Class13RNA-Seq analysis mini-project

Qihao Liu (U08901197)

Table of contents

Background

Here we work through a complete RNASeq Analysis project. The input data comes from a knock-down experiement of a HOX gene

Data Import

Reading the `counts` and `metaData` CSV file

```
counts <- read.csv("GSE37704_featurecounts.csv",row.names = 1)
metadata <- read.csv("GSE37704_metadata.csv")
```

Check on data structure

```
head(metadata)
```

```
         id     condition
1 SRR493366 control_sirna
2 SRR493367 control_sirna
3 SRR493368 control_sirna
4 SRR493369      hoxa1_kd
5 SRR493370      hoxa1_kd
6 SRR493371      hoxa1_kd
```

```
head(counts)
```

```
                length SRR493366 SRR493367 SRR493368 SRR493369 SRR493370
ENSG00000186092    918         0         0         0         0         0
ENSG00000279928    718         0         0         0         0         0
ENSG00000279457   1982        23        28        29        29        28
ENSG00000278566    939         0         0         0         0         0
ENSG00000273547    939         0         0         0         0         0
ENSG00000187634   3214       124       123       205       207       212
                SRR493371
ENSG00000186092         0
ENSG00000279928         0
ENSG00000279457        46
ENSG00000278566         0
ENSG00000273547         0
ENSG00000187634       258
```

```
metadata$id == colnames(counts)
```

```
Warning in metadata$id == colnames(counts): longer object length is not a
multiple of shorter object length
```

```
[1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

We see that the metadata ID does not match the column names of count data. Some book-keeping is required to make sure the two matches

> Q1. Complete the code below to remove the troublesome first column from countData

Getting rid of the length column

```
cleancounts <- counts[,-1]

head(cleancounts)
```

	SRR493366	SRR493367	SRR493368	SRR493369	SRR493370	SRR493371
ENSG00000186092	0	0	0	0	0	0
ENSG00000279928	0	0	0	0	0	0
ENSG00000279457	23	28	29	29	28	46
ENSG00000278566	0	0	0	0	0	0
ENSG00000273547	0	0	0	0	0	0
ENSG00000187634	124	123	205	207	212	258

```
all(metadata$id == colnames(cleancounts))
```

```
[1] TRUE
```

Remove zero count genes

Notice that there are lots of genes with zero counts. We can remove these from future analysis

> Q2. Complete the code below to filter countData to exclude genes (i.e. rows) where we have 0 read count across all samples (i.e. columns).

```
to.keep.inds <- rowSums(cleancounts)>0
nonzero_counts <- cleancounts[to.keep.inds,]
nrow(nonzero_counts)
```

```
[1] 15975
```

DESeq analysis

Load the package

```
library(DESeq2)
```

```
Warning: package 'matrixStats' was built under R version 4.5.2
```

```
metadata
```

```
         id     condition
1 SRR493366 control_sirna
2 SRR493367 control_sirna
3 SRR493368 control_sirna
4 SRR493369      hoxa1_kd
5 SRR493370      hoxa1_kd
6 SRR493371      hoxa1_kd
```

Setup DESeq object

```
dds <- DESeqDataSetFromMatrix(countData=nonzero_counts,
                              colData=metadata,
                              design=~condition)
```

```
Warning in DESeqDataSet(se, design = design, ignoreRank): some variables in
design formula are characters, converting to factors
```

run DESeq

```
dds <- DESeq(dds)
```

```
estimating size factors
```

```
estimating dispersions
```

```
gene-wise dispersion estimates
```

```
mean-dispersion relationship
```

```
final dispersion estimates


fitting model and testing
```

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

```
log2 fold change (MLE): condition hoxa1 kd vs control sirna
Wald test p-value: condition hoxa1 kd vs control sirna
DataFrame with 15975 rows and 6 columns
                 baseMean log2FoldChange     lfcSE       stat      pvalue
                <numeric>      <numeric> <numeric>  <numeric>   <numeric>
ENSG00000279457   29.9136      0.1792571 0.3248216   0.551863 5.81042e-01
ENSG00000187634  183.2296      0.4264571 0.1402658   3.040350 2.36304e-03
ENSG00000188976 1651.1881     -0.6927205 0.0548465 -12.630158 1.43990e-36
ENSG00000187961  209.6379      0.7297556 0.1318599   5.534326 3.12428e-08
ENSG00000187583   47.2551      0.0405765 0.2718928   0.149237 8.81366e-01
...                   ...            ...       ...        ...          ...
ENSG00000273748  35.30265       0.674387  0.303666   2.220817 2.63633e-02
ENSG00000278817   2.42302      -0.388988  1.130394  -0.344117 7.30758e-01
ENSG00000278384   1.10180       0.332991  1.660261   0.200565 8.41039e-01
ENSG00000276345  73.64496      -0.356181  0.207716  -1.714752 8.63908e-02
ENSG00000271254 181.59590      -0.609667  0.141320  -4.314071 1.60276e-05
                      padj
                 <numeric>
ENSG00000279457 6.86555e-01
ENSG00000187634 5.15718e-03
ENSG00000188976 1.76549e-35
ENSG00000187961 1.13413e-07
ENSG00000187583 9.19031e-01
...                     ...
ENSG00000273748 4.79091e-02
ENSG00000278817 8.09772e-01
ENSG00000278384 8.92654e-01
ENSG00000276345 1.39762e-01
ENSG00000271254 4.53648e-05
```

Q3. Call the summary() function on your results to get a sense of how many genes are up or down-regulated at the default 0.1 p-value cutoff.

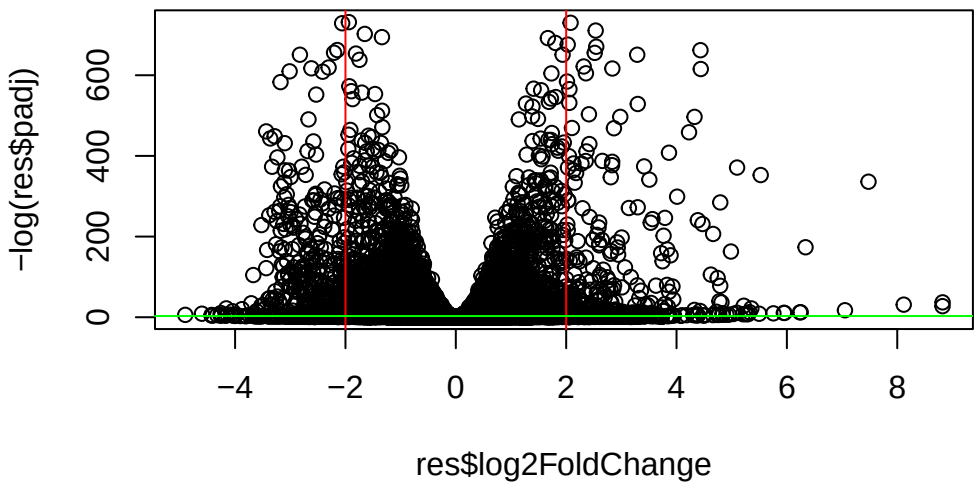
```
summary(res)
```

```
out of 15975 with nonzero total read count
adjusted p-value < 0.1
LFC > 0 (up)       : 4349, 27%
LFC < 0 (down)     : 4396, 28%
outliers [1]       : 0, 0%
low counts [2]     : 1237, 7.7%
(mean count < 0)
[1] see 'cooksCutoff' argument of ?results
[2] see 'independentFiltering' argument of ?results
```

Data Visualization with Volcano Plot

Volcano Plot

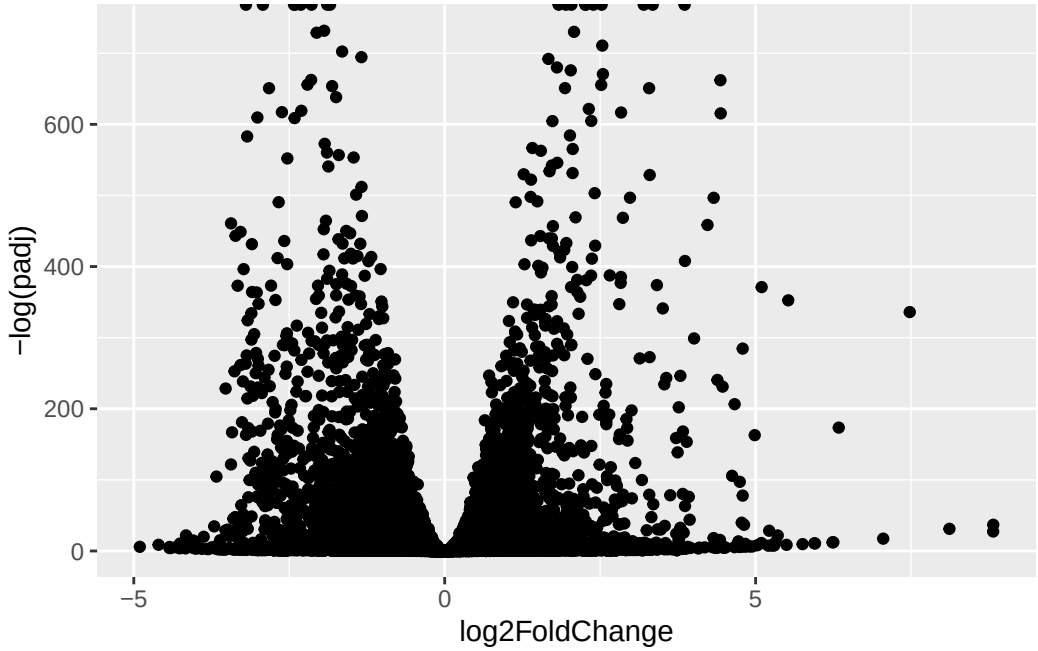
```
plot(res$log2FoldChange,-log(res$padj))
abline(v=c(-2,2), col="red")
abline(h=-log(0.05), col="green")
```



Using ggplot

```
library(ggplot2)

ggplot(res)+
  aes(log2FoldChange,-log(padj))+
  geom_point()
```

```
Warning: Removed 1237 rows containing missing values or values outside the scale range
(`geom_point()`).
```

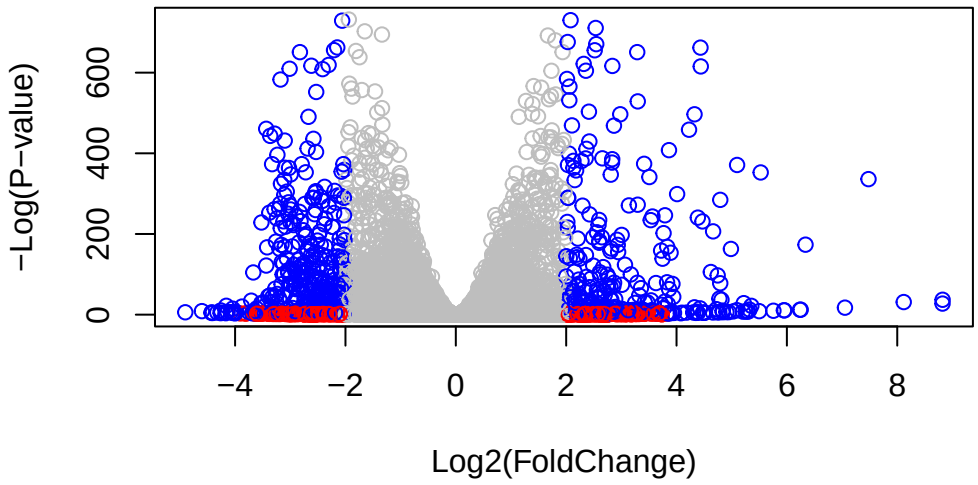


Add threshold lines for ofld change and p-value and color our subset genes that make these threshold cut-off in the plots

> Q4. Improve this plot by completing the below code, which adds color and axis labels

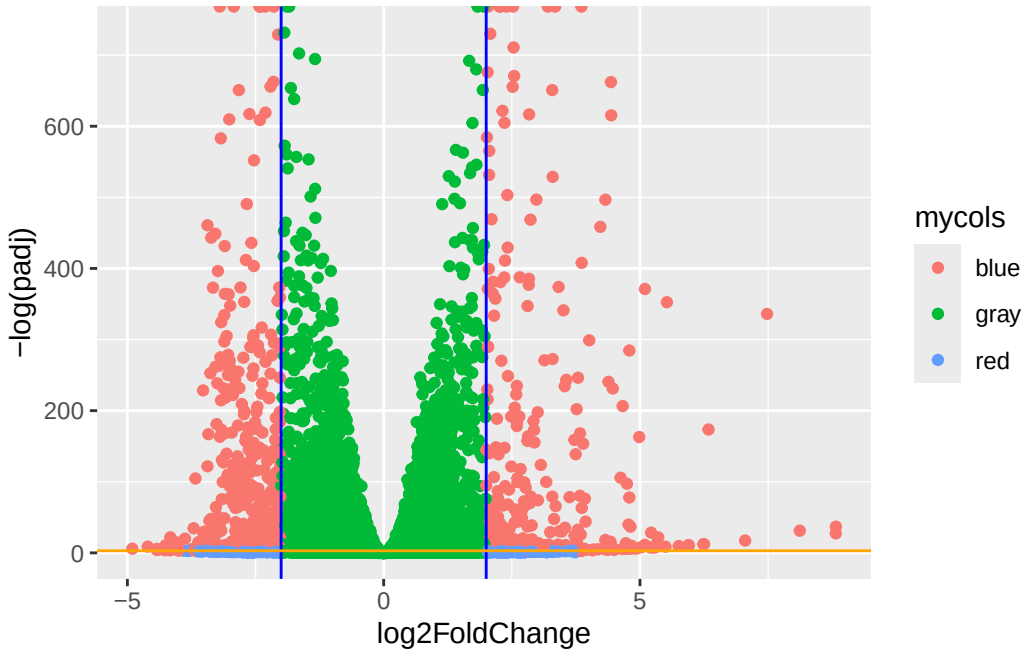
```
# Make a color vector for all genes
mycols <- rep("gray", nrow(res) )

# Color red the genes with absolute fold change above 2
mycols[ abs(res$log2FoldChange) > 2 ] <- "red"
```

```
# Color blue those with adjusted p-value less than 0.01
#  and absolute fold change more than 2
inds <- (res$padj < 0.05 ) & (abs(res$log2FoldChange) > 2 )
mycols[ inds ] <- "blue"

plot( res$log2FoldChange, -log(res$padj), col=mycols, xlab="Log2(FoldChange)", ylab="-Log(P-
```




```
mycols <- rep("gray", nrow(res))
mycols[ abs(res$log2FoldChange) > 2 ]  <- "red"

inds <- (res$padj < 0.05) & (abs(res$log2FoldChange) > 2 )
mycols[ inds ] <- "blue"

ggplot(res)+
  aes(log2FoldChange,-log(padj))+
  geom_point(aes(color = mycols))+
  geom_vline(xintercept = c(-2, 2),col = "blue")+
  geom_hline(yintercept = -log(0.05), col = "orange")
```

```
Warning: Removed 1237 rows containing missing values or values outside the scale range
(`geom_point()`).
```



Add anotation

Add gene symbols and IDs

> Q5. Use the mapIDs() function multiple times to add SYMBOL, ENTREZID and GENENAME annotation to our results by completing the code below.

```
library(AnnotationDbi)
library(org.Hs.eg.db)
```

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

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

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

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

```
res$genename <- mapIds(x=org.Hs.eg.db,
       keys = row.names(res),
       keytype = "ENSEMBL",
       column = "GENENAME"
       )
```

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

```
head(res,5)
```

```
log2 fold change (MLE): condition hoxa1 kd vs control sirna
Wald test p-value: condition hoxa1 kd vs control sirna
DataFrame with 5 rows and 9 columns
                 baseMean log2FoldChange     lfcSE       stat      pvalue
                <numeric>      <numeric> <numeric>  <numeric>   <numeric>
ENSG00000279457   29.9136      0.1792571 0.3248216   0.551863 5.81042e-01
ENSG00000187634  183.2296      0.4264571 0.1402658   3.040350 2.36304e-03
ENSG00000188976 1651.1881     -0.6927205 0.0548465 -12.630158 1.43990e-36
ENSG00000187961  209.6379      0.7297556 0.1318599   5.534326 3.12428e-08
ENSG00000187583   47.2551      0.0405765 0.2718928   0.149237 8.81366e-01
                      padj      symbol      entrez              genename
                 <numeric> <character> <character>           <character>
ENSG00000279457 6.86555e-01          NA          NA                    NA
ENSG00000187634 5.15718e-03      SAMD11      148398 sterile alpha motif ..
ENSG00000188976 1.76549e-35       NOC2L       26155 NOC2 like nucleolar ..
ENSG00000187961 1.13413e-07      KLHL17      339451 kelch like family me..
ENSG00000187583 9.19031e-01     PLEKHN1       84069 pleckstrin homology ..
```

Q6. Finally for this section let's reorder these results by adjusted p-value and save them to a CSV file in your current project directory.

```
res = res[order(res$pvalue),]
write.csv(res,file = "myresults.csv")
```

Pathway Analysis

Pathway analysis with KEGG

Perform Pathway Analysis

```
library(gage)
```

```
library(gageData)
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).
################################################################################
```

We need a named vector of fold change values for DEGs

```
foldchanges <- res$log2FoldChange
names(foldchanges) <- res$entrez
head(foldchanges)
```

```
     1266     54855      1465      2034      2150      6659
-2.422719  3.201955 -2.313738 -1.888019  3.344508  2.392288
```

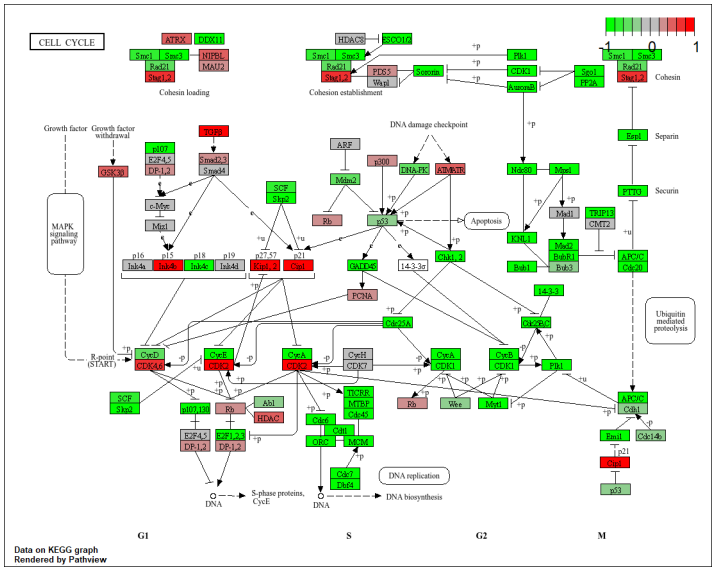
```
data(kegg.sets.hs)

keggres = gage(foldchanges, gsets=kegg.sets.hs)

#attributes(keggres)
head(keggres$less,5)
```

```
                                           p.geomean stat.mean
hsa04110 Cell cycle                        8.995727e-06 -4.378644
hsa03030 DNA replication                   9.424076e-05 -3.951803
hsa05130 Pathogenic Escherichia coli infection 1.405864e-04 -3.765330
hsa03013 RNA transport                     1.375901e-03 -3.028500
hsa03440 Homologous recombination          3.066756e-03 -2.852899
                                                  p.val       q.val
hsa04110 Cell cycle                        8.995727e-06 0.001889103
hsa03030 DNA replication                   9.424076e-05 0.009841047
hsa05130 Pathogenic Escherichia coli infection 1.405864e-04 0.009841047
hsa03013 RNA transport                     1.375901e-03 0.072234819
hsa03440 Homologous recombination          3.066756e-03 0.128803765
                                           set.size         exp1
hsa04110 Cell cycle                             121 8.995727e-06
hsa03030 DNA replication                         36 9.424076e-05
hsa05130 Pathogenic Escherichia coli infection   53 1.405864e-04
hsa03013 RNA transport                          144 1.375901e-03
hsa03440 Homologous recombination                28 3.066756e-03
```

To get the pathways and include it in our lab report

```
pathview(pathway.id = "hsa04110", gene.data = foldchanges)
```

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

Info: Working in directory C:/Users/Thoma/Desktop/BIMM_143_Assignment/Class13

Info: Writing image file hsa04110.pathview.png
```



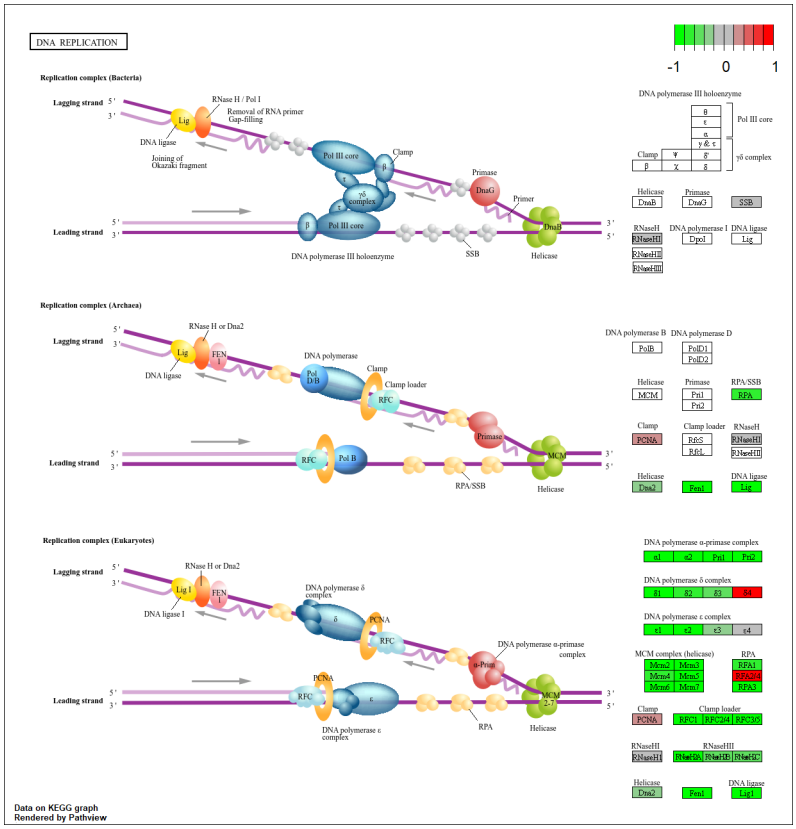
To get a different pathways and include it in our lab report

```
pathview(pathway.id = "hsa03030", gene.data = foldchanges)
```

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

```
Info: Working in directory C:/Users/Thoma/Desktop/BIMM_143_Assignment/Class13
```

```
Info: Writing image file hsa03030.pathview.png
```



Q7. Can you do the same procedure as above to plot the pathview figures for the top 5 down-reguled pathways?

```
## Focus on top 5 upregulated pathways here for demo purposes only
keggrespathways <- rownames(keggres$less)[1:5]

# Extract the 8 character long IDs part of each string
keggresids = substr(keggrespathways, start=1, stop=8)
keggresids
```

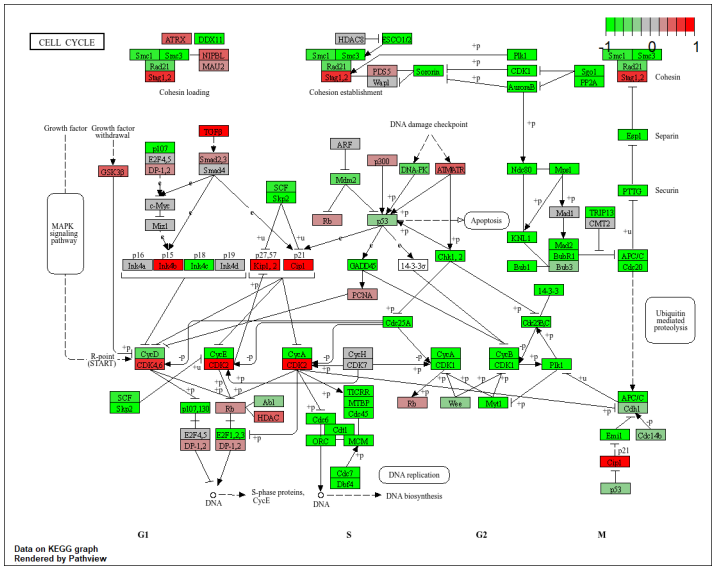
```
[1] "hsa04110" "hsa03030" "hsa05130" "hsa03013" "hsa03440"
```

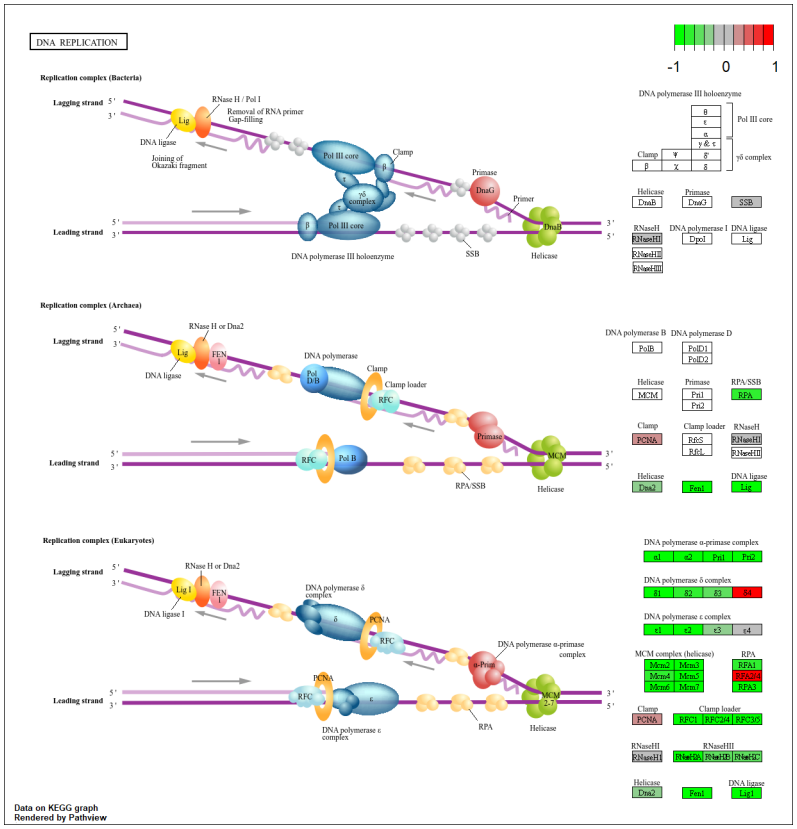
```
pathview(gene.data=foldchanges, pathway.id=keggresids, species="hsa")
```

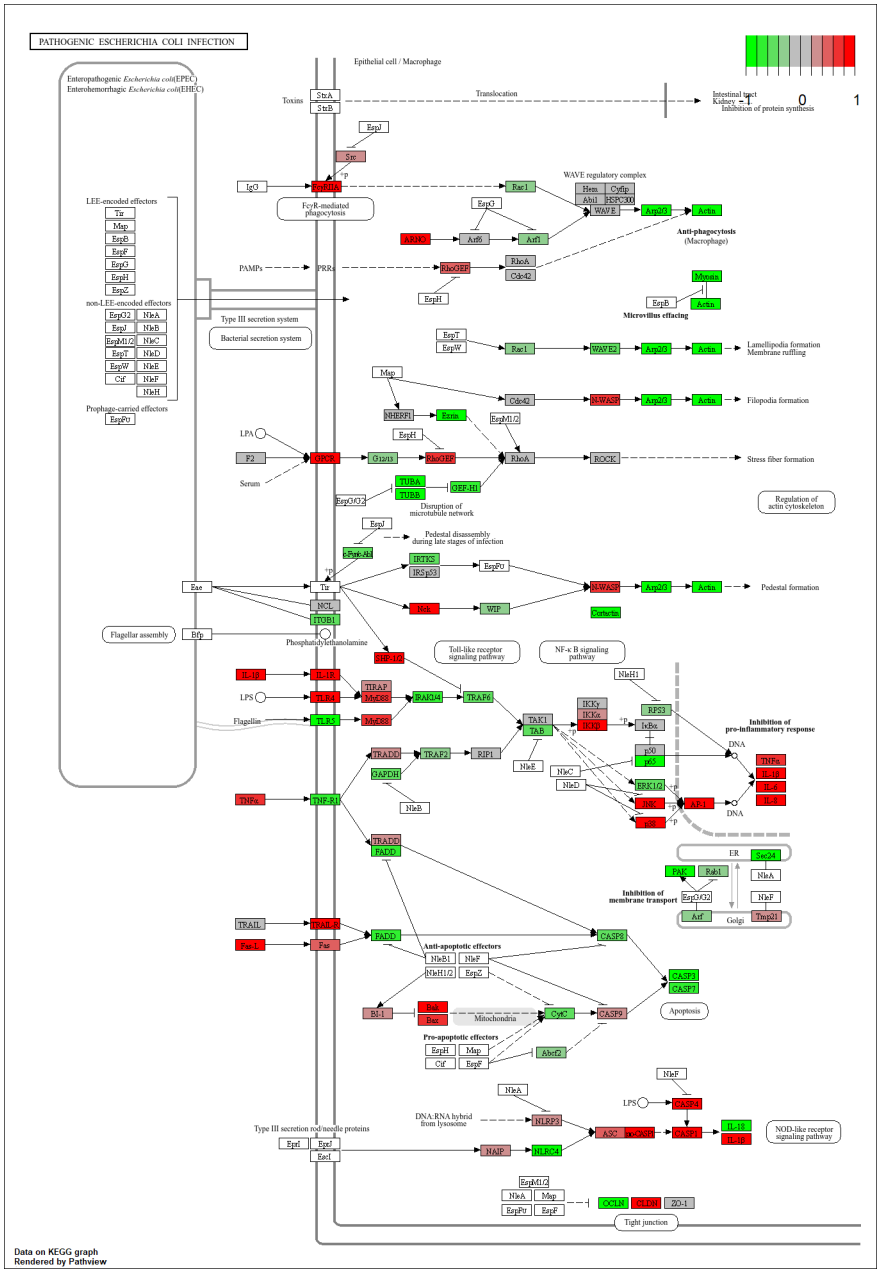
```
'select()' returned 1:1 mapping between keys and columns
```

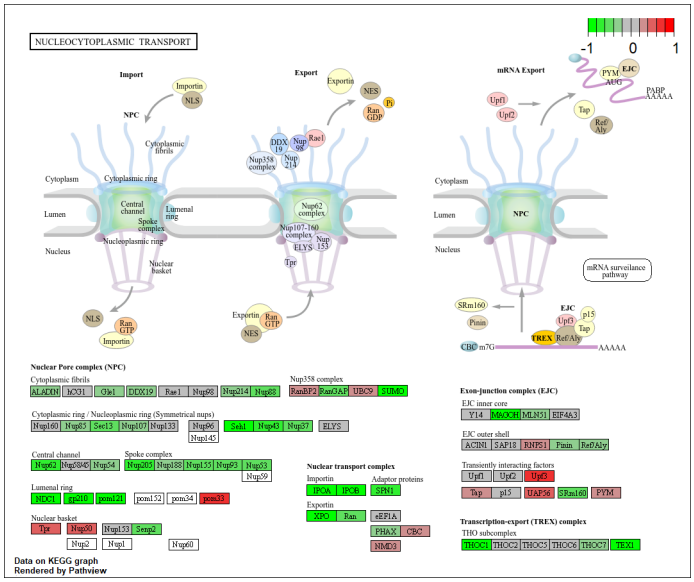
```
Info: Working in directory C:/Users/Thoma/Desktop/BIMM_143_Assignment/Class13
```

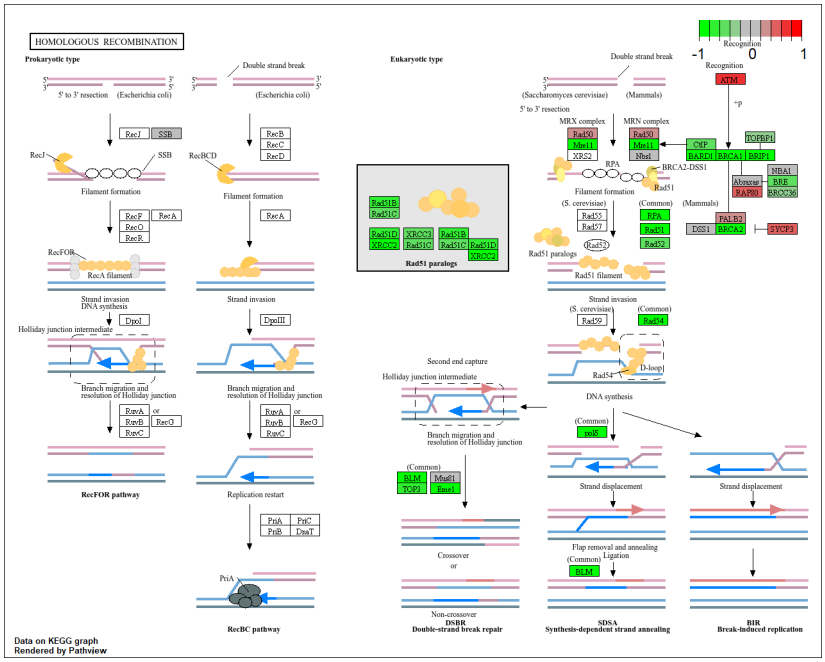
```
Info: Writing image file hsa04110.pathview.png

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

Info: Working in directory C:/Users/Thoma/Desktop/BIMM_143_Assignment/Class13

Info: Writing image file hsa03030.pathview.png

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

Info: Working in directory C:/Users/Thoma/Desktop/BIMM_143_Assignment/Class13

Info: Writing image file hsa05130.pathview.png

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

Info: Working in directory C:/Users/Thoma/Desktop/BIMM_143_Assignment/Class13

Info: Writing image file hsa03013.pathview.png

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

Info: Working in directory C:/Users/Thoma/Desktop/BIMM_143_Assignment/Class13

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




DNA REPLICATION
-1
0
1
Replication complex (Bacteria)
Lagging strand
Leading strand
RNase H / Pol I
Removal of RNA primer
Gap-filling
Lig
DNA ligase
Joining of
Okazaki fragment
Pol III core
Clamp
Primase
DnaG
Primer
DnaB
SSB
DNA polymerase III holoenzyme
DNA polymerase III holoenzyme
Pol III core
γδ complex
Clamp
Helicase
Primase
SSB
DnaB
DnaG
RNaseH
DNA polymerase I
DNA ligase
DpoI
Lig
Replication complex (Archaea)
RNase H or Dna2
FEN 1
DNA polymerase
Pol D/B
Clamp
Clamp loader
RFC
Primase
MCM
RPA/SSB
Helicase
DNA polymerase B
DNA polymerase D
PolB
PolD1
PolD2
Helicase
Primase
RPA/SSB
MCM
Pri1
Pri2
RPA
Clamp
Clamp loader
RNaseH
PCNA
RfcS
RfcL
Helicase
DNA ligase
Dna2
Pen1
Lig
Replication complex (Eukaryotes)
RNase H or Dna2
Lig I
DNA ligase I
DNA polymerase δ complex
PCNA
RFC
DNA polymerase α-primase complex
α-Prim
Pol ε
DNA polymerase ε complex
RPA
Helicase
DNA polymerase α-primase complex
DNA polymerase δ complex
DNA polymerase ε complex
MCM complex (helicase)
RPA
Clamp
Clamp loader
RNaseHI
RNaseHII
Helicase
DNA ligase
Dna2
Pen1
Lig1
Data on KEGG graph
Rendered by Pathview


PATHOGENIC ESCHERICHIA COLI INFECTION
Epithelial cell / Macrophage
Enteropathogenic Escherichia coli(EPEC)
Enterohemorrhagic Escherichia coli(EHEC)
Toxins
StcA
StcB
Translocation
Intestinal tract
Kidney
Inhibition of protein synthesis
0
1
EspJ
Src
IgG
FcgRIIA
FcγR-mediated phagocytosis
WAVE regulatory complex
Hem
Cyfip
Abi1
HSPC300
WAVE
Rac1
EspG
Arp2/3
Actin
ABI3BP
Arf6
Arf1
Anti-phagocytosis
(Macrophage)
PAMPs
PRRs
RhoGEF
RhoA
Cdc42
EspH
Myosin
EspB
Actin
Microvillus effacing
LEE-encoded effectors
Tir
Map
EspB
EspF
EspG
EspH
EspZ
non-LEE-encoded effectors
EspZ2
NleA
EspJ
NleB
EspM1/2
NleC
EspT
NleD
EspW
NleE
Cif
NleF
NleH
Prophage-carried effectors
EspFU
Type III secretion system
Bacterial secretion system
EspT
EspW
Rac1
WAVE2
Arp2/3
Actin
Lamellipodia formation
Membrane ruffling
Map
Cdc42
N-WASP
Arp2/3
Actin
Filopodia formation
NHERF1
Ezrin
EspM1/2
LPA
P2
GPCR
G12/13
RhoGEF
RhoA
ROCK
Stress fiber formation
Serum
EspH
TUBA
TUBB
OSF-H1
EspG/G2
Disruption of microtubule network
Regulation of actin cytoskeleton
EspJ
Pedestal disassembly during late stages of infection
c-Fgr-Kz
IRTKS
IRSp53
EspFU
Exo
Tir
N-WASP
Arp2/3
Actin
Pedestal formation
NCL
Nck
WIP
Cortactin
TTOB1
Flagellar assembly
Bfp
Phosphatidylethanolamine
SHP-1C
Toll-like receptor signaling pathway
NF-κ B signaling pathway
IL-1R
IL-1R
NleH1
LPS
TLR4
TRAP
MyD88
IRAK1/4
TRAF6
Flagellin
TLR5
MyD88
TAK1
TAB
IKKγ
IKKα
IKKβ
RIP3
IκBα
p50
p65
Inhibition of pro-inflammatory response
DNA
TNFα
IL-1β
IL-6
IL-8
TRADD
TRAF2
RIP1
GAPDH
NleE
NleC
NleD
ERK1/2
JNK
p38
AP-1
DNA
TNFα
TNF-R1
NleB
TRADD
FADD
ER
Sec24
FAK
Rab1
NleA
Inhibition of membrane transport
EspG/G2
NleF
Arf
Golgi
Tmp21
TRAIL
TRAIL-R
FasL
Fas
FADD
CASP8
Anti-apoptotic effectors
NleB1
NleF
NleH1/2
EspZ
CASP3
CASP7
Bcl-1
Bak
Bax
Mitochondria
Cyt C
CASP9
Apoptosis
Pro-apoptotic effectors
EspH
Map
Cif
EspF
Abcf2
NleF
NleA
LPS
CASP4
DNA/RNA hybrid from lysosome
NLRP3
ASC
pro-CASP1
CASP1
IL-1β
IL-18
NOD-like receptor signaling pathway
Type III secretion rod/needle proteins
EspI
EspJ
EscI
NAIP
NLRC4
EspM1/2
NleA
Map
EspFU
EspF
CYCLN
CLDN
ZO-1
Tight junction
Data on KEGG graph
Rendered by Pathview





Pathway analysis with GO (Geneontology)

Same analysis but using GO genesets rather than KEGG

```
data(go.sets.hs)
data(go.subs.hs)
```

```
# Focus on Biological Process subset of GO
gobpsets = go.sets.hs[go.subs.hs$BP]

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

head(gobpres$greater,5)
```

```
                                           p.geomean stat.mean        p.val
GO:0007156 homophilic cell adhesion       8.519724e-05  3.824205 8.519724e-05
GO:0002009 morphogenesis of an epithelium 1.396681e-04  3.653886 1.396681e-04
GO:0048729 tissue morphogenesis           1.432451e-04  3.643242 1.432451e-04
GO:0007610 behavior                       1.925222e-04  3.565432 1.925222e-04
GO:0060562 epithelial tube morphogenesis  5.932837e-04  3.261376 5.932837e-04
                                               q.val set.size         exp1
GO:0007156 homophilic cell adhesion       0.1951953      113 8.519724e-05
GO:0002009 morphogenesis of an epithelium 0.1951953      339 1.396681e-04
GO:0048729 tissue morphogenesis           0.1951953      424 1.432451e-04
GO:0007610 behavior                       0.1967577      426 1.925222e-04
GO:0060562 epithelial tube morphogenesis  0.3565320      257 5.932837e-04
```

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

```
$greater
                                           p.geomean stat.mean        p.val
GO:0007156 homophilic cell adhesion       8.519724e-05  3.824205 8.519724e-05
GO:0002009 morphogenesis of an epithelium 1.396681e-04  3.653886 1.396681e-04
GO:0048729 tissue morphogenesis           1.432451e-04  3.643242 1.432451e-04
GO:0007610 behavior                       1.925222e-04  3.565432 1.925222e-04
GO:0060562 epithelial tube morphogenesis  5.932837e-04  3.261376 5.932837e-04
GO:0035295 tube development               5.953254e-04  3.253665 5.953254e-04
                                               q.val set.size         exp1
GO:0007156 homophilic cell adhesion       0.1951953      113 8.519724e-05
GO:0002009 morphogenesis of an epithelium 0.1951953      339 1.396681e-04
GO:0048729 tissue morphogenesis           0.1951953      424 1.432451e-04
GO:0007610 behavior                       0.1967577      426 1.925222e-04
GO:0060562 epithelial tube morphogenesis  0.3565320      257 5.932837e-04
GO:0035295 tube development               0.3565320      391 5.953254e-04

$less
                                             p.geomean stat.mean        p.val
```

```
GO:0048285 organelle fission             1.536227e-15 -8.063910 1.536227e-15
GO:0000280 nuclear division              4.286961e-15 -7.939217 4.286961e-15
GO:0007067 mitosis                       4.286961e-15 -7.939217 4.286961e-15
GO:0000087 M phase of mitotic cell cycle 1.169934e-14 -7.797496 1.169934e-14
GO:0007059 chromosome segregation        2.028624e-11 -6.878340 2.028624e-11
GO:0000236 mitotic prometaphase          1.729553e-10 -6.695966 1.729553e-10
                                                q.val set.size         exp1
GO:0048285 organelle fission             5.841698e-12      376 1.536227e-15
GO:0000280 nuclear division              5.841698e-12      352 4.286961e-15
GO:0007067 mitosis                       5.841698e-12      352 4.286961e-15
GO:0000087 M phase of mitotic cell cycle 1.195672e-11      362 1.169934e-14
GO:0007059 chromosome segregation        1.658603e-08      142 2.028624e-11
GO:0000236 mitotic prometaphase          1.178402e-07       84 1.729553e-10

$stats
                                          stat.mean     exp1
GO:0007156 homophilic cell adhesion        3.824205 3.824205
GO:0002009 morphogenesis of an epithelium  3.653886 3.653886
GO:0048729 tissue morphogenesis            3.643242 3.643242
GO:0007610 behavior                        3.565432 3.565432
GO:0060562 epithelial tube morphogenesis   3.261376 3.261376
GO:0035295 tube development                3.253665 3.253665
```

Reactome Analysis

reactome analysis can be performed on a website or run as a R package. We are going to look at the website first

The input website asks for is a gene list (text file with one gene symbol perline), so we are going to generate it here. We want to include only the genes that are significant

```
sig_genes <- res[res$padj <= 0.05 & !is.na(res$padj), "symbol"]
write.table(sig_genes, file="significant_genes.txt", row.names=FALSE, col.names=FALSE, quote=
```

> Q8. What pathway has the most significant "Entities p-value"? Do the most significant pathways listed match your previous KEGG results? What factors could cause differences between the two methods?
>
> A: the pathway with the most significant entites p-value is cell cycle. This is consistent with my previous KEGG results. This is because both KEGG and reactome analysis looks at the pathways

GO Online

Q9. What pathway has the most significant "Entities p-value"? Do the most significant pathways listed match your previous KEGG results? What factors could cause differences between the two methods?

A: the pathway with the most significant entities p-value is metabolic process. This does not exactly match my previous KEGG results. This is because while the KEGG analysis looks at the pathways, GO term enrichment looks at broader and more generic biological functions.

Saving Our Results

```
write.csv(res,file = "myresults.csv")
```