Class 13: RNASeq mini project

Henry (A16354124)

Background

The authors report on differential analysis of lung fibroblasts in response to loss of the developmental transcription factor HOXA1. Their results and others indicate that HOXA1 is required for lung fibroblast and HeLa cell cycle progression. In particular their analysis show that "loss of HOXA1 results in significant expression level changes in thousands of individual transcripts, along with isoform switching events in key regulators of the cell cycle". For our session we have used their Sailfish gene-level estimated counts and hence are restricted to protein-coding genes only.

Data Import

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

# Import metadata and take a peak
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
```

```r
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
```

```
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
```

```
# Filter count data where you have 0 read count across all samples.
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
```

Setup for DESeq

```
library("DESeq2")
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
```

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

Run DESeq

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

```
using pre-existing size factors
```

```
estimating dispersions
```

```
found already estimated dispersions, replacing these
```

```
gene-wise dispersion estimates
```

```
mean-dispersion relationship
```

```
final dispersion estimates
```

```
fitting model and testing
```

Get results

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

```r
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
```

Add annotation

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

```r
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"
```

```r
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
```

```r
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
```

```r
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
```

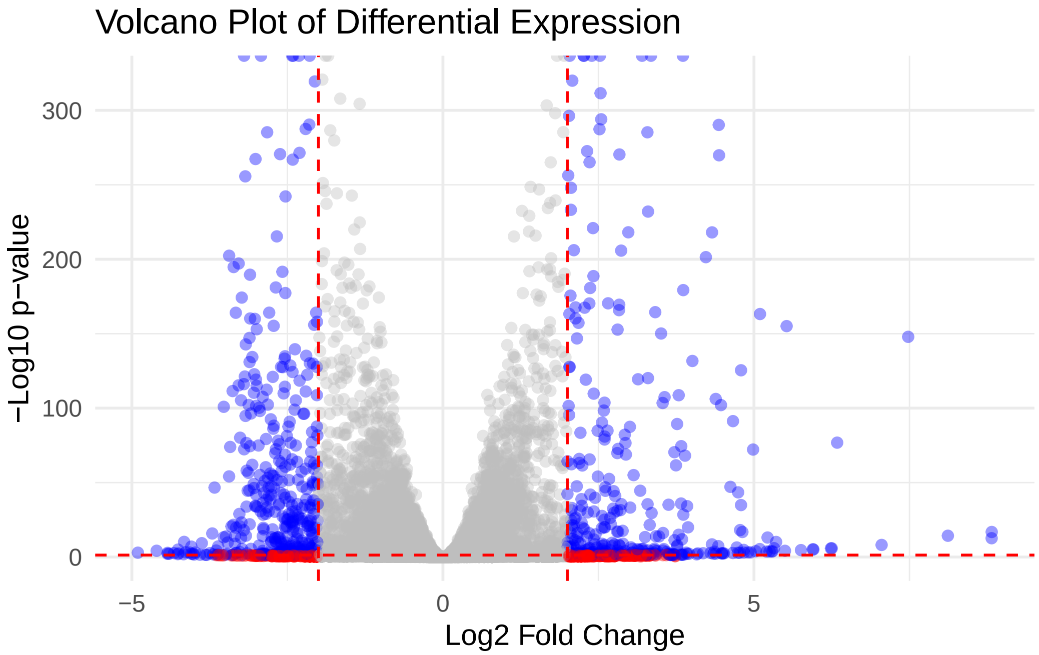
```r
head(res, 10)
```

```
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 0.3248215   0.551863 5.81042e-01
ENSG00000187634  183.229650      0.4264571 0.1402658   3.040350 2.36304e-03
ENSG00000188976 1651.188076     -0.6927205 0.0548465 -12.630156 1.43993e-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 1.51281e-25
ENSG00000187608  350.716868      0.2573837 0.1027266   2.505522 1.22271e-02
ENSG00000188157 9128.439422      0.3899088 0.0467164   8.346302 7.04333e-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 1.76553e-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 4.21970e-16        AGRN      375790                  agrin
ENSG00000237330          NA      RNF223      401934 ring finger protein ..
```

Visualize results

```r
library(ggplot2)

mycols <- rep("gray", nrow(res) )
mycols[ abs(res$log2FoldChange) > 2 ] <- "red"
inds <- which(res$pvalue < 0.05 & abs(res$log2FoldChange) > 2)
mycols[inds] <- "blue"

ggplot(res) +
  aes(x=log2FoldChange, y=-log10(pvalue)) +
  geom_point(alpha=0.4, color = mycols) +
  theme_minimal() +
  xlab("Log2 Fold Change") +
  ylab("-Log10 p-value") +
  geom_vline(xintercept=c(-2, 2), col="red", linetype="dashed") +
  geom_hline(yintercept=-log10(0.05), col="red", linetype="dashed") +
  ggtitle("Volcano Plot of Differential Expression")
```



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).
##############################################################################
```

```r
data(kegg.sets.hs)
data(sigmet.idx.hs)
kegg.sets.hs = kegg.sets.hs[sigmet.idx.hs]
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"  
```

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

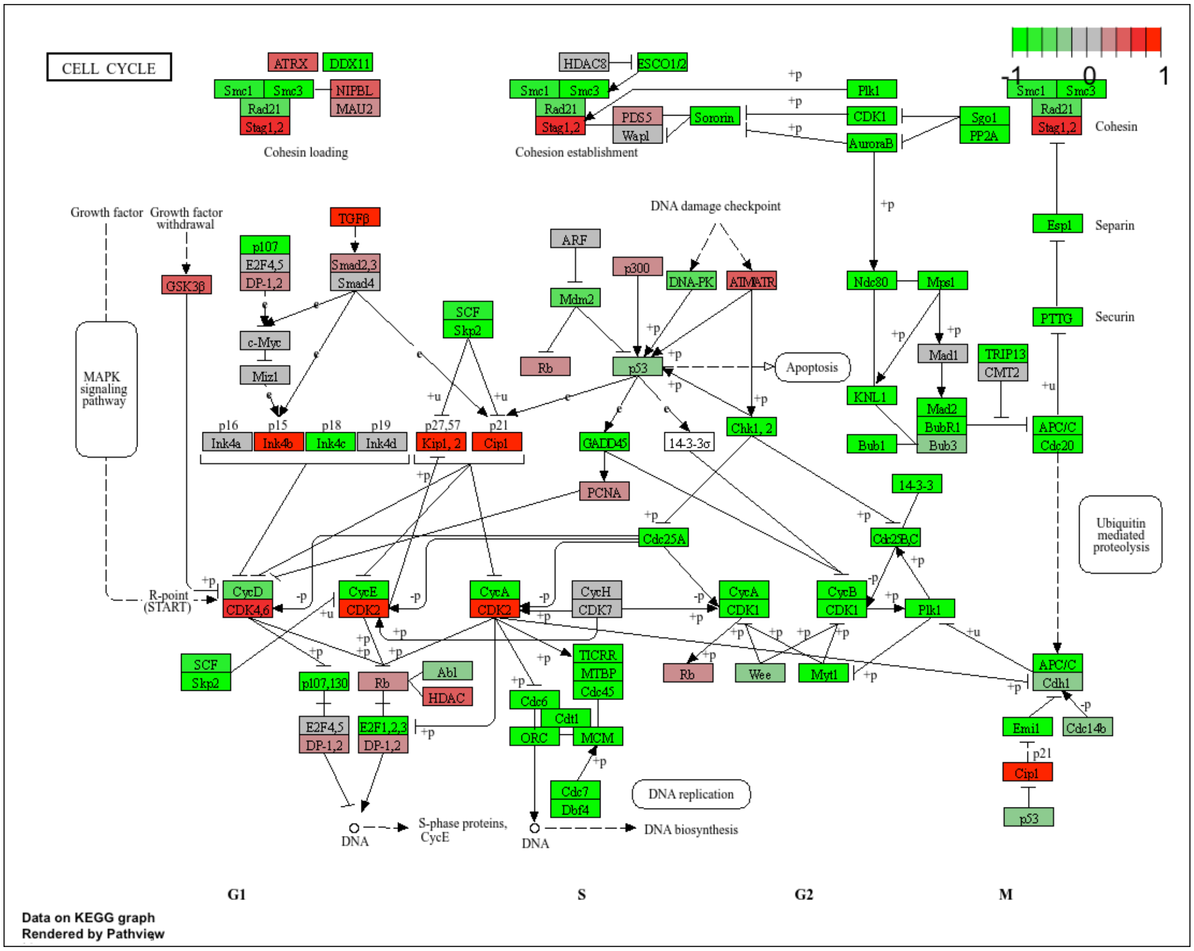
```
       <NA>      148398       26155      339451       84069       84808
 0.17925708  0.42645712 -0.69272046  0.72975561  0.04057653  0.54281049
```

```r
keggres = gage(foldchanges, gsets=kegg.sets.hs)
pathview(gene.data=foldchanges, pathway.id="hsa04110")
```

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

```
Info: Working in directory /Users/hix/Documents/BGGN213/class13
```

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



Another version

```
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 /Users/hix/Documents/BGGN213/class13
```

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

```
## 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"
```

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

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

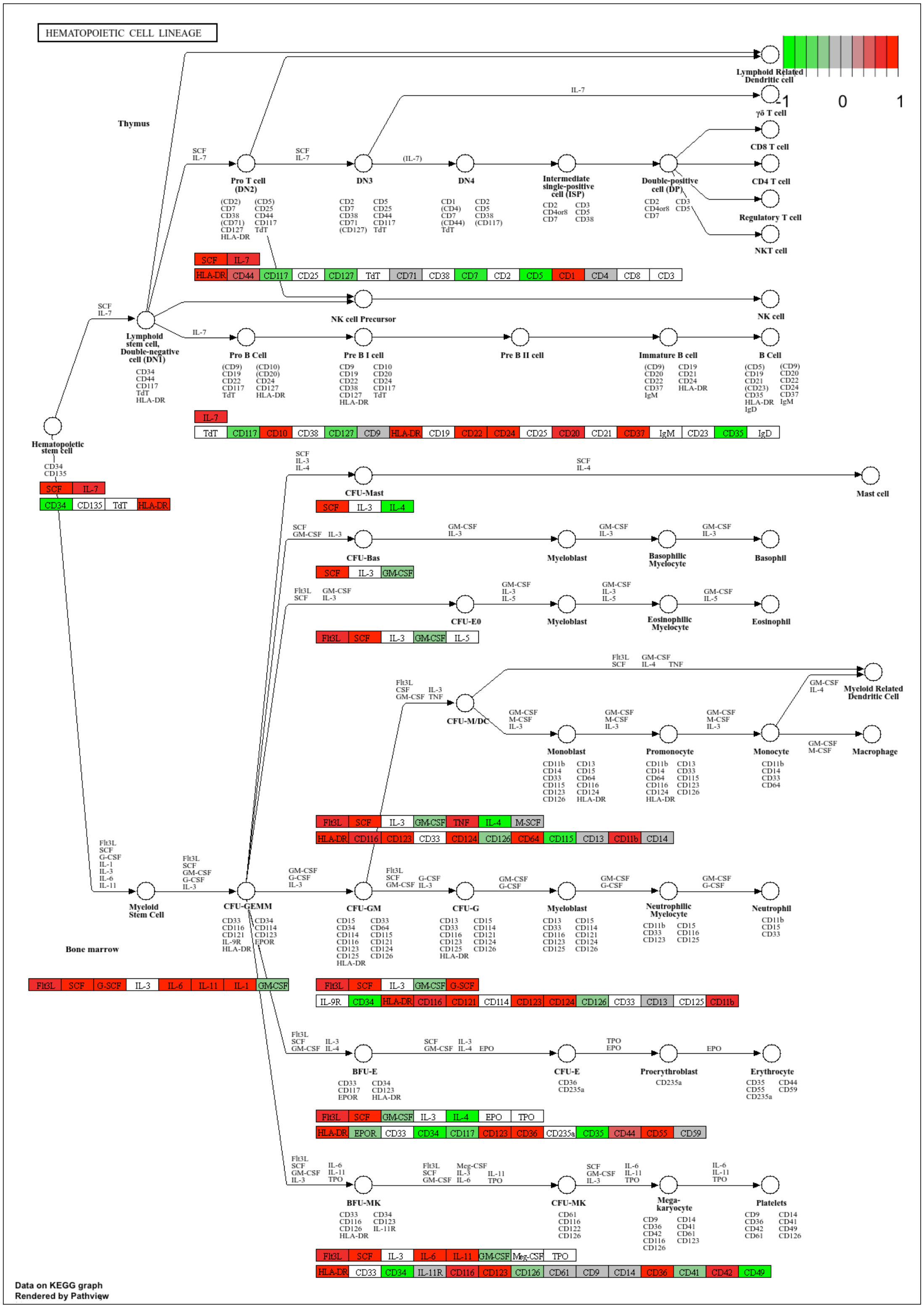
```
Info: Working in directory /Users/hix/Documents/BGGN213/class13
```

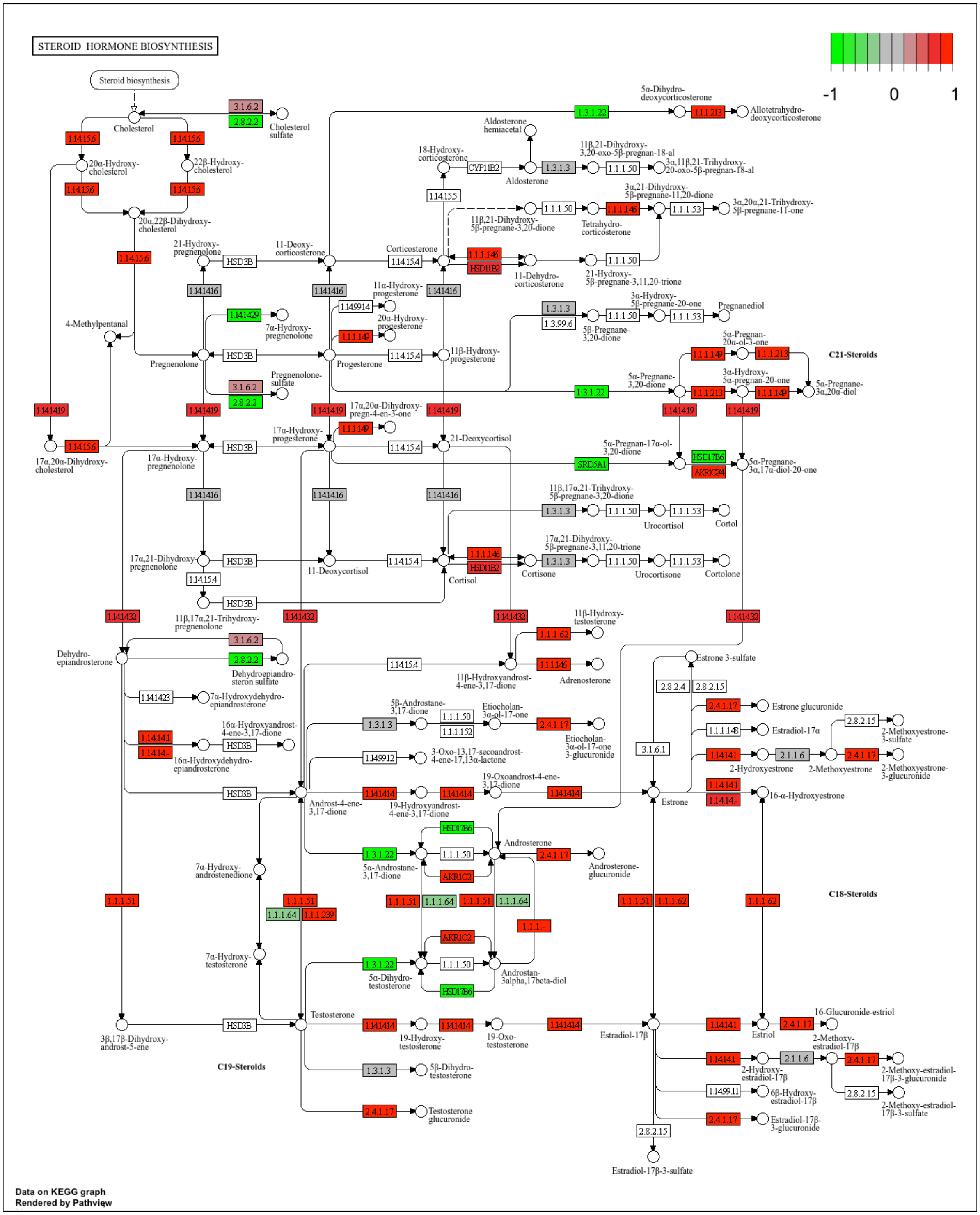
```
Info: Writing image file hsa04640.pathview.png
```

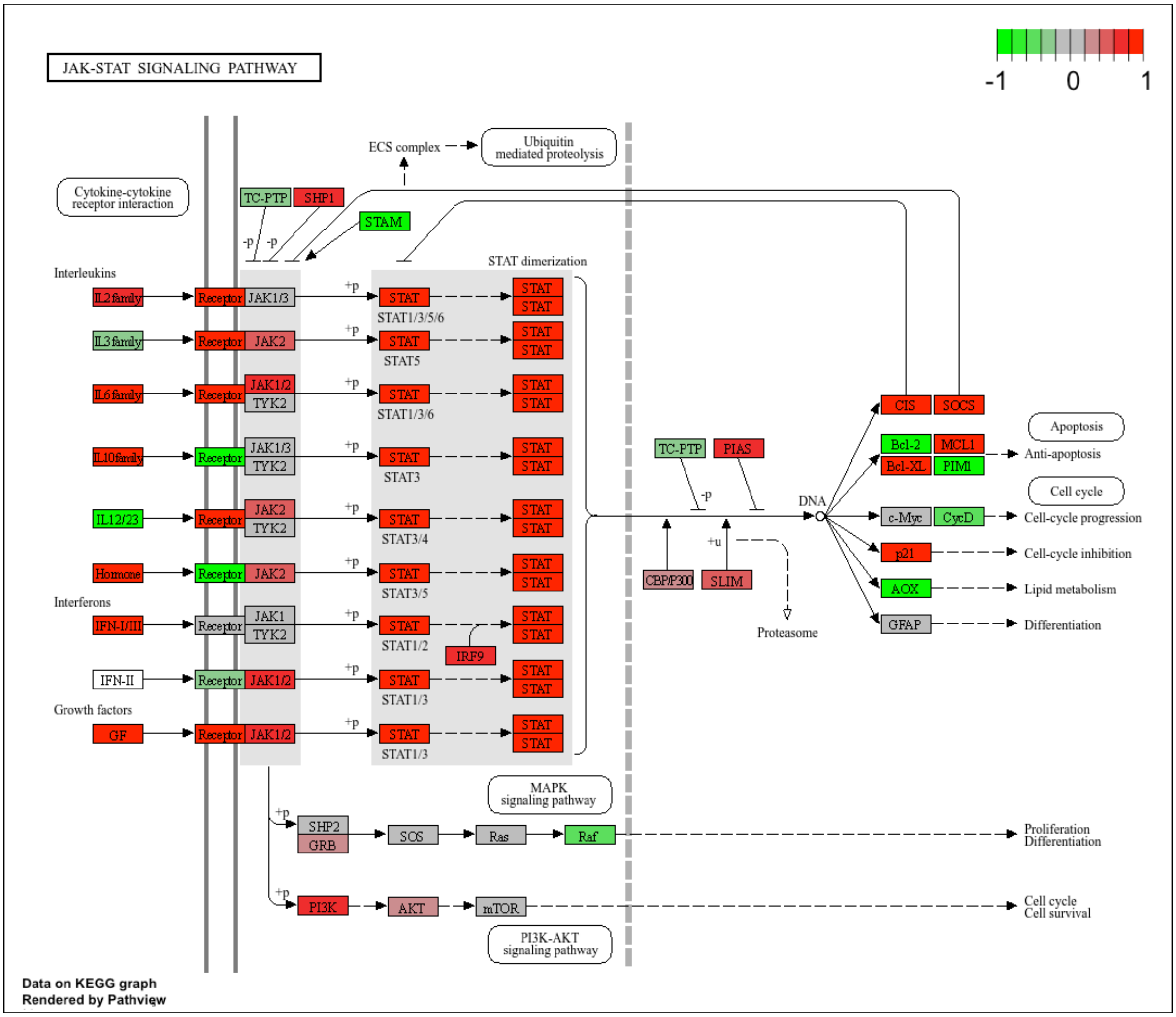
```
'select()' returned 1:1 mapping between keys and columns
```

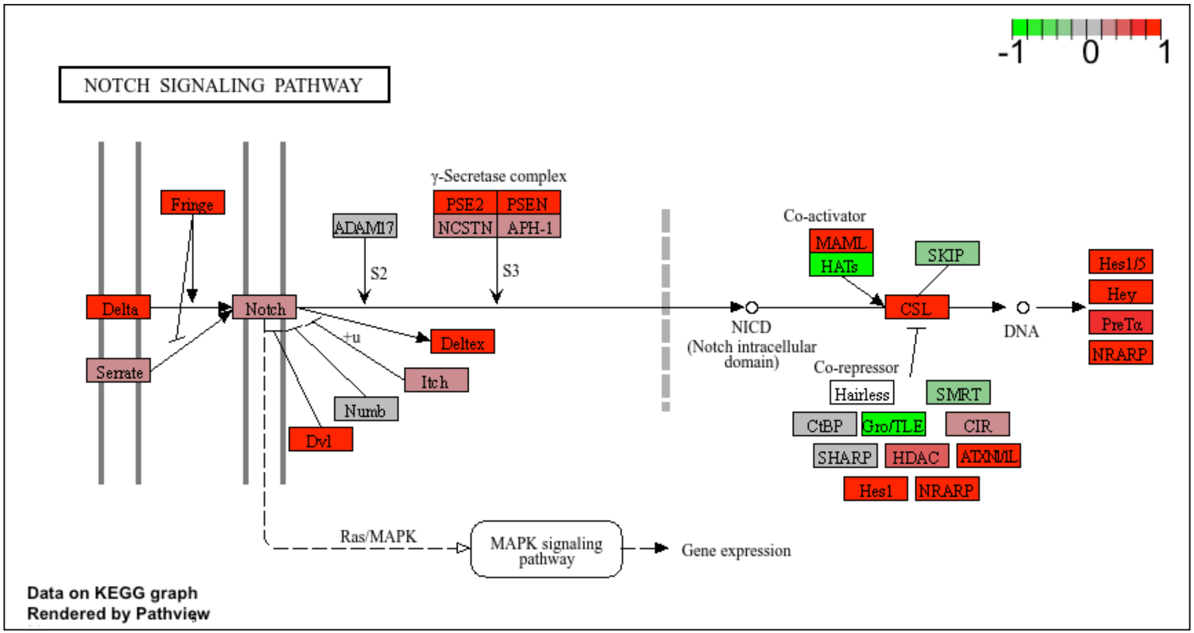
```
Info: Working in directory /Users/hix/Documents/BGGN213/class13
```

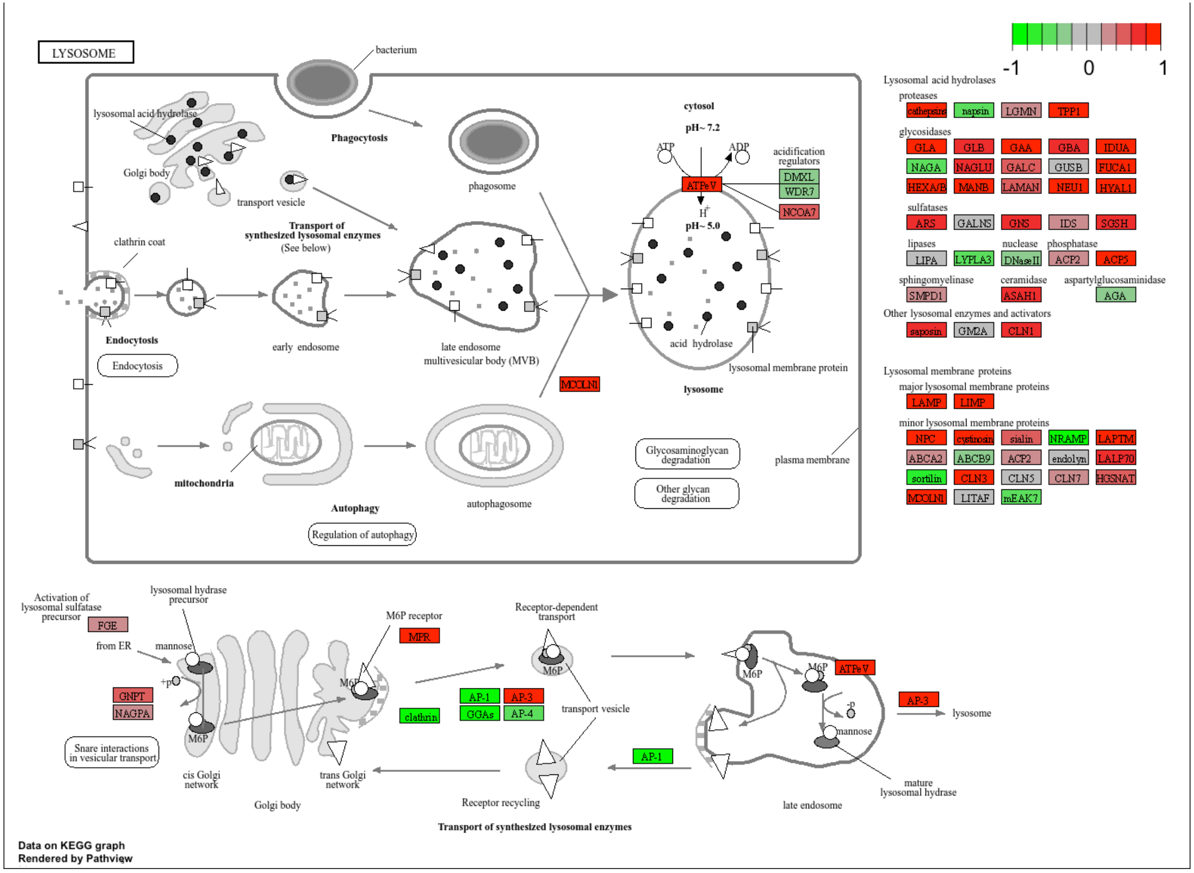
```
Info: Writing image file hsa04630.pathview.png

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

Info: Working in directory /Users/hix/Documents/BGGN213/class13

Info: Writing image file hsa00140.pathview.png

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

Info: Working in directory /Users/hix/Documents/BGGN213/class13

Info: Writing image file hsa04142.pathview.png

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

Info: Working in directory /Users/hix/Documents/BGGN213/class13

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


HEMATOPOIETIC CELL LINEAGE
-1
0
1
Thymus
Lymphoid Related Dendritic cell
IL-7
γδ T cell
CD8 T cell
CD4 T cell
Regulatory T cell
NKT cell
SCF
IL-7
Pro T cell (DN2)
(CD2) CD7 CD38 (CD71) CD127 HLA-DR
(CD5) CD25 CD44 CD117 TdT
DN3
CD2 CD7 CD38 CD71 (CD127)
CD5 CD25 CD44 CD117 TdT
(IL-7)
DN4
CD1 (CD4) CD7 (CD44) TdT
CD2 CD5 CD38 (CD117)
Intermediate single-positive cell (ISP)
CD2 CD4or8 CD7
CD3 CD5 CD38
Double-positive cell (DP)
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
(CD9) CD19 CD22 CD117 TdT
(CD10) (CD20) CD24 CD127 HLA-DR
Pre B I cell
CD9 CD19 CD22 CD38 CD127 HLA-DR
CD10 CD20 CD24 CD117 TdT
Pre B II cell
Immature B cell
(CD9) CD20 CD22 CD37 IgM
CD19 CD21 CD24 HLA-DR
B Cell
(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
GM-CSF IL-3
Basophilic Myelocyte
GM-CSF IL-3
Basophil
SCF
IL-3
GM-CSF
Flt3L SCF
GM-CSF IL-3
CFU-E0
GM-CSF IL-3 IL-5
Myeloblast
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
GM-CSF IL-4
Flt3L CSF GM-CSF
IL-3 TNF
CFU-M/DC
GM-CSF M-CSF IL-3
Monoblast
CD11b CD14 CD33 CD115 CD123 CD126
CD13 CD15 CD64 CD116 CD124 HLA-DR
GM-CSF M-CSF IL-3
Promonocyte
CD11b CD14 CD64 CD116 CD124 HLA-DR
CD13 CD33 CD115 CD123 CD126
GM-CSF M-CSF IL-3
Monocyte
CD11b CD14 CD33 CD64
GM-CSF M-CSF
Macrophage
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
CD15 CD34 CD114 CD116 CD123 CD125 HLA-DR
CD33 CD64 CD115 CD121 CD124 CD126
Flt3L SCF GM-CSF
G-CSF IL-3
CFU-G
CD13 CD33 CD116 CD123 CD125 HLA-DR
CD15 CD114 CD121 CD124 CD126
GM-CSF G-CSF
Myeloblast
CD13 CD33 CD116 CD123 CD125
CD15 CD114 CD121 CD124 CD126
GM-CSF G-CSF
Neutrophilic Myelocyte
CD11b CD33 CD123
CD15 CD116 CD125
GM-CSF G-CSF
Neutrophil
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
CD33 CD117 EPOR
CD34 CD123 HLA-DR
SCF GM-CSF
IL-3 IL-4
EPO
CFU-E
CD36 CD235a
TPO EPO
Proerythroblast
CD235a
EPO
Erythrocyte
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
CD33 CD116 CD126 HLA-DR
CD34 CD123 IL-11R
Flt3L SCF GM-CSF
Meg-CSF IL-3 IL-6
IL-11 TPO
CFU-MK
CD61 CD116 CD122 CD126
SCF GM-CSF IL-3
IL-6 IL-11 TPO
Mega-karyocyte
CD9 CD36 CD42 CD116 CD126
CD14 CD41 CD61 CD123
IL-6 IL-11 TPO
Platelets
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


STEROID HORMONE BIOSYNTHESIS
-1
0
1
Steroid biosynthesis
Cholesterol
3.1.6.2
2.8.2.2
Cholesterol sulfate
1.14.15.6
20α-Hydroxy-cholesterol
22β-Hydroxy-cholesterol
20α,22β-Dihydroxy-cholesterol
21-Hydroxy-pregnenolone
HSD3B
11-Deoxy-corticosterone
1.3.1.22
5α-Dihydro-deoxycorticosterone
1.1.1.213
Allotetrahydro-deoxycorticosterone
Aldosterone hemiacetal
18-Hydroxy-corticosterone
CYP11B2
Aldosterone
1.3.1.3
1.1.1.50
11β,21-Dihydroxy-3,20-oxo-5β-pregnan-18-al
3α,11β,21-Trihydroxy-20-oxo-5β-pregnan-18-al
1.14.15.5
3α,21-Dihydroxy-5β-pregnane-11,20-dione
1.1.1.146
1.1.1.53
3α,20α,21-Trihydroxy-5β-pregnan-11-one
11β,21-Dihydroxy-5β-pregnane-3,20-dione
Tetrahydro-corticosterone
Corticosterone
1.14.15.4
HSD11B2
11-Dehydro-corticosterone
21-Hydroxy-5β-pregnane-3,11,20-trione
1.14.14.16
11α-Hydroxy-progesterone
1.14.99.14
20α-Hydroxy-progesterone
1.14.14.29
7α-Hydroxy-pregnenolone
1.1.1.149
4-Methylpentanal
Pregnenolone
Progesterone
11β-Hydroxy-progesterone
1.3.99.6
5β-Pregnane-3,20-dione
3α-Hydroxy-5β-pregnane-20-one
Pregnanediol
5α-Pregnan-20α-ol-3-one
C21-Steroids
Pregnenolone-sulfate
5α-Pregnane-3,20-dione
3α-Hydroxy-5α-pregnan-20-one
5α-Pregnane-3α,20α-diol
1.14.14.19
17α,20α-Dihydroxy-pregn-4-en-3-one
17α-Hydroxy-progesterone
21-Deoxycortisol
17α,20α-Dihydroxy-cholesterol
17α-Hydroxy-pregnenolone
5α-Pregnan-17α-ol-3,20-dione
SRD5A1
HSD17B6
AKR1C4
5α-Pregnane-3α,17α-diol-20-one
11β,17α,21-Trihydroxy-5β-pregnane-3,20-dione
Urocortisol
Cortol
17α,21-Dihydroxy-5β-pregnane-3,11,20-trione
17α,21-Dihydroxy-pregnenolone
11-Deoxycortisol
Cortisol
Cortisone
Urocortisone
Cortolone
1.14.15.4
11β,17α,21-Trihydroxy-pregnenolone
1.14.14.32
11β-Hydroxy-testosterone
1.1.1.62
Dehydro-epiandrosterone
Dehydroepiandro-sterone sulfate
11β-Hydroxyandrost-4-ene-3,17-dione
Adrenosterone
Estrone 3-sulfate
2.8.2.4
2.8.2.15
1.14.14.23
7α-Hydroxydehydro-epiandrosterone
5β-Androstane-3,17-dione
Etiocholan-3α-ol-17-one
1.1.1.52
2.4.1.17
Estrone glucuronide
2.8.2.15
2-Methoxyestrone-3-sulfate
16α-Hydroxyandrost-4-ene-3,17-dione
HSD3B
1.14.14.1
Etiocholan-3α-ol-17-one 3-glucuronide
3.1.6.1
1.1.1.148
Estradiol-17α
16α-Hydroxydehydro-epiandrosterone
1.14.99.12
3-Oxo-13,17-secoandrost-4-ene-17,13α-lactone
2.1.1.6
2-Hydroxyestrone
2-Methoxyestrone
2-Methoxyestrone-3-glucuronide
19-Oxoandrost-4-ene-3,17-dione
Androst-4-ene-3,17-dione
19-Hydroxyandrost-4-ene-3,17-dione
Estrone
16-α-Hydroxyestrone
Androsterone
5α-Androstane-3,17-dione
AKR1C2
Androsterone-glucuronide
7α-Hydroxy-androstenedione
1.1.1.51
1.1.1.64
1.1.1.239
1.1.1.-
C18-Steroids
7α-Hydroxy-testosterone
5α-Dihydro-testosterone
Androstan-3alpha,17beta-diol
Testosterone
16-Glucuronide-estriol
3β,17β-Dihydroxy-androst-5-ene
HSD8B
19-Hydroxy-testosterone
19-Oxo-testosterone
Estradiol-17β
Estriol
2-Methoxy-estradiol-17β
2-Hydroxy-estradiol-17β
2-Methoxy-estradiol-17β-3-glucuronide
C19-Steroids
5β-Dihydro-testosterone
1.14.99.11
6β-Hydroxy-estradiol-17β
2-Methoxy-estradiol-17β-3-sulfate
Testosterone glucuronide
Estradiol-17β-3-glucuronide
Estradiol-17β-3-sulfate
Data on KEGG graph
Rendered by Pathview


JAK-STAT SIGNALING PATHWAY
-1
0
1
Cytokine-cytokine receptor interaction
ECS complex
Ubiquitin mediated proteolysis
TC-PTP
SHP1
STAM
-p
-p
STAT dimerization
Interleukins
IL2 family
Receptor
JAK1/3
+p
STAT
STAT1/3/5/6
IL3 family
Receptor
JAK2
STAT5
IL6 family
Receptor
JAK1/2
TYK2
STAT1/3/6
IL10family
Receptor
JAK1/3
TYK2
STAT3
IL12/23
Receptor
JAK2
TYK2
STAT3/4
Hormone
Receptor
JAK2
STAT3/5
Interferons
IFN-I/III
Receptor
JAK1
TYK2
STAT1/2
IRF9
IFN-II
Receptor
JAK1/2
STAT1/3
Growth factors
GF
Receptor
JAK1/2
STAT1/3
STAT
TC-PTP
PIAS
-p
CBP/P300
SLIM
+u
Proteasome
DNA
CIS
SOCS
Apoptosis
Bcl-2
MCL1
Bcl-XL
PIM1
Anti-apoptosis
Cell cycle
c-Myc
CycD
Cell-cycle progression
p21
Cell-cycle inhibition
AOX
Lipid metabolism
GFAP
Differentiation
MAPK signaling pathway
+p
SHP2
GRB
SOS
Ras
Raf
Proliferation Differentiation
+p
PI3K
AKT
mTOR
Cell cycle Cell survival
PI3K-AKT signaling pathway
Data on KEGG graph
Rendered by Pathview


-1
0
1
NOTCH SIGNALING PATHWAY
Fringe
Delta
Serrate
Notch
ADAM17
S2
γ-Secretase complex
PSE2
PSEN
NCSTN
APH-1
S3
+u
Deltex
Itch
Numb
Dvl
NICD
(Notch intracellular domain)
Co-activator
MAML
HATs
SKIP
CSL
Co-repressor
Hairless
SMRT
CtBP
Gro/TLE
CIR
SHARP
HDAC
ATXN1/L
Hes1
NRARP
DNA
Hes1/5
Hey
PreTα
NRARP
Ras/MAPK
MAPK signaling pathway
Gene expression
Data on KEGG graph
Rendered by Pathview


LYSOSOME
-1
0
1
bacterium
lysosomal acid hydrolase
Golgi body
Phagocytosis
phagosome
transport vesicle
Transport of synthesized lysosomal enzymes
(See below)
clathrin coat
Endocytosis
Endocytosis
early endosome
late endosome
multivesicular body (MVB)
MCOLN1
cytosol
pH~ 7.2
ATP
ADP
acidification regulators
ATPeV
DMXL
WDR7
NCOA7
H+
pH~ 5.0
acid hydrolase
lysosomal membrane protein
lysosome
mitochondria
Autophagy
Regulation of autophagy
autophagosome
Glycosaminoglycan degradation
Other glycan degradation
plasma membrane
Lysosomal acid hydrolases
proteases
cathepsins
napsin
LGMN
TPP1
glycosidases
GLA
GLB
GAA
GBA
IDUA
NAGA
NAGLU
GALC
GUSB
FUCA1
HEXA/B
MANB
LAMAN
NEU1
HYAL1
sulfatases
ARS
GALNS
GNS
IDS
SGSH
lipases
LIPA
LYPLA3
nuclease
DNaseII
phosphatase
ACP2
ACP5
sphingomyelinase
SMPD1
ceramidase
ASAH1
aspartylglucosaminidase
AGA
Other lysosomal enzymes and activators
saposin
GM2A
CLN1
Lysosomal membrane proteins
major lysosomal membrane proteins
LAMP
LIMP
minor lysosomal membrane proteins
NPC
cystinosin
sialin
NRAMP
LAPTM
ABCA2
ABCB9
ACP2
endolyn
LALP70
sortilin
CLN3
CLN5
CLN7
HGSNAT
MCOLN1
LITAF
mEAK7
Activation of lysosomal sulfatase precursor
FGE
lysosomal hydrase precursor
from ER
mannose
+p
GNPT
NAGPA
M6P
Snare interactions in vesicular transport
cis Golgi network
trans Golgi network
Golgi body
M6P receptor
MPR
M6P
clathrin
AP-1
AP-3
GGAs
AP-4
Receptor-dependent transport
M6P
transport vesicle
Receptor recycling
AP-1
M6P
M6P
ATPeV
-p
mannose
AP-3
lysosome
mature lysosomal hydrase
late endosome
Transport of synthesized lysosomal enzymes
Data on KEGG graph
Rendered by Pathview

GO Analyses

```r
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)

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
```

```
head(gobpres$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
```

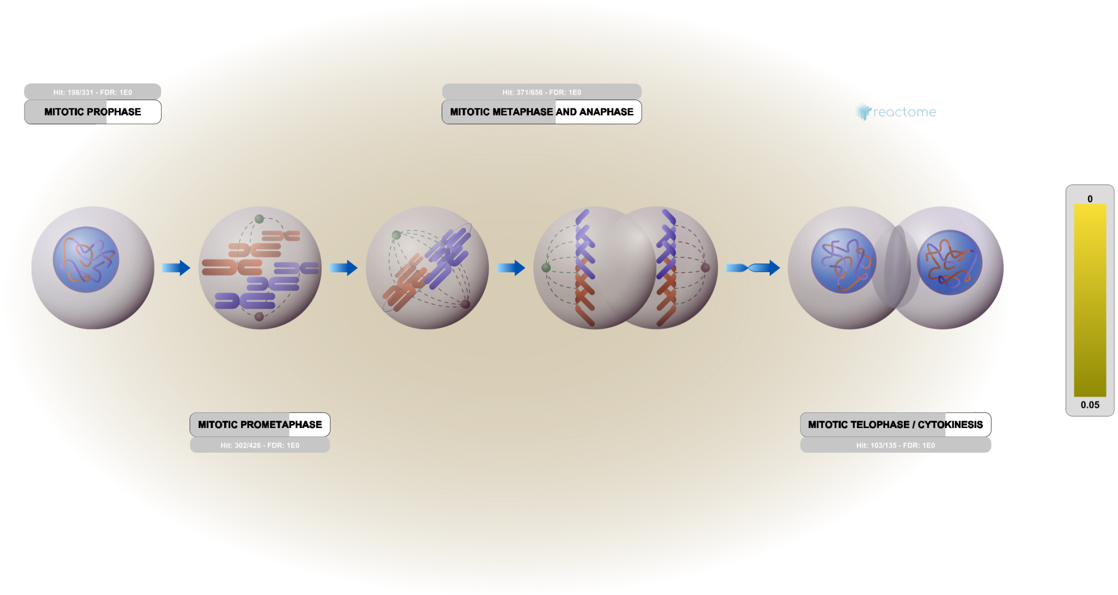
Reactome

Some folks like it, we can try it out here

```
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=
```



## Save results

```
save(res, file = "my_results.RData")
write.csv(res,file = "my_results.csv")
```