HSC1_MUT vs. HSC1_WT

NP

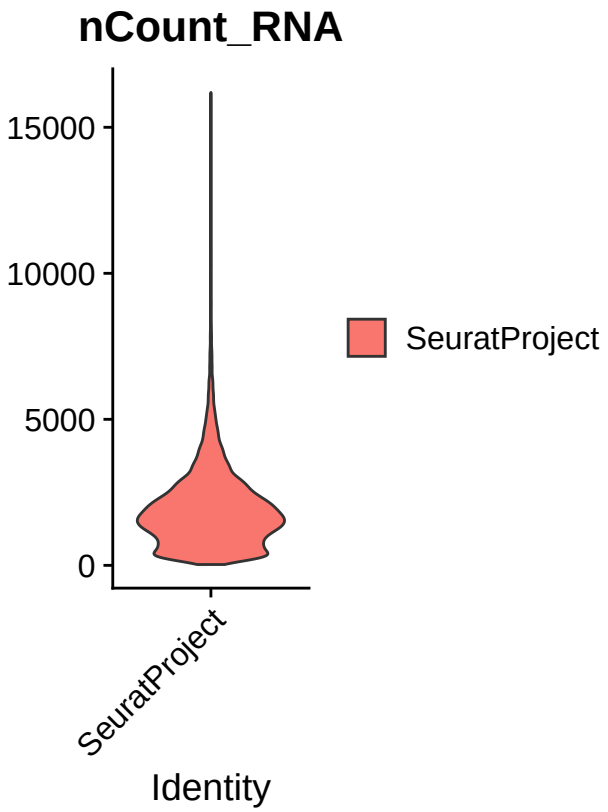
1/20/2022

```
# https://satijalab.org/signac/articles/pbmc_multiomic.html
counts <- Read10X_h5("filtered_feature_bc_matrix.h5")
fragpath <- "atac_fragments.tsv.gz"
```

```
# Get gene annotations for HG38

annotation <- GetGRangesFromEnsDb(ensdb = EnsDb.Hsapiens.v86)
genome(annotation) <- "hg38"
seqlevelsStyle(annotation) <- "UCSC"


# Create a Seurat object containing the RNA data

npm1 <- CreateSeuratObject(
        counts = counts$`Gene Expression`,
        assay = "RNA"
)

# npm1[["percent.mt"]] <- PercentageFeatureSet(npm1, pattern = "^MT-")
#
# # create ATAC assay and add it to the object
#
# npm1[["ATAC"]] <- CreateChromatinAssay(
#         counts = counts$Peaks,
#         sep = c(":", "-"),
#         fragments = fragpath,
#         annotation = annotation
#
# )
```

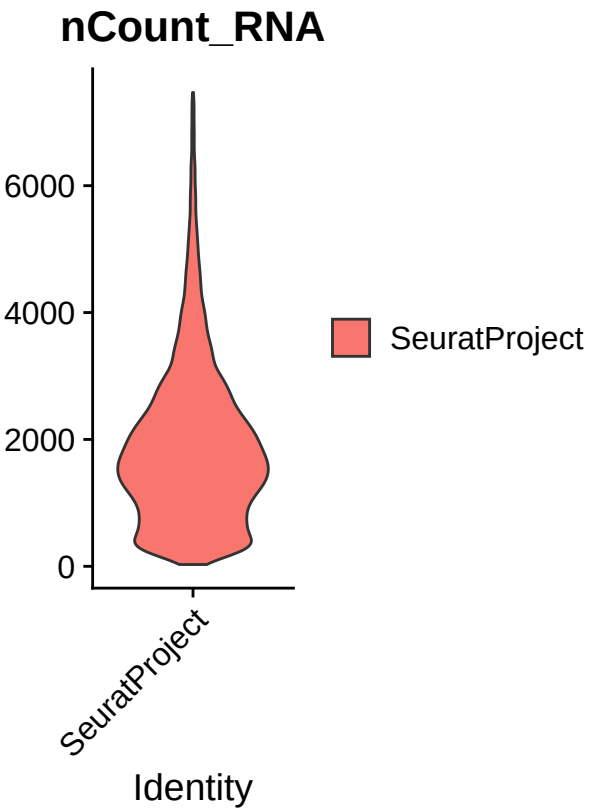
Violin plot of RNA counts

```
DefaultAssay(npm1) <- "RNA"


VlnPlot(
  object = npm1,
  features = c("nCount_RNA"),
  ncol = 4,
  pt.size = 0
)
```



Discard low count genes and display subsequent violin plot.

```
npm1 <- subset(
  x = npm1,
  subset = nCount_RNA < 7500
)

VlnPlot(
  object = npm1,
  features = c("nCount_RNA"),
  ncol = 4,
  pt.size = 0
)
```



Dimensionality reduction: SC transform and PCA

```
#npm1 <- SCTransform(npm1) %>% RunPCA() %>% RunUMAP(dims = 1:50, reduction.name = 'umap.rna', reduction
npm1 <- SCTransform(npm1)
```

```
##   |                                                                              |
##   |                                                                              |
```

```
npm1 <- RunPCA(npm1)
```

Add cell identity status from Seurat object to NPM1 Seurat metadata.

```
reference <- readRDS("~/Downloads/namlab/NPM1_seurat/NPM1_seurat.rds")

npm1 <- AddMetaData(
        object = npm1,
        metadata = reference@meta.data %>% select(Cell.Ident_Mutation.Status) %>% filter(rownames(.) %i

Idents(npm1) <- "Cell.Ident_Mutation.Status"
```

Find DEGs using FindMarkers(). Convert gene symbols to ENTREZ IDs and add to dataframe.

```
DefaultAssay(npm1) <- "SCT"
stem_cell_markers_1 <- FindMarkers(npm1, ident.1 = "HSC1_MUT", ident.2 = "HSC1_WT", only.pos = FALSE, l
stem_cell_markers_1$entrez = mapIds(org.Hs.eg.db, rownames(stem_cell_markers_1), 'ENTREZID', 'SYMBOL')
stem_cell_markers_1 = na.omit(stem_cell_markers_1)
```

Split DEG list into upregulated and downregulated genes.

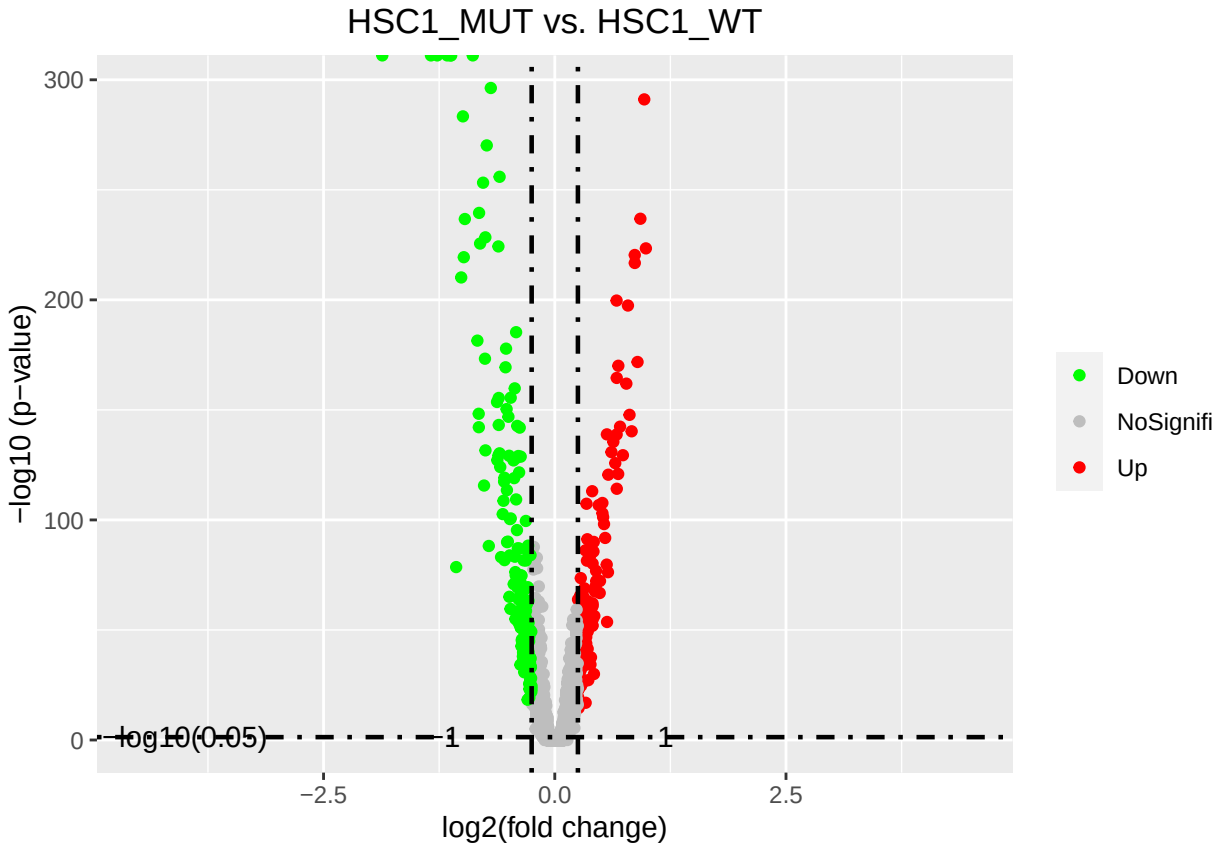
```
upreg = subset(stem_cell_markers_1, subset = avg_log2FC > 0.25 & p_val_adj < 0.01)
downreg = subset(stem_cell_markers_1, subset = avg_log2FC < -0.25 & p_val_adj < 0.01)
```

Create sorted list of upregulated and downregulated genes.

```
upreg_list <- sign(upreg$avg_log2FC)*(-log10(upreg$p_val_adj))
names(upreg_list) <- rownames(upreg)
upreg_list <- upreg_list[na.exclude(names(upreg_list))]
upreg_list <- sort(upreg_list, decreasing = T)

downreg_list <- sign(downreg$avg_log2FC)*(-log10(downreg$p_val_adj))
names(downreg_list) <- rownames(downreg)
downreg_list <- downreg_list[na.exclude(names(downreg_list))]
downreg_list <- sort(downreg_list, decreasing = T)
downreg_list = replace(downreg_list, c(which(downreg_list %in% -Inf)),-(.Machine$double.xmax/100))

full_list = sign(stem_cell_markers_1$avg_log2FC)*(-log10(stem_cell_markers_1$p_val_adj))
names(full_list) <- rownames(stem_cell_markers_1)
full_list <- full_list[na.exclude(names(full_list))]
full_list <- sort(full_list, decreasing = T)
full_list = replace(full_list, c(which(full_list %in% -Inf)),-(.Machine$double.xmax/100))
```

Volcano plot

```
Name2 = "HSC1_MUT vs. HSC1_WT"

stem_cell_markers_1$threshold <- as.factor(ifelse(stem_cell_markers_1$p_val_adj < 0.05 & abs(stem_cell_

ggplot(data=stem_cell_markers_1, aes(x=avg_log2FC, y=-log10(p_val_adj), colour=threshold)) +
geom_point(alpha=1, size=1.5) +
scale_color_manual(values=c("green", "grey", "red")) +
xlim(c(-4.5, 4.5)) +
geom_vline(xintercept=c(-.25, .25), lty=4,col="black",lwd=0.8) +
geom_hline(yintercept=-log10(0.05), lty=4,col="black",lwd=0.8) +
annotate("text", x=c(-1.2, 1.2), y=1.8, label=c("-1", "1")) +
annotate("text", x=-4, y=1.8, label="-log10(0.05)") +
labs(x="log2(fold change)", y="-log10 (p-value)", title=Name2) +
theme(plot.title=element_text(hjust=0.5), legend.position="right", legend.title=element_blank(
))
```




```
GO.title <- paste(Name2,"GO", collapse = " ")
KEGG.title <- paste(Name2,"KEGG", collapse = " ")
```

GSEA

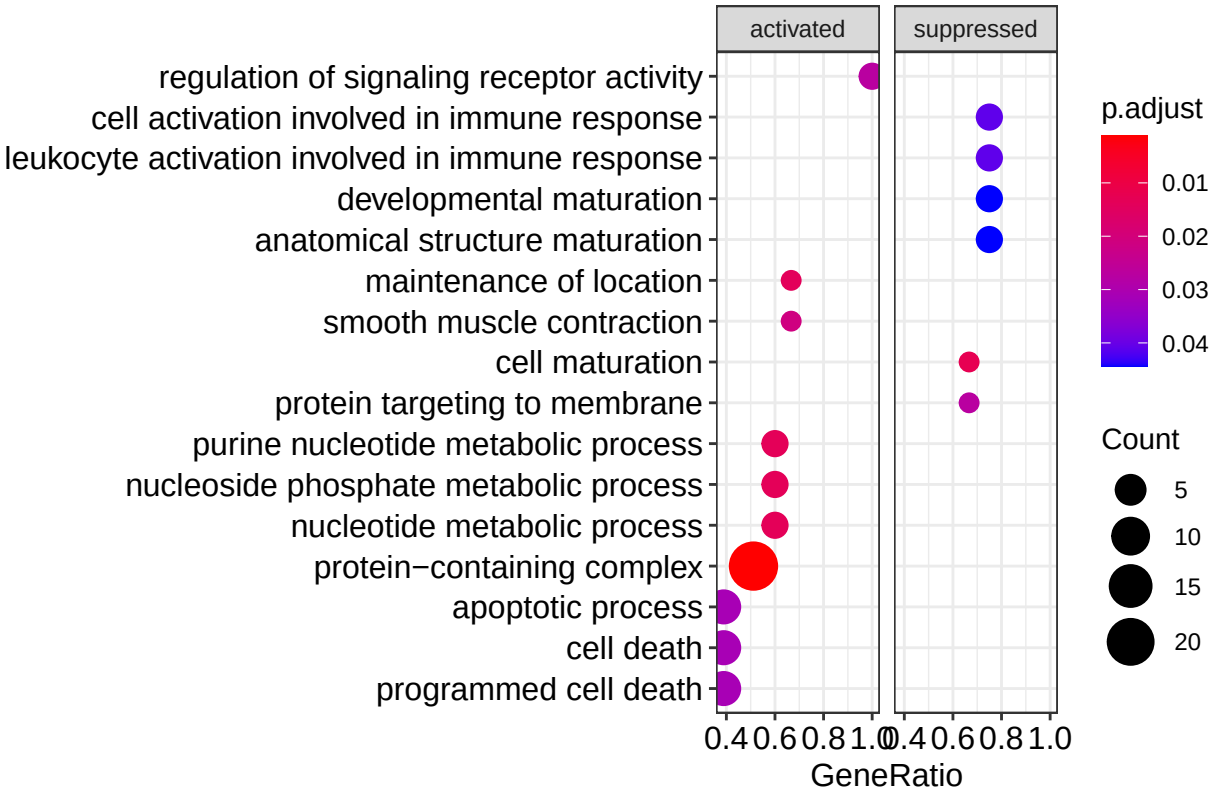
Perform GSEA on up- and downregulated genes.

```
gse_up <- gseGO(geneList=upreg_list,
              ont ="ALL",
              keyType = "SYMBOL",
              minGSSize = 3,
              maxGSSize = 800,
              pvalueCutoff = 0.05,
              verbose = TRUE,
              OrgDb = org.Hs.eg.db,
              pAdjustMethod = "none")

gse_down <- gseGO(geneList=downreg_list,
              ont ="ALL",
              keyType = "SYMBOL",
              minGSSize = 3,
              maxGSSize = 800,
              pvalueCutoff = 0.05,
              verbose = TRUE,
              OrgDb = org.Hs.eg.db,
              pAdjustMethod = "none")
```

Dot plots

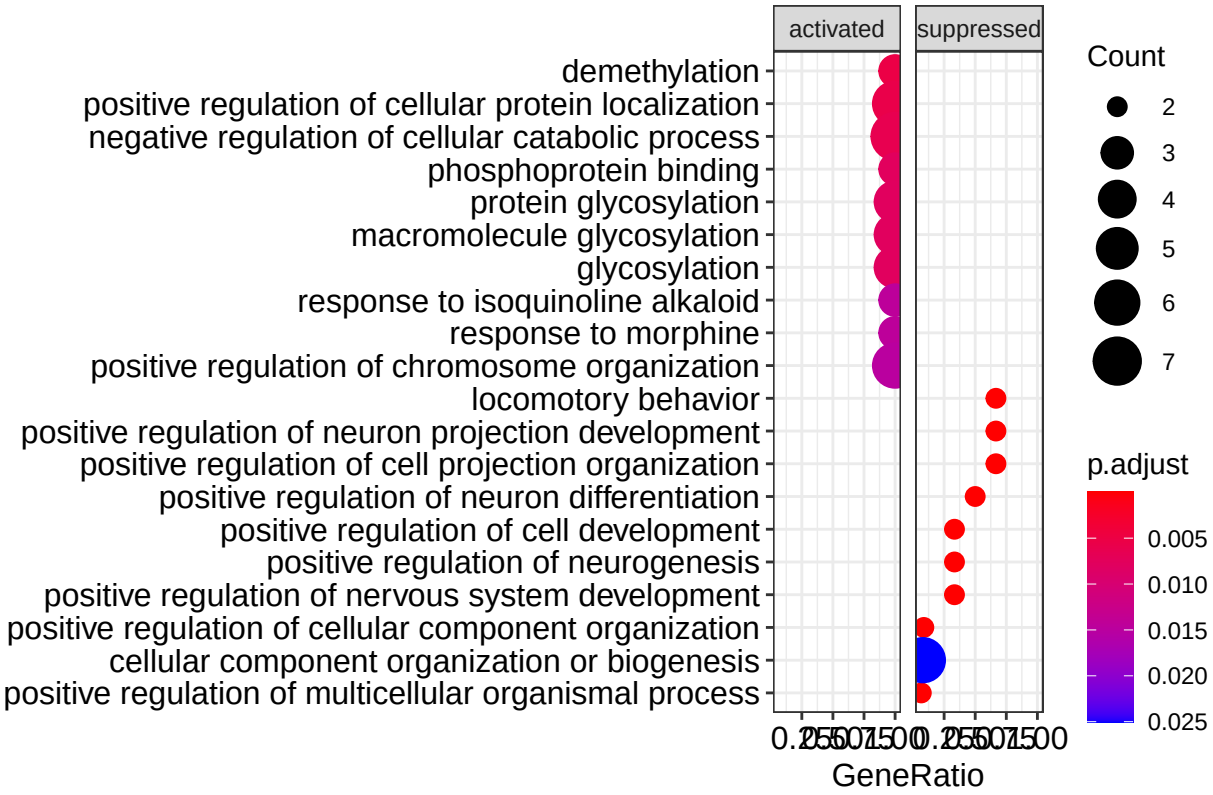
Upregulated

```
require(DOSE)
dotplot(gse_up, showCategory=10, split=".sign") + facet_grid(.~.sign)
```



Downregulated

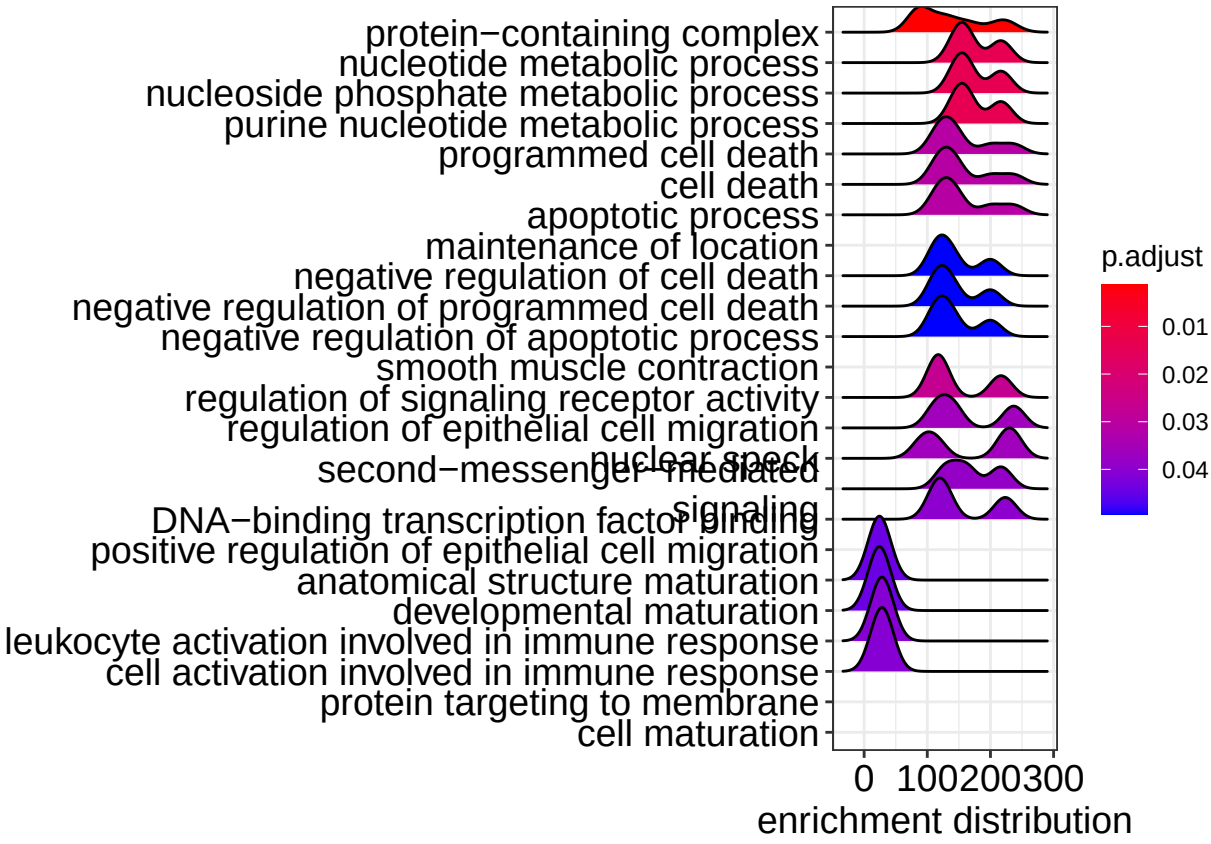
```
require(DOSE)
dotplot(gse_down, showCategory=10, split=".sign") + facet_grid(.~.sign)
```



Ridge plot (frequency of fold values per gene within each set)

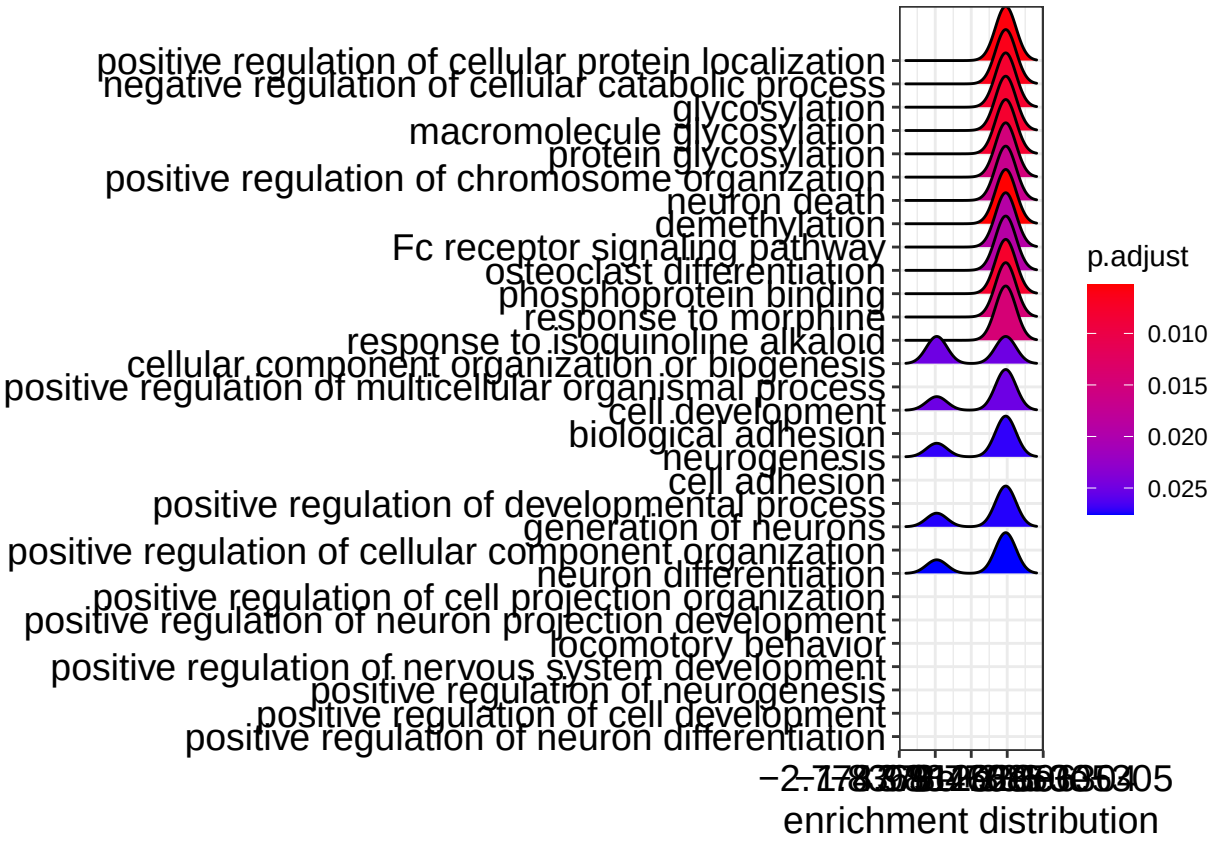
Upregulated

```
ridgeplot(gse_up) + labs(x = "enrichment distribution")
```



Downregulated

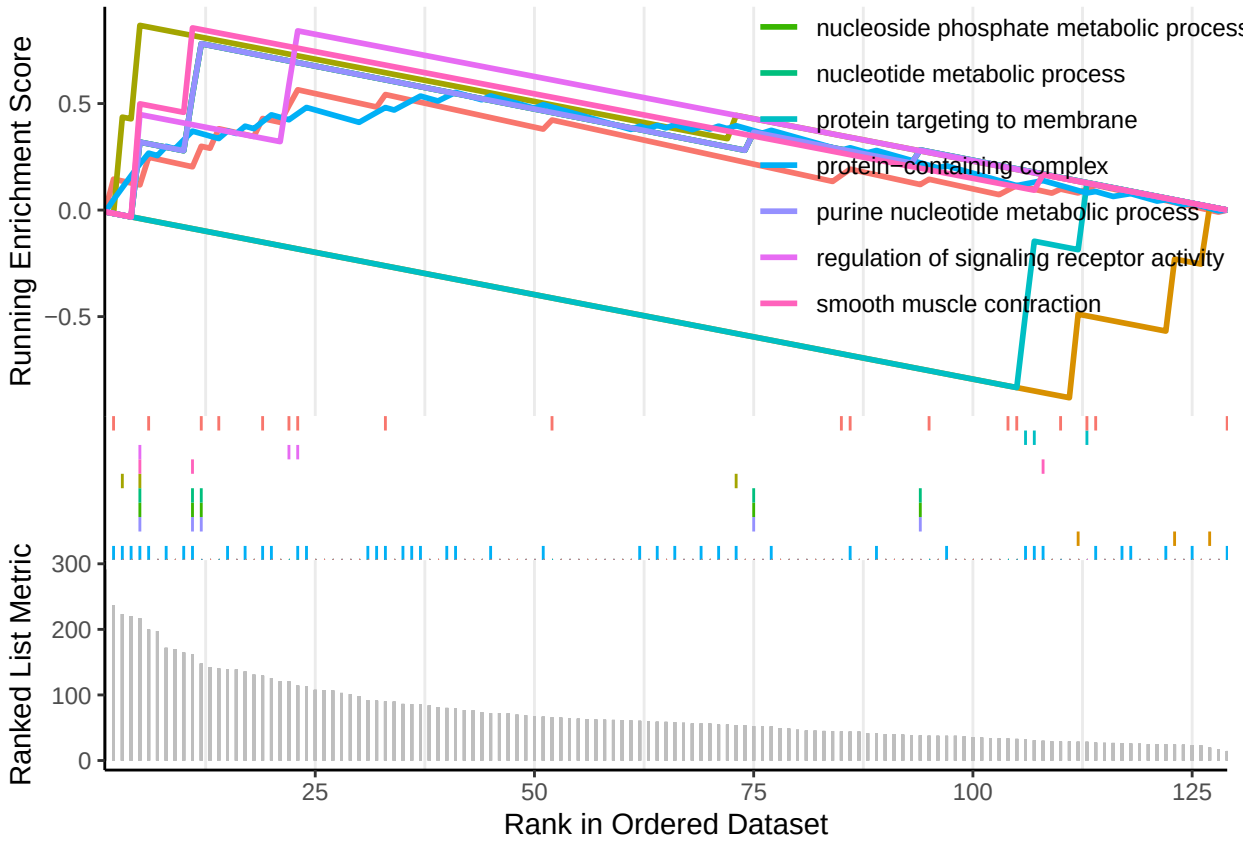
```
ridgeplot(gse_down) + labs(x = "enrichment distribution")
```



GSEA plot

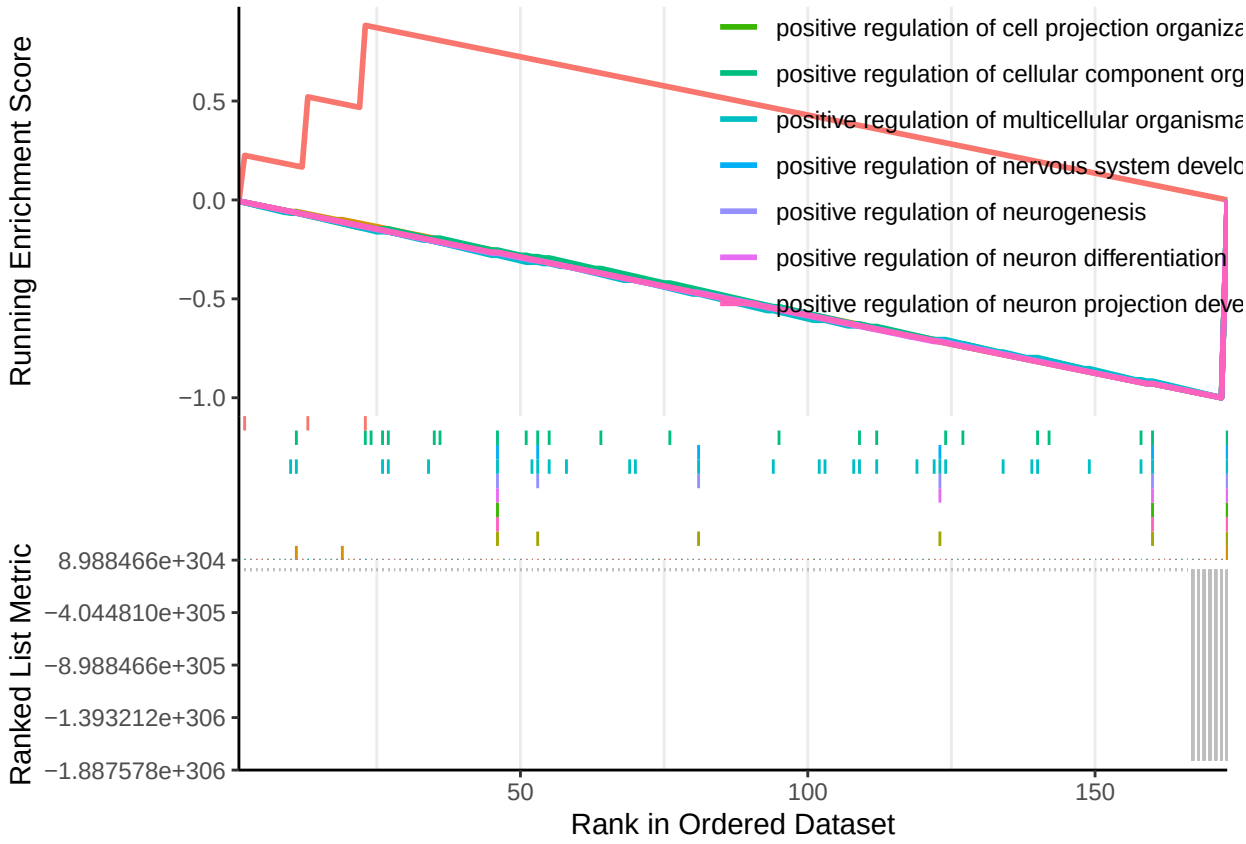
Upregulated

```
#gseaplot(gse, by = "all", title = gse$Description[3], geneSetID = 3)
gseaplot2(gse_up, geneSetID=1:10)
```



Downregulated

```
#gseaplot(gse, by = "all", title = gse$Description[3], geneSetID = 3)
gseaplot2(gse_down, geneSetID=1:10)
```



KEGG GSEA

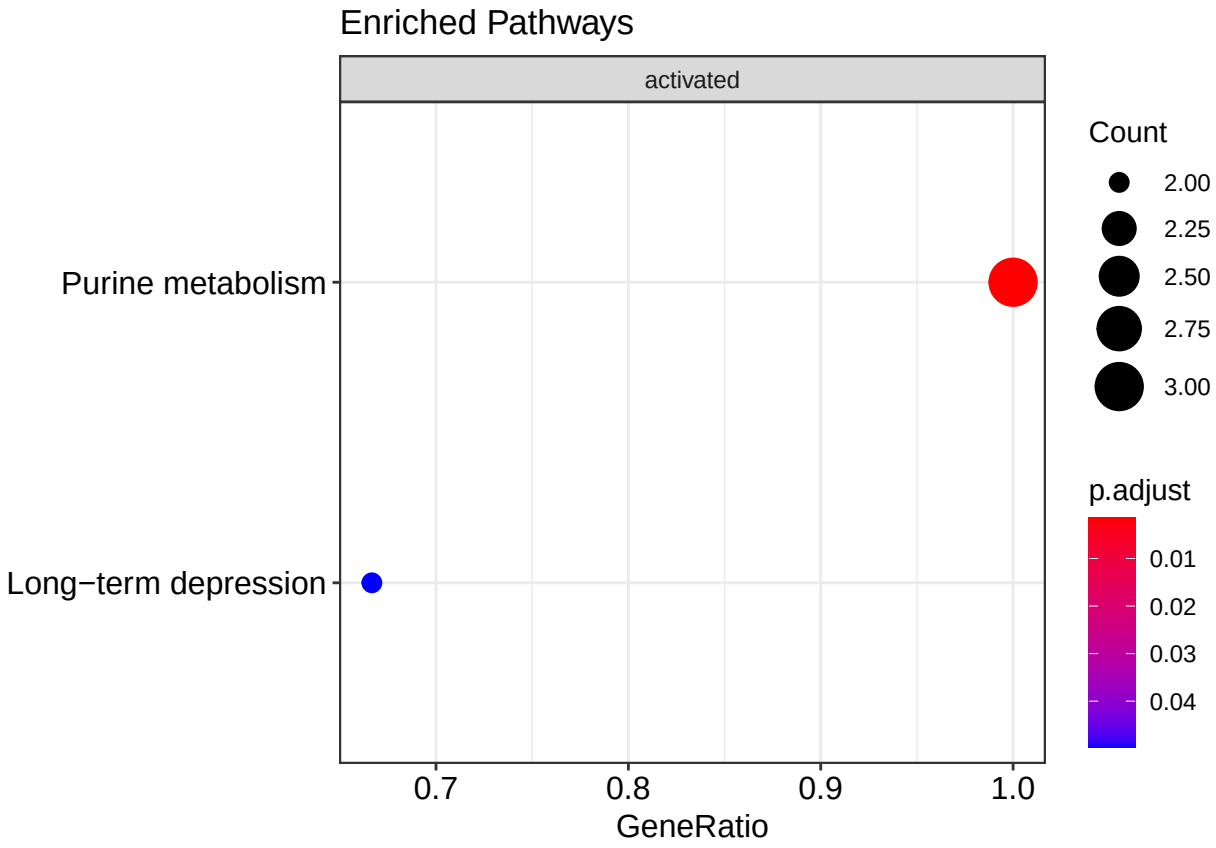
Create gseKEGG objects

```
kegg_organism = "hsa"

names(upreg_list) = mapIds(org.Hs.eg.db, names(upreg_list), 'ENTREZID', 'SYMBOL')
names(downreg_list) = mapIds(org.Hs.eg.db, names(downreg_list), 'ENTREZID', 'SYMBOL')

kk2_up <- gseKEGG(geneList  = upreg_list,
                  organism     = kegg_organism,
                  minGSSize    = 3,
                  maxGSSize    = 800,
                  pvalueCutoff = 0.05,
                  pAdjustMethod = "none",
                  keyType       = "ncbi-geneid")

kk2_down <- gseKEGG(geneList = downreg_list,
                  organism     = kegg_organism,
                  minGSSize    = 3,
                  maxGSSize    = 800,
                  pvalueCutoff = 0.05,
                  pAdjustMethod = "none",
                  keyType       = "ncbi-geneid")
```

Dot plot

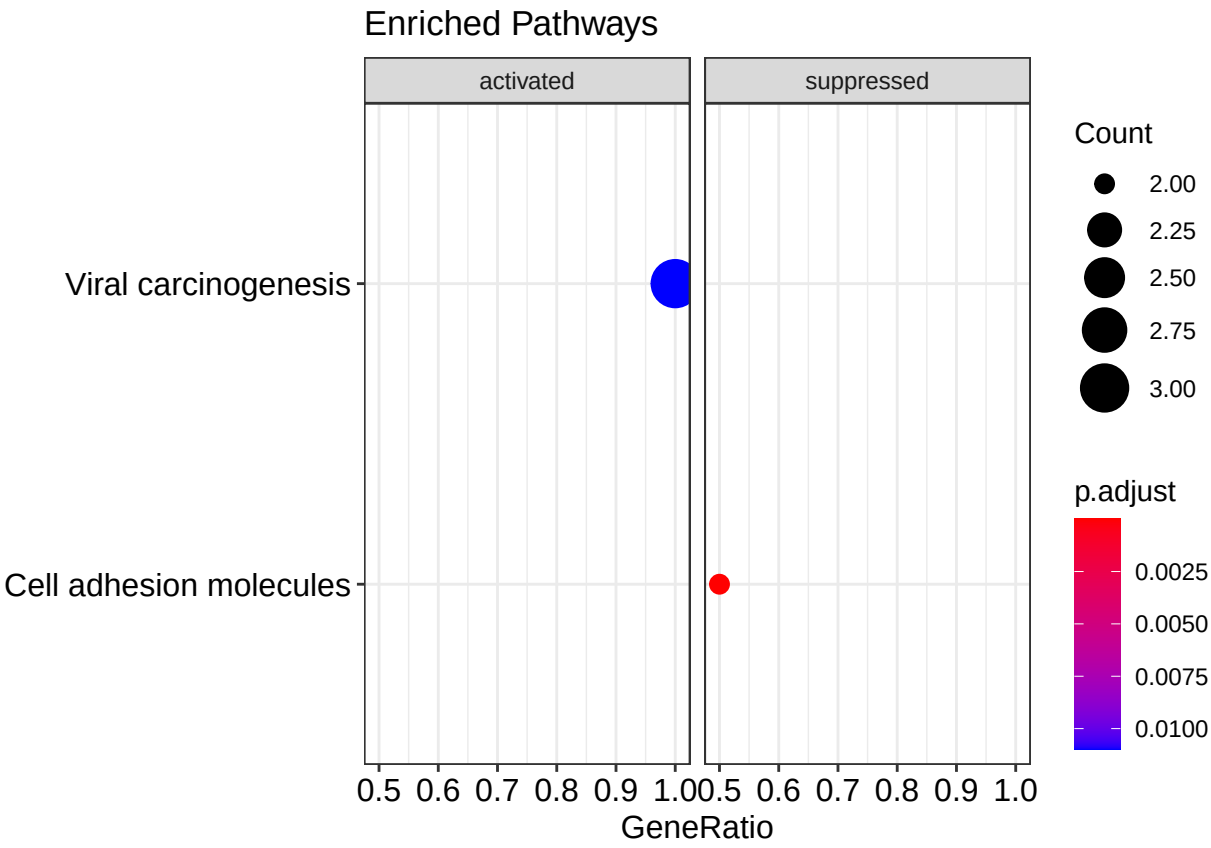
Upregulated

```
dotplot(kk2_up, showCategory = 10, title = "Enriched Pathways" , split=".sign") + facet_grid(.~.sign)
```



Downregulated

```
dotplot(kk2_down, showCategory = 10, title = "Enriched Pathways" , split=".sign") + facet_grid(.~.sign)
```



MSigDB

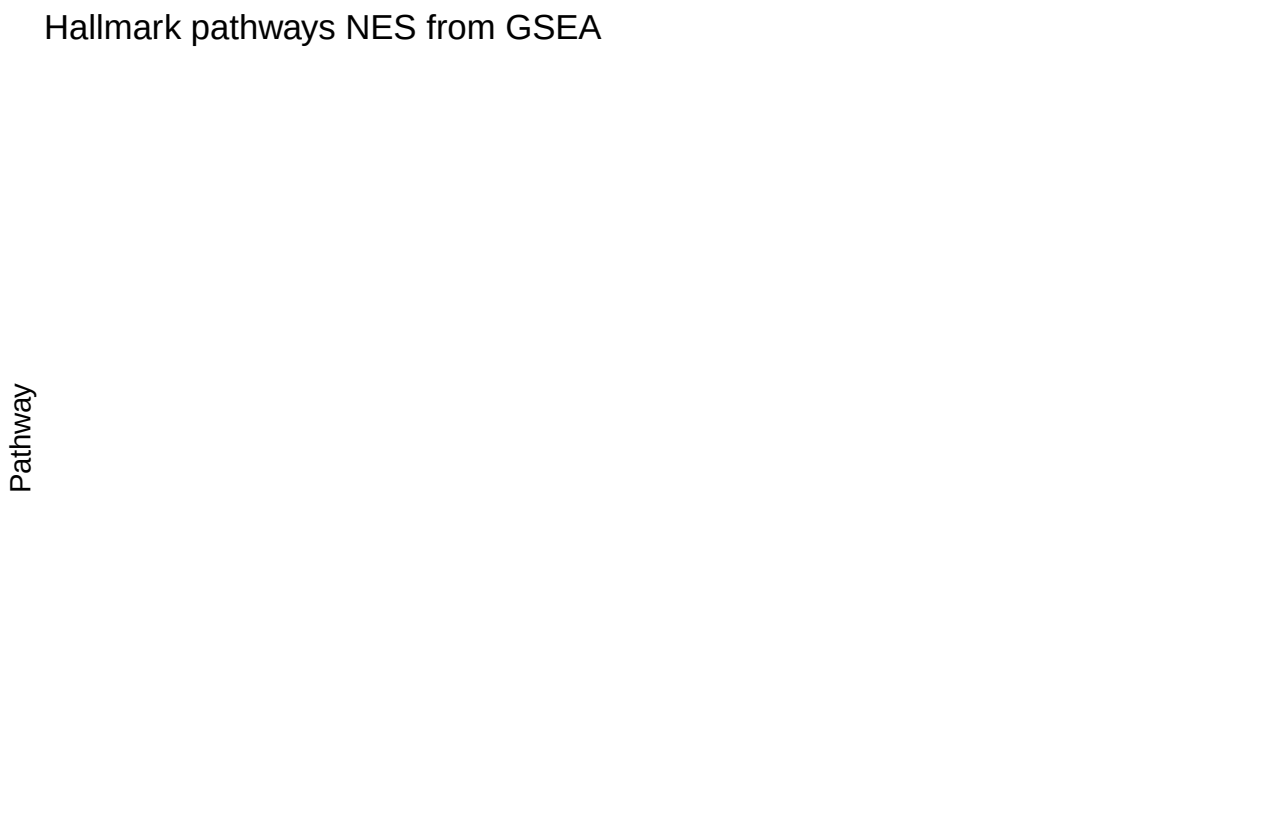
Hallmark

```
#all_gene_sets = msigdbr(species = "Homo sapiens")
h_gene_sets = msigdbr(species = "human", category = "H")
pathwaysH = split(x = h_gene_sets$entrez_gene, f = h_gene_sets$gs_name)
```

Upregulated - NONE with padj < 0.05

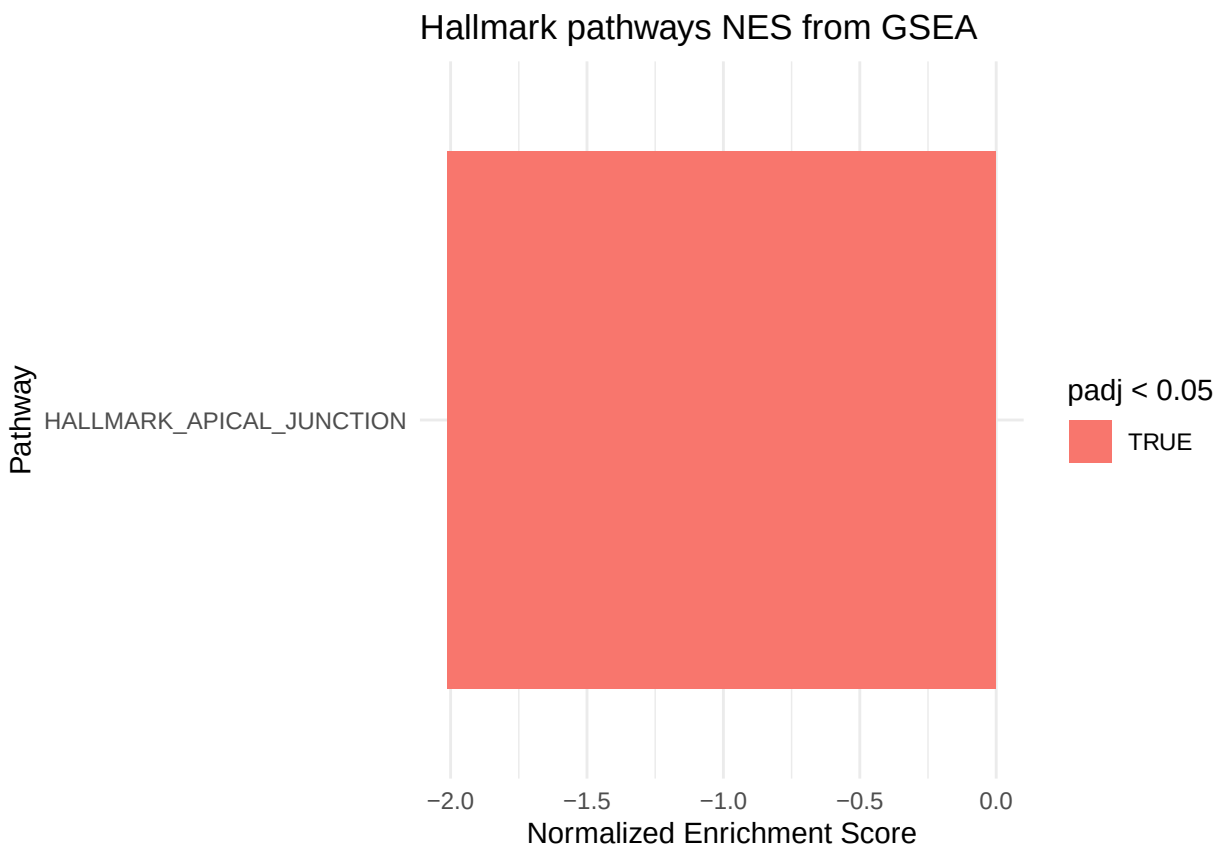
```
#names(upreg_list) = mapIds(org.Hs.eg.db, names(upreg_list), 'ENTREZID', 'SYMBOL')

fgseaRes <- fgseaMultilevel(pathways=pathwaysH, stats=upreg_list, scoreType = "pos")

fgseaResTidy <- fgseaRes %>%
  as_tibble() %>%
  arrange(desc(NES))

fgseaResTidy_sig = subset(fgseaResTidy, subset = padj < 0.05)

ggplot(fgseaResTidy_sig, aes(reorder(pathway, NES), NES)) +
  geom_col(aes(fill=padj<0.05)) +
  coord_flip() +
  labs(x="Pathway", y="Normalized Enrichment Score",
       title="Hallmark pathways NES from GSEA") +
```

```
  theme_minimal()
```

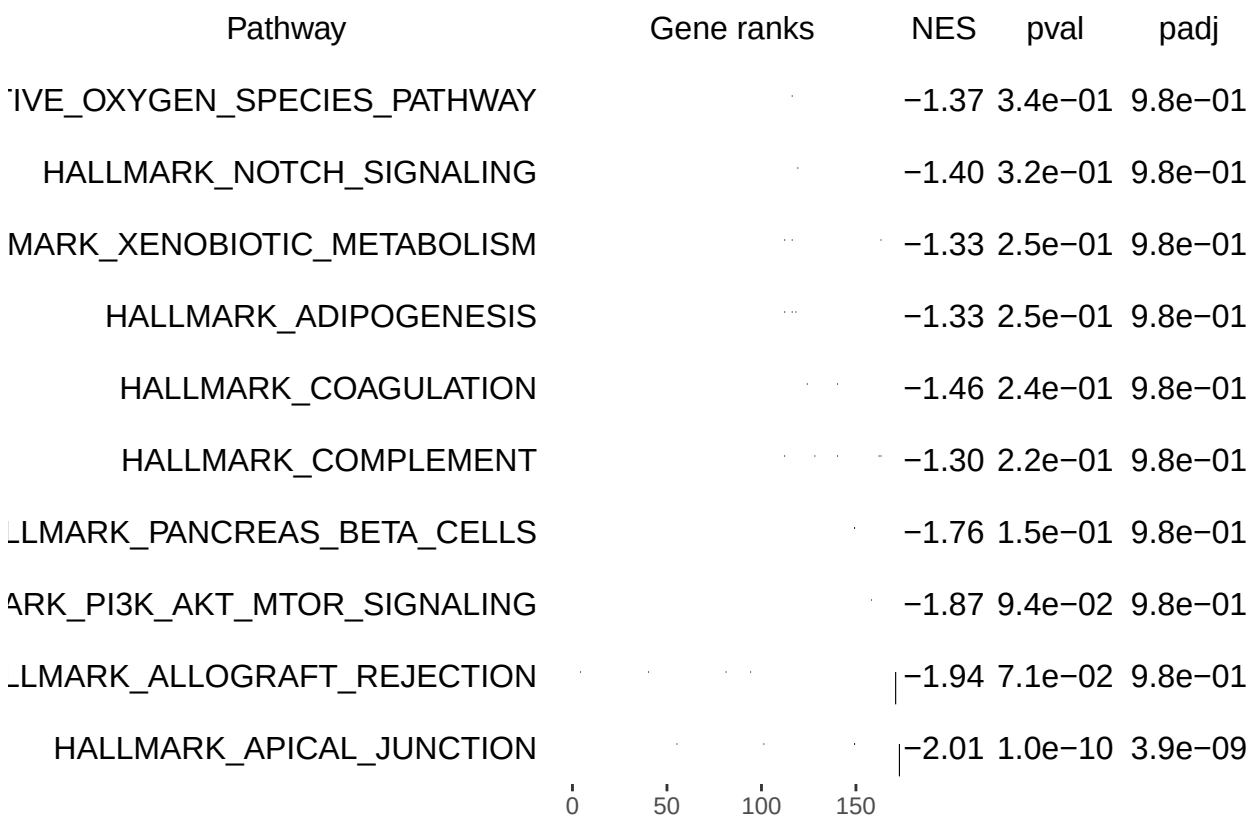


Downregulated

```
#names(downreg_list) = mapIds(org.Hs.eg.db, names(downreg_list), 'ENTREZID', 'SYMBOL')

fgseaRes <- fgseaMultilevel(pathways=pathwaysH, stats=downreg_list, scoreType = "neg")

fgseaResTidy <- fgseaRes %>%
  as_tibble() %>%
  arrange(desc(NES))

fgseaResTidy_sig = subset(fgseaResTidy, subset = padj < 0.05)

ggplot(fgseaResTidy_sig, aes(reorder(pathway, NES), NES)) +
  geom_col(aes(fill=padj<0.05)) +
  coord_flip() +
  labs(x="Pathway", y="Normalized Enrichment Score",
       title="Hallmark pathways NES from GSEA") +
  theme_minimal()
```

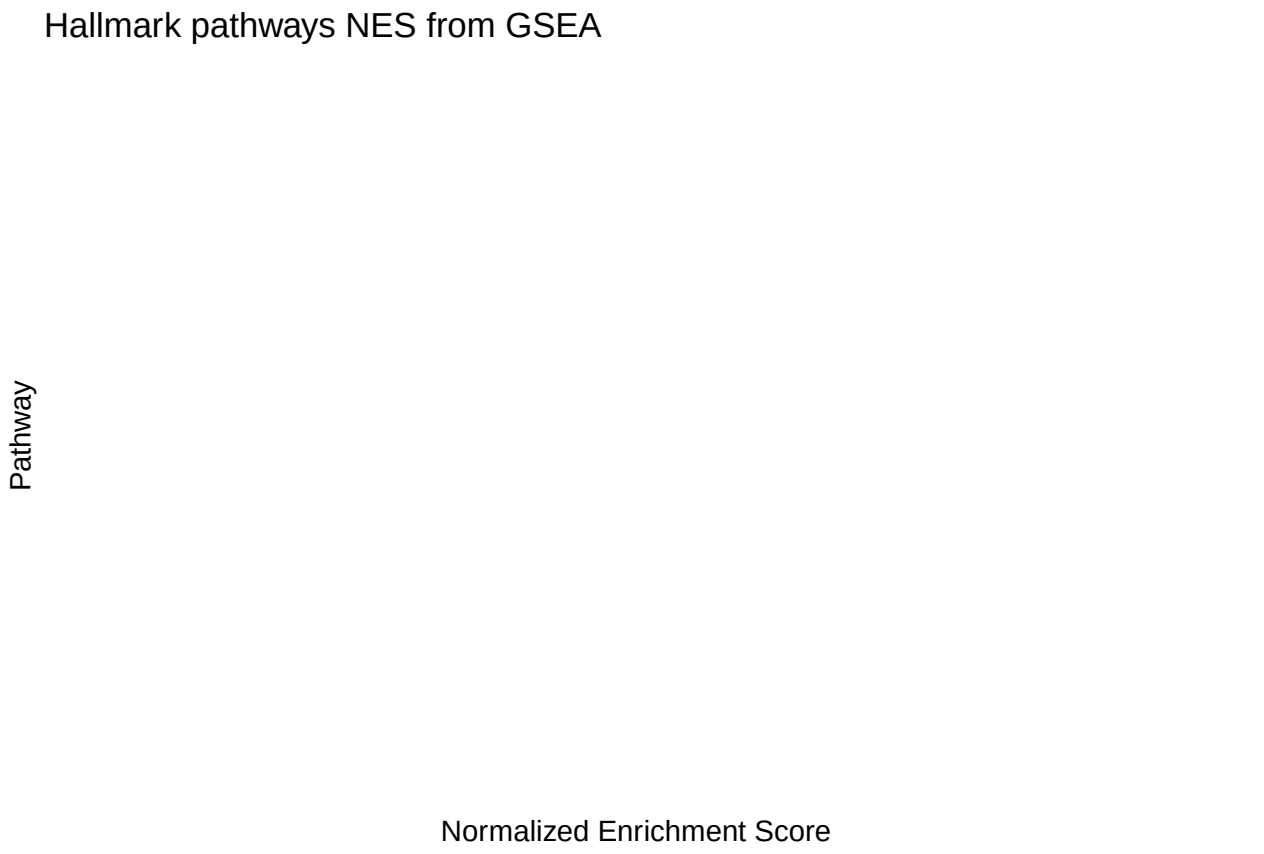



```
topPathwaysUp <- fgseaRes[ES > 0][head(order(pval), n=10), pathway]
topPathwaysDown <- fgseaRes[ES < 0][head(order(pval), n=10), pathway]
topPathways <- c(topPathwaysUp, rev(topPathwaysDown))
plotGseaTable(pathwaysH[topPathways], downreg_list, fgseaRes,
              gseaParam=0.5)
```



All

```
#names(full_list) = mapIds(org.Hs.eg.db, names(full_list), 'ENTREZID', 'SYMBOL')

fgseaRes <- fgseaMultilevel(pathways=pathwaysH, stats=full_list)

fgseaResTidy <- fgseaRes %>%
  as_tibble() %>%
  arrange(desc(NES))

fgseaResTidy_sig = subset(fgseaResTidy, subset = padj < 0.05)

ggplot(fgseaResTidy_sig, aes(reorder(pathway, NES), NES)) +
  geom_col(aes(fill=padj<0.05)) +
  coord_flip() +
  labs(x="Pathway", y="Normalized Enrichment Score",
       title="Hallmark pathways NES from GSEA") +
  theme_minimal()
```

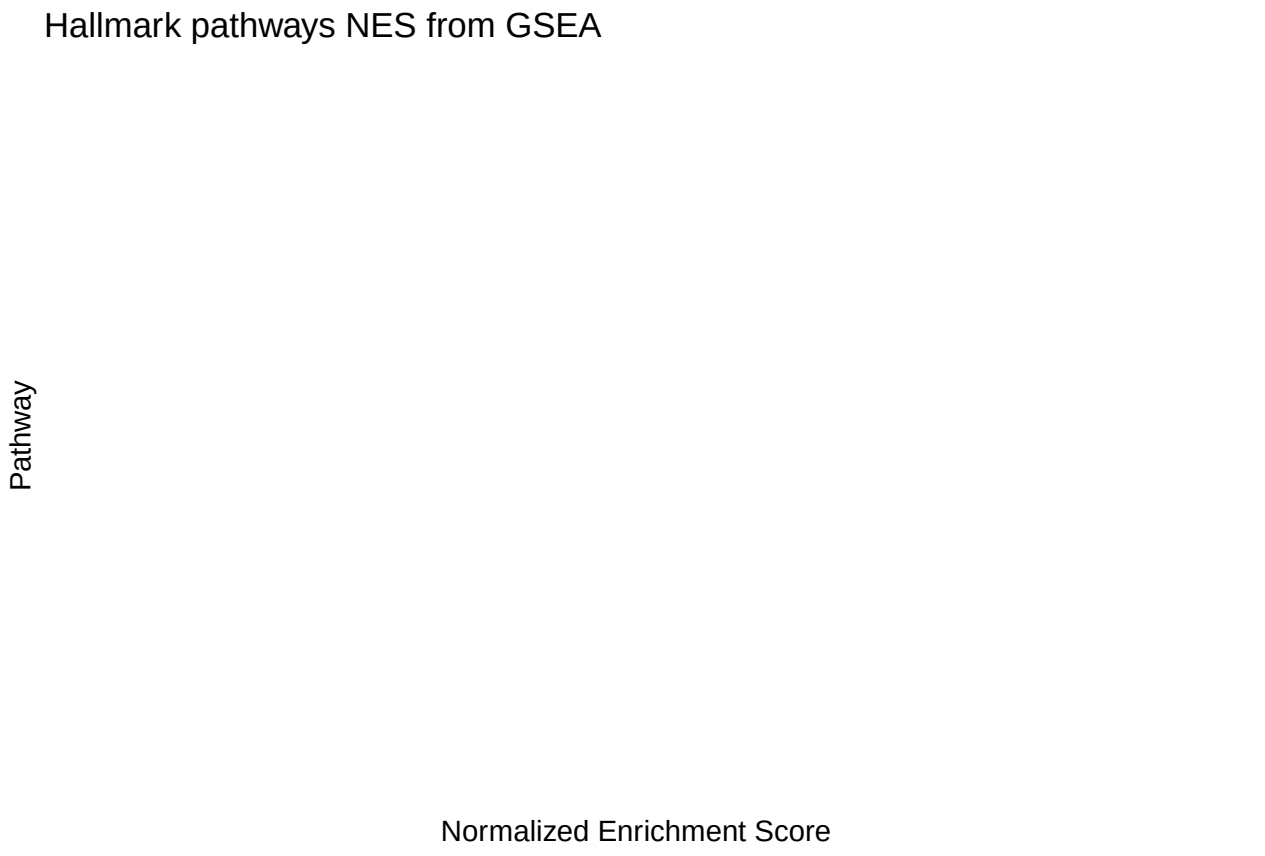


Biocarta

```
#all_gene_sets = msigdbr(species = "Homo sapiens")
b_gene_sets = msigdbr(species = "human", category = "C2", subcategory = "CP:BIOCARTA")
pathwaysB = split(x = b_gene_sets$entrez_gene, f = b_gene_sets$gs_name)
```

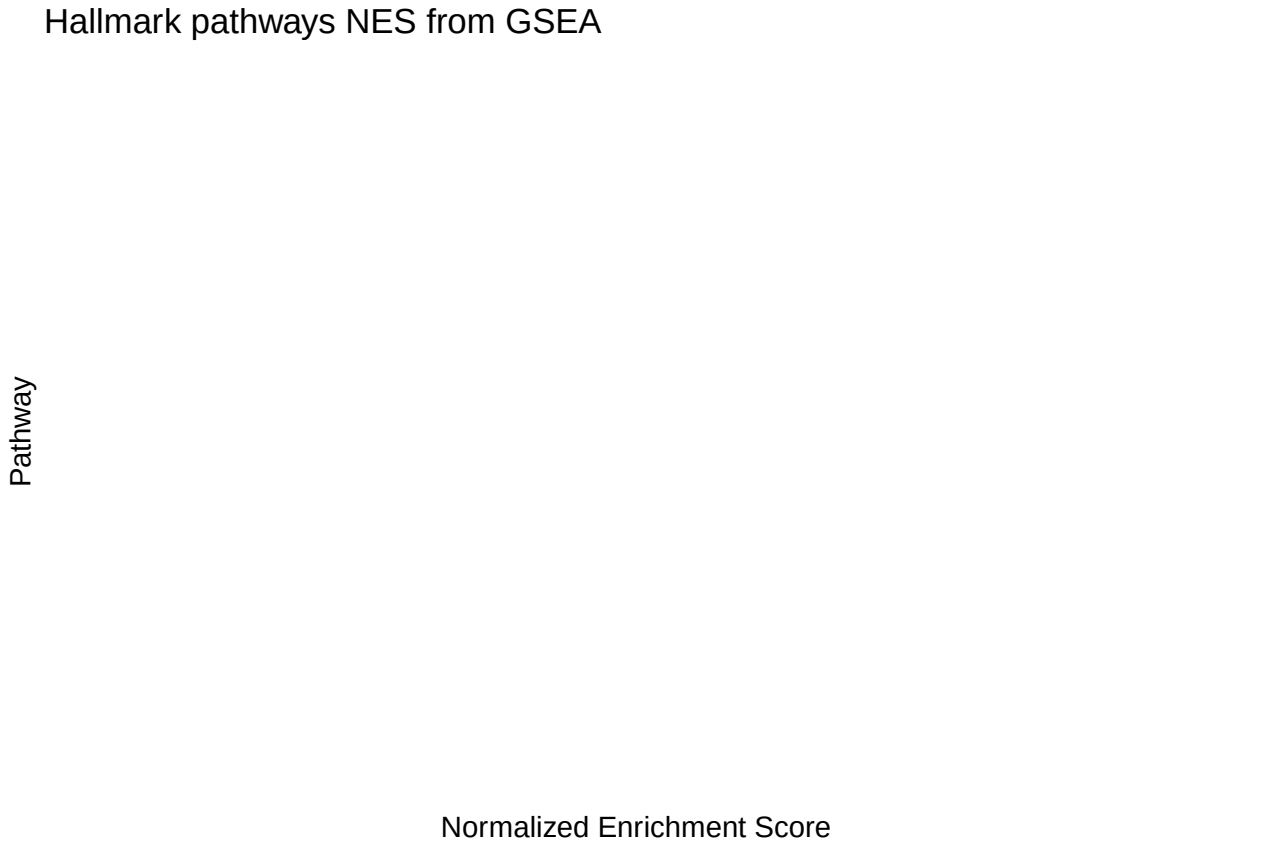
Upregulated - NONE with padj < 0.05

```
#names(upreg_list) = mapIds(org.Hs.eg.db, names(upreg_list), 'ENTREZID', 'SYMBOL')

fgseaRes <- fgseaMultilevel(pathways=pathwaysB, stats=upreg_list, scoreType = "pos")

fgseaResTidy <- fgseaRes %>%
  as_tibble() %>%
  arrange(desc(NES))

fgseaResTidy_sig = subset(fgseaResTidy, subset = padj < 0.05)

ggplot(fgseaResTidy_sig, aes(reorder(pathway, NES), NES)) +
  geom_col(aes(fill=padj<0.05)) +
  coord_flip() +
  labs(x="Pathway", y="Normalized Enrichment Score",
       title="Hallmark pathways NES from GSEA") +
  theme_minimal()
```



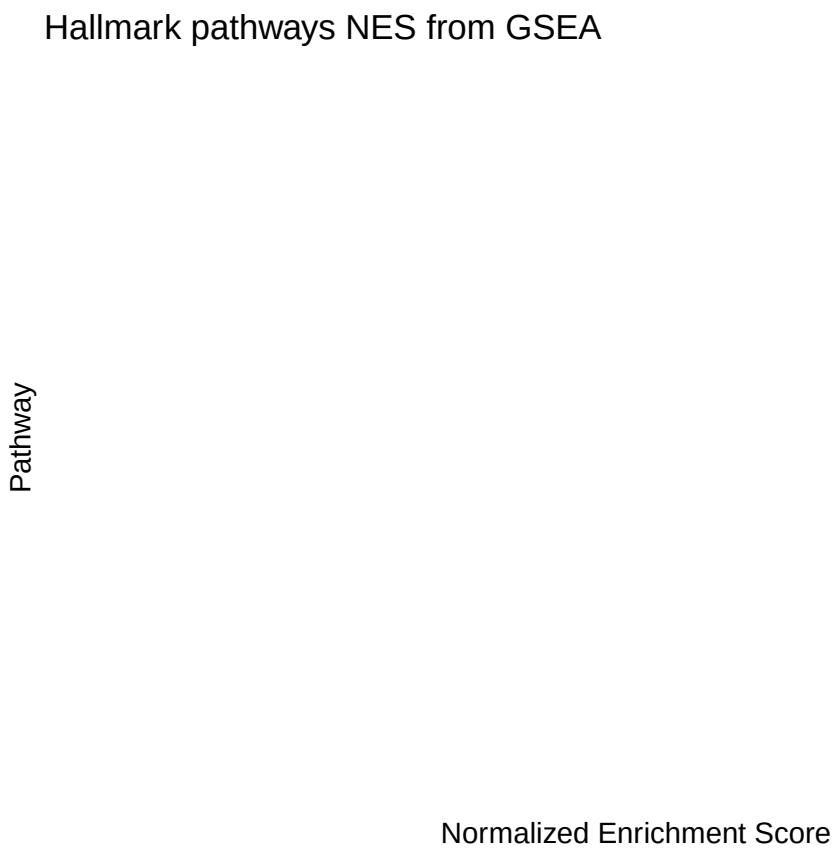
Downregulated - NONE with padj < 0.05

```
#names(downreg_list) = mapIds(org.Hs.eg.db, names(downreg_list), 'ENTREZID', 'SYMBOL')

fgseaRes <- fgseaMultilevel(pathways=pathwaysB, stats=downreg_list, scoreType = "neg")

fgseaResTidy <- fgseaRes %>%
  as_tibble() %>%
  arrange(desc(NES))

fgseaResTidy_sig = subset(fgseaResTidy, subset = padj < 0.05)

ggplot(fgseaResTidy_sig, aes(reorder(pathway, NES), NES)) +
  geom_col(aes(fill=padj<0.05)) +
  coord_flip() +
  labs(x="Pathway", y="Normalized Enrichment Score",
       title="Hallmark pathways NES from GSEA") +
  theme_minimal()
```



All - NONE with padj < 0.05

```
#names(full_list) = mapIds(org.Hs.eg.db, names(full_list), 'ENTREZID', 'SYMBOL')

fgseaRes <- fgseaMultilevel(pathways=pathwaysB, stats=full_list)

fgseaResTidy <- fgseaRes %>%
  as_tibble() %>%
  arrange(desc(NES))

fgseaResTidy_sig = subset(fgseaResTidy, subset = padj < 0.05)

ggplot(fgseaResTidy_sig, aes(reorder(pathway, NES), NES)) +
  geom_col(aes(fill=padj<0.05)) +
  coord_flip() +
  labs(x="Pathway", y="Normalized Enrichment Score",
       title="Hallmark pathways NES from GSEA") +
  theme_minimal()
```

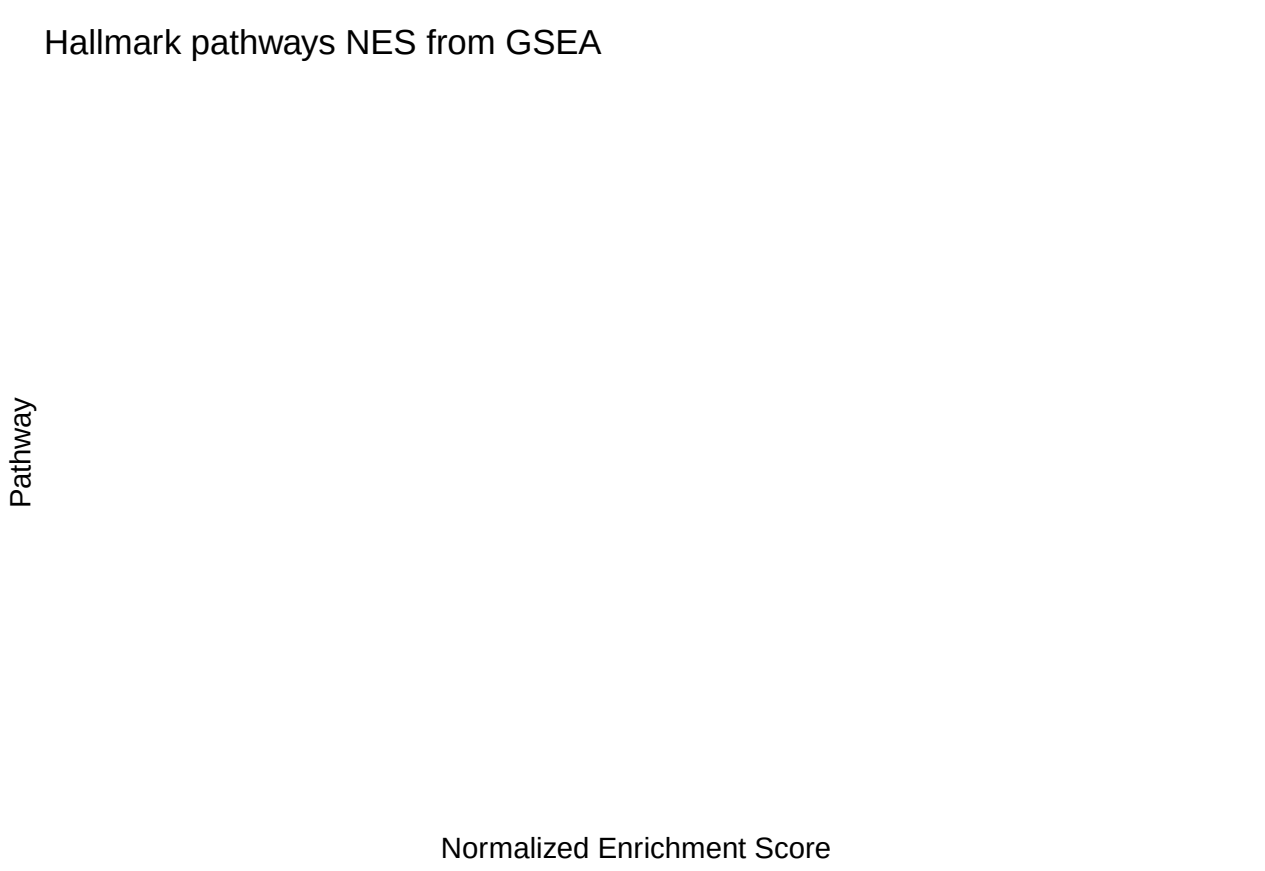


CGP

```
#all_gene_sets = msigdbr(species = "Homo sapiens")
c_gene_sets = msigdbr(species = "human", category = "C2", subcategory = "CGP")
pathwaysC = split(x = c_gene_sets$entrez_gene, f = c_gene_sets$gs_name)
```

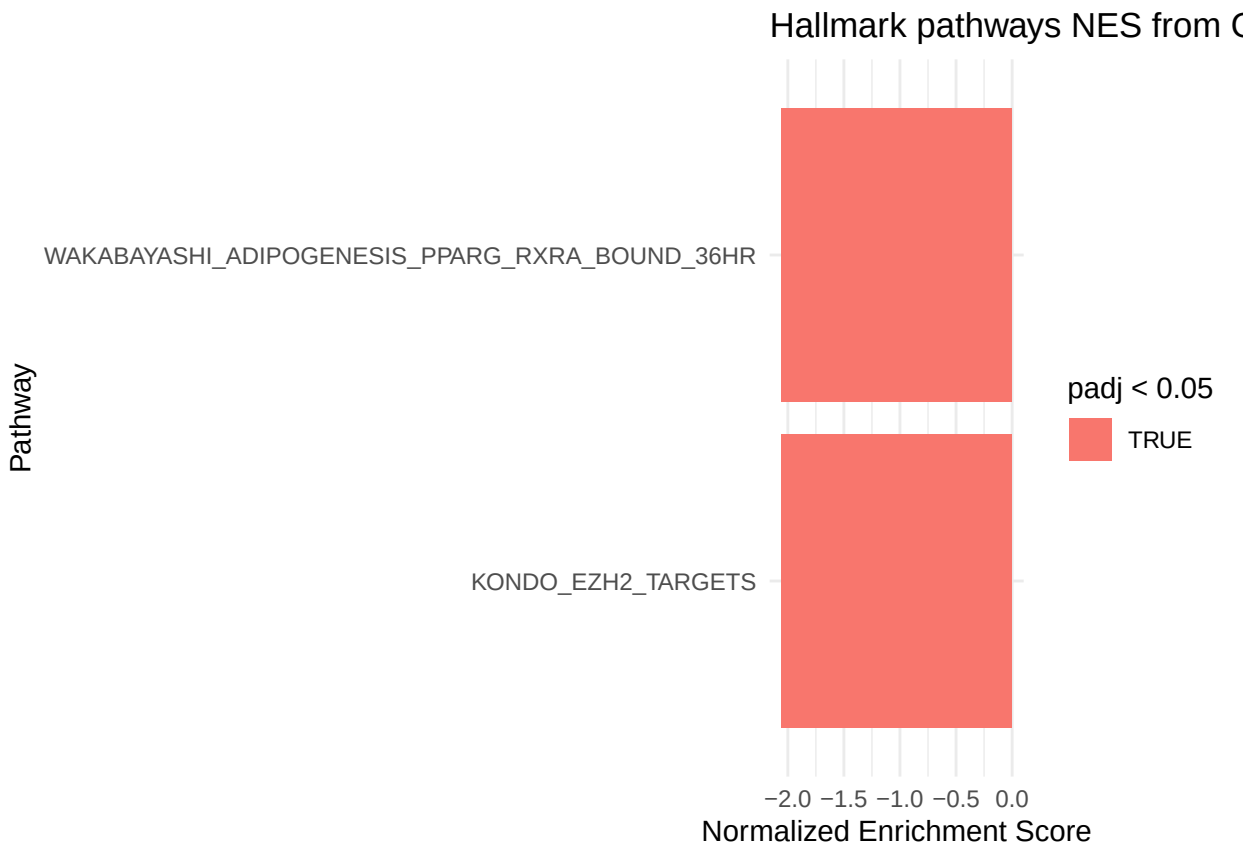
Upregulated - NONE with padj < 0.05

```
#names(upreg_list) = mapIds(org.Hs.eg.db, names(upreg_list), 'ENTREZID', 'SYMBOL')

fgseaRes <- fgseaMultilevel(pathways=pathwaysC, stats=upreg_list, scoreType = "pos")

fgseaResTidy <- fgseaRes %>%
  as_tibble() %>%
  arrange(desc(NES))

fgseaResTidy_sig = subset(fgseaResTidy, subset = padj < 0.05)

ggplot(fgseaResTidy_sig, aes(reorder(pathway, NES), NES)) +
  geom_col(aes(fill=padj<0.05)) +
  coord_flip() +
  labs(x="Pathway", y="Normalized Enrichment Score",
       title="Hallmark pathways NES from GSEA") +
  theme_minimal()
```



Downregulated

```
#names(downreg_list) = mapIds(org.Hs.eg.db, names(downreg_list), 'ENTREZID', 'SYMBOL')

fgseaRes <- fgseaMultilevel(pathways=pathwaysC, stats=downreg_list, scoreType = "neg")

fgseaResTidy <- fgseaRes %>%
  as_tibble() %>%
  arrange(desc(NES))

fgseaResTidy_sig = subset(fgseaResTidy, subset = padj < 0.05)

ggplot(fgseaResTidy_sig, aes(reorder(pathway, NES), NES)) +
  geom_col(aes(fill=padj<0.05)) +
  coord_flip() +
  labs(x="Pathway", y="Normalized Enrichment Score",
       title="Hallmark pathways NES from GSEA") +
  theme_minimal()
```



All

```
#names(full_list) = mapIds(org.Hs.eg.db, names(full_list), 'ENTREZID', 'SYMBOL')

fgseaRes <- fgseaMultilevel(pathways=pathwaysC, stats=full_list)

fgseaResTidy <- fgseaRes %>%
  as_tibble() %>%
  arrange(desc(NES))

fgseaResTidy_sig = subset(fgseaResTidy, subset = padj < 0.05)

ggplot(head(fgseaResTidy_sig, 30), aes(reorder(pathway, NES), NES)) +
  geom_col(aes(fill=padj<0.05)) +
  coord_flip() +
  labs(x="Pathway", y="Normalized Enrichment Score",
       title="Hallmark pathways NES from GSEA") +
  theme_minimal()
```


Hallmark pathways NES from GSEA
Pathway
Normalized Enrichment Score