

Dysregulated processes in NAFLD (BB103X)

Hong Yang

02/03/2022

3.1 Load necessary packages

- Rstudio

```
library(DESeq2)
library(pheatmap)
library(tidyverse)
library(xlsx)
library(readxl)
library(gplots)
library(ggbiplot)
library(piano)
library(venn)
library(clusterProfiler)
library(GEOquery)
library(openxlsx)
library(GEOquery)
library(ggrepel)
library(MAGeCKFlute)
library(DOSE)
library(org.Hs.eg.db)
```

3.2 Load data from differential analysis

```
deseq_data = read.xlsx('data/DEseq_results_NAFLvscontrol.xlsx')
head(deseq_data)
```

##	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
## 1	3.180934	-0.01490257	0.90839708	-0.01640535	0.986911009	0.99227207
## 2	486.025667	0.03269755	0.13727192	0.23819549	0.811729476	0.88807235
## 3	323.407339	-0.07917556	0.08499411	-0.93154178	0.351573381	0.49338748
## 4	88.722171	-0.48628837	0.18092777	-2.68774865	0.007193552	0.02032398
## 5	126.952163	-0.16389423	0.18132936	-0.90384832	0.366075834	0.50859810
## 6	48801.045715	0.39738106	0.15415678	2.57777211	0.009943955	0.02670424
##	Gene.name					
## 1	TNMD					
## 2	DPM1					
## 3	SCYL3					
## 4	C1orf112					
## 5	FGR					
## 6	CFH					

3.3 Basic statistic - differential expressed gene

- $\text{padj} < 0.05$
- $\log_2\text{FoldChange} > 1$ or $\log_2\text{FoldChange} < -1$

```
deseq_data_sig = deseq_data %>%
  mutate(sig = case_when((padj < 0.05 & log2FoldChange < -1) ~ "Down",
                        (padj < 0.05 & log2FoldChange > 1) ~ "Up",
                        TRUE ~ "No")) %>%
  filter(!(sig == 'No'))
head(deseq_data_sig)
```

##	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
## 1	90.817855	4.333222	0.5358942	8.085966	6.167344e-16	4.353204e-14
## 2	196.777211	1.090650	0.2045675	5.331489	9.741048e-08	1.009119e-06
## 3	73.983099	1.182610	0.2690149	4.396077	1.102249e-05	6.851220e-05
## 4	5377.125701	-1.656390	0.2377945	-6.965634	3.269275e-12	9.655954e-11
## 5	1.059186	-1.715800	0.5504674	-3.116987	1.827097e-03	6.217729e-03
## 6	27.245128	-1.240884	0.2435324	-5.095357	3.480854e-07	3.171995e-06
##	Gene.name	sig				
## 1	CYP51A1	Up				
## 2	MAD1L1	Up				
## 3	DBNDD1	Up				
## 4	PDK4	Down				
## 5	CALCR	Down				
## 6	MCUB	Down				

3.4 summary - differential expressed gene

```
table(deseq_data_sig$sig)
```

```
##
## Down   Up
##  827   726
```

3.4 GO enrichment - differential expressed gene

- GO ontology enrichment

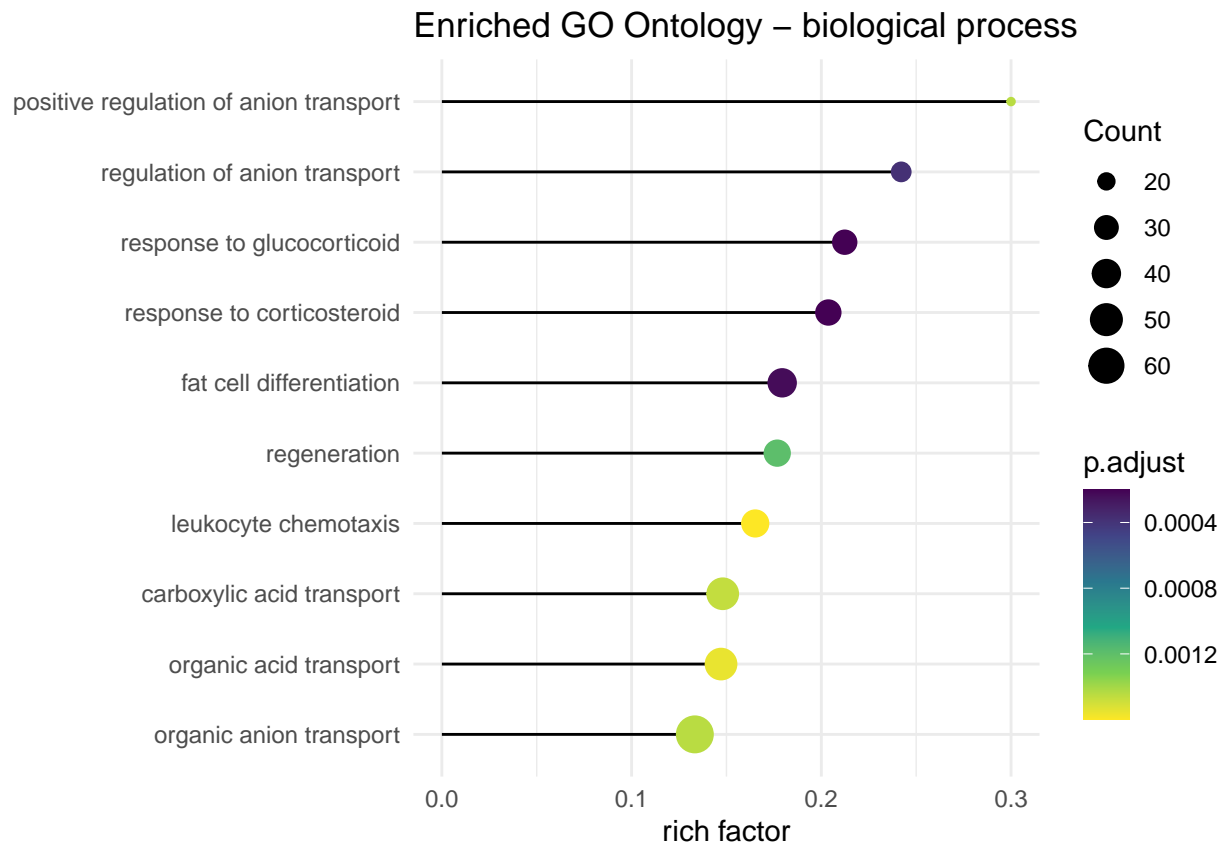
```
ego_bp <- enrichGO(gene      = deseq_data_sig$Gene.name,
                   OrgDb      = org.Hs.eg.db,
                   keyType     = 'SYMBOL',
                   ont         = "BP",
                   pAdjustMethod = "BH",
                   pvalueCutoff = 0.01,
                   qvalueCutoff = 0.05)
head(ego_bp@result)
```

##	ID	Description	GeneRatio	BgRatio
##	G0:0051384 G0:0051384	response to glucocorticoid	31/1373	146/18670
##	G0:0031960 G0:0031960	response to corticosteroid	33/1373	162/18670
##	G0:0045444 G0:0045444	fat cell differentiation	40/1373	223/18670
##	G0:0044070 G0:0044070	regulation of anion transport	23/1373	95/18670
##	G0:0031099 G0:0031099	regeneration	35/1373	198/18670
##	G0:0015711 G0:0015711	organic anion transport	66/1373	495/18670

```
##           pvalue      p.adjust      qvalue
## G0:0051384 6.465248e-08 0.0002004318 0.0001743143
## G0:0031960 7.051251e-08 0.0002004318 0.0001743143
## G0:0045444 1.215158e-07 0.0002302724 0.0002002666
## G0:0044070 2.690554e-07 0.0003823950 0.0003325667
## G0:0031099 1.047796e-06 0.0011913436 0.0010361044
## G0:0015711 1.749473e-06 0.0014438526 0.0012557100
##
## G0:0051384
## G0:0031960
## G0:0045444
## G0:0044070
## G0:0031099
## G0:0015711 SLC7A14/MPC1/SCARB1/EDN1/ABCB1/PITPNM2/SLC26A4/SLC52A3/OXT/ABCD1/SLC7A5/SLC16A6/PITPNM1/S
##           Count
## G0:0051384     31
## G0:0031960     33
## G0:0045444     40
## G0:0044070     23
## G0:0031099     35
## G0:0015711     66
```

```
y <- mutate(ego_bp, richFactor = Count / as.numeric(sub("/\\d+", "", BgRatio)))
```

```
ggplot(y, showCategory = 10,
  aes(richFactor, fct_reorder(Description, richFactor))) +
  geom_segment(aes(xend=0, yend = Description)) +
  geom_point(aes(color=p.adjust, size = Count)) +
  scale_color_viridis_c(guide=guide_colorbar(reverse=TRUE)) +
  scale_size_continuous(range=c(1, 6)) +
  theme_minimal() +
  xlab("rich factor") +
  ylab(NULL) +
  ggtitle("Enriched GO Ontology - biological process")
```



3.4 KEGG enrichment - differential expressed gene

- KEGG pathway enrichment

```

entrezID <- TransGeneID(deseq_data_sig$Gene.name,
                        fromType = "Symbol",
                        toType = "Entrez",
                        organism = "hsa")

entrezID = as.data.frame(entrezID) %>% mutate(Gene.name = rownames(.))
rownames(deseq_data_sig) = deseq_data_sig$Gene.name
deseq_data_sig_withID = merge(deseq_data_sig, entrezID, by = 'Gene.name')

geneList = deseq_data_sig_withID$log2FoldChange
names(geneList) = as.character(deseq_data_sig_withID$entrezID)
geneList = sort(geneList, decreasing = TRUE)
gene <- names(geneList)[abs(geneList) > 2]

kk <- enrichKEGG(gene      = gene,
                  organism = 'hsa',
                  pvalueCutoff = 0.05)

## Reading KEGG annotation online:
##
## Reading KEGG annotation online:
head(kk@result)

```

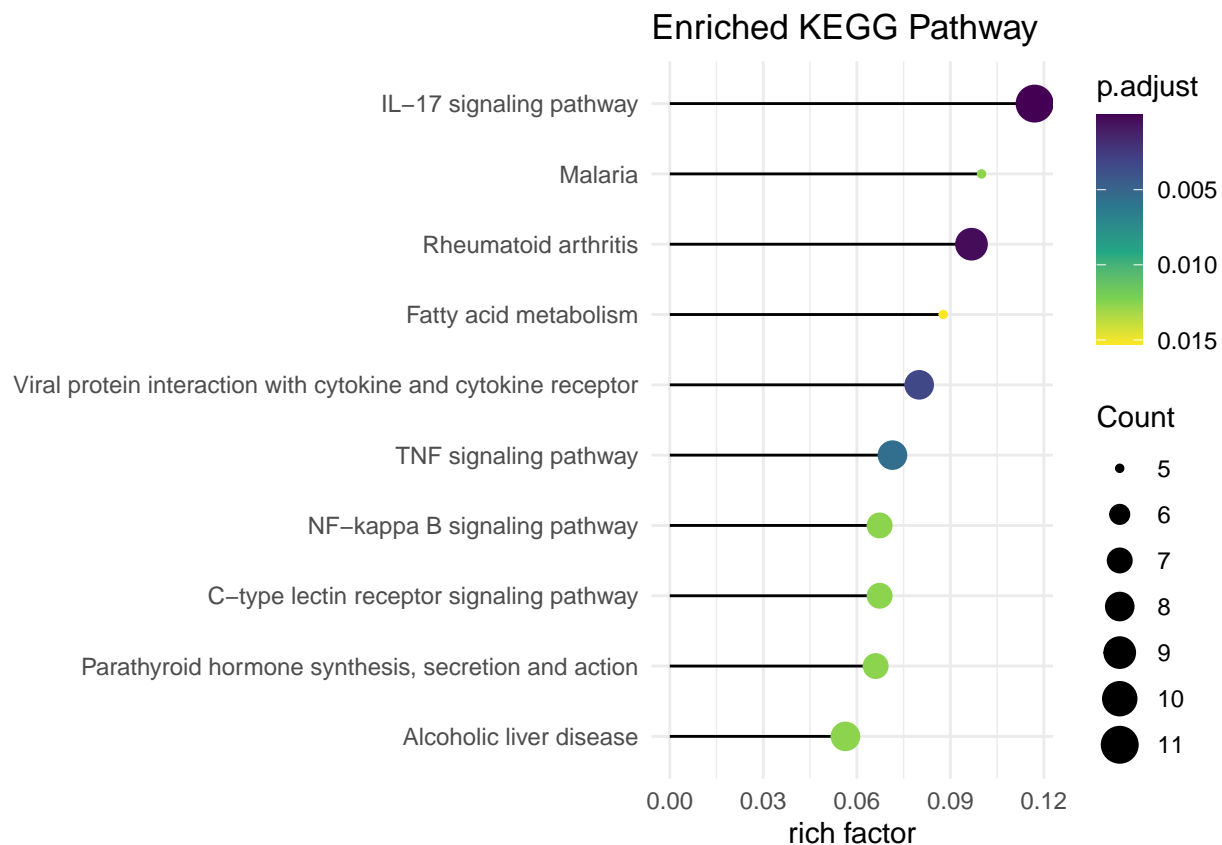
ID	Description
hsa04657	IL-17 signaling pathway
hsa05323	Rheumatoid arthritis
hsa04061	Viral protein interaction with cytokine and cytokine receptor
hsa04668	TNF signaling pathway
hsa04064	NF-kappa B signaling pathway
hsa04625	C-type lectin receptor signaling pathway

	GeneRatio	BgRatio	pvalue	p.adjust	qvalue
hsa04657	11/105	94/8114	2.832437e-08	6.656227e-06	5.336907e-06
hsa05323	9/105	93/8114	2.809322e-06	3.300954e-04	2.646677e-04
hsa04061	8/105	100/8114	4.183548e-05	3.277113e-03	2.627562e-03
hsa04668	8/105	112/8114	9.406940e-05	5.526577e-03	4.431164e-03
hsa04064	7/105	104/8114	3.775233e-04	1.259235e-02	1.009644e-02
hsa04625	7/105	104/8114	3.775233e-04	1.259235e-02	1.009644e-02

	geneID	Count
hsa04657	3725/6347/2921/3576/4312/2920/8061/5743/3569/2353/2354	11
hsa05323	3725/6347/1493/2921/3576/4312/2920/3569/2353	9
hsa04061	6369/9560/6347/2921/3576/2920/7852/3569	8
hsa04668	3725/6347/9021/2921/2920/5743/3569/2353	8
hsa04064	9560/4616/2921/3576/2920/5743/10912	7
hsa04625	810/3725/26253/1959/5743/1960/3569	7

```
y <- mutate(kk, richFactor = Count / as.numeric(sub("/\\d+", "", BgRatio)))
```

```
ggplot(y, showCategory = 10,
  aes(richFactor, fct_reorder(Description, richFactor))) +
  geom_segment(aes(xend=0, yend = Description)) +
  geom_point(aes(color=p.adjust, size = Count)) +
  scale_color_viridis_c(guide=guide_colorbar(reverse=TRUE)) +
  scale_size_continuous(range=c(1, 6)) +
  theme_minimal() +
  xlab("rich factor") +
  ylab(NULL) +
  ggtitle("Enriched KEGG Pathway")
```



3.4 disease enrichment - differential expressed gene

- disease enrichment

```
dd <- enrichDO(gene      = gene,
               ont        = "DO",
               pvalueCutoff = 0.05,
               pAdjustMethod = "BH",
               universe     = names(geneList),
               minGSSize    = 5,
               maxGSSize    = 500,
               qvalueCutoff = 0.05,
               readable      = FALSE)

head(dd)
```

##	ID	Description	GeneRatio	BgRatio	pvalue
##	DOID:65	connective tissue disease	34/113	120/691	0.0001752258
##	DOID:193	reproductive organ cancer	31/113	107/691	0.0002387158
##	DOID:4896	bile duct adenocarcinoma	8/113	13/691	0.0002555118
##	DOID:4947	cholangiocarcinoma	8/113	13/691	0.0002555118
##	DOID:3342	bone inflammation disease	28/113	96/691	0.0004662977
##	DOID:3451	skin carcinoma	7/113	11/691	0.0004920780
##		p.adjust	qvalue		
##	DOID:65	0.03015916	0.02156115		
##	DOID:193	0.03015916	0.02156115		
##	DOID:4896	0.03015916	0.02156115		
##	DOID:4947	0.03015916	0.02156115		

```
## D0ID:3342 0.03015916 0.02156115
## D0ID:3451 0.03015916 0.02156115
##
## D0ID:65 9294/6369/116085/7078/1311/4584/50943/3725/639/388/5806/9507/7538/1839/1490/5744/6347/1493/
## D0ID:193 57016/10642/2194/7078/4326/6319/2810/43/50943/247/3725/1839/1490/5744/6347/1
## D0ID:4896
## D0ID:4947
## D0ID:3342 9294/6369/116085/7078/1311/4584/50943/3725/639/388/9507/7538/
## D0ID:3451
## Count
## D0ID:65 34
## D0ID:193 31
## D0ID:4896 8
## D0ID:4947 8
## D0ID:3342 28
## D0ID:3451 7
```

```
y <- mutate(dd, richFactor = Count / as.numeric(sub("/\\d+", "", BgRatio)))
```

```
ggplot(y, showCategory = 20,
  aes(richFactor, fct_reorder(Description, richFactor))) +
  geom_segment(aes(xend=0, yend = Description)) +
  geom_point(aes(color=p.adjust, size = Count)) +
  scale_color_viridis_c(guide=guide_colorbar(reverse=TRUE)) +
  scale_size_continuous(range=c(1, 6)) +
  theme_minimal() +
  xlab("rich factor") +
  ylab(NULL) +
  ggtitle("Enriched Disease Ontology")
```

