DESeq Mini Project: Pathway Analysis from RNA-Seq Results

Andrew Kapinos

11/19/2021

Section 1. Differential Expression Analysis

```
library(DESeq2)
```

```
## Warning: package 'GenomicRanges' was built under R version 4.1.2
```

Let's load our count and metagene data files. We can then import our metadata.

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

# Import metadata and take a look
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
```

Let's import our count data as well.

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

```
##                 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.raw[,-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?

rowSums(countData)==0 will return the row numbers of rows will all 0s, which we want to remove. We can add the minus symbol to select rows that don't have all 0s.

```
# Filter count data where you have 0 read count across all samples.
countData = countData[-which(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
```

Optional: PCA Analysis

Let's perform PCA to ensure that the data looks well separated.

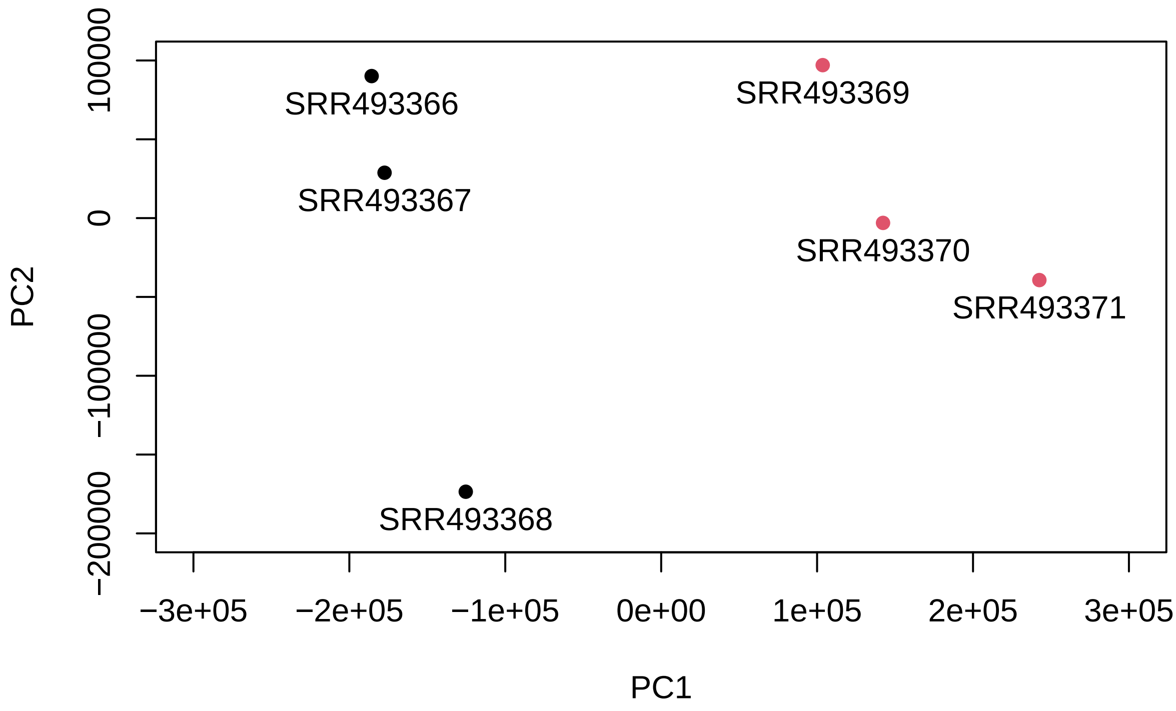
```
pca.counts <- prcomp(t(countData))
attributes(pca.counts)
```

```
## $names
## [1] "sdev"     "rotation" "center"   "scale"    "x"
##
## $class
## [1] "prcomp"
```

```
summary(pca.counts$x)
```

```
##       PC1               PC2               PC3               PC4        
##  Min.   :-185705   Min.   :-173585   Min.   :-33499   Min.   :-10518  
##  1st Qu.:-164392   1st Qu.: -30247   1st Qu.: -3696   1st Qu.: -4387  
##  Median : -10861   Median :  12872   Median : -1112   Median :  2204  
##  Mean   :      0   Mean   :      0   Mean   :     0   Mean   :     0  
##  3rd Qu.: 132618   3rd Qu.:  74778   3rd Qu.: 10853   3rd Qu.:  5270  
##  Max.   : 242550   Max.   :  97047   Max.   : 25439   Max.   :  6403  
##       PC5               PC6            
##  Min.   :-6734.7   Min.   :-3.045e-09  
##  1st Qu.:-3527.9   1st Qu.:-4.769e-10  
##  Median :  323.4   Median : 1.729e-10  
##  Mean   :    0.0   Mean   :-1.890e-12  
##  3rd Qu.: 2729.8   3rd Qu.: 1.178e-09  
##  Max.   : 7367.7   Max.   : 1.868e-09
```

```
plot(pca.counts$x[,1], pca.counts$x[,2],
     xlim=c(-300000,300000), ylim=c(-200000,100000),
     xlab="PC1",
     ylab="PC2",col=as.factor(colData$condition),pch=16)
text(pca.counts$x[,1], pca.counts$x[,2], colnames(countData), pos=1)
```



It looks like the data has some trends between the two groups based on their separation along PC1 in the PC plot. Let's continue to DESeq2.

Running DESEq2

> 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.

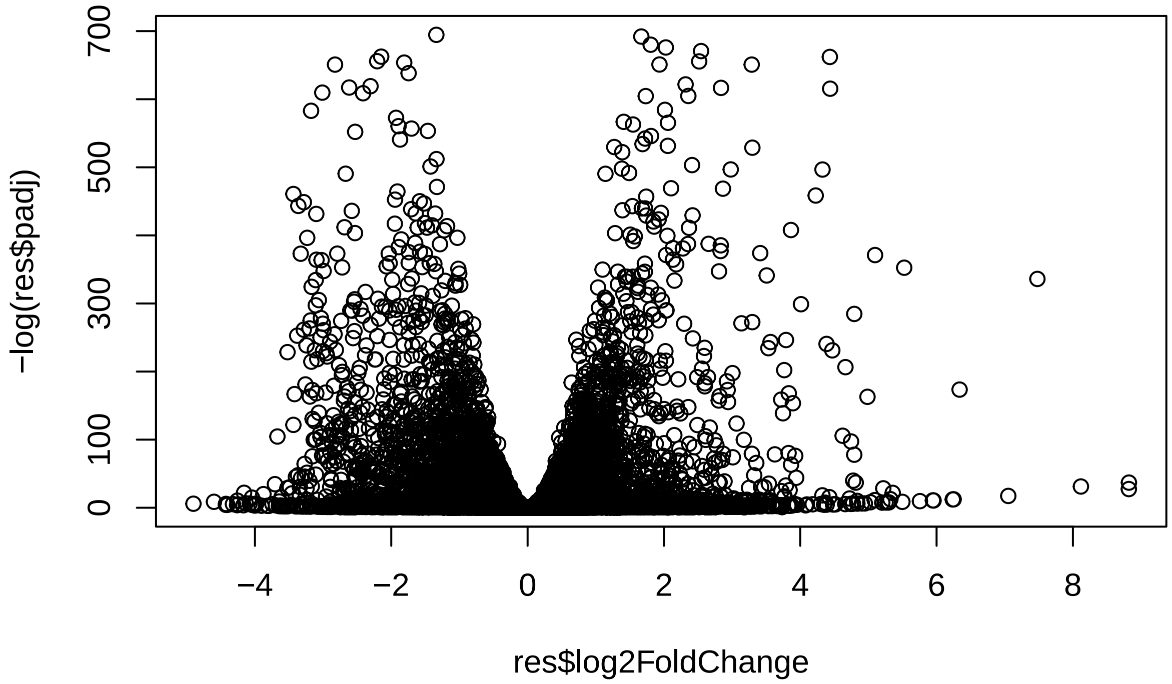
```
dds = DESeqDataSetFromMatrix(countData=countData,
                             colData=colData,
                             design=~condition)
dds = DESeq(dds)
res = results(dds)
summary(res)
```

```
##
## 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
```

27% are upregulated and 28% are downlregulated.

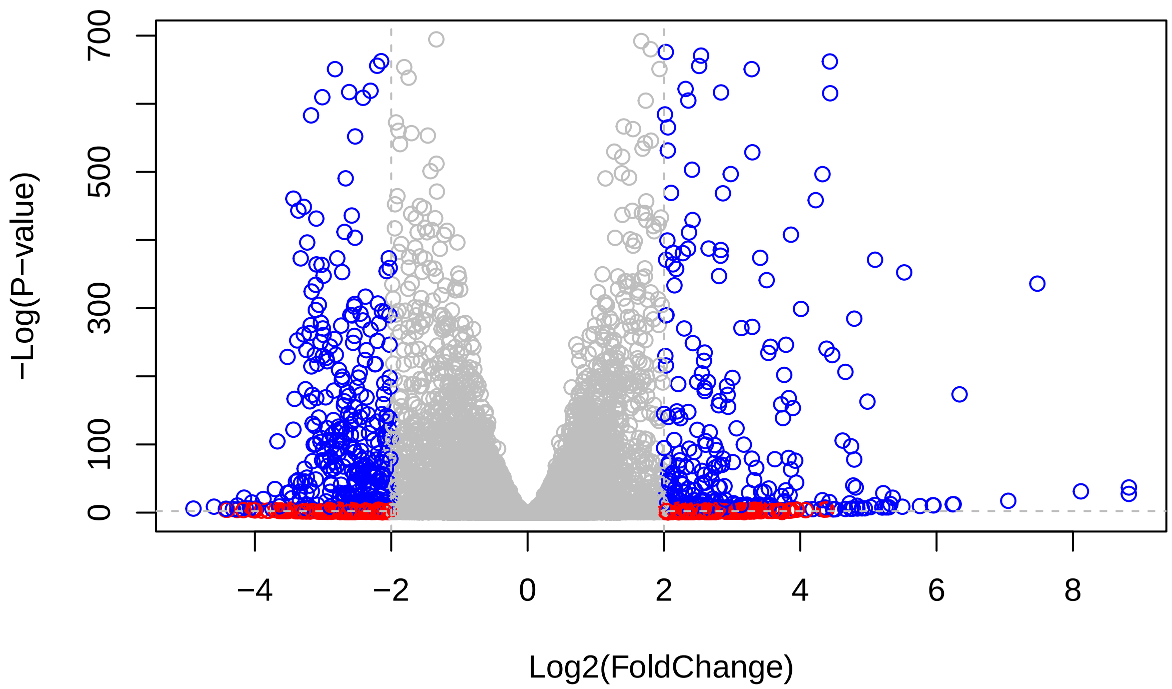
Volcano plot

```
plot(res$log2FoldChange, -log(res$padj))
```



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

```r
# 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$padj < 0.01) & (abs(res$log2FoldChange) > 2 )
mycols[ inds ] <- "blue"

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

# Cut-off lines
abline(v=c(-2,2), col="gray", lty=2)
abline(h=-log(0.1), col="gray", lty=2)
```



Adding gene annotation

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

```
library("AnnotationDbi")
```

```
## Warning: package 'AnnotationDbi' was built under R version 4.1.2
```

```
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")

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

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

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      WASH9P   102723897 WAS protein family h..
## 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$pvalue),]
write.csv(res, file="deseq_results.csv")
```

Section 2: Pathway Analysis

Let’s load the packages and data we’ll need.

```
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"
```

To use gage() we'll need a named vector of fold changes.

```
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
```

Now let's run the gage pathway analysis.

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

Let's take a look at the result structure and the first few down (less) pathways.

```
attributes(keggres)
```

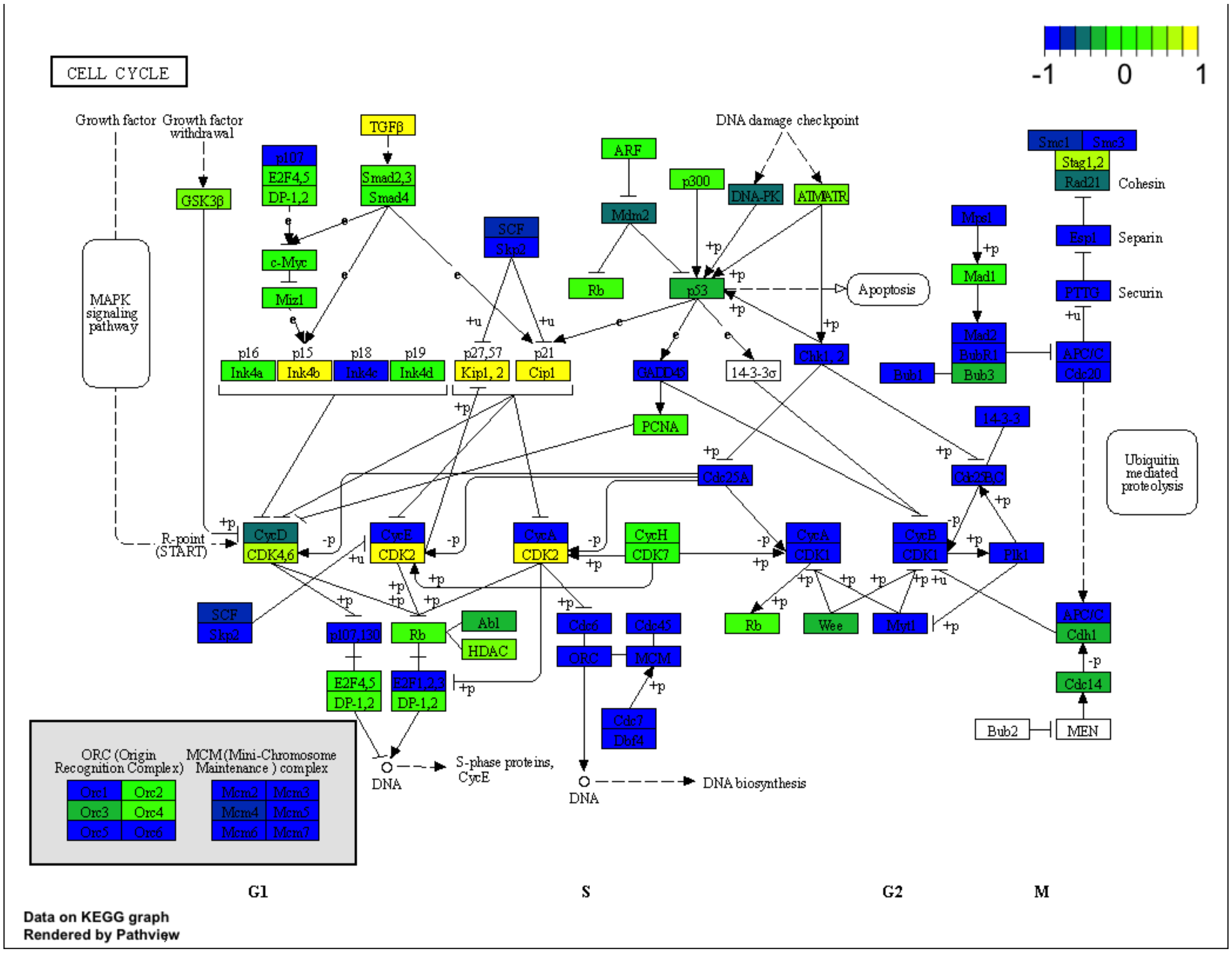
```
## $names
## [1] "greater" "less"    "stats"
```

```
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
```

And let's generate a pathview figure from KEGG data using our results.

```
pathview(gene.data=foldchanges, pathway.id="hsa04110", low="blue", mid="green", high="yellow")
```



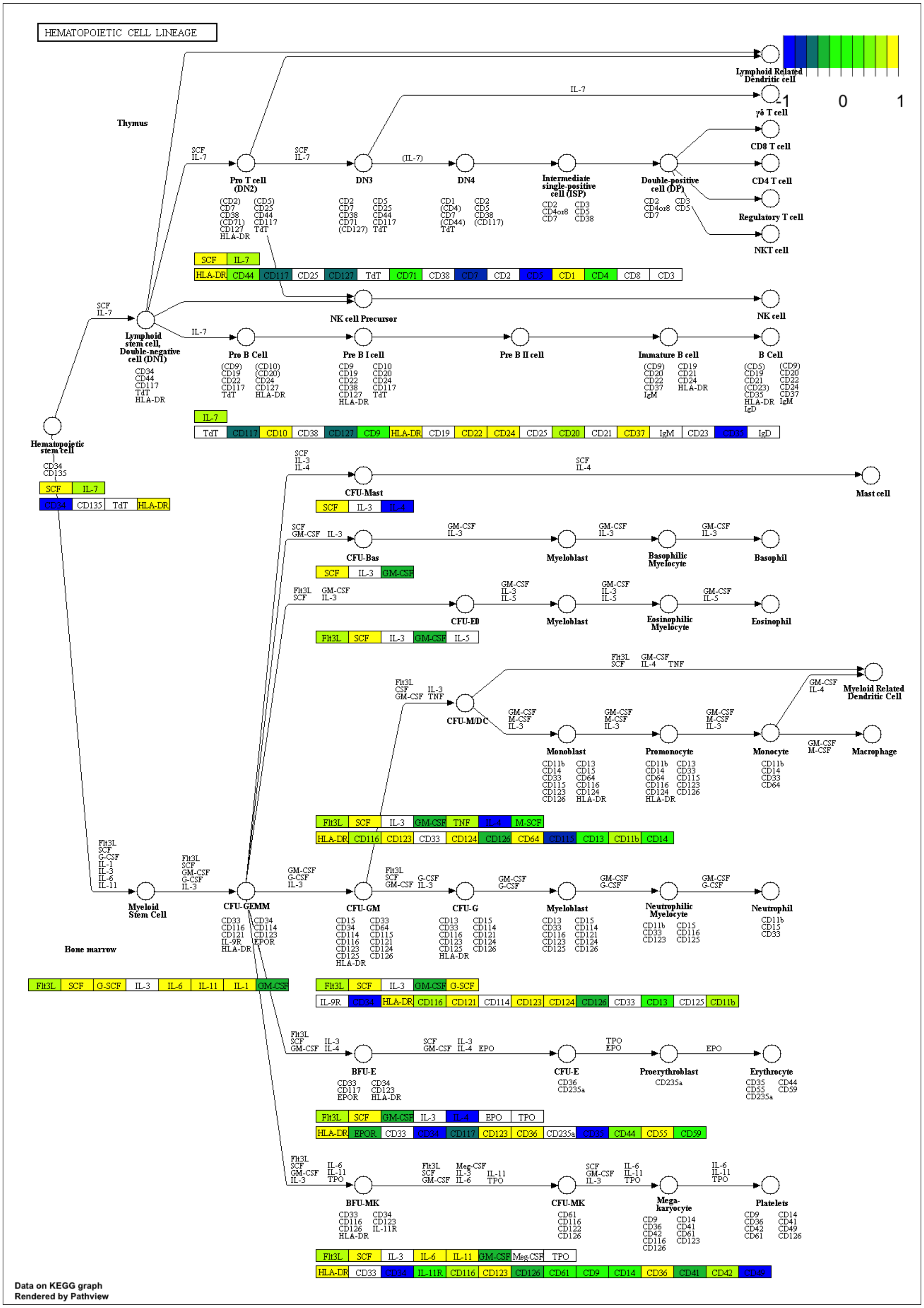
Let's take a look at the up (more) pathways, too.

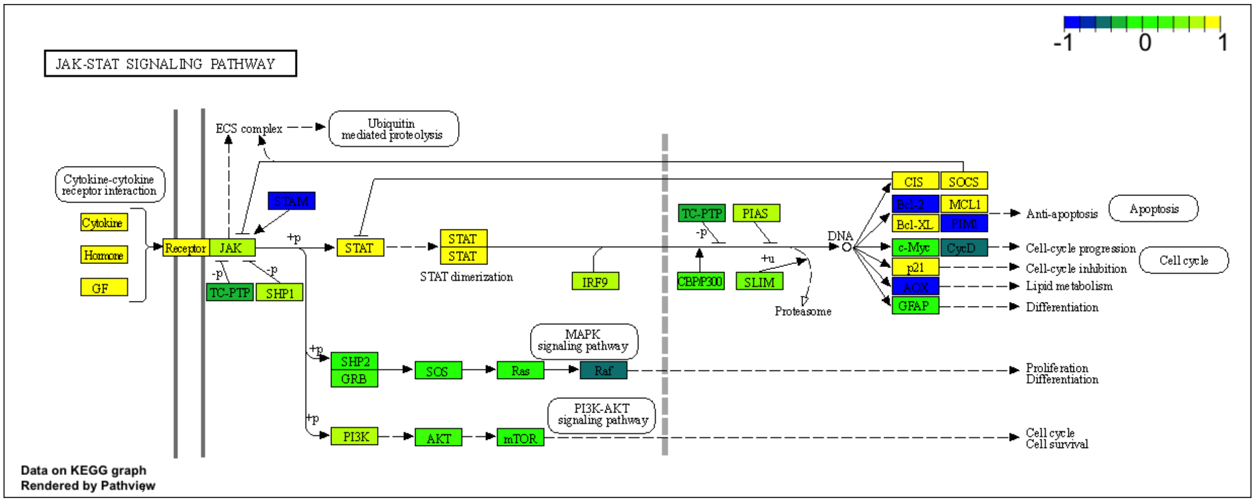
```
## Focus on top 5 upregulated pathways here for demo purposes only
keggrespathways <- rownames(keggres$greater)[1:5]

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

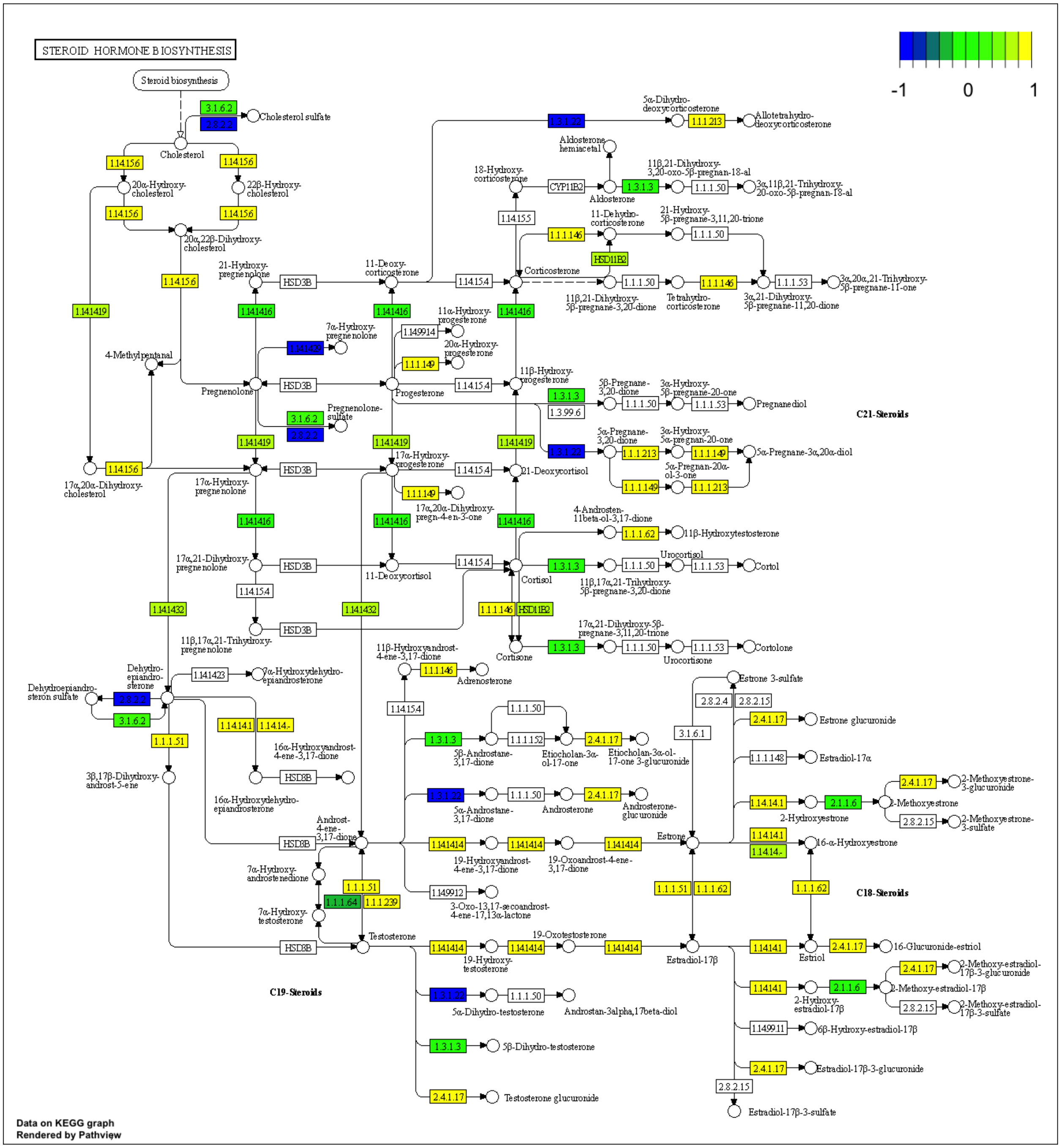
```
## [1] "hsa04640" "hsa04630" "hsa00140" "hsa04142" "hsa04330"
```

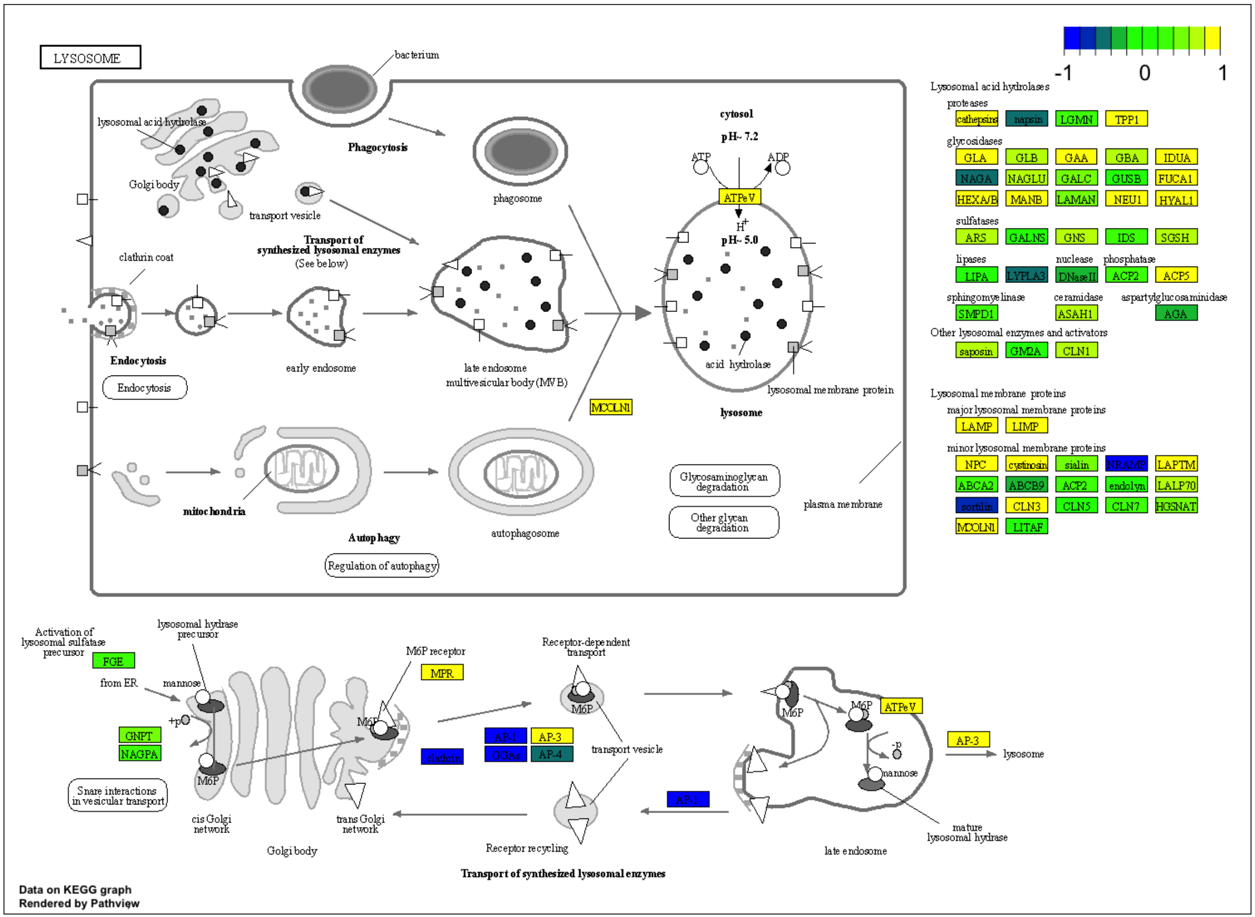
Let's pass these 5 IDs to the pathview() function, which will give us a combined output.

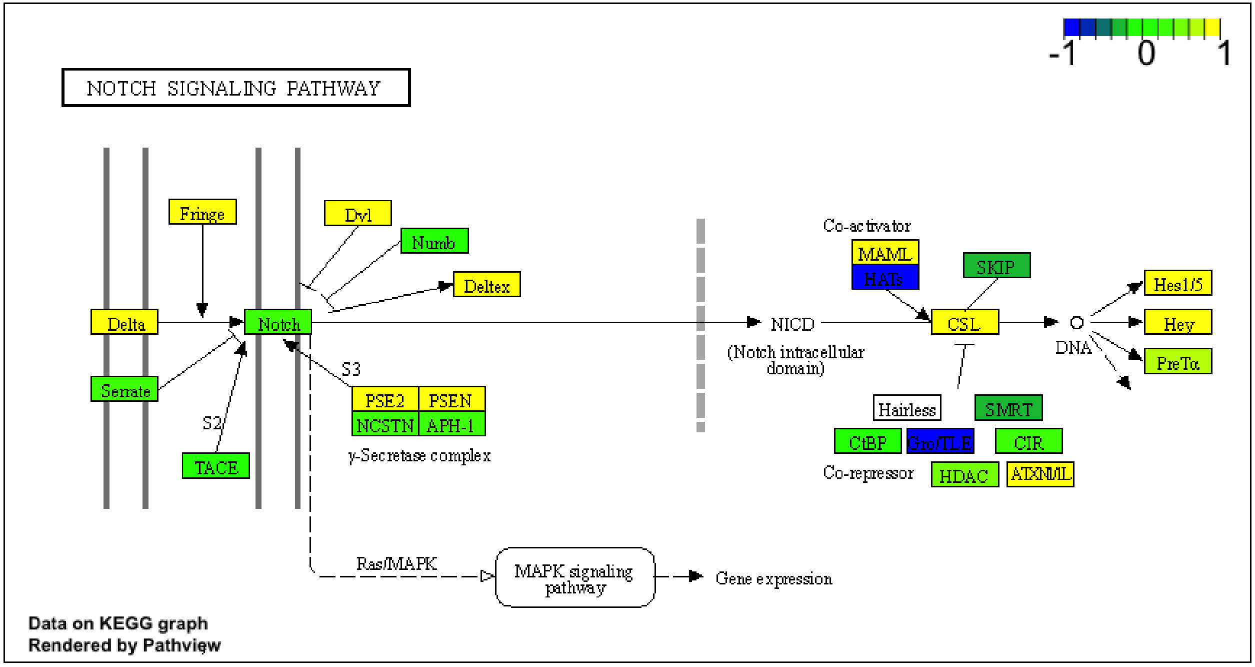
```
pathview(gene.data=foldchanges, pathway.id=keggresids, species="hsa", low="blue", mid="green", high="ye
```


HEMATOPOIETIC CELL LINEAGE
-1
0
1
Lymphoid Related Dendritic cell
IL-7
γδ T cell
Thymus
SCF IL-7
Pro T cell (DN2)
DN3
(IL-7)
DN4
Intermediate single-positive cell (ISP)
Double-positive cell (DP)
CD8 T cell
CD4 T cell
Regulatory T cell
NKT cell
(CD2) CD7 CD38 (CD71) CD127 HLA-DR (CD5) CD25 CD44 CD117 TdT
CD2 CD7 CD38 CD71 (CD127) CD5 CD25 CD44 CD117 TdT
CD1 (CD4) CD7 (CD44) TdT CD2 CD5 CD38 (CD117)
CD2 CD4or8 CD7 CD3 CD5 CD38
CD2 CD4or8 CD7 CD3 CD5
SCF IL-7
HLA-DR CD44 CD117 CD25 CD127 TdT CD71 CD38 CD7 CD2 CD5 CD1 CD4 CD8 CD3
NK cell Precursor
NK cell
Lymphoid stem cell, Double-negative cell (DN1)
CD34 CD44 CD117 TdT HLA-DR
Pro B Cell
Pre B I cell
Pre B II cell
Immature B cell
B Cell
(CD9) CD19 CD22 CD117 TdT (CD10) (CD20) CD24 CD127 HLA-DR
CD9 CD19 CD22 CD38 CD127 HLA-DR CD10 CD20 CD24 CD117 TdT
(CD9) CD20 CD22 CD37 IgM CD19 CD21 CD24 HLA-DR
(CD5) CD19 CD21 (CD23) CD35 HLA-DR IgD (CD9) CD20 CD22 CD24 CD37 IgM
IL-7
TdT CD117 CD10 CD38 CD127 CD9 HLA-DR CD19 CD22 CD24 CD25 CD20 CD21 CD37 IgM CD23 CD35 IgD
Hematopoietic stem cell
CD34 CD135
SCF IL-7
CD34 CD135 TdT HLA-DR
SCF IL-3 IL-4
CFU-Mast
SCF IL-4
Mast cell
SCF IL-3 IL-4
SCF GM-CSF IL-3
CFU-Bas
GM-CSF IL-3
Myeloblast
Basophilic Myelocyte
Basophil
SCF IL-3 GM-CSF
Flt3L SCF GM-CSF IL-3
CFU-EO
GM-CSF IL-3 IL-5
Eosinophilic Myelocyte
GM-CSF IL-5
Eosinophil
Flt3L SCF IL-3 GM-CSF IL-5
Flt3L SCF GM-CSF IL-4 TNF
Myeloid Related Dendritic Cell
Flt3L CSF GM-CSF IL-3 TNF
CFU-M/DC
GM-CSF M-CSF IL-3
Monoblast
Promonocyte
Monocyte
GM-CSF IL-4
GM-CSF M-CSF
Macrophage
CD11b CD14 CD33 CD115 CD123 CD126 CD13 CD15 CD64 CD116 CD124 HLA-DR
CD11b CD14 CD64 CD116 CD124 HLA-DR CD13 CD33 CD115 CD123 CD126
CD11b CD14 CD33 CD64
Flt3L SCF IL-3 GM-CSF TNF IL-4 M-SCF
HLA-DR CD116 CD123 CD33 CD124 CD126 CD64 CD115 CD13 CD11b CD14
Flt3L SCF G-CSF IL-1 IL-3 IL-6 IL-11
Myeloid Stem Cell
Flt3L SCF GM-CSF G-CSF IL-3
CFU-GEMM
CD33 CD116 CD121 IL-9R HLA-DR CD34 CD114 CD123 EPOR
GM-CSF G-CSF IL-3
CFU-GM
Flt3L SCF GM-CSF G-CSF IL-3
CFU-G
GM-CSF G-CSF
Myeloblast
Neutrophilic Myelocyte
Neutrophil
CD15 CD34 CD114 CD116 CD123 CD125 HLA-DR CD33 CD64 CD115 CD121 CD124 CD126
CD13 CD33 CD116 CD123 CD125 HLA-DR CD15 CD114 CD121 CD124 CD126
CD13 CD33 CD116 CD123 CD125 CD15 CD114 CD121 CD124 CD126
CD11b CD33 CD123 CD15 CD116 CD125
CD11b CD15 CD33
Bone marrow
Flt3L SCF G-SCF IL-3 IL-6 IL-11 IL-1 GM-CSF
Flt3L SCF IL-3 GM-CSF G-SCF
IL-9R CD34 HLA-DR CD116 CD121 CD114 CD123 CD124 CD126 CD33 CD13 CD125 CD11b
Flt3L SCF GM-CSF IL-3 IL-4
BFU-E
SCF GM-CSF IL-3 IL-4 EPO
CFU-E
TPO EPO
Proerythroblast
EPO
Erythrocyte
CD33 CD117 EPOR CD34 CD123 HLA-DR
CD36 CD235a
CD235a
CD35 CD55 CD235a CD44 CD59
Flt3L SCF GM-CSF IL-3 IL-4 EPO TPO
HLA-DR EPOR CD33 CD34 CD117 CD123 CD36 CD235a CD35 CD44 CD55 CD59
Flt3L SCF GM-CSF IL-3 IL-6 IL-11 TPO
BFU-MK
Flt3L SCF GM-CSF Meg-CSF IL-3 IL-6 IL-11 TPO
CFU-MK
SCF GM-CSF IL-3 IL-6 IL-11 TPO
Mega-karyocyte
IL-6 IL-11 TPO
Platelets
CD33 CD116 CD126 HLA-DR CD34 CD123 IL-11R
CD36 CD61 CD116 CD122 CD126
CD9 CD36 CD42 CD116 CD126 CD14 CD41 CD61 CD123
CD9 CD36 CD42 CD61 CD14 CD41 CD49 CD126
Flt3L SCF IL-3 IL-6 IL-11 GM-CSF Meg-CSF TPO
HLA-DR CD33 CD34 IL-11R CD116 CD123 CD126 CD61 CD9 CD14 CD36 CD41 CD42 CD49
Data on KEGG graph
Rendered by Pathview


-1
0
1
JAK-STAT SIGNALING PATHWAY
ECS complex
Ubiquitin mediated proteolysis
Cytokine-cytokine receptor interaction
STAM
Cytokine
Hormone
GF
Receptor
JAK
+p
-p
-p
TC-PTP
SHP1
STAT
STAT
STAT
STAT dimerization
IRF9
TC-PTP
PIAS
-p
CBP/P300
SLIM
+u
Proteasome
DNA
CIS
SOCS
Bcl-2
MCL1
Bcl-XL
PIM1
c-Myc
CycD
p21
AOX
GFAP
Anti-apoptosis
Apoptosis
Cell-cycle progression
Cell-cycle inhibition
Cell cycle
Lipid metabolism
Differentiation
+p
SHP2
GRB
SOS
Ras
Raf
MAPK signaling pathway
Proliferation
Differentiation
+p
PI3K
AKT
mTOR
PI3K-AKT signaling pathway
Cell cycle
Cell survival
Data on KEGG graph
Rendered by Pathview


STEROID HORMONE BIOSYNTHESIS
-1
0
1
Steroid biosynthesis
3.1.6.2
2.8.2.2
Cholesterol sulfate
Cholesterol
1.14.15.6
1.14.15.6
20α-Hydroxy-cholesterol
22β-Hydroxy-cholesterol
1.14.15.6
1.14.15.6
20α,22β-Dihydroxy-cholesterol
1.14.15.6
1.14.14.19
4-Methylpentanal
21-Hydroxy-pregnenolone
HSD3B
11-Deoxy-corticosterone
1.14.15.4
Corticosterone
1.14.14.16
1.14.14.16
1.14.14.16
1.14.14.29
7α-Hydroxy-pregnenolone
11α-Hydroxy-progesterone
1.14.99.14
20α-Hydroxy-progesterone
1.1.1.149
11β-Hydroxy-progesterone
Pregnenolone
HSD3B
Progesterone
1.14.15.4
3.1.6.2
2.8.2.2
Pregnenolone-sulfate
1.14.14.19
1.14.14.19
1.14.14.19
17α,20α-Dihydroxy-cholesterol
1.14.15.6
17α-Hydroxy-pregnenolone
HSD3B
17α-Hydroxy-progesterone
1.14.15.4
21-Deoxycortisol
1.1.1.149
17α,20α-Dihydroxy-pregn-4-en-3-one
1.14.14.16
1.14.14.16
1.14.14.16
17α,21-Dihydroxy-pregnenolone
HSD3B
11-Deoxycortisol
1.14.15.4
Cortisol
1.14.15.4
11β,17α,21-Trihydroxy-pregnenolone
HSD3B
1.14.14.32
1.14.14.32
1.1.1.146
HSD11B2
Cortisone
5α-Dihydro-deoxycorticosterone
1.3.1.22
1.1.1.213
Allotetrahydro-deoxycorticosterone
Aldosterone hemiacetal
18-Hydroxy-corticosterone
CYP11B2
Aldosterone
1.3.1.3
11β,21-Dihydroxy-3,20-oxo-5β-pregnan-18-al
1.1.1.50
3α,11β,21-Trihydroxy-20-oxo-5β-pregnan-18-al
1.14.15.5
1.1.1.146
11-Dehydro-corticosterone
21-Hydroxy-5β-pregnane-3,11,20-trione
1.1.1.50
HSD11B2
1.1.1.50
1.1.1.146
1.1.1.53
3α,20α,21-Trihydroxy-5β-pregnane-11-one
11β,21-Dihydroxy-5β-pregnane-3,20-dione
Tetrahydro-corticosterone
3α,21-Dihydroxy-5β-pregnane-11,20-dione
1.3.1.3
1.3.99.6
5β-Pregnane-3,20-dione
1.1.1.50
3α-Hydroxy-5β-pregnane-20-one
1.1.1.53
Pregnanediol
C21-Steroids
1.3.1.22
5α-Pregnane-3,20-dione
1.1.1.213
3α-Hydroxy-5α-pregnan-20-one
1.1.1.149
5α-Pregnane-3α,20α-diol
5α-Pregnan-20α-ol-3-one
1.1.1.149
1.1.1.213
4-Androsten-11beta-ol-3,17-dione
1.1.1.62
11β-Hydroxytestosterone
1.3.1.3
1.1.1.50
Urocortisol
1.1.1.53
Cortol
11β,17α,21-Trihydroxy-5β-pregnane-3,20-dione
17α,21-Dihydroxy-5β-pregnane-3,11,20-trione
1.3.1.3
1.1.1.50
1.1.1.53
Cortolone
Urocortisone
Dehydroepiandro-sterone
Dehydroepiandro-steron sulfate
2.8.2.2
3.1.6.2
1.14.14.23
7α-Hydroxydehydro-epiandrosterone
11β-Hydroxyandrost-4-ene-3,17-dione
1.1.1.146
Adrenosterone
1.14.15.4
1.14.14.1
1.14.14.-
16α-Hydroxyandrost-4-ene-3,17-dione
1.1.1.51
3β,17β-Dihydroxy-androst-5-ene
HSD3B
16α-Hydroxydehydro-epiandrosterone
1.3.1.3
1.1.1.50
1.1.1.152
2.4.1.17
5β-Androstane-3,17-dione
Etiocholan-3α-ol-17-one
Etiocholan-3α-ol-17-one 3-glucuronide
1.3.1.22
1.1.1.50
2.4.1.17
5α-Androstane-3,17-dione
Androsterone
Androsterone-glucuronide
Androst-4-ene-3,17-dione
HSD3B
1.14.14.14
1.14.14.14
1.14.14.14
19-Hydroxyandrost-4-ene-3,17-dione
19-Oxoandrost-4-ene-3,17-dione
Estrone
7α-Hydroxy-androstenedione
1.1.1.51
1.1.1.64
1.1.1.239
1.14.99.12
3-Oxo-13,17-secoandrost-4-ene-17,13α-lactone
7α-Hydroxy-testosterone
HSD3B
Testosterone
1.14.14.14
19-Hydroxy-testosterone
1.14.14.14
19-Oxotestosterone
1.14.14.14
1.1.1.51
1.1.1.62
Estradiol-17β
C19-Steroids
1.3.1.22
5α-Dihydro-testosterone
1.1.1.50
Androstan-3alpha,17beta-diol
1.3.1.3
5β-Dihydro-testosterone
2.4.1.17
Testosterone glucuronide
Estrone 3-sulfate
2.8.2.4
2.8.2.15
3.1.6.1
2.4.1.17
Estrone glucuronide
1.1.1.148
Estradiol-17α
1.14.14.1
2.1.1.6
2-Hydroxyestrone
2.4.1.17
2-Methoxyestrone-3-glucuronide
2-Methoxyestrone
2.8.2.15
2-Methoxyestrone-3-sulfate
1.14.14.1
1.14.14.-
16-α-Hydroxyestrone
1.1.1.62
C18-Steroids
1.14.14.1
Estriol
2.4.1.17
16-Glucuronide-estriol
1.14.14.1
2.1.1.6
2-Hydroxy-estradiol-17β
2.4.1.17
2-Methoxy-estradiol-17β-3-glucuronide
2-Methoxy-estradiol-17β
2.8.2.15
2-Methoxy-estradiol-17β-3-sulfate
1.14.99.11
6β-Hydroxy-estradiol-17β
2.4.1.17
Estradiol-17β-3-glucuronide
2.8.2.15
Estradiol-17β-3-sulfate
Data on KEGG graph
Rendered by Pathview



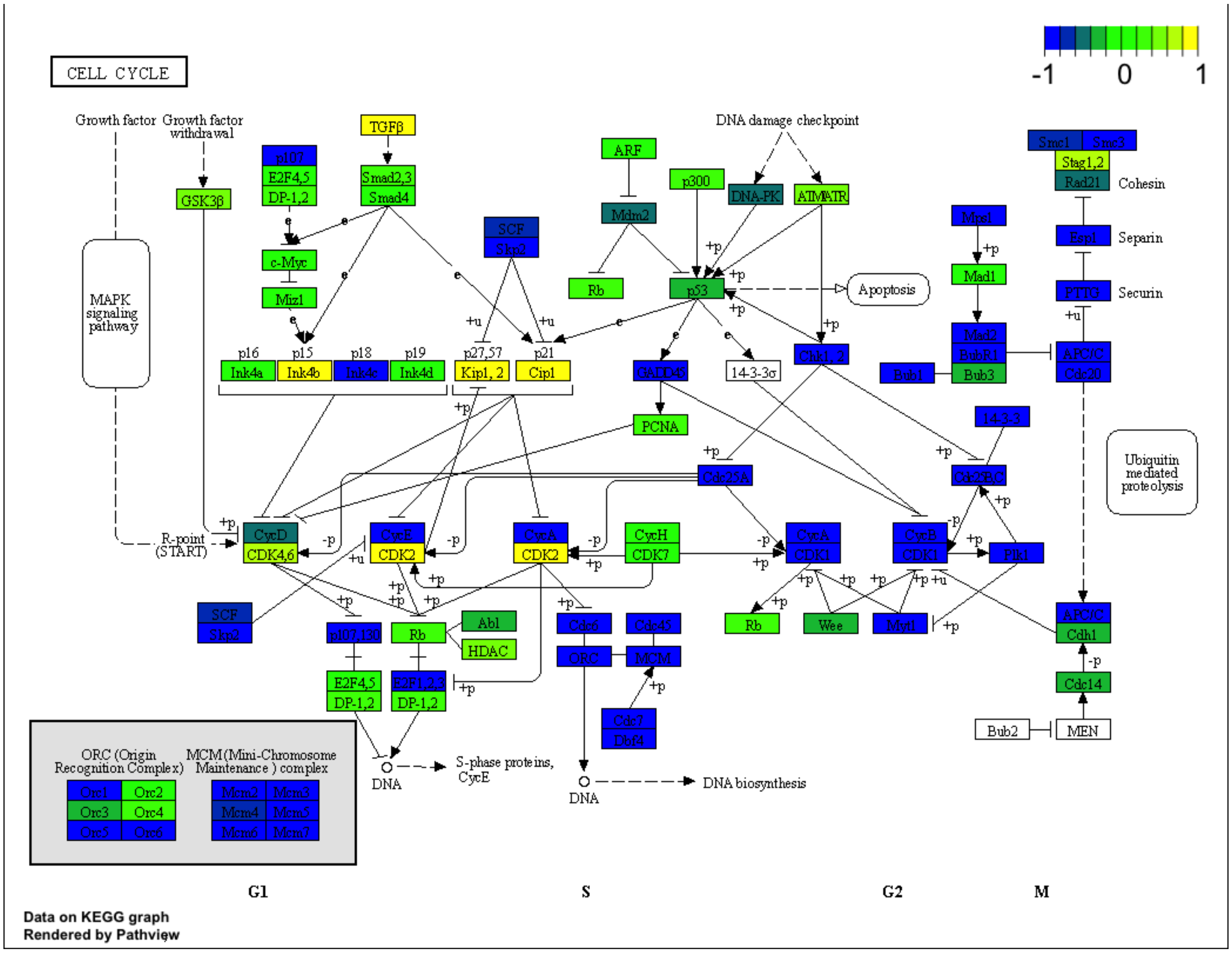


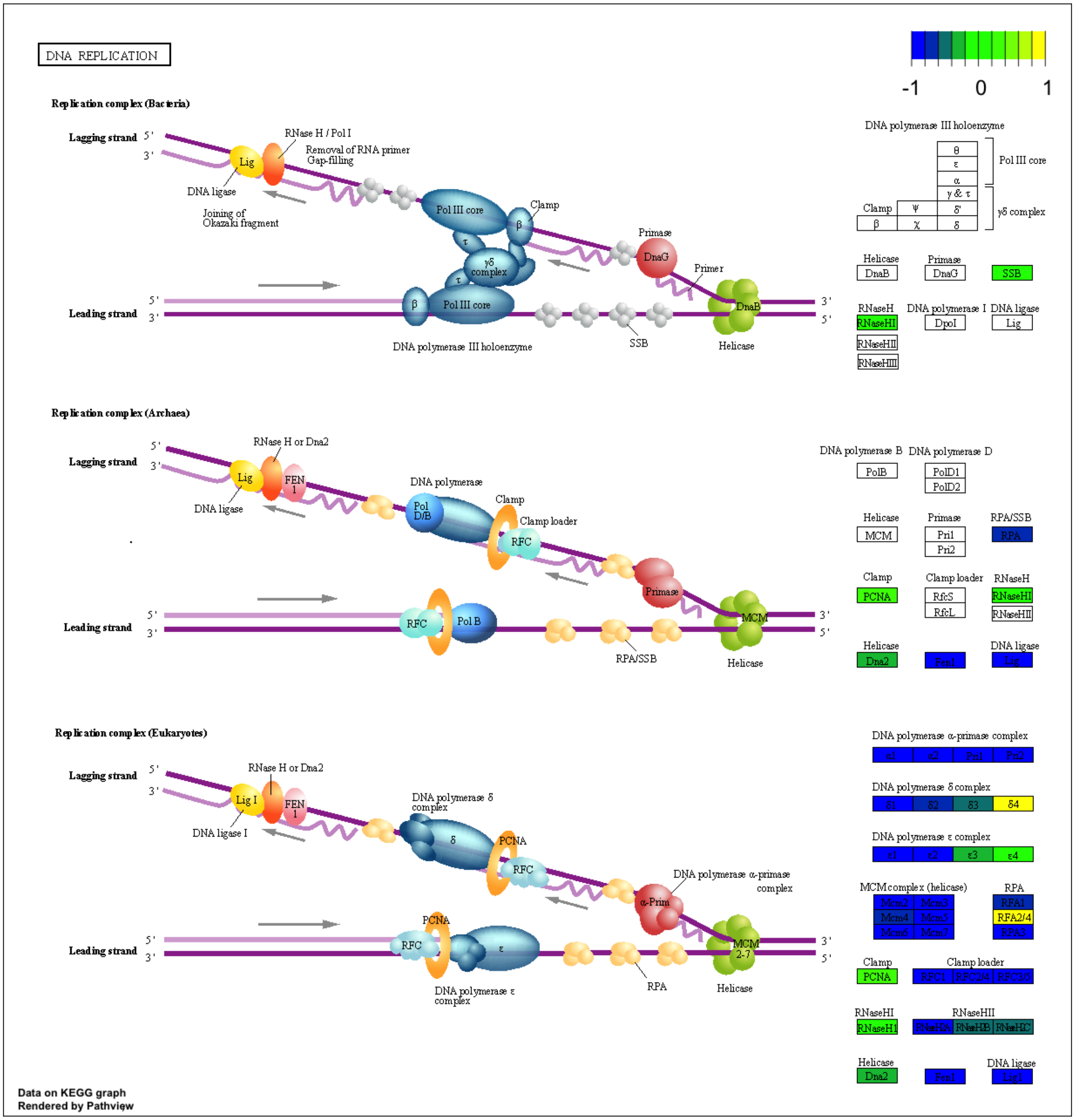
Q. Can you do the same procedure as above to plot the pathview figures for the top 5 down-reguled pathways?

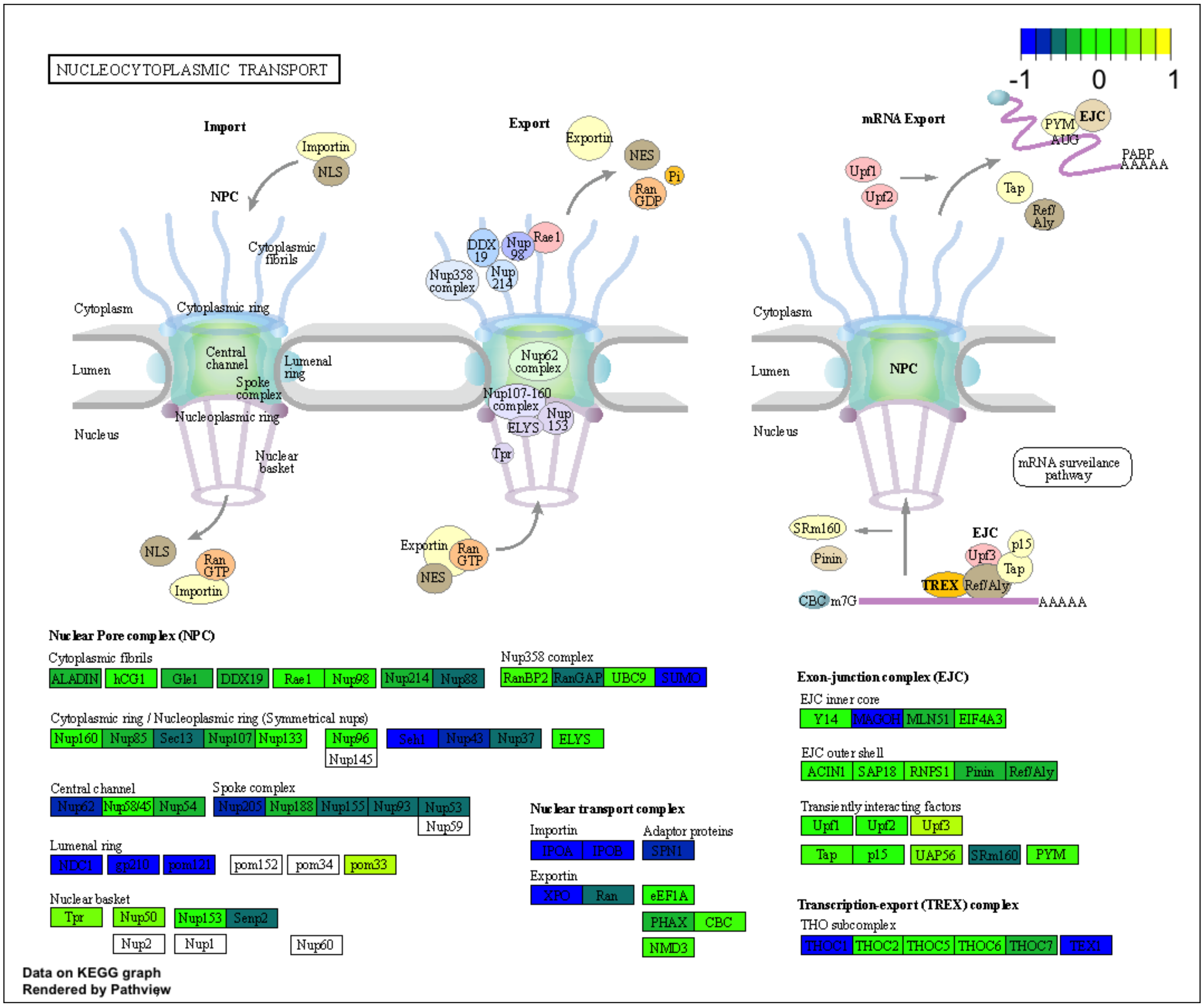
```r
## Focus on top 5 downregulated pathways here for demo purposes only
keggrespathways.down <- rownames(keggres$less)[1:5]

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

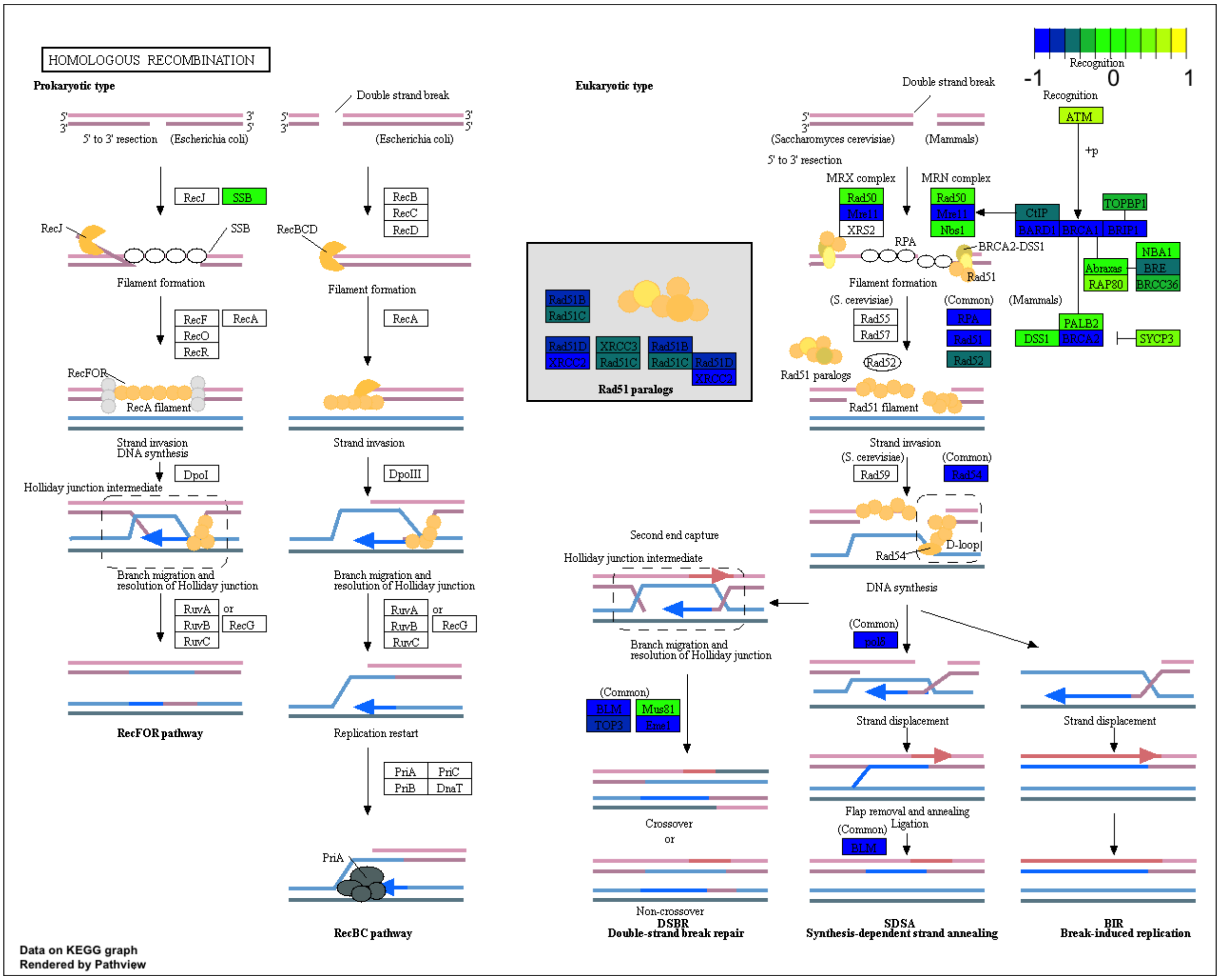
```
## [1] "hsa04110" "hsa03030" "hsa03013" "hsa03440" "hsa04114"
```

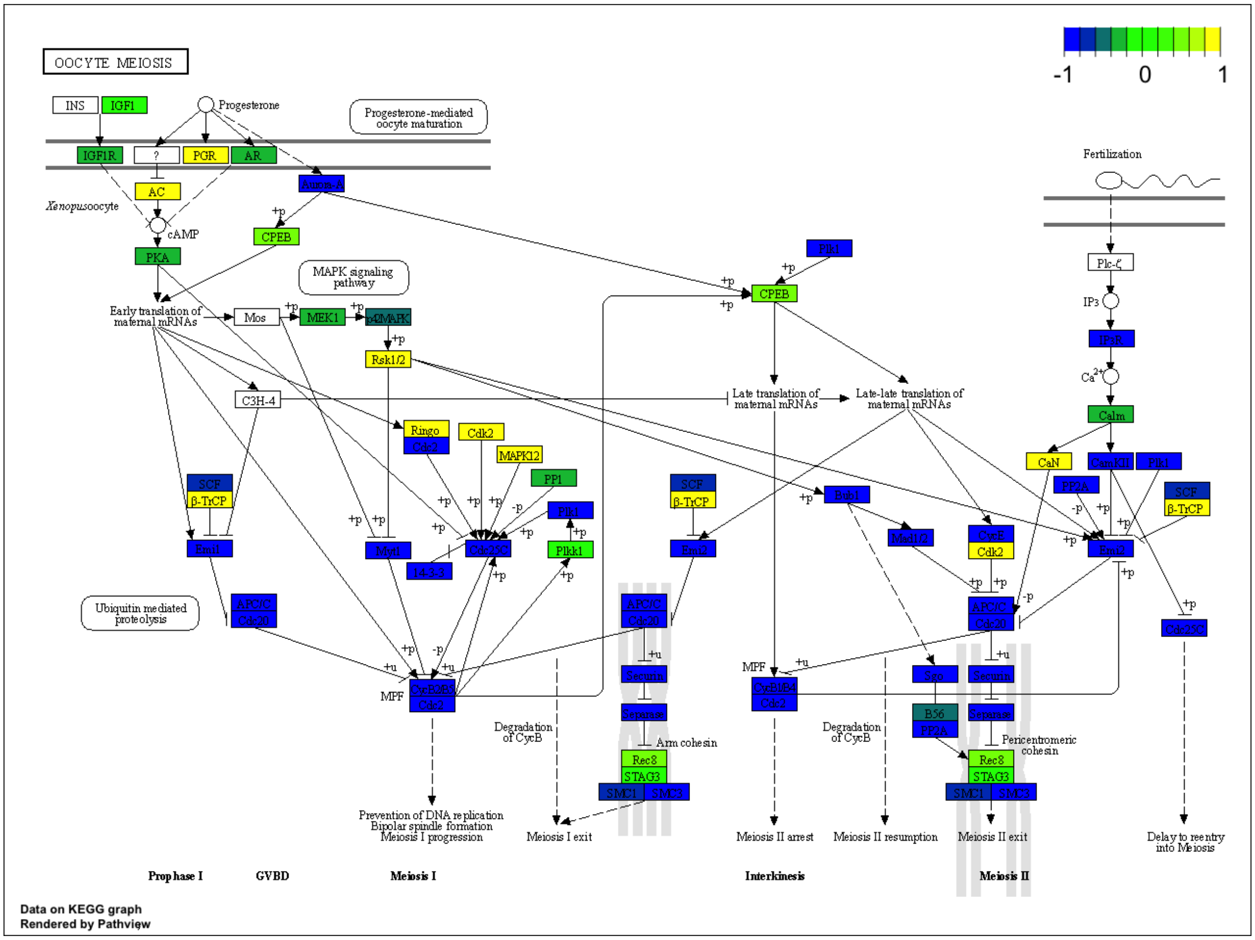
```r
pathview(gene.data=foldchanges, pathway.id=keggresids.down, species="hsa", low="blue", mid="green", hig
```




DNA REPLICATION
-1
0
1
Replication complex (Bacteria)
Lagging strand
5'
3'
Lig
RNase H / Pol I
Removal of RNA primer
Gap-filling
DNA ligase
Joining of
Okazaki fragment
Pol III core
Clamp
β
τ
γδ
complex
τ
Primase
DnaG
Primer
5'
3'
Leading strand
β
Pol III core
DnaB
3'
5'
DNA polymerase III holoenzyme
SSB
Helicase
DNA polymerase III holoenzyme
θ
ε
α
γ & τ
Pol III core
Clamp
ψ
δ'
β
χ
δ
γδ complex
Helicase
DnaB
Primase
DnaG
SSB
RNaseH
RNaseHI
RNaseHII
RNaseHIII
DNA polymerase I
DpoI
DNA ligase
Lig
Replication complex (Archaea)
5'
Lagging strand
3'
RNase H or Dna2
Lig
FEN 1
DNA ligase
DNA polymerase
Pol D/B
Clamp
Clamp loader
RFC
Primase
5'
Leading strand
3'
RFC
Pol B
MCM
3'
5'
RPA/SSB
Helicase
DNA polymerase B
PolB
DNA polymerase D
PolD1
PolD2
Helicase
MCM
Primase
Pri1
Pri2
RPA/SSB
RPA
Clamp
PCNA
Clamp loader
RfcS
RfcL
RNaseH
RNaseHI
RNaseHII
Helicase
Dna2
Fen1
DNA ligase
Lig
Replication complex (Eukaryotes)
Lagging strand
5'
3'
RNase H or Dna2
Lig I
FEN 1
DNA ligase I
DNA polymerase δ
complex
δ
PCNA
RFC
DNA polymerase α-primase
complex
α-Prim
PCNA
5'
Leading strand
3'
RFC
ε
MCM
2-7
3'
5'
DNA polymerase ε
complex
RPA
Helicase
DNA polymerase α-primase complex
α1
α2
Pri1
Pri2
DNA polymerase δ complex
δ1
δ2
δ3
δ4
DNA polymerase ε complex
ε1
ε2
ε3
ε4
MCM complex (helicase)
Mcm2
Mcm3
Mcm4
Mcm5
Mcm6
Mcm7
RPA
RFA1
RFA2/4
RPA3
Clamp
PCNA
Clamp loader
RFC1
RFC2/4
RFC3/5
RNaseHI
RNaseH1
RNaseHII
RNaseH2A
RNaseH2B
RNaseH2C
Helicase
Dna2
Fen1
DNA ligase
Lig1
Data on KEGG graph
Rendered by Pathview


NUCLEOCYTOPLASMIC TRANSPORT
-1
0
1
Import
Importin
NLS
NPC
Cytoplasmic fibrils
Cytoplasm
Cytoplasmic ring
Lumen
Central channel
Lumenal ring
Spoke complex
Nucleoplasmic ring
Nucleus
Nuclear basket
NLS
Ran GTP
Importin
Export
Exportin
NES
Pi
Ran GDP
DDX 19
Nup 98
Rae1
Nup358 complex
Nup 214
Nup62 complex
Nup107-160 complex
Nup 153
ELYS
Tpr
Exportin
Ran GTP
NES
mRNA Export
PYM
EJC
AUG
PABP
AAAAA
Upf1
Upf2
Tap
Ref/ Aly
Cytoplasm
Lumen
NPC
Nucleus
mRNA surveilance pathway
SRm160
Pinin
EJC
p15
Upf3
Tap
TREX
Ref/Aly
CBC
m7G
AAAAA
Nuclear Pore complex (NPC)
Cytoplasmic fibrils
ALADIN
hCG1
Gle1
DDX19
Rae1
Nup98
Nup214
Nup88
Nup358 complex
RanBP2
RanGAP
UBC9
SUMO
Cytoplasmic ring / Nucleoplasmic ring (Symmetrical nups)
Nup160
Nup85
Sec13
Nup107
Nup133
Nup96
Nup145
Seh1
Nup43
Nup37
ELYS
Central channel
Nup62
Nup58/45
Nup54
Spoke complex
Nup205
Nup188
Nup155
Nup93
Nup53
Nup59
Lumenal ring
NDC1
gp210
pom121
pom152
pom34
pom33
Nuclear basket
Tpr
Nup50
Nup153
Senp2
Nup2
Nup1
Nup60
Nuclear transport complex
Importin
IPOA
IPOB
Adaptor proteins
SPN1
Exportin
XPO
Ran
eEF1A
PHAX
CBC
NMD3
Exon-junction complex (EJC)
EJC inner core
Y14
MAGOH
MLN51
EIF4A3
EJC outer shell
ACIN1
SAP18
RNPS1
Pinin
Ref/Aly
Transiently interacting factors
Upf1
Upf2
Upf3
Tap
p15
UAP56
SRm160
PYM
Transcription-export (TREX) complex
THO subcomplex
THOC1
THOC2
THOC5
THOC6
THOC7
TEX1
Data on KEGG graph
Rendered by Pathview


HOMOLOGOUS RECOMBINATION
Prokaryotic type
5'
3'
3'
5'
5' to 3' resection
(Escherichia coli)
RecJ
SSB
RecJ
SSB
Filament formation
RecF
RecO
RecR
RecA
RecFOR
RecA filament
Strand invasion
DNA synthesis
DpoI
Holliday junction intermediate
Branch migration and
resolution of Holliday junction
RuvA
RuvB
RuvC
or
RecG
RecFOR pathway
Double strand break
5'
3'
3'
5'
(Escherichia coli)
RecB
RecC
RecD
RecBCD
Filament formation
RecA
Strand invasion
DpoIII
Branch migration and
resolution of Holliday junction
RuvA
RuvB
RuvC
or
RecG
Replication restart
PriA
PriC
PriB
DnaT
PriA
RecBC pathway
Eukaryotic type
Rad51B
Rad51C
Rad51D
XRCC2
XRCC3
Rad51C
Rad51B
Rad51C
Rad51D
XRCC2
Rad51 paralogs
Double strand break
5'
3'
(Saccharomyces cerevisiae)
(Mammals)
5' to 3' resection
MRX complex
Rad50
Mre11
XRS2
MRN complex
Rad50
Mre11
Nbs1
RPA
BRCA2-DSS1
Rad51
Filament formation
(S. cerevisiae)
Rad55
Rad57
(Common)
RPA
Rad51
Rad52
Rad52
Rad51 paralogs
Rad51 filament
Strand invasion
(S. cerevisiae)
Rad59
(Common)
Rad54
Rad54
D-loop
DNA synthesis
(Common)
polδ
-1
0
1
Recognition
Recognition
ATM
+p
CtIP
TOPBP1
BARD1
BRCA1
BRIP1
NBA1
Abraxas
BRE
RAP80
BRCC36
(Mammals)
PALB2
DSS1
BRCA2
SYCP3
Second end capture
Holliday junction intermediate
Branch migration and
resolution of Holliday junction
(Common)
BLM
TOP3
Mus81
Eme1
Crossover
or
Non-crossover
DSBR
Double-strand break repair
Strand displacement
Flap removal and annealing
Ligation
(Common)
BLM
SDSA
Synthesis-dependent strand annealing
Strand displacement
BIR
Break-induced replication
Data on KEGG graph
Rendered by Pathview


OOCYTE MEIOSIS
-1
0
1
INS
IGF1
Progesterone
Progesterone-mediated oocyte maturation
IGF1R
?
PGR
AR
AC
Aurora-A
Xenopus oocyte
cAMP
CPEB
PKA
MAPK signaling pathway
Early translation of maternal mRNAs
Mos
MEK1
p42MAPK
Rsk1/2
C3H-4
Ringo
Cdc2
Cdk2
MAPK12
PP1
SCF
β-TrCP
Plk1
Emi1
Myt1
Cdc25C
Plkk1
14-3-3
Ubiquitin mediated proteolysis
APC/C
Cdc20
MPF
CycB2/B5
Cdc2
Prevention of DNA replication
Bipolar spindle formation
Meiosis I progression
Degradation of CycB
Meiosis I exit
Securin
Separase
Arm cohesin
Rec8
STAG3
SMC1
SMC3
Plk1
CPEB
Late translation of maternal mRNAs
Late-late translation of maternal mRNAs
SCF
β-TrCP
Emi2
Bub1
Mad1/2
CycE
Cdk2
APC/C
Cdc20
MPF
CycB1/B4
Cdc2
Degradation of CycB
Sgo
B56
PP2A
Securin
Separase
Pericentromeric cohesin
Rec8
STAG3
SMC1
SMC3
Meiosis II arrest
Meiosis II resumption
Meiosis II exit
Fertilization
Plc-ζ
IP3
IP3R
Ca2+
Calm
CaN
CamKII
Plk1
PP2A
SCF
β-TrCP
Emi2
Cdc25C
Delay to reentry into Meiosis
Prophase I
GVBD
Meiosis I
Interkinesis
Meiosis II
Data on KEGG graph
Rendered by Pathview