Class14

Sabrina Koldinger (A16368238)

Class 14

```
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, setdiff, sort,
    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

Loading required package: GenomicRanges

Loading required package: GenomeInfoDb

Loading required package: SummarizedExperiment

Loading required package: MatrixGenerics

Loading required package: matrixStats


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
```

Load Data and Check

Load in the data:

```
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 <- 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).

```
to.rm.ind=rowSums(countData[,1:6]==0)> 0
   countData= countData[!to.rm.ind,]
   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

```
colnames(countData)==row.names(colData)
```

```
[1] TRUE TRUE TRUE TRUE TRUE TRUE
```

Run DESeq

```
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)
```

```
estimating size factors
```

```
estimating dispersions
```

```
gene-wise dispersion estimates
```

```
mean-dispersion relationship
```

```
final dispersion estimates
```

```
fitting model and testing
```

```
res = results(dds, contrast=c("condition", "hoxa1_kd", "control_sirna"))
head(res)
```

```
log2 fold change (MLE): condition hoxa1_kd vs control_sirna
Wald test p-value: condition hoxa1 kd vs control sirna
DataFrame with 6 rows and 6 columns
                 baseMean log2FoldChange     lfcSE       stat      pvalue
                <numeric>      <numeric> <numeric>  <numeric>   <numeric>
ENSG00000279457   29.9136      0.1803039 0.3121566   0.577607 5.63529e-01
ENSG00000187634  183.2296      0.4258966 0.1355303   3.142446 1.67543e-03
ENSG00000188976 1651.1881     -0.6927118 0.0549876 -12.597612 2.17635e-36
ENSG00000187961  209.6379      0.7299597 0.1277613   5.713463 1.10700e-08
ENSG00000187583   47.2551      0.0392549 0.2606192   0.150622 8.80274e-01
ENSG00000187642   11.9798      0.5395082 0.5001355   1.078724 2.80711e-01
                      padj
                 <numeric>
ENSG00000279457 6.47026e-01
ENSG00000187634 3.34029e-03
ENSG00000188976 2.35969e-35
ENSG00000187961 3.69612e-08
ENSG00000187583 9.10931e-01
ENSG00000187642 3.61174e-01
```

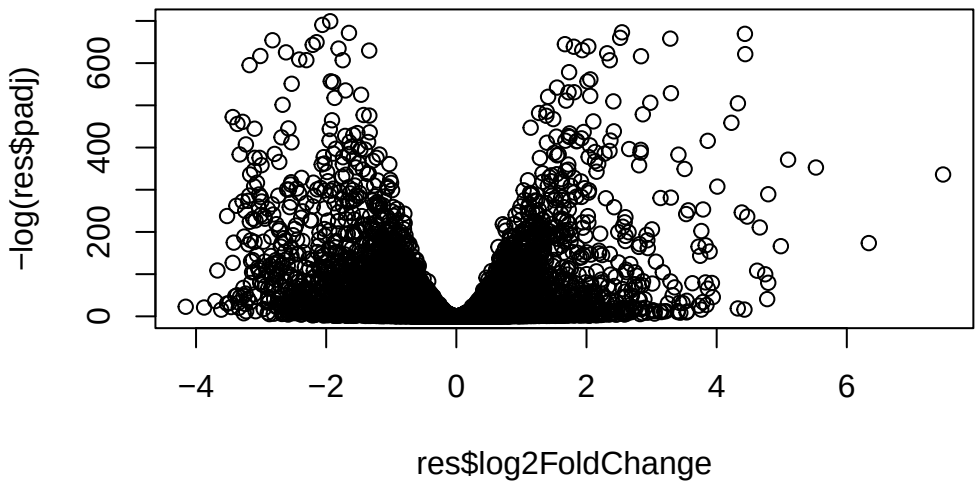
Q. Call the summary() function on your results to get a sense of how many genes are up or down-regulated at the default 0.1 p-value cutoff.

```
summary(res)
```

```
out of 13282 with nonzero total read count
adjusted p-value < 0.1
LFC > 0 (up)       : 4333, 33%
LFC < 0 (down)     : 4400, 33%
outliers [1]       : 0, 0%
low counts [2]     : 0, 0%
(mean count < 1)
[1] see 'cooksCutoff' argument of ?results
[2] see 'independentFiltering' argument of ?results
```

Volcano plot

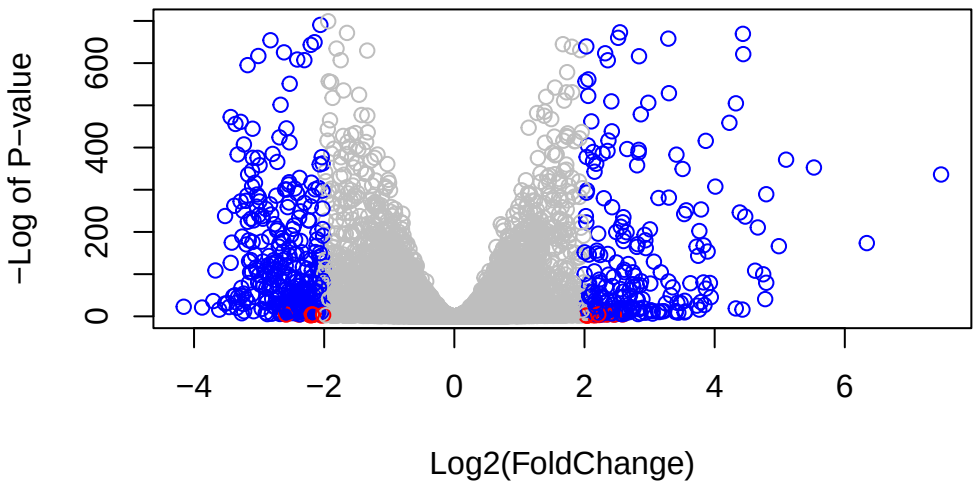
```
plot( res$log2FoldChange, -log(res$padj) )
```



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

Make a color vector for all genes, Color red the genes with absolute fold change above 2, Color blue those with adjusted p-value less than 0.01 and absolute fold change more than 2

```
mycols <- rep("gray", nrow(res) )
mycols[ abs(res$log2FoldChange) > 2 ] <- "red"
inds <- (res$padj<0.01) & (abs(res$log2FoldChange) > 2 )
mycols[ inds ] <- "blue"
plot( res$log2FoldChange, -log(res$padj), col=mycols, xlab="Log2(FoldChange)", ylab="-Log
```



Adding Gene annotation

```
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   29.9136      0.1803039 0.3121566   0.577607 5.63529e-01
ENSG00000187634  183.2296      0.4258966 0.1355303   3.142446 1.67543e-03
ENSG00000188976 1651.1881     -0.6927118 0.0549876 -12.597612 2.17635e-36
ENSG00000187961  209.6379      0.7299597 0.1277613   5.713463 1.10700e-08
ENSG00000187583   47.2551      0.0392549 0.2606192   0.150622 8.80274e-01
ENSG00000187642   11.9798      0.5395082 0.5001355   1.078724 2.80711e-01
ENSG00000188290  108.9221      2.0562855 0.1910714  10.761870 5.21018e-27
ENSG00000187608  350.7169      0.2570251 0.0999769   2.570845 1.01451e-02
ENSG00000188157 9128.4394      0.3899096 0.0482214   8.085827 6.17439e-16
```

```
ENSG00000131591  156.4791      0.1968918 0.1406800   1.399572 1.61641e-01
                      padj      symbol      entrez                   name
                 <numeric> <character> <character>            <character>
ENSG00000279457 6.47026e-01          NA          NA                     NA
ENSG00000187634 3.34029e-03      SAMD11      148398 sterile alpha motif ..
ENSG00000188976 2.35969e-35       NOC2L       26155 NOC2 like nucleolar ..
ENSG00000187961 3.69612e-08      KLHL17      339451 kelch like family me..
ENSG00000187583 9.10931e-01     PLEKHN1       84069 pleckstrin homology ..
ENSG00000187642 3.61174e-01       PERM1       84808 PPARGC1 and ESRR ind..
ENSG00000188290 4.17884e-26        HES4       57801 hes family bHLH tran..
ENSG00000187608 1.79950e-02       ISG15        9636 ISG15 ubiquitin like..
ENSG00000188157 3.15902e-15        AGRN      375790                  agrin
ENSG00000131591 2.23894e-01    C1orf159       54991 chromosome 1 open re..
```

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")
```

Pathway analysis

KEGG

```
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)
```

signaling and metabolic pathways

```
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"
```

Gage pathway anlaysis of our data:

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

```
     1266     54855      1465      2034      2150      6659
-2.422683  3.201858 -2.313713 -1.887999  3.344480  2.392257
```

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

Check the attributes and results of KEGG.

```
attributes(keggres)
```

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

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

```
                                      p.geomean stat.mean        p.val
hsa04110 Cell cycle                3.548176e-06 -4.604234 3.548176e-06
hsa03030 DNA replication           3.992330e-05 -4.191094 3.992330e-05
hsa04114 Oocyte meiosis            2.332810e-04 -3.564509 2.332810e-04
hsa03440 Homologous recombination  2.248158e-03 -2.967340 2.248158e-03
hsa03013 RNA transport             4.162613e-03 -2.662235 4.162613e-03
hsa00670 One carbon pool by folate 8.202725e-03 -2.535331 8.202725e-03
                                           q.val set.size         exp1
hsa04110 Cell cycle                0.0005535155      118 3.548176e-06
hsa03030 DNA replication           0.0031140177       36 3.992330e-05
hsa04114 Oocyte meiosis            0.0121306145       95 2.332810e-04
hsa03440 Homologous recombination  0.0876781678       28 2.248158e-03
hsa03013 RNA transport             0.1298735381      140 4.162613e-03
hsa00670 One carbon pool by folate 0.2115248982       17 8.202725e-03
```

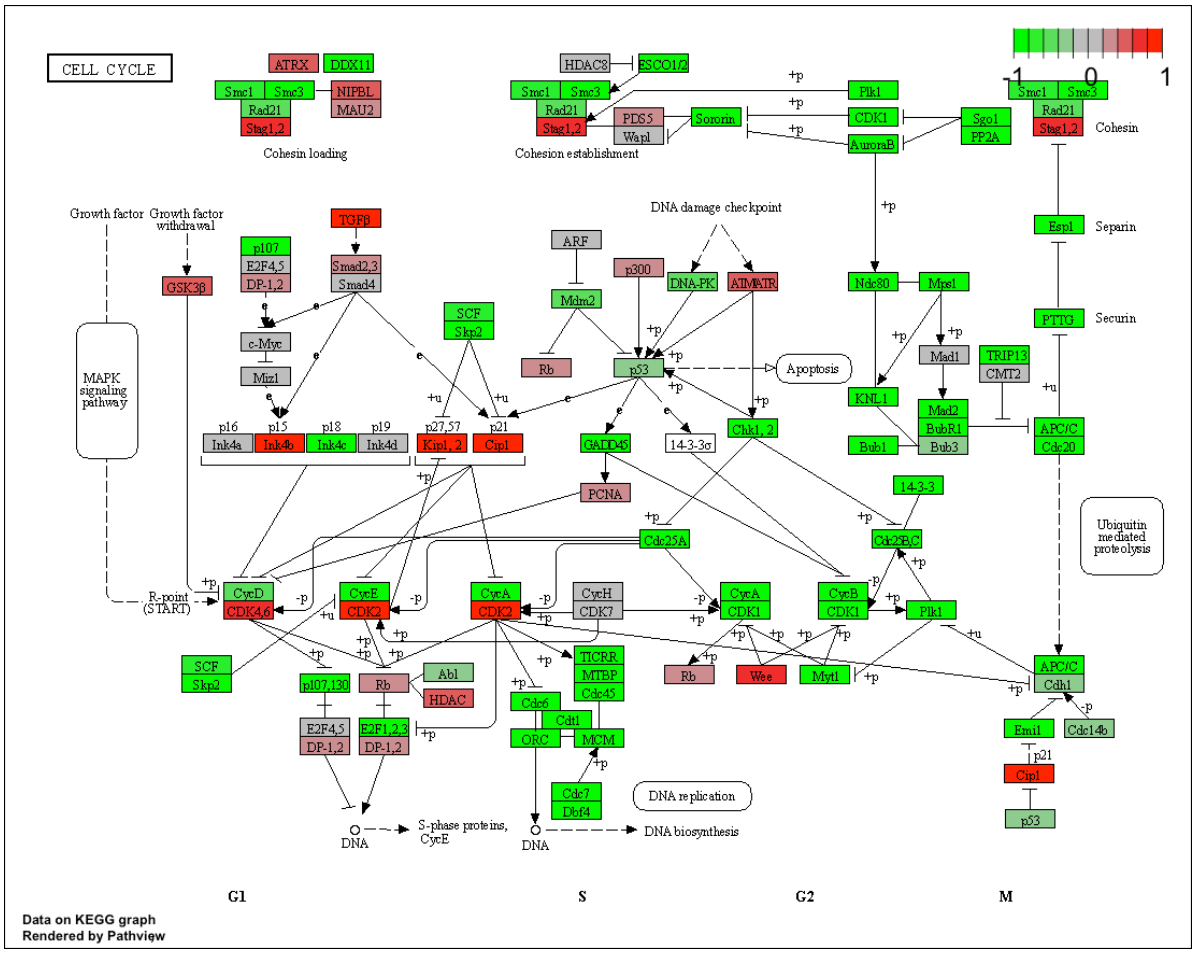
Now create a pathview of cell cycle:

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

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

```
Info: Working in directory /Users/sabrinakoldinger/Desktop/BIMM 143 Bioinformatics/Class14
```

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



Create a PDF:

```
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/sabrinakoldinger/Desktop/BIMM 143 Bioinformatics/Class14
```

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

Focusing on top 5 upregulated pathways

```
keggrespathways <- rownames(keggres$greater)[1:5]
keggresids = substr(keggrespathways, start=1, stop=8)
keggresids
```

```
[1] "hsa04142" "hsa04640" "hsa04974" "hsa00603" "hsa04380"
```

Have KEGG create pathviews for all of these:

```
pathview(gene.data=foldchanges, pathway.id="hsa04142")
```

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

```
Info: Working in directory /Users/sabrinakoldinger/Desktop/BIMM 143 Bioinformatics/Class14
```

```
Info: Writing image file hsa04142.pathview.png
```

```
Info: some node width is different from others, and hence adjusted!
```

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

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

```
Info: Working in directory /Users/sabrinakoldinger/Desktop/BIMM 143 Bioinformatics/Class14
```

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

```
  pathview(gene.data=foldchanges, pathway.id="hsa04974")
```

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

```
Info: Working in directory /Users/sabrinakoldinger/Desktop/BIMM 143 Bioinformatics/Class14
```

```
Info: Writing image file hsa04974.pathview.png
```

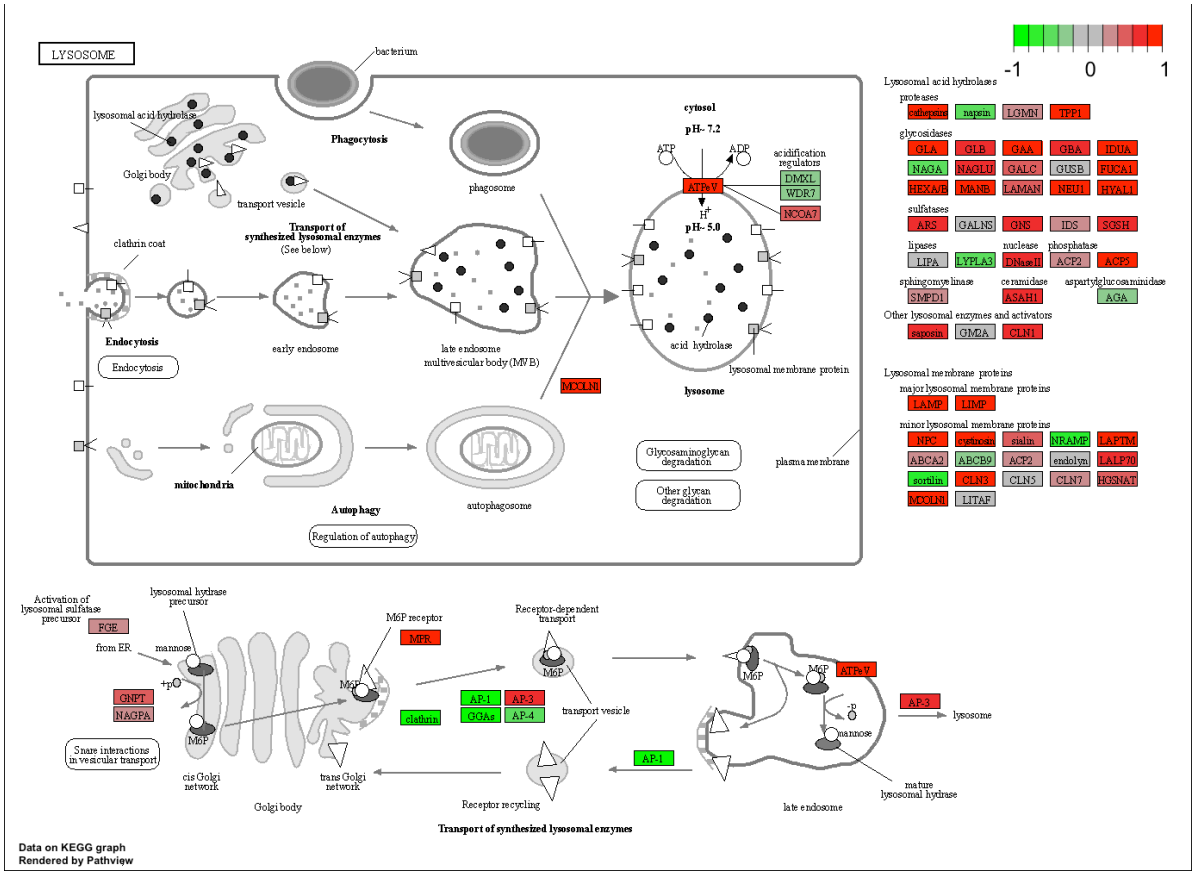
```
Info: some node width is different from others, and hence adjusted!
```

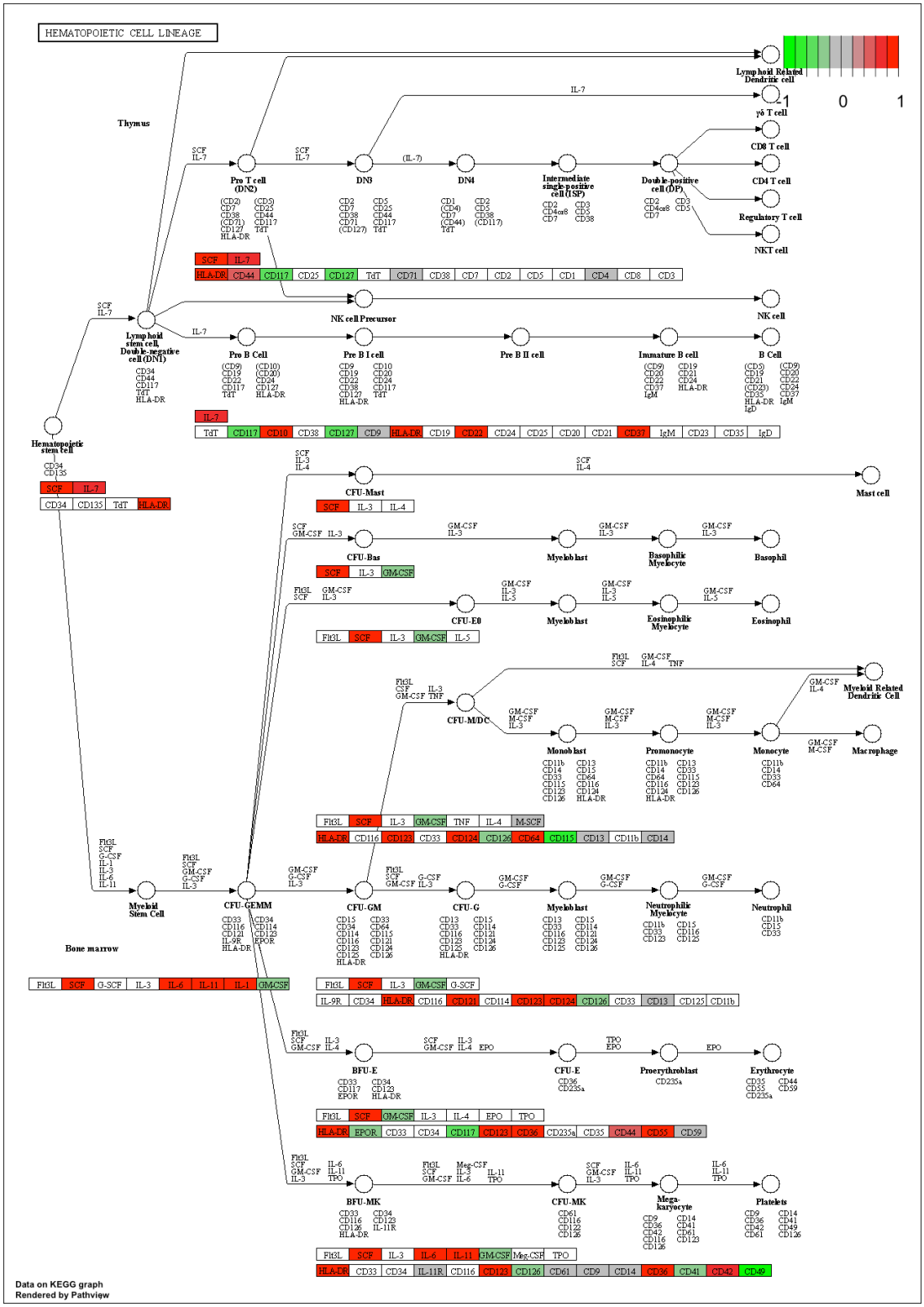
```
  pathview(gene.data=foldchanges, pathway.id="hsa04380")
```

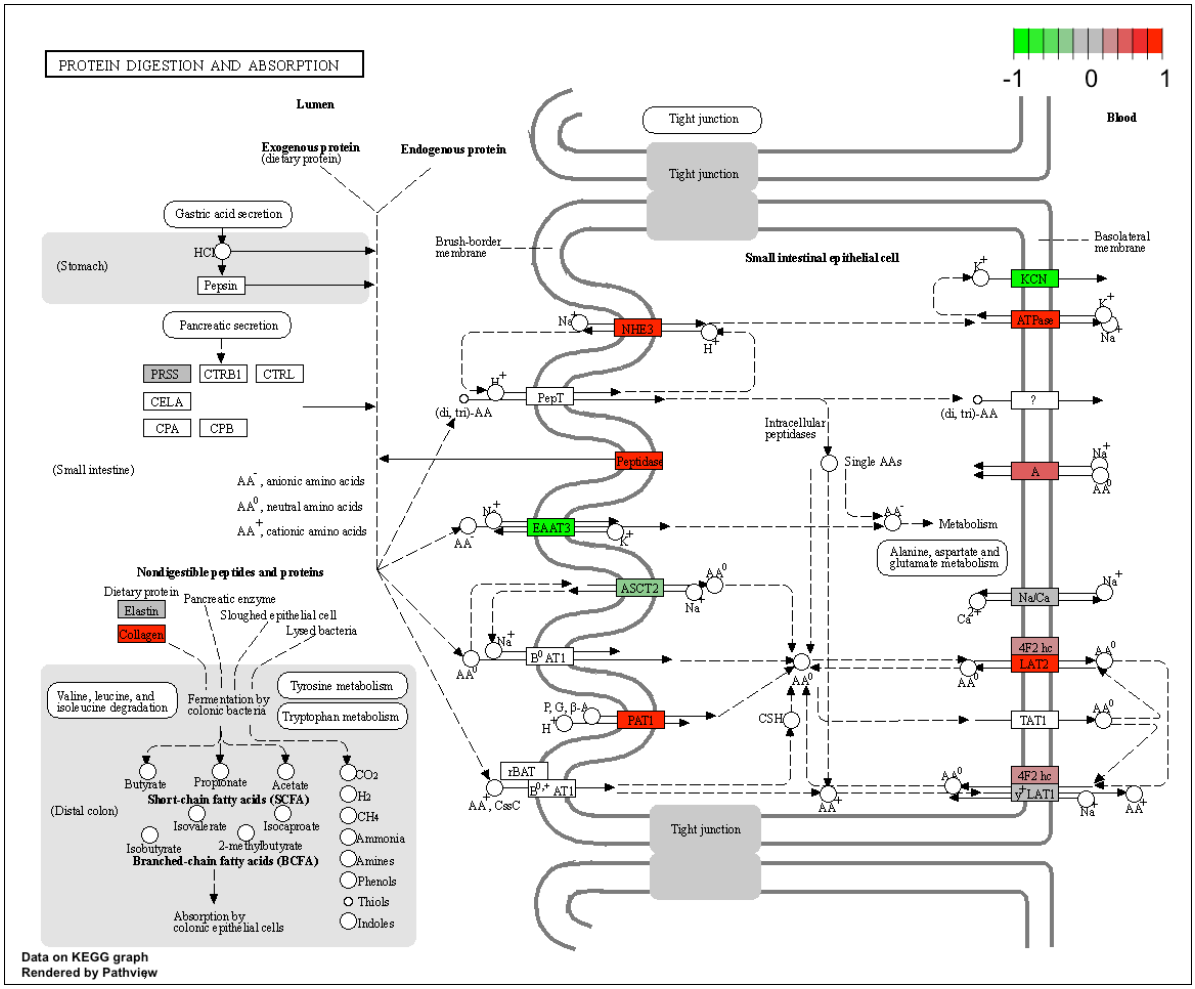
```
'select()' returned 1:1 mapping between keys and columns
```

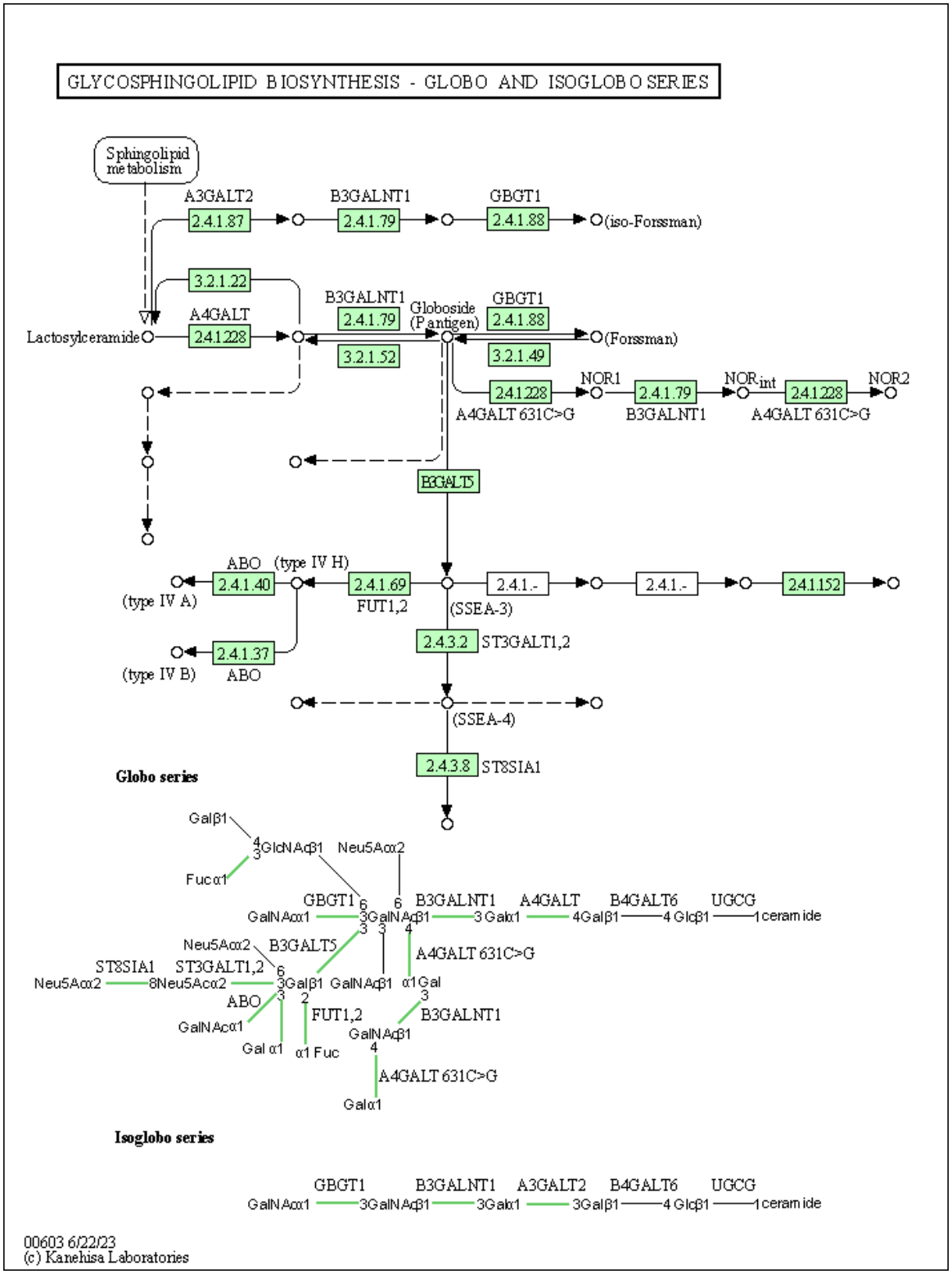
```
Info: Working in directory /Users/sabrinakoldinger/Desktop/BIMM 143 Bioinformatics/Class14
```

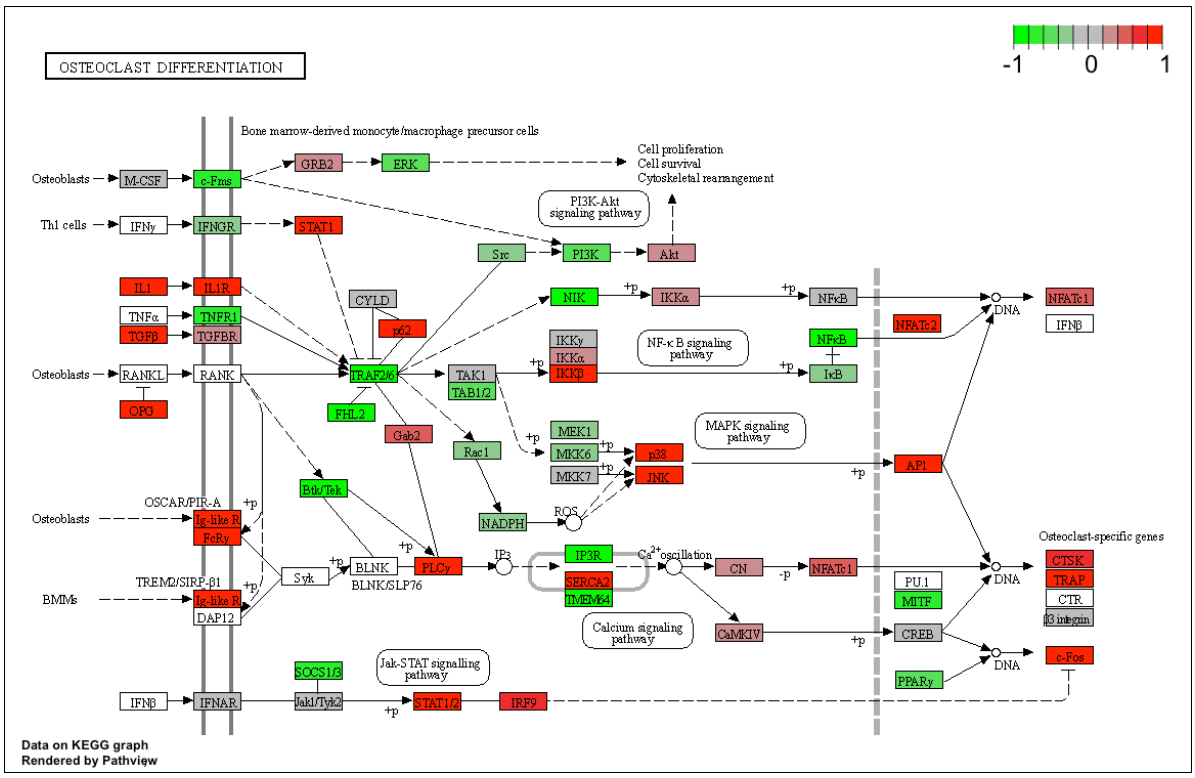
```
Info: Writing image file hsa04380.pathview.png
```


LYSOSOME
bacterium
lysosomal acid hydrolase
Golgi body
transport vesicle
Phagocytosis
phagosome
cytosol
pH~ 7.2
ATP
ADP
acidification regulators
DMXL
WDR7
NCOA7
ATPeV
H+
pH~ 5.0
Transport of synthesized lysosomal enzymes
(See below)
clathrin coat
Endocytosis
Endocytosis
early endosome
late endosome
multivesicular body (MVB)
MCOLN1
acid hydrolase
lysosomal membrane protein
lysosome
mitochondria
Autophagy
autophagosome
Regulation of autophagy
Glycosaminoglycan degradation
Other glycan degradation
plasma membrane
-1
0
1
Lysosomal acid hydrolases
proteases
cathepsin
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
nuclease
phosphatase
LIPA
LYPLA3
DNaseII
ACP2
ACP5
sphingomyelinase
ceramidase
aspartylglucosaminidase
SMPD1
ASAH1
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
LAL/P70
sortilin
CLN3
CLN5
CLN7
HGSNAT
MCOLN1
LITAF
Activation of lysosomal sulfatase precursor
FGE
lysosomal hydrase precursor
from ER
mannose
+pO
GNPT
NAGPA
M6P
Snare interactions in vesicular transport
cis Golgi network
trans Golgi network
Golgi body
M6P receptor
MPR
clathrin
AP-1
GGAs
AP-3
AP-4
Receptor-dependent transport
M6P
transport vesicle
Receptor recycling
AP-1
M6P
M6P
ATPeV
-P
mannose
late endosome
AP-3
lysosome
mature lysosomal hydrase
Transport of synthesized lysosomal enzymes
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
Pre T cell (DN2)
SCF
IL-7
DN3
(IL-7)
DN4
Intermediate single-positive cell (ISP)
Double-positive cell (DP)
(CD2) (CD5) CD7 CD25 CD38 CD44 (CD7dim) CD117 CD127 HLA-DR
CD2 CD5 CD7 CD25 CD38 CD44 CD71 CD117 (CD127) TdT
CD1 (CD4) CD7 (CD44) TdT
CD2 CD5
CD38 (CD117)
CD2 CD4cor8 CD7
CD3 CD5 CD38
CD2 CD4cor8 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
SCF
IL-7
Lymphoid stem cell, Double-negative cell (DN1)
IL-7
Pro B Cell
Pre B I cell
Pre B II cell
Immature B cell
B Cell
CD34 CD44 CD117 TdT HLA-DR
(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 (CD22) CD25 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
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
Flt3L CSF GM-CSF IL-3 TNF
CFU-M/DC
GM-CSF IL-4
Myeloid Related Dendritic Cell
GM-CSF M-CSF IL-3
Monoblast
GM-CSF M-CSF IL-3
Promonocyte
GM-CSF M-CSF IL-3
Monocyte
GM-CSF M-CSF
Macrophage
CD11b CD14 CD33 CD13 CD15 CD64 CD116 CD124 HLA-DR
CD11b CD14 CD64 CD13 CD33 CD115 CD123 CD126 HLA-DR
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
Flt3L SCF GM-CSF G-CSF IL-3
Myeloid Stem Cell
GM-CSF G-CSF IL-3
CFU-GEMM
GM-CSF G-CSF IL-3
CFU-GM
Flt3L SCF GM-CSF G-CSF IL-3
CFU-G
GM-CSF G-CSF
Myeloblast
GM-CSF G-CSF
Neutrophilic Myelocyte
GM-CSF G-CSF
Neutrophil
CD33 CD116 CD121 IL-9R HLA-DR
CD34 CD114 CD123 EPOR
CD15 CD34 CD94 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-CSF IL-3 IL-6 IL-11 IL-1 GM-CSF
Flt3L SCF IL-3 GM-CSF G-CSF
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 IL-3 GM-CSF IL-4 EPO
CFU-E
TPO EPO
Proerythroblast
EPO
Erythrocyte
CD33 CD117 EPOR 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 IL-6 Meg-CSF 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 CD34 CD123 IL-11R HLA-DR
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


PROTEIN DIGESTION AND ABSORPTION
-1
0
1
Lumen
Blood
Exogenous protein
(dietary protein)
Endogenous protein
Tight junction
Tight junction
Gastric acid secretion
(Stomach)
HCl
Pepsin
Brush-border membrane
Small intestinal epithelial cell
Basolateral membrane
Pancreatic secretion
PRSS
CTRB1
CTRL
CELA
CPA
CPB
(Small intestine)
AA-, anionic amino acids
AA0, neutral amino acids
AA+, cationic amino acids
NHE3
KCN
ATPase
PepT
(di, tri)-AA
(di, tri)-AA
?
Peptidase
Intracellular peptidases
Single AAs
A
EAAT3
Metabolism
Alanine, aspartate and glutamate metabolism
ASCT2
Na/Ca
B0AT1
4F2 hc
LAT2
PAT1
TAT1
rBAT
b0,+AT1
4F2 hc
y+LAT1
Tight junction
Nondigestible peptides and proteins
Dietary protein
Elastin
Collagen
Pancreatic enzyme
Sloughed epithelial cell
Lysed bacteria
Valine, leucine, and isoleucine degradation
Fermentation by colonic bacteria
Tyrosine metabolism
Tryptophan metabolism
Butyrate
Propionate
Acetate
Short-chain fatty acids (SCFA)
(Distal colon)
Isovalerate
Isocaproate
Isobutyrate
2-methylbutyrate
Branched-chain fatty acids (BCFA)
Absorption by colonic epithelial cells
CO2
H2
CH4
Ammonia
Amines
Phenols
Thiols
Indoles
Data on KEGG graph
Rendered by Pathview


GLYCOSPHINGOLIPID BIOSYNTHESIS - GLOBO AND ISOGLOBO SERIES
Sphingolipid metabolism
A3GALT2
2.4.1.87
B3GALNT1
2.4.1.79
GBGT1
2.4.1.88
(iso-Forssman)
3.2.1.22
A4GALT
2.4.1.228
Lactosylceramide
B3GALNT1
2.4.1.79
Globoside (P antigen)
GBGT1
2.4.1.88
(Forssman)
3.2.1.52
3.2.1.49
NOR1
2.4.1.228
A4GALT 631C>G
2.4.1.79
B3GALNT1
NOR int
2.4.1.228
A4GALT 631C>G
NOR2
B3GALT5
ABO (type IV H)
2.4.1.40
(type IV A)
2.4.1.69
FUT1,2
2.4.1.-
(SSEA-3)
2.4.1.-
2.4.1.152
2.4.1.37
(type IV B)
ABO
2.4.3.2
ST3GALT1,2
(SSEA-4)
2.4.3.8
ST8SIA1
Globo series
Galβ1
4
3
GlcNAcβ1
Neu5Acα2
Fucα1
GBGT1
6
6
B3GALNT1
A4GALT
B4GALT6
UGCG
GalNAcα1
3GalNAcβ1
3 Galα1
4Galβ1
4 Glcβ1
1ceramide
3
3
4
Neu5Acα2
B3GALT5
A4GALT 631C>G
ST8SIA1
ST3GALT1,2
6
Neu5Acα2
8Neu5Acα2
3Galβ1
GalNAcβ1
α1Gal
3
3
2
ABO
FUT1,2
B3GALNT1
GalNAcα1
GalNAcβ1
4
Gal α1
α1 Fuc
A4GALT 631C>G
Galα1
Isoglobo series
GBGT1
B3GALNT1
A3GALT2
B4GALT6
UGCG
GalNAcα1
3GalNAcβ1
3Galα1
3Galβ1
4 Glcβ1
1ceramide
00603 6/22/23
(c) Kanehisa Laboratories



Genotology

```
data(go.sets.hs)
data(go.subs.hs)
gobpsets = go.sets.hs[go.subs.hs$BP]

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

lapply(gobpres, head)
```

```
$greater
                                                     p.geomean stat.mean
GO:0007156 homophilic cell adhesion                 7.523307e-05  3.873939
GO:0016339 calcium-dependent cell-cell adhesion     8.556504e-04  3.340855
GO:0010817 regulation of hormone levels             1.058523e-03  3.091986
GO:0048729 tissue morphogenesis                     1.389102e-03  3.002504
GO:0008285 negative regulation of cell proliferation 1.443571e-03  2.989717
GO:0051047 positive regulation of secretion         1.877703e-03  2.927781
```

```
                                                        p.val     q.val
GO:0007156 homophilic cell adhesion                 7.523307e-05 0.2796413
GO:0016339 calcium-dependent cell-cell adhesion     8.556504e-04 0.5718590
GO:0010817 regulation of hormone levels             1.058523e-03 0.5718590
GO:0048729 tissue morphogenesis                     1.389102e-03 0.5718590
GO:0008285 negative regulation of cell proliferation 1.443571e-03 0.5718590
GO:0051047 positive regulation of secretion         1.877703e-03 0.5718590
                                                    set.size         exp1
GO:0007156 homophilic cell adhesion                       90 7.523307e-05
GO:0016339 calcium-dependent cell-cell adhesion           24 8.556504e-04
GO:0010817 regulation of hormone levels                  225 1.058523e-03
GO:0048729 tissue morphogenesis                          347 1.389102e-03
GO:0008285 negative regulation of cell proliferation     386 1.443571e-03
GO:0051047 positive regulation of secretion              130 1.877703e-03

$less
                                          p.geomean stat.mean        p.val
GO:0000279 M phase                     6.451975e-18 -8.738701 6.451975e-18
GO:0048285 organelle fission           1.832907e-16 -8.369971 1.832907e-16
GO:0000280 nuclear division            2.627088e-16 -8.340038 2.627088e-16
GO:0007067 mitosis                     2.627088e-16 -8.340038 2.627088e-16
GO:0000087 M phase of mitotic cell cycle 9.244549e-16 -8.166584 9.244549e-16
GO:0007059 chromosome segregation      2.502912e-12 -7.264756 2.502912e-12
                                              q.val set.size         exp1
GO:0000279 M phase                     2.398199e-14      467 6.451975e-18
GO:0048285 organelle fission           2.441221e-13      360 1.832907e-16
GO:0000280 nuclear division            2.441221e-13      338 2.627088e-16
GO:0007067 mitosis                     2.441221e-13      338 2.627088e-16
GO:0000087 M phase of mitotic cell cycle 6.872398e-13      348 9.244549e-16
GO:0007059 chromosome segregation      1.550554e-09      135 2.502912e-12

$stats
                                                    stat.mean     exp1
GO:0007156 homophilic cell adhesion                  3.873939 3.873939
GO:0016339 calcium-dependent cell-cell adhesion      3.340855 3.340855
GO:0010817 regulation of hormone levels              3.091986 3.091986
GO:0048729 tissue morphogenesis                      3.002504 3.002504
GO:0008285 negative regulation of cell proliferation 2.989717 2.989717
GO:0051047 positive regulation of secretion          2.927781 2.927781
```

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: 8186"
```

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

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? mitotic cell cycle has the most significant entities p-value. They do match in some ways. They have different ways of analyzing and compiling the data.