24289.459
```

2.2 Log normalized data

```
logDf <- log(df+1)
summary.data.frame(logDf)
```

```
##    control_1          control_2         control_3          Hoxa1KN_1
##  Min.   : 0.00000   Min.   : 0.0000   Min.   : 0.00000   Min.   : 0.0000
##  1st Qu.: 0.06789   1st Qu.: 0.1162   1st Qu.: 0.09309   1st Qu.: 0.1127
##  Median : 2.02851   Median : 2.0110   Median : 2.00013   Median : 1.9578
##  Mean   : 2.07267   Mean   : 2.0759   Mean   : 2.07095   Mean   : 2.0668
##  3rd Qu.: 3.61940   3rd Qu.: 3.6255   3rd Qu.: 3.61614   3rd Qu.: 3.6100
##  Max.   :10.25018   Max.   :10.2052   Max.   :10.21345   Max.   :10.1459
##    Hoxa1KN_2         Hoxa1KN_3
##  Min.   : 0.0000   Min.   : 0.00000
##  1st Qu.: 0.1011   1st Qu.: 0.08806
##  Median : 1.9627   Median : 1.95661
##  Mean   : 2.0686   Mean   : 2.07094
```

```
##  3rd Qu.: 3.6028   3rd Qu.: 3.61813
##  Max.   :10.1860   Max.   :10.09784
```

2.3 cut data

```
s <- df %>%
  dplyr::add_rownames() %>%
  dplyr::filter( (control_1 >= 5) | (control_2 >= 5) | (control_3 >= 5)
         | (Hoxa1KN_1 >= 5) | (Hoxa1KN_2 >= 5) | (Hoxa1KN_3 >= 5)) %>%
  dplyr::select(rowname)
```

```
## Warning: Deprecated, use tibble::rownames_to_column() instead.
```

```
cuttedData <- data %>%
  dplyr::add_rownames() %>%
  dplyr::inner_join(s, by='rowname')
```

```
## Warning: Deprecated, use tibble::rownames_to_column() instead.
```

```
cuttedData <- as.data.frame(cuttedData)
rownames(cuttedData) <- cuttedData[,1]
cuttedData <- cuttedData[,-1]
print("")
```

```
## [1] ""
```

```
nrow(cuttedData)
```

```
## [1] 10859
```

```
summary(cuttedData)
```

```
##    control_1        control_2        control_3        Hoxa1KN_1
##  Min.   :     0   Min.   :     0   Min.   :     0   Min.   :     0
##  1st Qu.:   196   1st Qu.:   225   1st Qu.:   282   1st Qu.:   218
##  Median :   445   Median :   510   Median :   642   Median :   520
##  Mean   :  1305   Mean   :  1488   Mean   :  1880   Mean   :  1537
##  3rd Qu.:  1018   3rd Qu.:  1172   3rd Qu.:  1466   3rd Qu.:  1234
##  Max.   :402701   Max.   :438811   Max.   :558984   Max.   :427295
##    Hoxa1KN_2        Hoxa1KN_3
##  Min.   :     0   Min.   :     4
##  1st Qu.:   245   1st Qu.:   284
##  Median :   587   Median :   683
##  Mean   :  1723   Mean   :  1992
##  3rd Qu.:  1379   3rd Qu.:  1624
##  Max.   :498587   Max.   :527743
```

3 Deseq 2

3.1 Generate result

```
colna <- colnames(cuttedData)
coldata <- as.data.frame(c('control','control','control','Hoxa1KN','Hoxa1KN','Hoxa1KN'))
colnames(coldata) <- c('condition')
```

```r
rownames(coldata) <- colna

head(coldata)
```

```
##           condition
## control_1   control
## control_2   control
## control_3   control
## Hoxa1KN_1   Hoxa1KN
## Hoxa1KN_2   Hoxa1KN
## Hoxa1KN_3   Hoxa1KN
```

```r
head(cuttedData)
```

```
##                 control_1 control_2 control_3 Hoxa1KN_1 Hoxa1KN_2
## ENSG00000198888     17528     23007     30241     24418     29152
## ENSG00000198763     21264     26720     35550     28878     32416
## ENSG00000198804    130975    151207    195514    178130    196727
## ENSG00000198712     49769     61906     78608     66478     69758
## ENSG00000228253      9304     11160     12830     12608     13041
## ENSG00000198899     45401     51260     66851     63433     66123
##                 Hoxa1KN_3
## ENSG00000198888     34416
## ENSG00000198763     38422
## ENSG00000198804    244670
## ENSG00000198712     86808
## ENSG00000228253     16063
## ENSG00000198899     79215
```

```r
dds <- DESeq2::DESeqDataSetFromMatrix(countData = as.matrix(cuttedData), colData = coldata, design = ~ c


dds
```

```
## class: DESeqDataSet
## dim: 10859 6
## metadata(1): version
## assays(1): counts
## rownames(10859): ENSG00000198888 ENSG00000198763 ...
##   ENSG00000103222 ENSG00000048828
## rowData names(0):
## colnames(6): control_1 control_2 ... Hoxa1KN_2 Hoxa1KN_3
## colData names(1): condition
```

```r
dds <- DESeq2::DESeq(dds)
```

```
## estimating size factors

## estimating dispersions

## gene-wise dispersion estimates

## mean-dispersion relationship

## final dispersion estimates

## fitting model and testing
```

```
res <- DESeq2::results(dds)
res
```

```
## log2 fold change (MLE): condition Hoxa1KN vs control
## Wald test p-value: condition Hoxa1KN vs control
## DataFrame with 10859 rows and 6 columns
##                         baseMean     log2FoldChange             lfcSE
##                        <numeric>          <numeric>         <numeric>
## ENSG00000198888 25904.6473430736  0.161124364094056 0.0549351104971018
## ENSG00000198763 29989.1626645948  0.103849055295708 0.0473593152192017
## ENSG00000198804 179554.339318949  0.212049138544963 0.0363782524606112
## ENSG00000198712 67814.0020428773 0.0731078325330875 0.0432079292898051
## ENSG00000228253 12341.1879305687  0.158626546883101 0.0441781784779912
## ...                          ...                ...               ...
## ENSG00000173545 837.551361122154  0.257404803543496 0.0727546063210207
## ENSG00000177119 2715.69337174232  0.324855096616299  0.050828982456208
## ENSG00000205808 139.203734703101 -0.789342411447881  0.163975329910848
## ENSG00000103222 1833.02533526154  0.419437629571403 0.0680043534358781
## ENSG00000048828 3471.06222283727 -0.462607929416545 0.0439829337436671
##                             stat              pvalue
##                        <numeric>           <numeric>
## ENSG00000198888  2.93299426607244  0.00335710091955791
## ENSG00000198763  2.19279047458867   0.0283224785505033
## ENSG00000198804  5.82900838281225 5.57577029318664e-09
## ENSG00000198712   1.6920003743465    0.090645902249144
## ENSG00000228253  3.59060858432917 0.000329906754191281
## ...                          ...                 ...
## ENSG00000173545  3.53798634285407 0.000403190912006971
## ENSG00000177119  6.39113908873112 1.64654522353325e-10
## ENSG00000205808 -4.81378760986205 1.48096204018437e-06
## ENSG00000103222  6.16780556507893 6.92442143716582e-10
## ENSG00000048828 -10.5178961483704  7.1451187336472e-26
##                               padj
##                          <numeric>
## ENSG00000198888  0.00562209543168907
## ENSG00000198763   0.0412415704966221
## ENSG00000198804  1.6130066099954e-08
## ENSG00000198712    0.120590927670501
## ENSG00000228253 0.000624224055464401
## ...                            ...
## ENSG00000173545 0.000754471515280883
## ENSG00000177119 5.26399926145481e-10
## ENSG00000205808 3.53691264194494e-06
## ENSG00000103222 2.13696542192465e-09
## ENSG00000048828 4.63685320329992e-25
```

```
summary(res)
```

```
##
## out of 10859 with nonzero total read count
## adjusted p-value < 0.1
## LFC > 0 (up)       : 4007, 37%
## LFC < 0 (down)     : 4009, 37%
## outliers [1]       : 2, 0.018%
## low counts [2]     : 0, 0%
```

```
## (mean count < 26)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results
```

3.2 Cut result by FDR = 0.05 and abs(lfc) > 1

```
res05 <- DESeq2::results(dds, alpha=0.05, lfcThreshold=1, altHypothesis = 'greaterAbs')
res05inc <- DESeq2::results(dds, alpha=0.05, lfcThreshold=1, altHypothesis = 'greater')
res05decs <- DESeq2::results(dds, alpha=0.05, lfcThreshold=1, altHypothesis = 'less')

summary(res05)
```

```
##
## out of 10859 with nonzero total read count
## adjusted p-value < 0.05
## LFC > 1.00 (up)    : 391, 3.6%
## LFC < -1.00 (down) : 516, 4.8%
## outliers [1]       : 2, 0.018%
## low counts [2]     : 0, 0%
## (mean count < 26)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results
```

```
summary(res05inc)
```

```
##
## out of 10859 with nonzero total read count
## adjusted p-value < 0.05
## LFC > 1.00 (up)    : 390, 3.6%
## LFC < -1.00 (down) : 0, 0%
## outliers [1]       : 2, 0.018%
## low counts [2]     : 0, 0%
## (mean count < 26)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results
```

```
summary(res05decs)
```

```
##
## out of 10859 with nonzero total read count
## adjusted p-value < 0.05
## LFC > 1.00 (up)    : 0, 0%
## LFC < -1.00 (down) : 518, 4.8%
## outliers [1]       : 2, 0.018%
## low counts [2]     : 0, 0%
## (mean count < 26)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results
```

3.3 Get Gene List, sort by lfc

```
UpGeneFilter <- which(  !is.na(res$padj) &
                                    !is.na(res$log2FoldChange) &
```

```
                                    res$padj < 0.05 &
                                    res$log2FoldChange > 1)

UpGeneslfc <- res$log2FoldChange[UpGeneFilter]
UpGeneList <- rownames(res)[UpGeneFilter][order(-UpGeneslfc)]

DownGeneFilter <- which(  !is.na(res$padj) &
                          !is.na(res$log2FoldChange) &
                          res$padj < 0.05 &
                          res$log2FoldChange < -1)
DownGeneslfc <- res$log2FoldChange[DownGeneFilter]
DownGeneList <- rownames(res)[DownGeneFilter][order(DownGeneslfc)]

AllGeneList <- rownames(res)


length(AllGeneList)
```

```
## [1] 10859
```

```
length(UpGeneList)
```

```
## [1] 665
```

```
length(DownGeneList)
```

```
## [1] 882
```

3.4 Refine hierarchical clustering

Use all differentially expressed gene

```
vsd <- vst(dds, blind=FALSE)
ntd <- normTransform(dds)
resLFC <- DESeq2::lfcShrink(dds, coef="condition_Hoxa1KN_vs_control", type="apeglm")
```

```
## using 'apeglm' for LFC shrinkage. If used in published research, please cite:
##     Zhu, A., Ibrahim, J.G., Love, M.I. (2018) Heavy-tailed prior distributions for
##     sequence count data: removing the noise and preserving large differences.
##     bioRxiv. https://doi.org/10.1101/303255
```

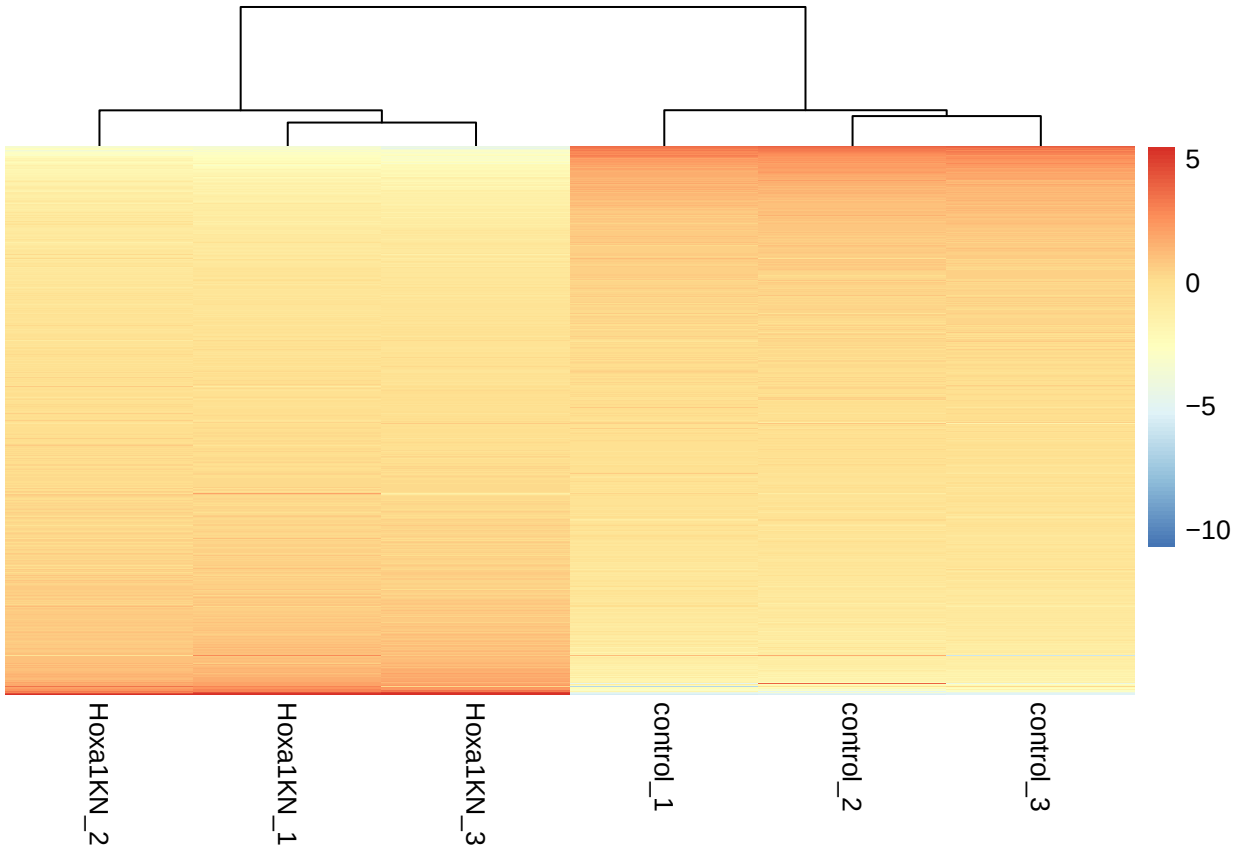
```
select2 <- order(abs(resLFC$log2FoldChange[which(resLFC$padj<0.05&abs(resLFC$log2FoldChange)>1)]),
                 decreasing=TRUE)

select_diff <- assay(ntd)[select2,]-rowMeans(assay(ntd)[select2,])

select3 <- order(rowMeans(select_diff[,1:3]), decreasing = TRUE)
#df <- as.data.frame(colData(dds)[,c("condition","type")])
index <- select2[select3]
length(index)
```

```
## [1] 1494
```

```
mySd <- sd( assay(ntd)[index,]-rowMeans(assay(ntd)[index,]) )
pheatmap(( assay(ntd)[index,]-rowMeans(assay(ntd)[index,]) )/mySd, cluster_rows=FALSE, show_rownames=FA
         cluster_cols=TRUE)
```




```
library(biomaRt)
library(ALL)
library(gage)
library("org.Hs.eg.db")
```

```
## Loading required package: AnnotationDbi
```

```
##
## Attaching package: 'AnnotationDbi'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
##
```

```
library(pathview)
```

```
## ##############################################################################
## Pathview is an open source software package distributed under GNU General
## Public License version 3 (GPLv3). Details of GPLv3 is available at
## http://www.gnu.org/licenses/gpl-3.0.html. Particullary, users are required to
## formally cite the original Pathview paper (not just mention it) in publications
## or products. For details, do citation("pathview") within R.
##
## The pathview downloads and uses KEGG data. Non-academic uses may require a KEGG
## license agreement (details at http://www.kegg.jp/kegg/legal.html).
## ##############################################################################
```

```
library(png)
library(GO.db)
```

```
##
```

```
library(GOstats)
```

```
## Loading required package: Category
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following object is masked from 'package:S4Vectors':
##
##     expand
```

```
## Loading required package: graph
```

```
##
## Attaching package: 'GOstats'
```

```
## The following object is masked from 'package:AnnotationDbi':
##
##     makeGOGraph
```

4. Enrichment: TopGo Solution

I failed to use topGo package. Instead, I use GOStats package. ## 4.1 Load Ensmbl to Go_id table

```
if(!file.exists('EG2GO.RDS')){
  grch37 = useEnsembl(biomart="ensembl", dataset = "hsapiens_gene_ensembl")
  EG2GO <- getBM(mart=grch37, attributes=c('ensembl_gene_id','go_id'))
  saveRDS(EG2GO, file='EG2GO.RDS')
}else{
  EG2GO <- readRDS('EG2GO.RDS')
}
```

4.2 Convert Ensembl Id into Go Id (or name)

```
EG2GO <- EG2GO[EG2GO$go_id != '',]

upGoId <- EG2GO %>%
  filter(ensembl_gene_id %in% UpGeneList) %>%
  dplyr::select(go_id)
upGoId <- unique(upGoId)

downGoId <- EG2GO %>%
  filter(ensembl_gene_id %in% DownGeneList) %>%
  dplyr::select(go_id)
downGoId <- unique(downGoId)

allGoId <- EG2GO %>%
  filter(ensembl_gene_id %in% AllGeneList) %>%
```

```
  dplyr::select(go_id)
allGoId <- unique(allGoId)
```

4.3 Using topGo package to do enrichment tests

```
xx <- annFUN.org("BP", mapping = "org.Hs.eg.db", ID = "symbol")
head(xx)

geneList <- factor(as.integer( allGoId[,1] %in% upGoId[,1]))
names(geneList) <- allGoId

str(geneList)

upGoData <- new("topGOdata",
                description='GO data for increased genes',
                ontology = "BP", allGenes = geneList, geneSel = upGoId, nodeSize = 5,
                annot = annFUN.db, mapping = "org.Hs.eg.db", ID = "symbol")


resultFisher <- runTest(upGoData, algorithm = "classic", statistic = "fisher")
resultFisher

resultKS <- runTest(sampleGOdata, algorithm = "classic", statistic = "ks")
resultKS.elim <- runTest(sampleGOdata, algorithm = "elim", statistic = "ks")

showSigOfNodes(upGoData, score(resultKS.elim), firstSigNodes = 5, useInfo = 'all')
```

5. Using GOStat package to do enrichment tests

5.1 Map Ids

```
res$symbol = mapIds(org.Hs.eg.db,
                    keys=row.names(res),
                    column="SYMBOL",
                    keytype="ENSEMBL",
                    multiVals="first")
```

```
## 'select()' returned 1:many mapping between keys and columns
```

```
res$entrez = mapIds(org.Hs.eg.db,
                    keys=row.names(res),
                    column="ENTREZID",
                    keytype="ENSEMBL",
                    multiVals="first")
```

```
## 'select()' returned 1:many mapping between keys and columns
```

```
res$goids = mapIds(org.Hs.eg.db,
                    keys=row.names(res),
                    column="GO",
```

```
                    keytype="ENSEMBL",
                    multiVals="first")
```

```
## 'select()' returned 1:many mapping between keys and columns
```

```
res$name =   mapIds(org.Hs.eg.db,
                    keys=row.names(res),
                    column="GENENAME",
                    keytype="ENSEMBL",
                    multiVals="first")
```

```
## 'select()' returned 1:many mapping between keys and columns
```

5.2 Up regulated list and Down regulated list

Up: filter by P.adj < 0.05 and LFC > 1
Down: filter by P.adj < 0.05 and LFC < -1

```
up.idx <- which(res$padj<0.05 & res$log2FoldChange > 1)
up.entrez.id <- unique(res[up.idx, 'entrez'])

down.idx <- which(res$padj<0.05 & res$log2FoldChange < -1)
down.entrez.id <- unique(res[down.idx, 'entrez'])

universeGenes <-  unique(res$entrez)
```

5.3 Up gene GO analysis

BP

```
params <- new("GOHyperGParams",
     geneIds=up.entrez.id,
     universeGeneIds=universeGenes,
     annotation="org.Hs.eg.db",
     ontology="BP",
     pvalueCutoff=0.001,
     conditional=FALSE,
     testDirection="over")

hgOver <- hyperGTest(params)
summary(hgOver)[1:10,c("GOBPID","Pvalue","Term")]
```

```
##        GOBPID       Pvalue                                    Term
## 1  GO:0023052 1.425185e-21                                 signaling
## 2  GO:0032501 1.928764e-21        multicellular organismal process
## 3  GO:0007154 7.758182e-21                      cell communication
## 4  GO:0007166 2.384079e-19 cell surface receptor signaling pathway
## 5  GO:0007165 1.127884e-17                    signal transduction
## 6  GO:0007275 2.855824e-17      multicellular organism development
## 7  GO:0042127 6.335171e-17        regulation of cell proliferation
## 8  GO:0048856 3.210409e-16        anatomical structure development
## 9  GO:0032502 6.211288e-16                  developmental process
## 10 GO:0040011 8.888323e-16                             locomotion
```

CC

```
params <- new("GOHyperGParams",
     geneIds=up.entrez.id,
     universeGeneIds=universeGenes,
     annotation="org.Hs.eg.db",
     ontology="CC",
     pvalueCutoff=0.001,
     conditional=FALSE,
     testDirection="over")

hgOver <- hyperGTest(params)
summary(hgOver)[1:10,c("GOCCID","Pvalue","Term")]
```

```
##        GOCCID       Pvalue                                  Term
## 1  GO:0005887 1.080787e-23  integral component of plasma membrane
## 2  GO:0031226 1.978680e-23 intrinsic component of plasma membrane
## 3  GO:0044459 9.522246e-23                   plasma membrane part
## 4  GO:0005886 6.930665e-21                        plasma membrane
## 5  GO:0071944 7.512664e-20                         cell periphery
## 6  GO:0031224 1.402582e-15        intrinsic component of membrane
## 7  GO:0016021 3.441904e-15         integral component of membrane
## 8  GO:0044425 1.238801e-12                          membrane part
## 9  GO:0031012 4.567027e-09                   extracellular matrix
## 10 GO:0005578 3.775583e-08     proteinaceous extracellular matrix
```

MF

```
params <- new("GOHyperGParams",
     geneIds=up.entrez.id,
     universeGeneIds=universeGenes,
     annotation="org.Hs.eg.db",
     ontology="MF",
     pvalueCutoff=0.001,
     conditional=FALSE,
     testDirection="over")

hgOver <- hyperGTest(params)
summary(hgOver)[1:10,c("GOMFID","Pvalue","Term")]
```

```
##        GOMFID       Pvalue                                     Term
## 1  GO:0048018 4.413064e-18                 receptor ligand activity
## 2  GO:0030545 1.697980e-17               receptor regulator activity
## 3  GO:0038023 3.446304e-17               signaling receptor activity
## 4  GO:0060089 3.446304e-17             molecular transducer activity
## 5  GO:0004871 3.476208e-16                signal transducer activity
## 6  GO:0004888 1.525144e-14 transmembrane signaling receptor activity
## 7  GO:0005102 9.265250e-13                signaling receptor binding
## 8  GO:0005125 2.345639e-12                          cytokine activity
## 9  GO:0008083 1.054373e-11                    growth factor activity
## 10 GO:0022804 2.797723e-07 active transmembrane transporter activity
```

5.4 Down gene GO analysis

BP

```
params <- new("GOHyperGParams",
     geneIds=down.entrez.id,
     universeGeneIds=universeGenes,
     annotation="org.Hs.eg.db",
     ontology="BP",
     pvalueCutoff=0.001,
     conditional=FALSE,
     testDirection="over")

hgOver <- hyperGTest(params)
summary(hgOver)[1:10,c("GOBPID","Pvalue","Term")]
```

```
##        GOBPID       Pvalue                          Term
## 1  GO:0007049 2.487452e-30                     cell cycle
## 2  GO:0022402 8.589428e-30             cell cycle process
## 3  GO:0000278 4.309951e-28             mitotic cell cycle
## 4  GO:0051301 8.941152e-28                  cell division
## 5  GO:0007059 9.179667e-26         chromosome segregation
## 6  GO:0000280 4.115387e-25               nuclear division
## 7  GO:0098813 6.598595e-24 nuclear chromosome segregation
## 8  GO:1903047 7.673398e-24     mitotic cell cycle process
## 9  GO:0000819 8.080195e-24   sister chromatid segregation
## 10 GO:0048285 1.308955e-22              organelle fission
```

CC

```
params <- new("GOHyperGParams",
     geneIds=down.entrez.id,
     universeGeneIds=universeGenes,
     annotation="org.Hs.eg.db",
     ontology="CC",
     pvalueCutoff=0.001,
     conditional=FALSE,
     testDirection="over")

hgOver <- hyperGTest(params)
summary(hgOver)[1:10,c("GOCCID","Pvalue","Term")]
```

```
##        GOCCID       Pvalue                                    Term
## 1  GO:0000793 3.869670e-22                   condensed chromosome
## 2  GO:0099512 4.027250e-21                   supramolecular fiber
## 3  GO:0099081 8.837715e-21                 supramolecular polymer
## 4  GO:0099080 1.032632e-20                 supramolecular complex
## 5  GO:0044430 4.042598e-20                      cytoskeletal part
## 6  GO:0005856 8.327907e-19                           cytoskeleton
## 7  GO:0000779 1.463614e-18 condensed chromosome, centromeric region
## 8  GO:0098687 5.894188e-18                       chromosomal region
## 9  GO:0000775 1.190760e-17           chromosome, centromeric region
## 10 GO:0005819 6.624982e-17                                  spindle
```

MF

```
params <- new("GOHyperGParams",
     geneIds=down.entrez.id,
     universeGeneIds=universeGenes,
     annotation="org.Hs.eg.db",
     ontology="MF",
     pvalueCutoff=0.001,
     conditional=FALSE,
     testDirection="over")

hgOver <- hyperGTest(params)
summary(hgOver)[1:10,c("GOMFID","Pvalue","Term")]
```

```
##        GOMFID       Pvalue                                     Term
## 1  GO:0008092 1.925953e-13               cytoskeletal protein binding
## 2  GO:0003779 3.703376e-10                             actin binding
## 3  GO:0008017 4.690700e-10                       microtubule binding
## 4  GO:0015631 5.564481e-08                           tubulin binding
## 5  GO:0035639 1.754166e-07 purine ribonucleoside triphosphate binding
## 6  GO:0005524 2.208051e-07                               ATP binding
## 7  GO:0000166 2.483132e-07                        nucleotide binding
## 8  GO:1901265 2.601452e-07               nucleoside phosphate binding
## 9  GO:0008144 2.984813e-07                             drug binding
## 10 GO:0032555 4.954504e-07             purine ribonucleotide binding
```

6. Pathway: Gage

6.1 Gage anaylsis using kegg

```
library(gageData)
resLFC = res$log2FoldChange
names(resLFC) = res$entrez
data(kegg.sets.hs)

kres = gage(resLFC, gsets=kegg.sets.hs, same.dir=TRUE)


k_pathways_Up = data.frame(id=rownames(kres$greater), kres$greater) %>%
  tbl_df() %>%
  filter(row_number()<=5) %>%
  .$id %>%
  as.character()

k_pathways_Down = data.frame(id=rownames(kres$less), kres$less) %>%
  tbl_df() %>%
  filter(row_number()<=5) %>%
  .$id %>%
  as.character()

k_Up_ids = substr(k_pathways_Up, start=1, stop=8)
```

```
k_Down_ids = substr(k_pathways_Down, start=1, stop=8)
```

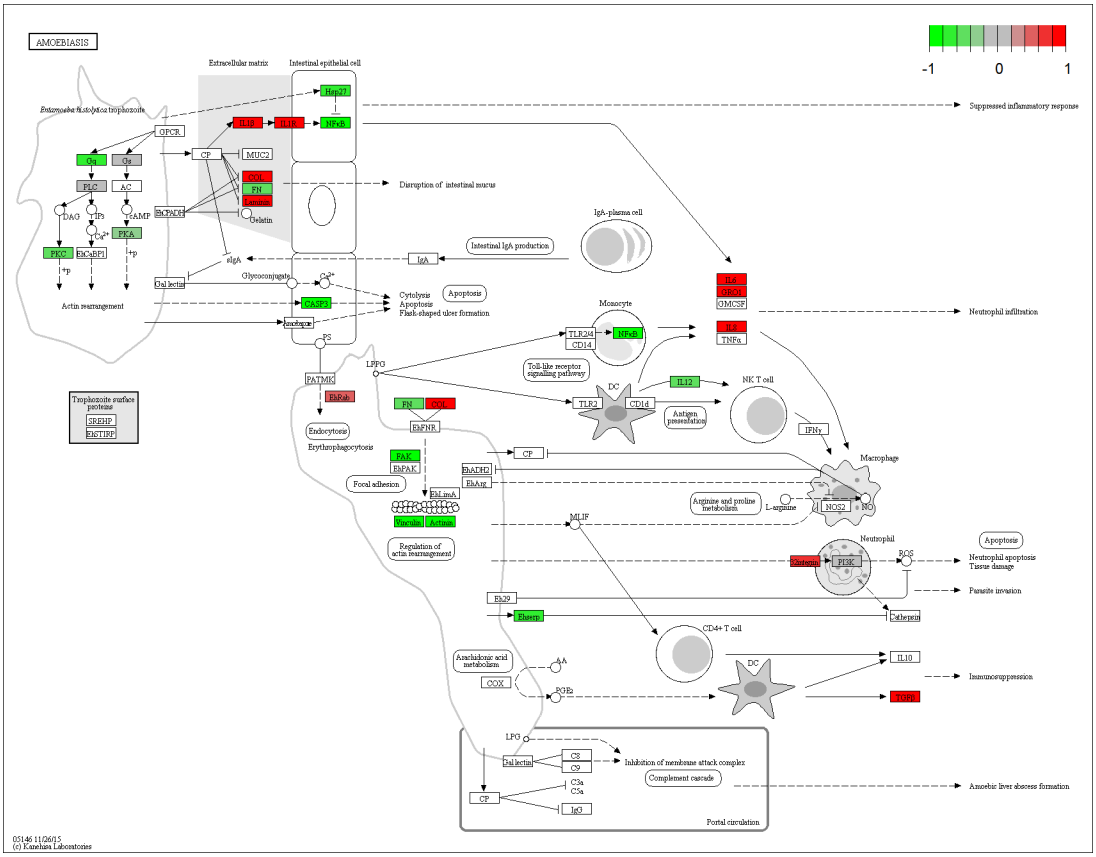
Up regulated pathways

```
pathview(gene.data=resLFC, pathway.id=k_Up_ids, species="hsa", kegg.dir = "PathviewResult/Up/", new.sig
```

```
## 'select()' returned 1:1 mapping between keys and columns
## Info: Working in directory C:/GitRepo/STATBioinformatics
## Info: Writing image file hsa04060.pathview.png
## 'select()' returned 1:1 mapping between keys and columns
## Info: Working in directory C:/GitRepo/STATBioinformatics
## Info: Writing image file hsa04142.pathview.png
## Info: some node width is different from others, and hence adjusted!
## 'select()' returned 1:1 mapping between keys and columns
## Info: Working in directory C:/GitRepo/STATBioinformatics
## Info: Writing image file hsa05323.pathview.png
## 'select()' returned 1:1 mapping between keys and columns
## Info: Working in directory C:/GitRepo/STATBioinformatics
## Info: Writing image file hsa04974.pathview.png
## Info: some node width is different from others, and hence adjusted!
## 'select()' returned 1:1 mapping between keys and columns
## Info: Working in directory C:/GitRepo/STATBioinformatics
## Info: Writing image file hsa05146.pathview.png
```

```
for(f in k_Up_ids){
  f <- paste0(f,'.pathview.png')
  knitr::include_graphics(f)
}

knitr::include_graphics(f)
```



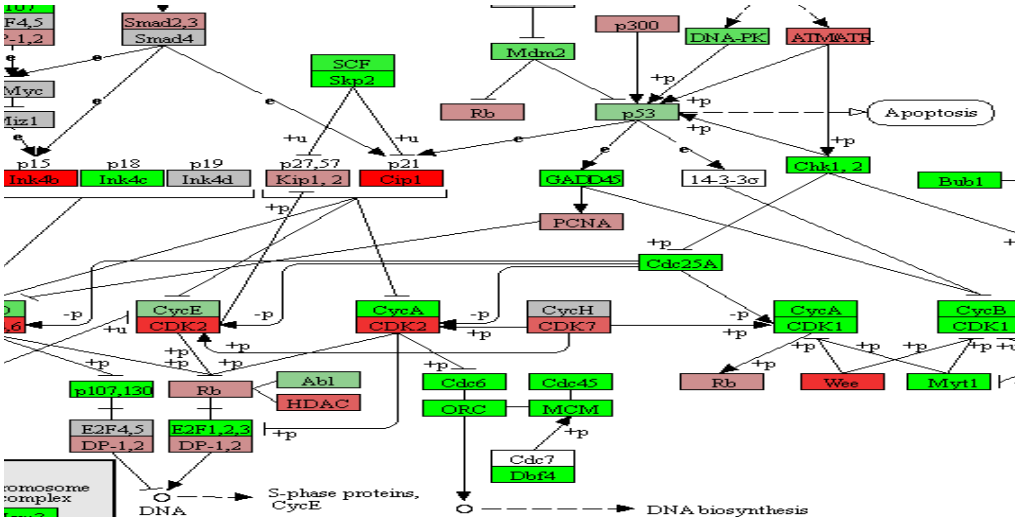
Down regulated pathways

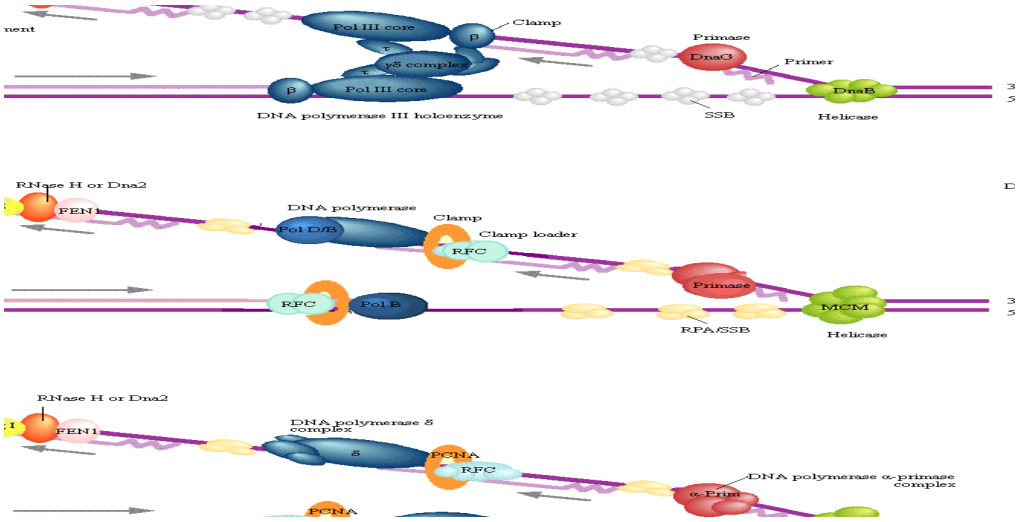
```
pathview(gene.data=resLFC, pathway.id=k_Down_ids, species="hsa", kegg.dir = "PathviewResult/Down/", new
```

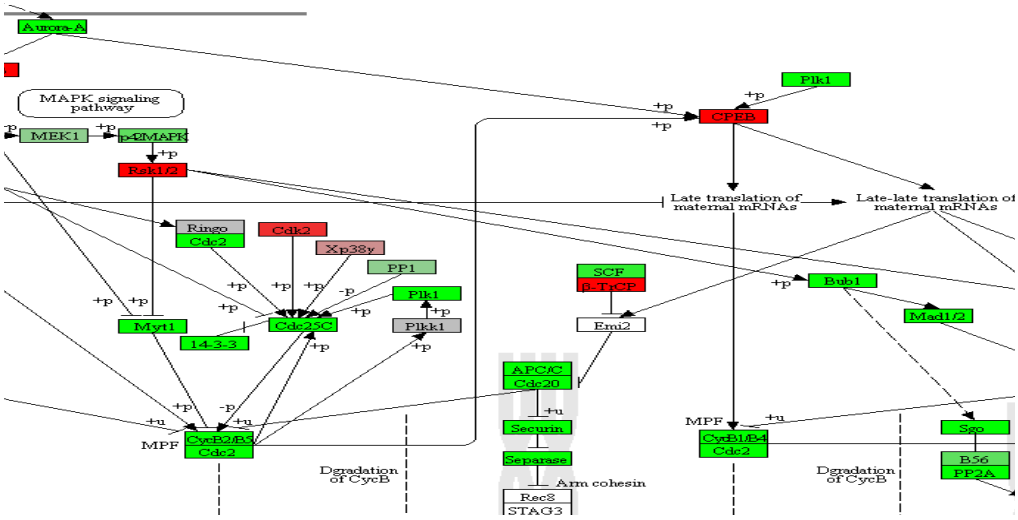
```
## 'select()' returned 1:1 mapping between keys and columns
## Info: Working in directory C:/GitRepo/STATBioinformatics
## Info: Writing image file hsa04110.pathview.png
## 'select()' returned 1:1 mapping between keys and columns
## Info: Working in directory C:/GitRepo/STATBioinformatics
## Info: Writing image file hsa03030.pathview.png
## 'select()' returned 1:1 mapping between keys and columns
## Info: Working in directory C:/GitRepo/STATBioinformatics
## Info: Writing image file hsa04114.pathview.png
## 'select()' returned 1:1 mapping between keys and columns
## Info: Working in directory C:/GitRepo/STATBioinformatics
## Info: Writing image file hsa05130.pathview.png
## 'select()' returned 1:1 mapping between keys and columns
## Info: Working in directory C:/GitRepo/STATBioinformatics
```

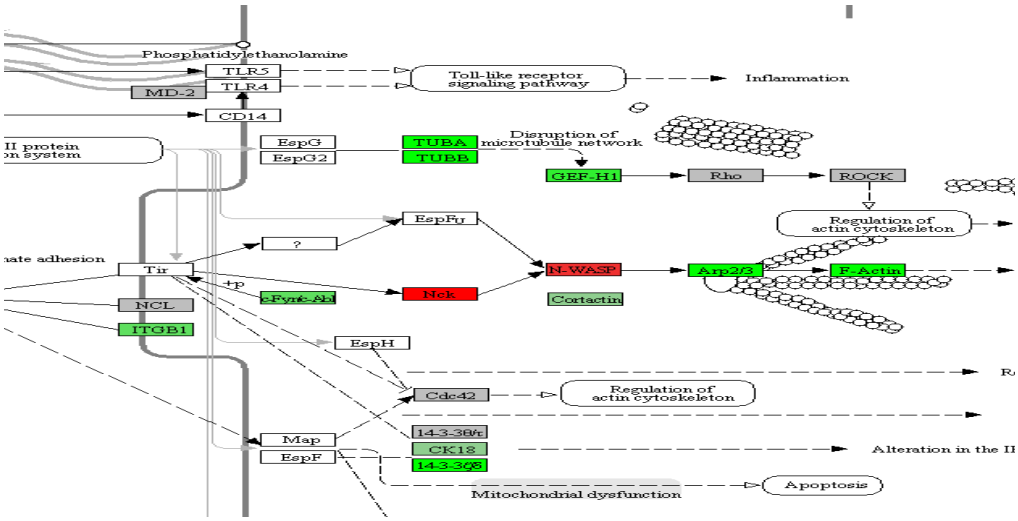
```
## Info: Writing image file hsa03440.pathview.png
```

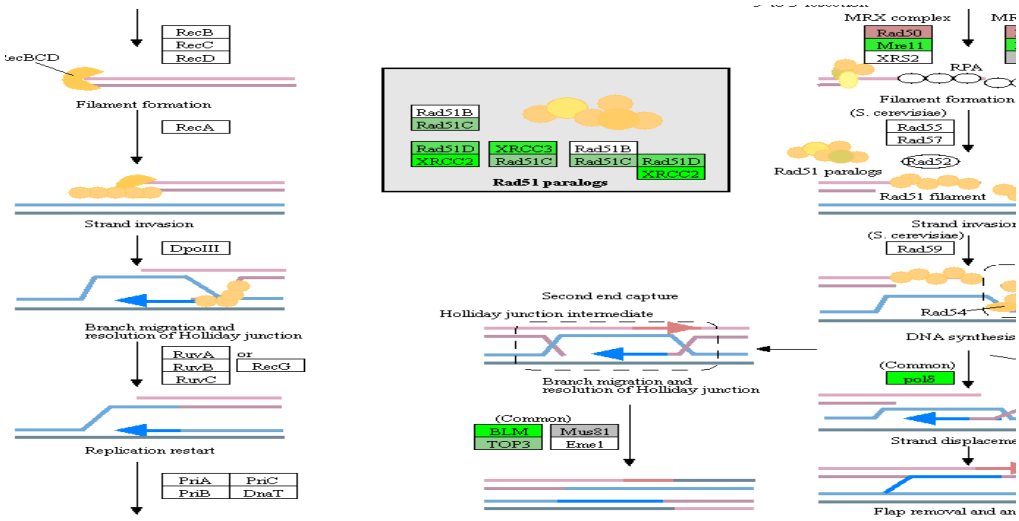
```
for(f in k_Down_ids){
  f <- paste0(f,'.pathview.png')
  img <- readPNG(f)
  plot.new()
  rasterImage(img,-0.5, -0.3, 1.5, 1.3)
}
```




nent
Pol III core
Clamp
β
ε
γδ complex
τ
β
Pol III core
DNA polymerase III holoenzyme
Primase
DnaG
Primer
DnaB
SSB
Helicase
3
5
RNase H or Dna2
FEN1
DNA polymerase
Pol Dδ
Clamp
Clamp loader
RFC
RFC
Pol δ
Primase
MCM
RPA/SSB
Helicase
3
5
RNase H or Dna2
FEN1
DNA polymerase δ
complex
δ
PCNA
RFC
DNA polymerase α-primase
complex
α-Prim
PCNA


Aurora-A
MAPK signaling pathway
MEK1
p-42MAPK
Rsk1/2
Ringo
Cdc2
Cdk2
Xp38γ
PP1
Plk1
Myt1
14-3-3
Cdc25C
Plkk1
MPF
CycB2/B3
Cdc2
Degradation of CycB
APC/C
Cdc20
Securin
Separase
Arm cohesin
Rec8
STAG3
SCF
β-TrCP
Emi2
Plk1
CPEB
Late translation of maternal mRNAs
Late-late translation of maternal mRNAs
Bub1
Mad1/2
MPF
CycB1/B4
Cdc2
Sgo
B56
PP2A
Degradation of CycB


Phosphatidylethanolamine
TLR5
TLR4
MD-2
CD14
Toll-like receptor signaling pathway
Inflammation
EspG
EspG2
TUBA
TUBB
Disruption of microtubule network
GEF-H1
Rho
ROCK
Regulation of actin cytoskeleton
EspFu
?
Tir
N-WASP
Arp2/3
F-Actin
Nck
Cortactin
NCL
ITGB1
EspH
Cdc42
Regulation of actin cytoskeleton
Map
EspF
14-3-3θ
CK18
14-3-3β
Alteration in the
Mitochondrial dysfunction
Apoptosis



6.2 Gage anaylsis using GO

```
data(go.sets.hs)
data(go.subs.hs)
resLFC = res$log2FoldChange

gobpsets = go.sets.hs[go.subs.hs$BP]
goccsets = go.sets.hs[go.subs.hs$CC]
gomfsets = go.sets.hs[go.subs.hs$MF]

GoLFC <- resLFC
names(GoLFC) <- res$goids

gobpres <- gage(GoLFC, gsets=gobpsets, same.dir=TRUE)
goccres = gage(GoLFC, gsets=goccsets, same.dir=TRUE)
gomfres = gage(GoLFC, gsets=gomfsets, same.dir=TRUE)


gobpres = gage(resLFC, gsets=gobpsets, same.dir=TRUE)
```

bp result

```
lapply(gobpres, head)
```

```
## $greater
##                                                p.geomean stat.mean p.val
## GO:0000002 mitochondrial genome maintenance           NA       NaN    NA
## GO:0000003 reproduction                               NA       NaN    NA
## GO:0000012 single strand break repair                 NA       NaN    NA
## GO:0000018 regulation of DNA recombination            NA       NaN    NA
## GO:0000019 regulation of mitotic recombination        NA       NaN    NA
## GO:0000022 mitotic spindle elongation                 NA       NaN    NA
##                                                q.val set.size exp1
## GO:0000002 mitochondrial genome maintenance       NA        0   NA
## GO:0000003 reproduction                           NA        0   NA
## GO:0000012 single strand break repair             NA        0   NA
## GO:0000018 regulation of DNA recombination        NA        0   NA
## GO:0000019 regulation of mitotic recombination    NA        0   NA
## GO:0000022 mitotic spindle elongation             NA        0   NA
##
## $less
##                                                p.geomean stat.mean p.val
## GO:0000002 mitochondrial genome maintenance           NA       NaN    NA
## GO:0000003 reproduction                               NA       NaN    NA
## GO:0000012 single strand break repair                 NA       NaN    NA
## GO:0000018 regulation of DNA recombination            NA       NaN    NA
## GO:0000019 regulation of mitotic recombination        NA       NaN    NA
## GO:0000022 mitotic spindle elongation                 NA       NaN    NA
##                                                q.val set.size exp1
## GO:0000002 mitochondrial genome maintenance       NA        0   NA
## GO:0000003 reproduction                           NA        0   NA
## GO:0000012 single strand break repair             NA        0   NA
## GO:0000018 regulation of DNA recombination        NA        0   NA
## GO:0000019 regulation of mitotic recombination    NA        0   NA
## GO:0000022 mitotic spindle elongation             NA        0   NA
##
## $stats
##                                                stat.mean exp1
## GO:0000002 mitochondrial genome maintenance          NaN   NA
## GO:0000003 reproduction                              NaN   NA
## GO:0000012 single strand break repair                NaN   NA
## GO:0000018 regulation of DNA recombination           NaN   NA
## GO:0000019 regulation of mitotic recombination       NaN   NA
## GO:0000022 mitotic spindle elongation                NaN   NA
```

cc result

```
lapply(goccres, head)
```

```
## $greater
##                                                        p.geomean
## GO:0000015 phosphopyruvate hydratase complex                   NA
## GO:0000109 nucleotide-excision repair complex                  NA
```

```
## GO:0000110 nucleotide-excision repair factor 1 complex          NA
## GO:0000118 histone deacetylase complex                          NA
## GO:0000120 RNA polymerase I transcription factor complex        NA
## GO:0000123 histone acetyltransferase complex                    NA
##                                                          stat.mean p.val
## GO:0000015 phosphopyruvate hydratase complex                   NaN    NA
## GO:0000109 nucleotide-excision repair complex                  NaN    NA
## GO:0000110 nucleotide-excision repair factor 1 complex         NaN    NA
## GO:0000118 histone deacetylase complex                         NaN    NA
## GO:0000120 RNA polymerase I transcription factor complex       NaN    NA
## GO:0000123 histone acetyltransferase complex                   NaN    NA
##                                                          q.val set.size
## GO:0000015 phosphopyruvate hydratase complex                NA        0
## GO:0000109 nucleotide-excision repair complex               NA        0
## GO:0000110 nucleotide-excision repair factor 1 complex      NA        0
## GO:0000118 histone deacetylase complex                      NA        0
## GO:0000120 RNA polymerase I transcription factor complex    NA        0
## GO:0000123 histone acetyltransferase complex                NA        0
##                                                          exp1
## GO:0000015 phosphopyruvate hydratase complex               NA
## GO:0000109 nucleotide-excision repair complex              NA
## GO:0000110 nucleotide-excision repair factor 1 complex     NA
## GO:0000118 histone deacetylase complex                     NA
## GO:0000120 RNA polymerase I transcription factor complex   NA
## GO:0000123 histone acetyltransferase complex               NA
##
## $less
##                                                          p.geomean
## GO:0000015 phosphopyruvate hydratase complex                    NA
## GO:0000109 nucleotide-excision repair complex                   NA
## GO:0000110 nucleotide-excision repair factor 1 complex          NA
## GO:0000118 histone deacetylase complex                          NA
## GO:0000120 RNA polymerase I transcription factor complex        NA
## GO:0000123 histone acetyltransferase complex                    NA
##                                                          stat.mean p.val
## GO:0000015 phosphopyruvate hydratase complex                   NaN    NA
## GO:0000109 nucleotide-excision repair complex                  NaN    NA
## GO:0000110 nucleotide-excision repair factor 1 complex         NaN    NA
## GO:0000118 histone deacetylase complex                         NaN    NA
## GO:0000120 RNA polymerase I transcription factor complex       NaN    NA
## GO:0000123 histone acetyltransferase complex                   NaN    NA
##                                                          q.val set.size
## GO:0000015 phosphopyruvate hydratase complex                NA        0
## GO:0000109 nucleotide-excision repair complex               NA        0
## GO:0000110 nucleotide-excision repair factor 1 complex      NA        0
## GO:0000118 histone deacetylase complex                      NA        0
## GO:0000120 RNA polymerase I transcription factor complex    NA        0
## GO:0000123 histone acetyltransferase complex                NA        0
##                                                          exp1
## GO:0000015 phosphopyruvate hydratase complex               NA
## GO:0000109 nucleotide-excision repair complex              NA
## GO:0000110 nucleotide-excision repair factor 1 complex     NA
## GO:0000118 histone deacetylase complex                     NA
## GO:0000120 RNA polymerase I transcription factor complex   NA
```

```
## GO:0000123 histone acetyltransferase complex               NA
##
## $stats
##                                                          stat.mean exp1
## GO:0000015 phosphopyruvate hydratase complex                   NaN   NA
## GO:0000109 nucleotide-excision repair complex                  NaN   NA
## GO:0000110 nucleotide-excision repair factor 1 complex         NaN   NA
## GO:0000118 histone deacetylase complex                         NaN   NA
## GO:0000120 RNA polymerase I transcription factor complex       NaN   NA
## GO:0000123 histone acetyltransferase complex                   NaN   NA
```

mf result

```
lapply(gomfres, head)
```

```
## $greater
##                                                                    p.geomean
## GO:0000009 alpha-1,6-mannosyltransferase activity                         NA
## GO:0000010 trans-hexaprenyltranstransferase activity                      NA
## GO:0000014 single-stranded DNA specific endodeoxyribonuclease activity    NA
## GO:0000016 lactase activity                                               NA
## GO:0000026 alpha-1,2-mannosyltransferase activity                         NA
## GO:0000030 mannosyltransferase activity                                   NA
##                                                                    stat.mean
## GO:0000009 alpha-1,6-mannosyltransferase activity                        NaN
## GO:0000010 trans-hexaprenyltranstransferase activity                     NaN
## GO:0000014 single-stranded DNA specific endodeoxyribonuclease activity   NaN
## GO:0000016 lactase activity                                              NaN
## GO:0000026 alpha-1,2-mannosyltransferase activity                        NaN
## GO:0000030 mannosyltransferase activity                                  NaN
##                                                                    p.val
## GO:0000009 alpha-1,6-mannosyltransferase activity                     NA
## GO:0000010 trans-hexaprenyltranstransferase activity                  NA
## GO:0000014 single-stranded DNA specific endodeoxyribonuclease activity NA
## GO:0000016 lactase activity                                           NA
## GO:0000026 alpha-1,2-mannosyltransferase activity                     NA
## GO:0000030 mannosyltransferase activity                               NA
##                                                                    q.val
## GO:0000009 alpha-1,6-mannosyltransferase activity                     NA
## GO:0000010 trans-hexaprenyltranstransferase activity                  NA
## GO:0000014 single-stranded DNA specific endodeoxyribonuclease activity NA
## GO:0000016 lactase activity                                           NA
## GO:0000026 alpha-1,2-mannosyltransferase activity                     NA
## GO:0000030 mannosyltransferase activity                               NA
##                                                                    set.size
## GO:0000009 alpha-1,6-mannosyltransferase activity                         0
## GO:0000010 trans-hexaprenyltranstransferase activity                      0
## GO:0000014 single-stranded DNA specific endodeoxyribonuclease activity    0
## GO:0000016 lactase activity                                               0
## GO:0000026 alpha-1,2-mannosyltransferase activity                         0
## GO:0000030 mannosyltransferase activity                                   0
##                                                                    exp1
## GO:0000009 alpha-1,6-mannosyltransferase activity                    NA
```

```
## GO:0000010 trans-hexaprenyltranstransferase activity                     NA
## GO:0000014 single-stranded DNA specific endodeoxyribonuclease activity   NA
## GO:0000016 lactase activity                                              NA
## GO:0000026 alpha-1,2-mannosyltransferase activity                        NA
## GO:0000030 mannosyltransferase activity                                  NA
##
## $less
##                                                                      p.geomean
## GO:0000009 alpha-1,6-mannosyltransferase activity                           NA
## GO:0000010 trans-hexaprenyltranstransferase activity                        NA
## GO:0000014 single-stranded DNA specific endodeoxyribonuclease activity      NA
## GO:0000016 lactase activity                                                 NA
## GO:0000026 alpha-1,2-mannosyltransferase activity                           NA
## GO:0000030 mannosyltransferase activity                                     NA
##                                                                     stat.mean
## GO:0000009 alpha-1,6-mannosyltransferase activity                         NaN
## GO:0000010 trans-hexaprenyltranstransferase activity                      NaN
## GO:0000014 single-stranded DNA specific endodeoxyribonuclease activity    NaN
## GO:0000016 lactase activity                                               NaN
## GO:0000026 alpha-1,2-mannosyltransferase activity                         NaN
## GO:0000030 mannosyltransferase activity                                   NaN
##                                                                        p.val
## GO:0000009 alpha-1,6-mannosyltransferase activity                         NA
## GO:0000010 trans-hexaprenyltranstransferase activity                      NA
## GO:0000014 single-stranded DNA specific endodeoxyribonuclease activity    NA
## GO:0000016 lactase activity                                               NA
## GO:0000026 alpha-1,2-mannosyltransferase activity                         NA
## GO:0000030 mannosyltransferase activity                                   NA
##                                                                       q.val
## GO:0000009 alpha-1,6-mannosyltransferase activity                         NA
## GO:0000010 trans-hexaprenyltranstransferase activity                      NA
## GO:0000014 single-stranded DNA specific endodeoxyribonuclease activity    NA
## GO:0000016 lactase activity                                               NA
## GO:0000026 alpha-1,2-mannosyltransferase activity                         NA
## GO:0000030 mannosyltransferase activity                                   NA
##                                                                    set.size
## GO:0000009 alpha-1,6-mannosyltransferase activity                         0
## GO:0000010 trans-hexaprenyltranstransferase activity                      0
## GO:0000014 single-stranded DNA specific endodeoxyribonuclease activity    0
## GO:0000016 lactase activity                                               0
## GO:0000026 alpha-1,2-mannosyltransferase activity                         0
## GO:0000030 mannosyltransferase activity                                   0
##                                                                       exp1
## GO:0000009 alpha-1,6-mannosyltransferase activity                       NA
## GO:0000010 trans-hexaprenyltranstransferase activity                    NA
## GO:0000014 single-stranded DNA specific endodeoxyribonuclease activity  NA
## GO:0000016 lactase activity                                             NA
## GO:0000026 alpha-1,2-mannosyltransferase activity                       NA
## GO:0000030 mannosyltransferase activity                                 NA
##
## $stats
##                                                                     stat.mean
## GO:0000009 alpha-1,6-mannosyltransferase activity                         NaN
## GO:0000010 trans-hexaprenyltranstransferase activity                      NaN
```

```
## GO:0000014 single-stranded DNA specific endodeoxyribonuclease activity       NaN
## GO:0000016 lactase activity                                                  NaN
## GO:0000026 alpha-1,2-mannosyltransferase activity                            NaN
## GO:0000030 mannosyltransferase activity                                      NaN
##                                                                         exp1
## GO:0000009 alpha-1,6-mannosyltransferase activity                         NA
## GO:0000010 trans-hexaprenyltranstransferase activity                      NA
## GO:0000014 single-stranded DNA specific endodeoxyribonuclease activity    NA
## GO:0000016 lactase activity                                               NA
## GO:0000026 alpha-1,2-mannosyltransferase activity                         NA
## GO:0000030 mannosyltransferase activity                                   NA
```