Pathway Analysis from RNA-Seq Results

Kenny Dang (PID: A18544481)

Table of contents

Section 1: Differential Expression Analysis 1
Section 2. Pathway Analysis 11
Section 3 Gene Ontology (GO) 23
Section 4. Reactome Analysis 24

Section 1: Differential Expression Analysis

```
library(DESeq2)
```

```
Loading required package: S4Vectors
```

```
Loading required package: stats4
```

```
Loading required package: BiocGenerics
```

```

Attaching package: 'BiocGenerics'
```

```
The following objects are masked from 'package:stats':

    IQR, mad, sd, var, xtabs
```

```
The following objects are masked from 'package:base':

    anyDuplicated, aperm, append, as.data.frame, basename, cbind,
    colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,
    get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,
    match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
    Position, rank, rbind, Reduce, rownames, sapply, saveRDS, setdiff,
    table, tapply, union, unique, unsplit, which.max, which.min


Attaching package: 'S4Vectors'

The following object is masked from 'package:utils':

    findMatches

The following objects are masked from 'package:base':

    expand.grid, I, unname

Loading required package: IRanges

Warning: package 'IRanges' was built under R version 4.4.2


Attaching package: 'IRanges'

The following object is masked from 'package:grDevices':

    windows

Loading required package: GenomicRanges

Loading required package: GenomeInfoDb

Warning: package 'GenomeInfoDb' was built under R version 4.4.2

Loading required package: SummarizedExperiment
```

```
Loading required package: MatrixGenerics

Warning: package 'MatrixGenerics' was built under R version 4.4.2

Loading required package: matrixStats

Warning: package 'matrixStats' was built under R version 4.4.3


Attaching package: 'MatrixGenerics'

The following objects are masked from 'package:matrixStats':

    colAlls, colAnyNAs, colAnys, colAvgsPerRowSet, colCollapse,
    colCounts, colCummaxs, colCummins, colCumprods, colCumsums,
    colDiffs, colIQRDiffs, colIQRs, colLogSumExps, colMadDiffs,
    colMads, colMaxs, colMeans2, colMedians, colMins, colOrderStats,
    colProds, colQuantiles, colRanges, colRanks, colSdDiffs, colSds,
    colSums2, colTabulates, colVarDiffs, colVars, colWeightedMads,
    colWeightedMeans, colWeightedMedians, colWeightedSds,
    colWeightedVars, rowAlls, rowAnyNAs, rowAnys, rowAvgsPerColSet,
    rowCollapse, rowCounts, rowCummaxs, rowCummins, rowCumprods,
    rowCumsums, rowDiffs, rowIQRDiffs, rowIQRs, rowLogSumExps,
    rowMadDiffs, rowMads, rowMaxs, rowMeans2, rowMedians, rowMins,
    rowOrderStats, rowProds, rowQuantiles, rowRanges, rowRanks,
    rowSdDiffs, rowSds, rowSums2, rowTabulates, rowVarDiffs, rowVars,
    rowWeightedMads, rowWeightedMeans, rowWeightedMedians,
    rowWeightedSds, rowWeightedVars

Loading required package: Biobase

Welcome to Bioconductor

    Vignettes contain introductory material; view with
    'browseVignettes()'. To cite Bioconductor, see
    'citation("Biobase")', and for packages 'citation("pkgname")'.


Attaching package: 'Biobase'
```

```
The following object is masked from 'package:MatrixGenerics':

    rowMedians

The following objects are masked from 'package:matrixStats':

    anyMissing, rowMedians
```

```
metaFile <- "GSE37704_metadata.csv"
countFile <- "GSE37704_featurecounts.csv"

colData = read.csv(metaFile, row.names=1)
head(colData)
```

```
             condition
SRR493366 control_sirna
SRR493367 control_sirna
SRR493368 control_sirna
SRR493369      hoxa1_kd
SRR493370      hoxa1_kd
SRR493371      hoxa1_kd
```

```
countData = read.csv(countFile, row.names=1)
head(countData)
```

```
                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
```

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

```
countData <- as.matrix(countData[,-1])
head(countData)
```

	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

Q. 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).

Tip: What will rowSums() of countData return and how could you use it in this context?

```
countData = countData[rowSums(countData) !=0,]
head(countData)
```

	SRR493366	SRR493367	SRR493368	SRR493369	SRR493370	SRR493371
ENSG00000279457	23	28	29	29	28	46
ENSG00000187634	124	123	205	207	212	258
ENSG00000188976	1637	1831	2383	1226	1326	1504
ENSG00000187961	120	153	180	236	255	357
ENSG00000187583	24	48	65	44	48	64
ENSG00000187642	4	9	16	14	16	16

```
dds = DESeqDataSetFromMatrix(countData = countData ,
                             colData= colData  ,
                             design =~ condition)
```

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

```
dds1 = DESeq(dds)
```

```
estimating size factors
```

```
estimating dispersions
```

```
gene-wise dispersion estimates
```

```
mean-dispersion relationship
```

```
final dispersion estimates
```

```
fitting model and testing
```

```
dds1
```

```
class: DESeqDataSet
dim: 15975 6
metadata(1): version
assays(4): counts mu H cooks
rownames(15975): ENSG00000279457 ENSG00000187634 ... ENSG00000276345
  ENSG00000271254
rowData names(22): baseMean baseVar ... deviance maxCooks
colnames(6): SRR493366 SRR493367 ... SRR493370 SRR493371
colData names(2): condition sizeFactor
```

```
res = results (dds1)
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
```

Q. 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)
```

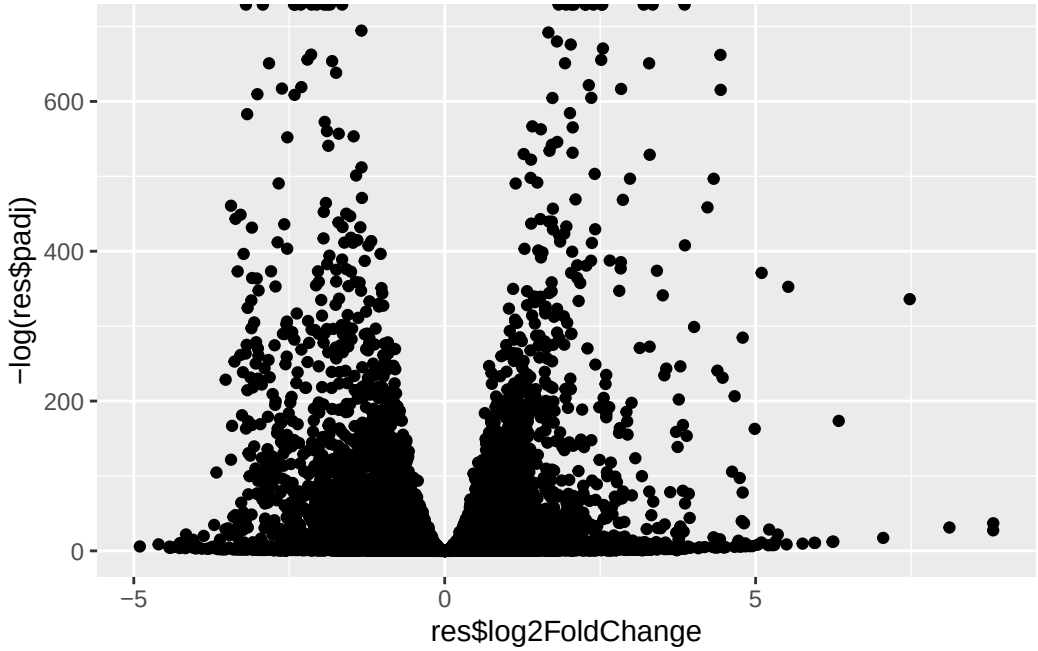
```
out of 15975 with nonzero total read count
adjusted p-value < 0.1
LFC > 0 (up)       : 4349, 27%
LFC < 0 (down)     : 4396, 28%
outliers [1]       : 0, 0%
low counts [2]     : 1237, 7.7%
(mean count < 0)
[1] see 'cooksCutoff' argument of ?results
[2] see 'independentFiltering' argument of ?results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

```
library (ggplot2)
```

```
Warning: package 'ggplot2' was built under R version 4.4.3
```

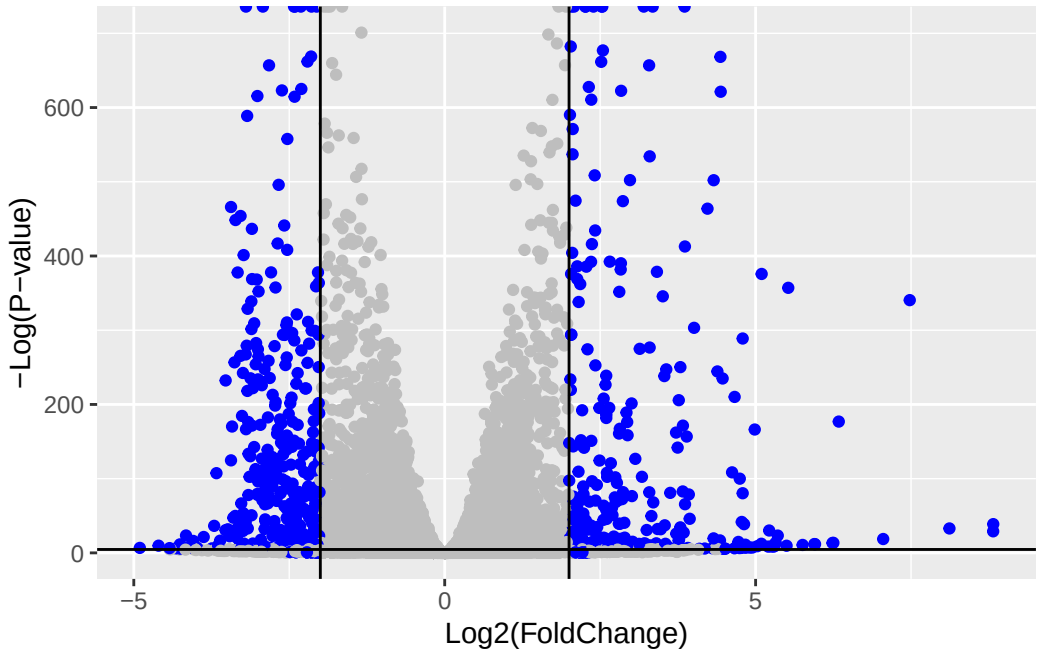
```
ggplot(res) +
  aes(res$log2FoldChange,
  y= -log(res$padj)) +
  geom_point()
```

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



Q. Improve this plot by completing the below code, which adds color, axis labels and cutoff lines:

```
mycols <- rep("gray", nrow(res))

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

# Color gray those with adjusted p-value more than 0.01
mycols[ res$padj > 0.01 ] <- "gray"

ggplot(res) +
  aes(x = log2FoldChange,
      y= -log(pvalue)) +
  geom_point(col = mycols) +
  xlab("Log2(FoldChange)") +
  ylab("-Log(P-value)") +
  geom_vline(xintercept = c(-2,2)) +
  geom_hline(yintercept = -log(0.01))
```



Q. 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")
```

```
columns(org.Hs.eg.db)
```

```
 [1] "ACCNUM"       "ALIAS"        "ENSEMBL"      "ENSEMBLPROT"  "ENSEMBLTRANS"
 [6] "ENTREZID"     "ENZYME"       "EVIDENCE"     "EVIDENCEALL"  "GENENAME"
[11] "GENETYPE"     "GO"           "GOALL"        "IPI"          "MAP"
[16] "OMIM"         "ONTOLOGY"     "ONTOLOGYALL"  "PATH"         "PFAM"
[21] "PMID"         "PROSITE"      "REFSEQ"       "SYMBOL"       "UCSCKG"
[26] "UNIPROT"
```

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

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

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

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

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

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

```
head(res, 10)
```

```
log2 fold change (MLE): condition hoxa1 kd vs control sirna
Wald test p-value: condition hoxa1 kd vs control sirna
DataFrame with 10 rows and 9 columns
                   baseMean log2FoldChange     lfcSE       stat      pvalue
                  <numeric>      <numeric> <numeric>  <numeric>   <numeric>
ENSG00000279457   29.913579      0.1792571 0.3248216   0.551863 5.81042e-01
ENSG00000187634  183.229650      0.4264571 0.1402658   3.040350 2.36304e-03
ENSG00000188976 1651.188076     -0.6927205 0.0548465 -12.630158 1.43990e-36
ENSG00000187961  209.637938      0.7297556 0.1318599   5.534326 3.12428e-08
ENSG00000187583   47.255123      0.0405765 0.2718928   0.149237 8.81366e-01
ENSG00000187642   11.979750      0.5428105 0.5215598   1.040744 2.97994e-01
ENSG00000188290  108.922128      2.0570638 0.1969053  10.446970 1.51282e-25
ENSG00000187608  350.716868      0.2573837 0.1027266   2.505522 1.22271e-02
ENSG00000188157 9128.439422      0.3899088 0.0467163   8.346304 7.04321e-17
ENSG00000237330    0.158192      0.7859552 4.0804729   0.192614 8.47261e-01
                       padj      symbol      entrez                   name
                  <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 ..
ENSG00000187642 4.03379e-01       PERM1       84808 PPARGC1 and ESRR ind..
ENSG00000188290 1.30538e-24        HES4       57801 hes family bHLH tran..
ENSG00000187608 2.37452e-02       ISG15        9636 ISG15 ubiquitin like..
ENSG00000188157 4.21963e-16        AGRN      375790                  agrin
ENSG00000237330          NA      RNF223      401934 ring finger protein ..
```

Q. 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$padj),]
write.csv(res, file  ="deseq_results.csv")
```

Section 2. Pathway Analysis

```
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(gage)
```

```
library(gageData)

data(kegg.sets.hs)
data(sigmet.idx.hs)

# Focus on signaling and metabolic pathways only
```

```
kegg.sets.hs = kegg.sets.hs[sigmet.idx.hs]

# Examine the first 3 pathways
head(kegg.sets.hs, 3)
```

```
$`hsa00232 Caffeine metabolism`
[1] "10"   "1544" "1548" "1549" "1553" "7498" "9"

$`hsa00983 Drug metabolism - other enzymes`
 [1] "10"     "1066"   "10720"  "10941"  "151531" "1548"   "1549"   "1551"
 [9] "1553"   "1576"   "1577"   "1806"   "1807"   "1890"   "221223" "2990"
[17] "3251"   "3614"   "3615"   "3704"   "51733"  "54490"  "54575"  "54576"
[25] "54577"  "54578"  "54579"  "54600"  "54657"  "54658"  "54659"  "54963"
[33] "574537" "64816"  "7083"   "7084"   "7172"   "7363"   "7364"   "7365"
[41] "7366"   "7367"   "7371"   "7372"   "7378"   "7498"   "79799"  "83549"
[49] "8824"   "8833"   "9"      "978"

$`hsa00230 Purine metabolism`
  [1] "100"    "10201"  "10606"  "10621"  "10622"  "10623"  "107"    "10714"
  [9] "108"    "10846"  "109"    "111"    "11128"  "11164"  "112"    "113"
 [17] "114"    "115"    "122481" "122622" "124583" "132"    "158"    "159"
 [25] "1633"   "171568" "1716"   "196883" "203"    "204"    "205"    "221823"
 [33] "2272"   "22978"  "23649"  "246721" "25885"  "2618"   "26289"  "270"
 [41] "271"    "27115"  "272"    "2766"   "2977"   "2982"   "2983"   "2984"
 [49] "2986"   "2987"   "29922"  "3000"   "30833"  "30834"  "318"    "3251"
 [57] "353"    "3614"   "3615"   "3704"   "377841" "471"    "4830"   "4831"
 [65] "4832"   "4833"   "4860"   "4881"   "4882"   "4907"   "50484"  "50940"
 [73] "51082"  "51251"  "51292"  "5136"   "5137"   "5138"   "5139"   "5140"
 [81] "5141"   "5142"   "5143"   "5144"   "5145"   "5146"   "5147"   "5148"
 [89] "5149"   "5150"   "5151"   "5152"   "5153"   "5158"   "5167"   "5169"
 [97] "51728"  "5198"   "5236"   "5313"   "5315"   "53343"  "54107"  "5422"
[105] "5424"   "5425"   "5426"   "5427"   "5430"   "5431"   "5432"   "5433"
[113] "5434"   "5435"   "5436"   "5437"   "5438"   "5439"   "5440"   "5441"
[121] "5471"   "548644" "55276"  "5557"   "5558"   "55703"  "55811"  "55821"
[129] "5631"   "5634"   "56655"  "56953"  "56985"  "57804"  "58497"  "6240"
[137] "6241"   "64425"  "646625" "654364" "661"    "7498"   "8382"   "84172"
[145] "84265"  "84284"  "84618"  "8622"   "8654"   "87178"  "8833"   "9060"
[153] "9061"   "93034"  "953"    "9533"   "954"    "955"    "956"    "957"
[161] "9583"   "9615"
```

```
foldchanges = res$log2FoldChange
names(foldchanges) = res$entrez
head(foldchanges)
```

```
     1266     54855      1465     51232      2034      2317
-2.422719  3.201955 -2.313738 -2.059631 -1.888019 -1.649792
```

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

```
attributes(keggres)
```

```
$names
[1] "greater" "less"    "stats"
```

```
head(keggres$less)
```

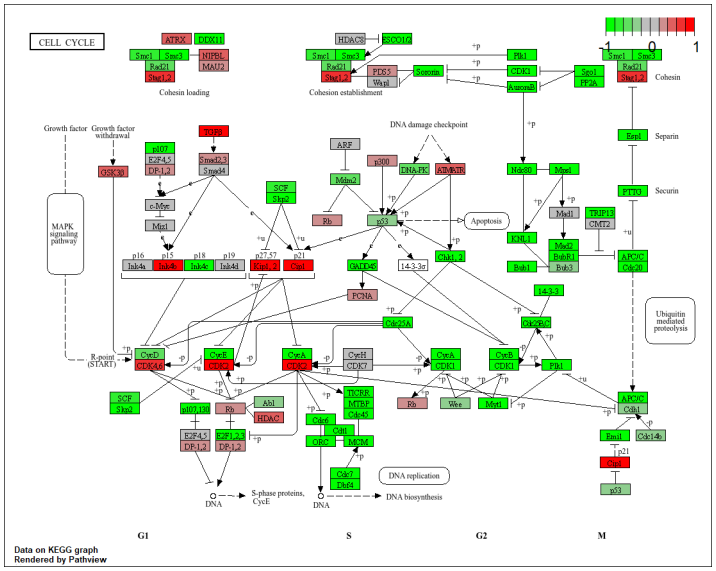
```
                                      p.geomean stat.mean        p.val
hsa04110 Cell cycle                   8.995727e-06 -4.378644 8.995727e-06
hsa03030 DNA replication              9.424076e-05 -3.951803 9.424076e-05
hsa03013 RNA transport                1.375901e-03 -3.028500 1.375901e-03
hsa03440 Homologous recombination     3.066756e-03 -2.852899 3.066756e-03
hsa04114 Oocyte meiosis               3.784520e-03 -2.698128 3.784520e-03
hsa00010 Glycolysis / Gluconeogenesis 8.961413e-03 -2.405398 8.961413e-03
                                            q.val set.size         exp1
hsa04110 Cell cycle                   0.001448312      121 8.995727e-06
hsa03030 DNA replication              0.007586381       36 9.424076e-05
hsa03013 RNA transport                0.073840037      144 1.375901e-03
hsa03440 Homologous recombination     0.121861535       28 3.066756e-03
hsa04114 Oocyte meiosis               0.121861535      102 3.784520e-03
hsa00010 Glycolysis / Gluconeogenesis 0.212222694       53 8.961413e-03
```

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

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

```
Info: Working in directory C:/Users/kenny/OneDrive/Documents/BIMM 143 W25/class04/class014/cl
```

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




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

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

```
Warning: reconcile groups sharing member nodes!
```

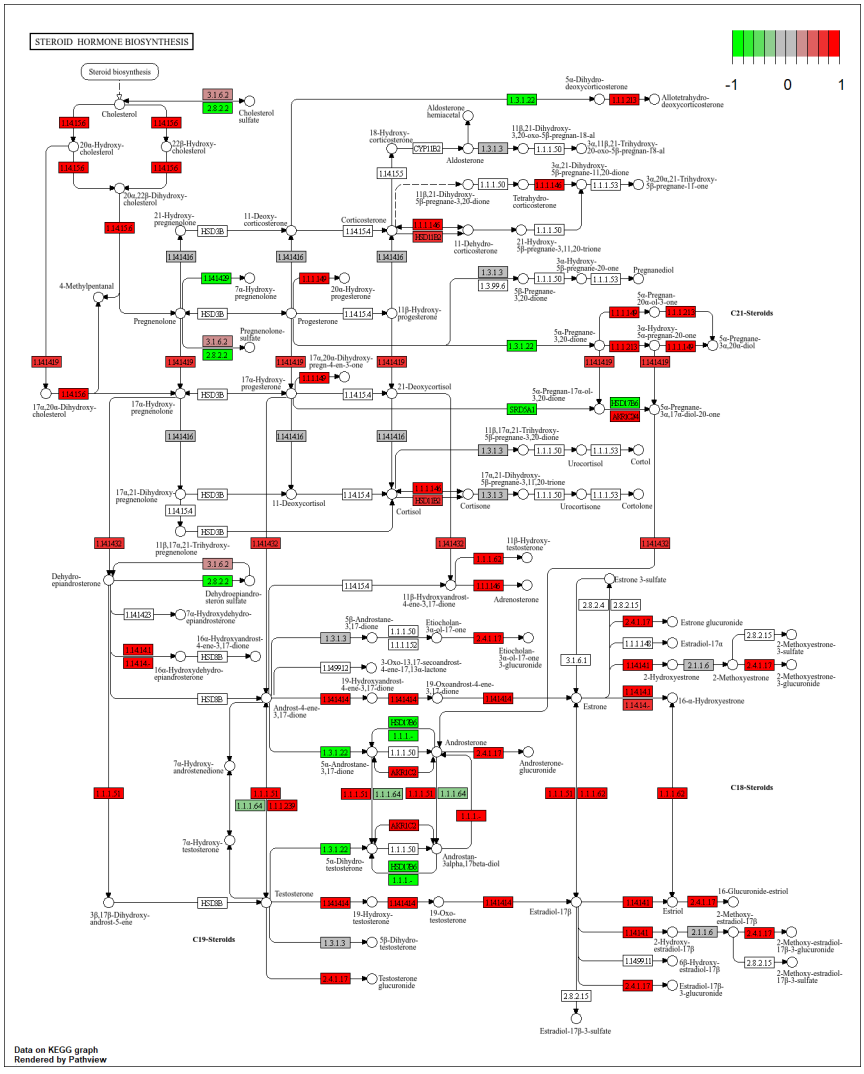
```
     [,1] [,2]
[1,] "9"  "300"
[2,] "9"  "306"
```

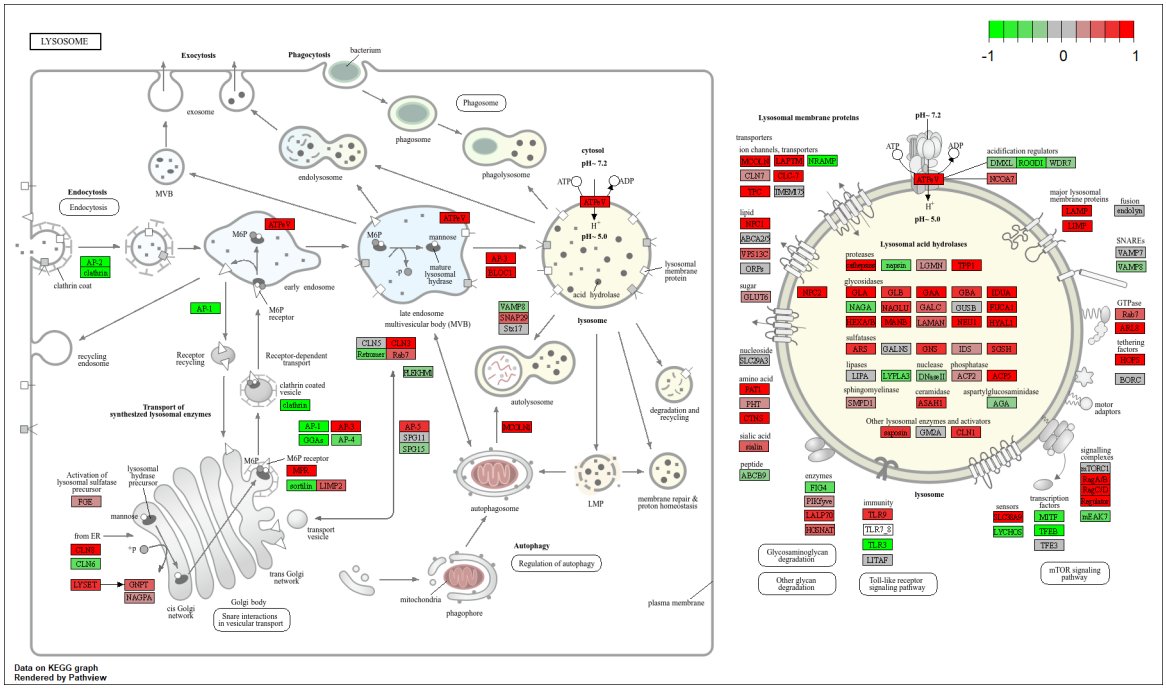
```
Info: Working in directory C:/Users/kenny/OneDrive/Documents/BIMM 143 W25/class04/class014/cl
```

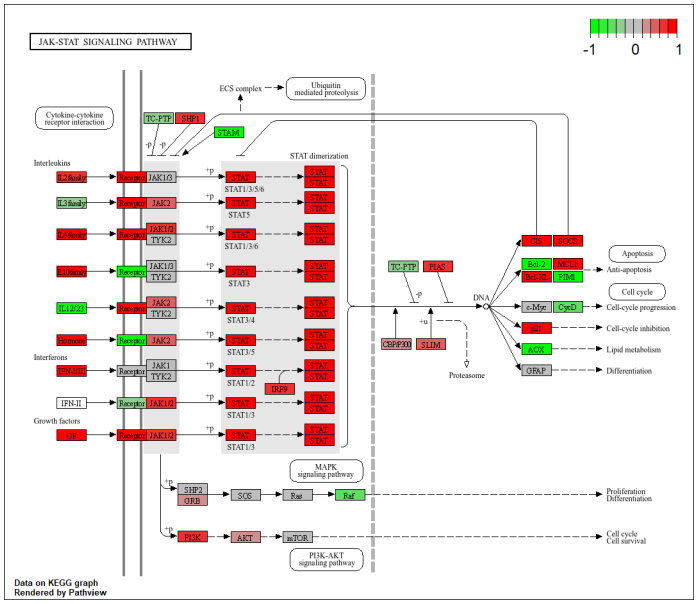
```
Info: Writing image file hsa04110.pathview.pdf
```

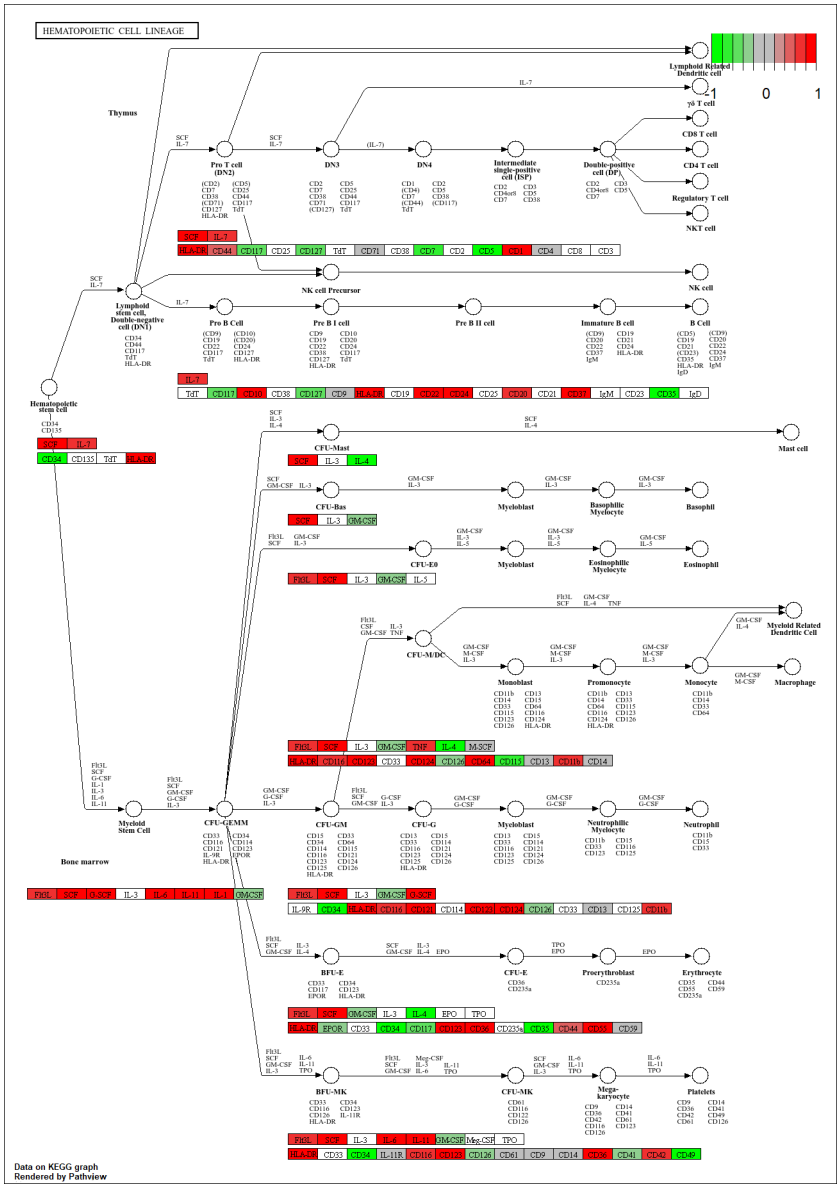
```
keggrespathways <- rownames(keggres$greater)[1:5]

keggresids = substr(keggrespathways, start=1, stop=8)
keggresids
```

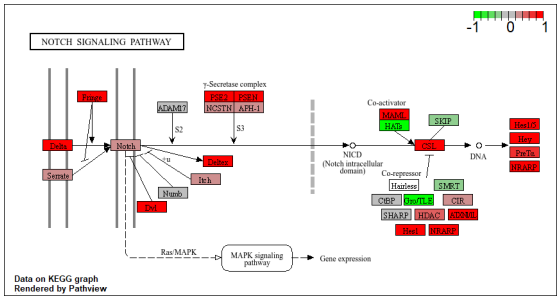
```
[1] "hsa04640" "hsa04630" "hsa00140" "hsa04142" "hsa04330"
```


STEROID HORMONE BIOSYNTHESIS
-1
0
1
Steroid biosynthesis
Cholesterol
Cholesterol sulfate
20α-Hydroxy-cholesterol
22β-Hydroxy-cholesterol
20α,22β-Dihydroxy-cholesterol
21-Hydroxy-pregnenolone
11-Deoxy-corticosterone
Corticosterone
11-Dehydro-corticosterone
Aldosterone hemiacetal
Aldosterone
18-Hydroxy-corticosterone
5α-Dihydro-deoxycorticosterone
Allotetrahydro-deoxycorticosterone
11β,21-Dihydroxy-3,20-oxo-5β-pregnan-18-al
3α,11β,21-Trihydroxy-20-oxo-5β-pregnan-18-al
3α,21-Dihydroxy-5β-pregnane-11,20-dione
3α,20α,21-Trihydroxy-5β-pregnane-11-one
Tetrahydro-corticosterone
11β,21-Dihydroxy-5β-pregnane-3,20-dione
21-Hydroxy-5β-pregnane-3,11,20-trione
4-Methylpentanal
Pregnenolone
7α-Hydroxy-pregnenolone
20α-Hydroxy-progesterone
Progesterone
11β-Hydroxy-progesterone
Pregnenolone-sulfate
3α-Hydroxy-5β-pregnane-20-one
Pregnanediol
5β-Pregnane-3,20-dione
5α-Pregnan-20α-ol-3-one
5α-Pregnane-3,20-dione
3α-Hydroxy-5α-pregnan-20-one
5α-Pregnane-3α,20α-diol
C21-Steroids
17α,20α-Dihydroxy-pregn-4-en-3-one
17α,20α-Dihydroxy-cholesterol
17α-Hydroxy-pregnenolone
17α-Hydroxy-progesterone
21-Deoxycortisol
5α-Pregnan-17α-ol-3,20-dione
5α-Pregnane-3α,17α-diol-20-one
11β,17α,21-Trihydroxy-5β-pregnane-3,20-dione
Urocortisol
Cortol
17α,21-Dihydroxy-pregnenolone
11-Deoxycortisol
17α,21-Dihydroxy-5β-pregnane-3,11,20-trione
Cortisone
Cortisol
Urocortisone
Cortolone
11β,17α,21-Trihydroxy-pregnenolone
Dehydro-epiandrosterone
Dehydroepiandro-sterone sulfate
7α-Hydroxydehydro-epiandrosterone
16α-Hydroxyandrost-4-ene-3,17-dione
16α-Hydroxydehydro-epiandrosterone
5β-Androstane-3,17-dione
3-Oxo-13,17-secoandrost-4-ene-17,13α-lactone
19-Hydroxyandrost-4-ene-3,17-dione
19-Oxoandrost-4-ene-3,17-dione
Androst-4-ene-3,17-dione
11β-Hydroxy-testosterone
Adrenosterone
11β-Hydroxyandrost-4-ene-3,17-dione
Etiocholan-3α-ol-17-one
Etiocholan-3α-ol-17-one 3-glucuronide
Estrone 3-sulfate
Estrone glucuronide
Estradiol-17α
2-Methoxyestrone-3-sulfate
2-Hydroxyestrone
2-Methoxyestrone
2-Methoxyestrone-3-glucuronide
Estrone
16-α-Hydroxyestrone
Androsterone
Androsterone-glucuronide
7α-Hydroxy-androstenedione
5α-Androstane-3,17-dione
7α-Hydroxy-testosterone
5α-Dihydro-testosterone
Androstan-3alpha,17beta-diol
C18-Steroids
Testosterone
3β,17β-Dihydroxy-androst-5-ene
19-Hydroxy-testosterone
19-Oxo-testosterone
Estradiol-17β
16-Glucuronide-estriol
Estriol
2-Methoxy-estradiol-17β
2-Hydroxy-estradiol-17β
6β-Hydroxy-estradiol-17β
2-Methoxy-estradiol-17β-3-glucuronide
Estradiol-17β-3-glucuronide
2-Methoxy-estradiol-17β-3-sulfate
Estradiol-17β-3-sulfate
C19-Steroids
5β-Dihydro-testosterone
Testosterone glucuronide
Data on KEGG graph
Rendered by Pathview


LYSOSOME
-1
0
1
Exocytosis
Phagocytosis
bacterium
exosome
Phagosome
phagosome
endolysosome
phagolysosome
MVB
cytosol
pH~ 7.2
ATP
ADP
ATPase
H+
pH~ 5.0
Endocytosis
Endocytosis
clathrin coat
AP-2
clathrin
M6P
M6P receptor
early endosome
late endosome
multivesicular body (MVB)
mannose
mature lysosomal hydrase
acid hydrolase
lysosomal membrane protein
lysosome
recycling endosome
Receptor recycling
Receptor-dependent transport
clathrin coated vesicle
clathrin
Transport of synthesized lysosomal enzymes
autolysosome
degradation and recycling
Activation of lysosomal sulfatase precursor
lysosomal hydrase precursor
M6P receptor
sortilin
LIMP2
transport vesicle
autophagosome
LMP
membrane repair & proton homeostasis
mannose
from ER
trans Golgi network
cis Golgi network
Golgi body
Snare interactions in vesicular transport
mitochondria
phagophore
Autophagy
Regulation of autophagy
plasma membrane
Lysosomal membrane proteins
transporters
ion channels, transporters
acidification regulators
major lysosomal membrane proteins
fusion
endolyn
lipid
sugar
SNAREs
GTPase
tethering factors
nucleoside
amino acid
sialic acid
peptide
Lysosomal acid hydrolases
proteases
glycosidases
sulfatases
lipases
nuclease
phosphatase
sphingomyelinase
ceramidase
aspartylglucosaminidase
Other lysosomal enzymes and activators
motor adaptors
signalling complexes
enzymes
immunity
sensors
transcription factors
Glycosaminoglycan degradation
Other glycan degradation
Toll-like receptor signaling pathway
mTOR signaling pathway
Data on KEGG graph
Rendered by Pathview


JAK-STAT SIGNALING PATHWAY
-1
0
1
ECS complex
Ubiquitin mediated proteolysis
Cytokine-cytokine receptor interaction
TC-PTP
SHP1
STAM
STAT dimerization
Interleukins
IL2Rab
IL3Rab
IL6Rab
IL10Rab
IL12/23
Hormone
Interferons
IFN-I/II
IFN-III
Growth factors
GH
Receptor
JAK1/3
JAK2
JAK1/2
TYK2
STAT1/3/5/6
STAT5
STAT1/3/6
STAT3
STAT3/4
STAT3/5
STAT1/2
IRF9
STAT1/3
P
MAPK signaling pathway
SHP2
GRB
SOS
Ras
Raf
Proliferation Differentiation
PI3K
AKT
mTOR
PI3K-AKT signaling pathway
Cell cycle Cell survival
PIAS
DNA
CBP/P300
SUMO
Proteasome
CIS
SOCS
Bcl-2
Mcl-1
Bcl-XL
Pim1
Apoptosis
Anti-apoptosis
c-Myc
CycD
Cell cycle
Cell-cycle progression
p21
Cell-cycle inhibition
ACOX
Lipid metabolism
GFAP
Differentiation
Data on KEGG graph
Rendered by Pathview


HEMATOPOIETIC CELL LINEAGE
Lymphoid Related Dendritic cell
-1
0
1
Thymus
IL-7
γδ T cell
CD8 T cell
CD4 T cell
Regulatory T cell
NKT cell
SCF
IL-7
Pro T cell (DN2)
DN3
DN4
Intermediate single-positive cell (ISP)
Double-positive cell (DP)
NK cell Precursor
NK cell
Lymphoid stem cell, Double-negative cell (DN1)
Pro B Cell
Pre B I cell
Pre B II cell
Immature B cell
B Cell
Hematopoietic stem Cell
CFU-Mast
Mast cell
CFU-Bas
Myeloblast
Basophilic Myelocyte
Basophil
CFU-E0
Eosinophilic Myelocyte
Eosinophil
CFU-MDC
Monoblast
Promonocyte
Monocyte
Myeloid Related Dendritic Cell
Macrophage
Myeloid Stem Cell
CFU-GEMM
CFU-GM
CFU-G
Neutrophilic Myelocyte
Neutrophil
Bone marrow
BFU-E
CFU-E
Proerythroblast
Erythrocyte
BFU-MK
CFU-MK
Mega-karyocyte
Platelets
Data on KEGG graph
Rendered by Pathview


NOTCH SIGNALING PATHWAY
-1
0
1
Ligand
Delta
Serrate
Notch
ADAM17
γ-Secretase complex
PSEN1
PSEN2
NCSTN
APH-1
S2
S3
Fringe
Deltex
Itch
Numb
Dvl
NICD
(Notch intracellular domain)
Co-activator
MAML
HAT
SKIP
CSL
DNA
Co-repressor
Hairless
SMRT
CtBP
Dsp/TLE
CIR
SHARP
HDAC
CDK8
Hes1
Hey
FoxL1
NRARP
Ras/MAPK
MAPK signaling pathway
Gene expression
Data on KEGG graph
Rendered by Pathview

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

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

Info: Working in directory C:/Users/kenny/OneDrive/Documents/BIMM 143 W25/class04/class014/cl

Info: Writing image file hsa04640.pathview.png

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

Info: Working in directory C:/Users/kenny/OneDrive/Documents/BIMM 143 W25/class04/class014/cl

Info: Writing image file hsa04630.pathview.png

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

Info: Working in directory C:/Users/kenny/OneDrive/Documents/BIMM 143 W25/class04/class014/cl

Info: Writing image file hsa00140.pathview.png

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

Info: Working in directory C:/Users/kenny/OneDrive/Documents/BIMM 143 W25/class04/class014/cl

Info: Writing image file hsa04142.pathview.png

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

Info: Working in directory C:/Users/kenny/OneDrive/Documents/BIMM 143 W25/class04/class014/cl

Info: Writing image file hsa04330.pathview.png
```

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

Yes.

```
head(keggres$greater)
```

```
                                      p.geomean stat.mean       p.val
hsa04640 Hematopoietic cell lineage   0.002822776  2.833362 0.002822776
hsa04630 Jak-STAT signaling pathway   0.005202070  2.585673 0.005202070
hsa00140 Steroid hormone biosynthesis 0.007255099  2.526744 0.007255099
hsa04142 Lysosome                     0.010107392  2.338364 0.010107392
hsa04330 Notch signaling pathway      0.018747253  2.111725 0.018747253
hsa04916 Melanogenesis                0.019399766  2.081927 0.019399766
                                          q.val set.size        exp1
hsa04640 Hematopoietic cell lineage   0.3893570       55 0.002822776
hsa04630 Jak-STAT signaling pathway   0.3893570      109 0.005202070
hsa00140 Steroid hormone biosynthesis 0.3893570       31 0.007255099
hsa04142 Lysosome                     0.4068225      118 0.010107392
hsa04330 Notch signaling pathway      0.4391731       46 0.018747253
hsa04916 Melanogenesis                0.4391731       90 0.019399766
```

```
pathview(gene.data=foldchanges, pathway.id="hsa04640")
```

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

```
Info: Working in directory C:/Users/kenny/OneDrive/Documents/BIMM 143 W25/class04/class014/cl
```

```
Info: Writing image file hsa04640.pathview.png
```

```
pathview(gene.data=foldchanges, pathway.id="hsa04640", kegg.native=T)
```

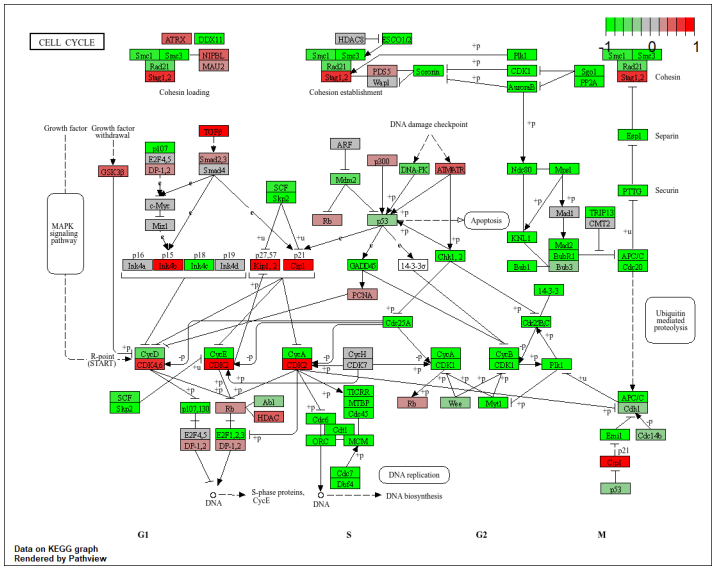
```
'select()' returned 1:1 mapping between keys and columns
```

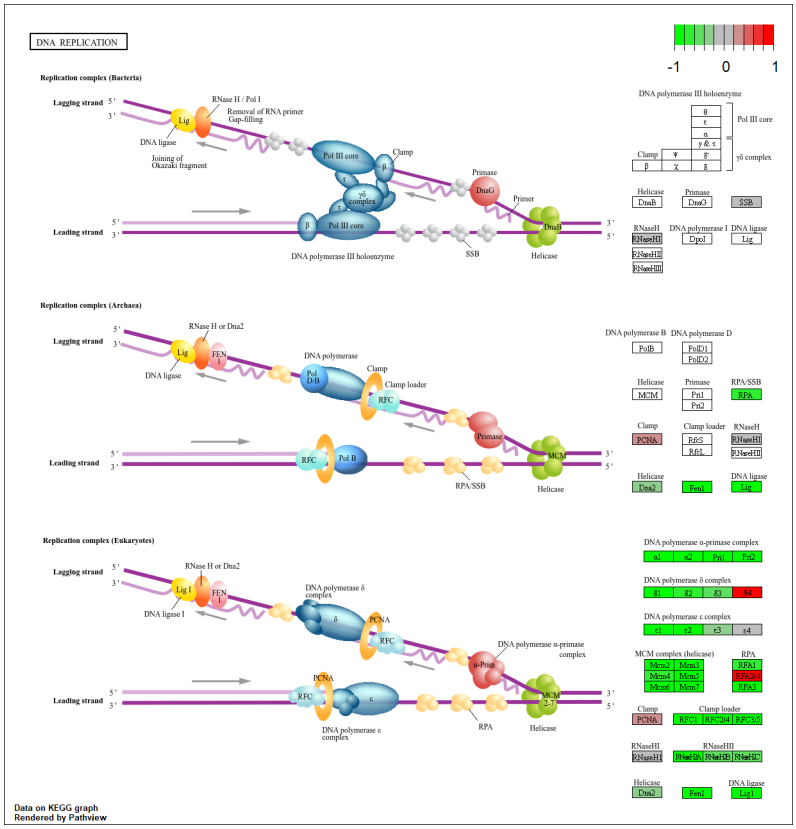
```
Info: Working in directory C:/Users/kenny/OneDrive/Documents/BIMM 143 W25/class04/class014/cl
```

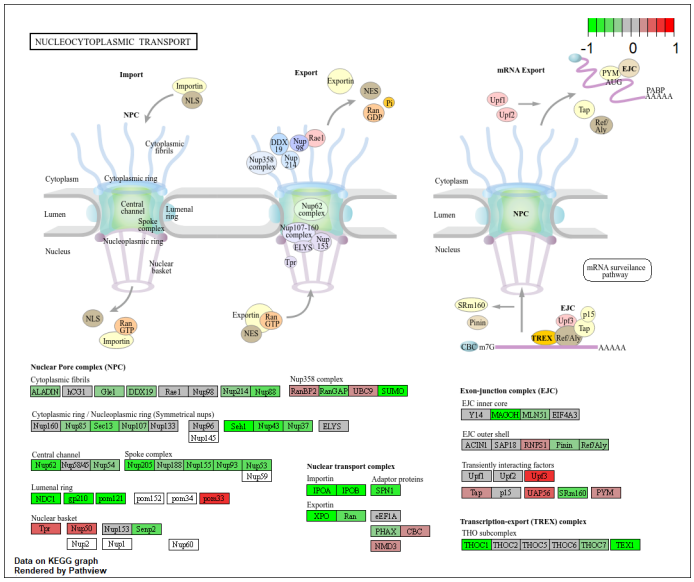
```
Info: Writing image file hsa04640.pathview.png
```

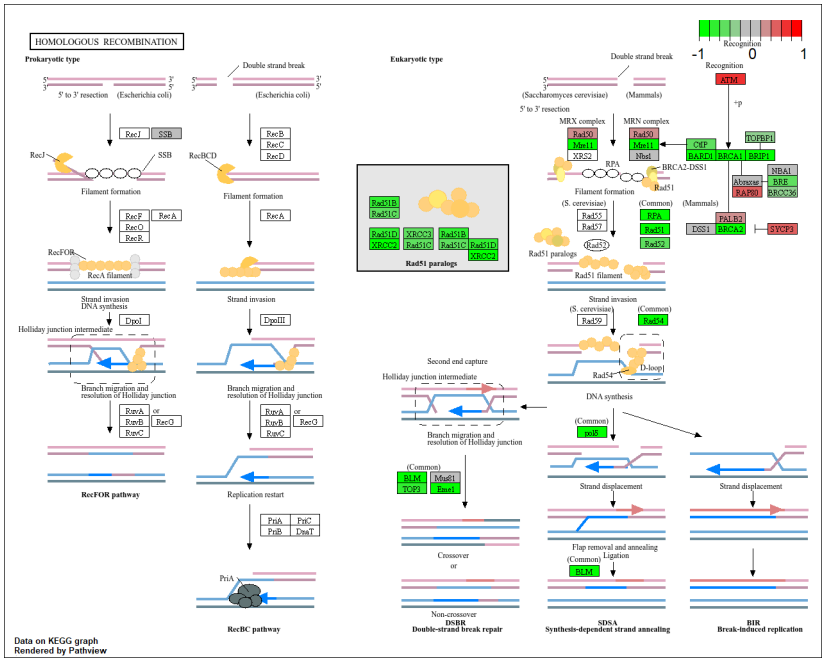
```
keggrespathways1 <- rownames(keggres$less)[1:5]

keggresids1 = substr(keggrespathways1, start=1, stop=8)
keggresids1
```

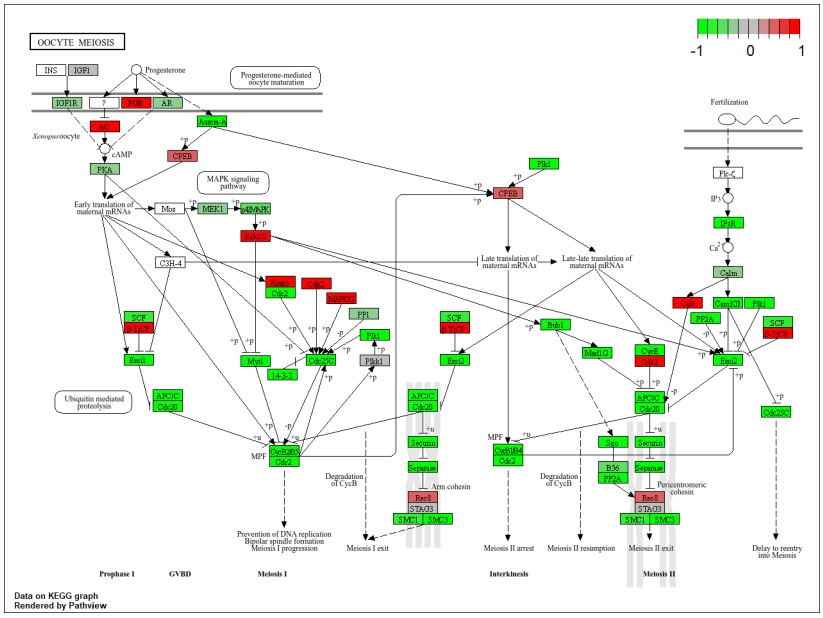
```
[1] "hsa04110" "hsa03030" "hsa03013" "hsa03440" "hsa04114"
```






NUCLEOCYTOPLASMIC TRANSPORT
Import
Export
mRNA Export
NPC
Cytoplasm
Lumen
Nucleus
Nuclear Pore complex (NPC)
Nup358 complex
Exon-junction complex (EJC)
Cytoplasmic ring / Nucleoplasmic ring (Symmetrical nups)
Central channel
Spoke complex
Lumenal ring
Nuclear basket
Nuclear transport complex
Importin
Exportin
Adaptor proteins
EJC inner core
EJC outer shell
Transiently interacting factors
Transcription-export (TREX) complex
THO subcomplex
mRNA surveilance pathway
Data on KEGG graph
Rendered by Pathview


HOMOLOGOUS RECOMBINATION
Prokaryotic type
Eukaryotic type
5' to 3' resection
(Escherichia coli)
Double strand break
RecJ
SSB
RecBCD
RecA
RecFOR
RecA filament
Filament formation
Strand invasion
DNA synthesis
DpoI
DpoIII
Holliday junction intermediate
Branch migration and resolution of Holliday junction
RuvA
RuvB
RuvC
or
RecO
RecFOR pathway
Replication restart
PriA
PriB
PriC
DnaT
RecBC pathway
Rad51 paralogs
Rad51B
Rad51C
Rad51D
XRCC2
XRCC3
Rad51 filament
(Saccharomyces cerevisiae)
(Mammals)
MRX complex
MRN complex
Rad50
Mre11
XRS2
Nbs1
RPA
BRCA2-DSS1
Rad51
(S. cerevisiae)
(Common)
(Mammals)
Rad55
Rad57
Rad52
RFA
Rad51
Rad54
D-loop
Second end capture
Branch migration and resolution of Holliday junction
(Common)
RLM
Mus81
TOP3
Eme1
Crossover
or
Non-crossover
DSBR
Double-strand break repair
Strand displacement
Flap removal and annealing
Ligation
SDSA
Synthesis-dependent strand annealing
BIR
Break-induced replication
Recognition
ATM
+p
TOPBP1
CtIP
RAD51
BRCA1
BRIP1
MRAI
BRE
Abraxas
RAP80
BRCC36
PALB2
BRCA2
DSS1
SYCP3
Data on KEGG graph
Rendered by Pathview




```
pathview(gene.data=foldchanges, pathway.id=keggresids1, species="hsa")
```

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

```
Info: Working in directory C:/Users/kenny/OneDrive/Documents/BIMM 143 W25/class04/class014/cl
```

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

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

```
Info: Working in directory C:/Users/kenny/OneDrive/Documents/BIMM 143 W25/class04/class014/cl
```

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

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

```
Info: Working in directory C:/Users/kenny/OneDrive/Documents/BIMM 143 W25/class04/class014/cl
```

```
Info: Writing image file hsa03013.pathview.png
```

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

```
Info: Working in directory C:/Users/kenny/OneDrive/Documents/BIMM 143 W25/class04/class014/cl

Info: Writing image file hsa03440.pathview.png

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

Info: Working in directory C:/Users/kenny/OneDrive/Documents/BIMM 143 W25/class04/class014/cl

Info: Writing image file hsa04114.pathview.png
```

Section 3 Gene Ontology (GO)

```
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)

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
```

Section 4. Reactome Analysis

```
sig_genes <- res[res$padj <= 0.05 & !is.na(res$padj), "symbol"]
print(paste("Total number of significant genes:", length(sig_genes)))
```

```
[1] "Total number of significant genes: 8147"
```

```
write.table(sig_genes, file="significant_genes.txt", row.names=FALSE, col.names=FALSE, quote=
```

> Q. 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?

The pathway that has the most significant "Entities p-value" is "Cell cycle, mitotic" as its value is 2E-5. Yes, the most significant pathways listed match my previous KEGG results as they both show cell-cycle related pathways as strongly enriched and they overlap. Some differences

that could cause differences between the two methods are that they use different pathway annotations/gene groupings and they use different statistical models for testing enrichment.

```
sessionInfo()
```

```
R version 4.4.1 (2024-06-14 ucrt)
Platform: x86_64-w64-mingw32/x64
Running under: Windows 11 x64 (build 26100)

Matrix products: default


locale:
[1] LC_COLLATE=English_United States.utf8
[2] LC_CTYPE=English_United States.utf8
[3] LC_MONETARY=English_United States.utf8
[4] LC_NUMERIC=C
[5] LC_TIME=English_United States.utf8

time zone: America/Los_Angeles
tzcode source: internal

attached base packages:
[1] stats4    stats     graphics  grDevices utils     datasets  methods
[8] base

other attached packages:
 [1] gageData_2.44.0             gage_2.56.0
 [3] pathview_1.46.0             org.Hs.eg.db_3.20.0
 [5] AnnotationDbi_1.68.0        ggplot2_4.0.2
 [7] DESeq2_1.46.0               SummarizedExperiment_1.36.0
 [9] Biobase_2.66.0              MatrixGenerics_1.18.1
[11] matrixStats_1.5.0           GenomicRanges_1.58.0
[13] GenomeInfoDb_1.42.3         IRanges_2.40.1
[15] S4Vectors_0.44.0            BiocGenerics_0.52.0

loaded via a namespace (and not attached):
 [1] KEGGREST_1.46.0         gtable_0.3.6            xfun_0.56
 [4] lattice_0.22-6          bitops_1.0-9            vctrs_0.7.1
 [7] tools_4.4.1             generics_0.1.4          parallel_4.4.1
[10] tibble_3.3.1            RSQLite_2.4.6           blob_1.3.0
[13] pkgconfig_2.0.3         Matrix_1.7-0            RColorBrewer_1.1-3
```

```
[16] S7_0.2.1                 graph_1.84.1            lifecycle_1.0.5
[19] GenomeInfoDbData_1.2.13 compiler_4.4.1          farver_2.1.2
[22] Biostrings_2.74.1        codetools_0.2-20        htmltools_0.5.9
[25] RCurl_1.98-1.17          yaml_2.3.12             GO.db_3.20.0
[28] pillar_1.11.1            crayon_1.5.3            BiocParallel_1.40.2
[31] cachem_1.1.0             DelayedArray_0.32.0     abind_1.4-8
[34] tidyselect_1.2.1         locfit_1.5-9.12         digest_0.6.39
[37] dplyr_1.2.0              labeling_0.4.3          fastmap_1.2.0
[40] grid_4.4.1               colorspace_2.1-2        cli_3.6.5
[43] SparseArray_1.6.2        magrittr_2.0.4          S4Arrays_1.6.0
[46] XML_3.99-0.22            withr_3.0.2             scales_1.4.0
[49] UCSC.utils_1.2.0         bit64_4.6.0-1           rmarkdown_2.30
[52] XVector_0.46.0           httr_1.4.8              bit_4.6.0
[55] otel_0.2.0               png_0.1-8               memoise_2.0.1
[58] evaluate_1.0.5           knitr_1.51              rlang_1.1.7
[61] Rcpp_1.1.1               glue_1.8.0              DBI_1.2.3
[64] Rgraphviz_2.50.0         KEGGgraph_1.66.0        rstudioapi_0.18.0
[67] jsonlite_2.0.0           R6_2.6.1                zlibbioc_1.52.0
```