

Balloon_plot_with_GSEA_result

Read the data for balloon plot. The data was exported after gene set enrichment analysis (<http://software.broadinstitute.org/gsea/index.jsp> (<http://software.broadinstitute.org/gsea/index.jsp>)). There are three variables: NAME (names of gene sets), NES (normalized enrichment score), and sig_genes (number of genes significantly changes in corresponding gene sets).

```
dt <- read.csv("data_for_balloon_plot.csv")
head(dt)
```

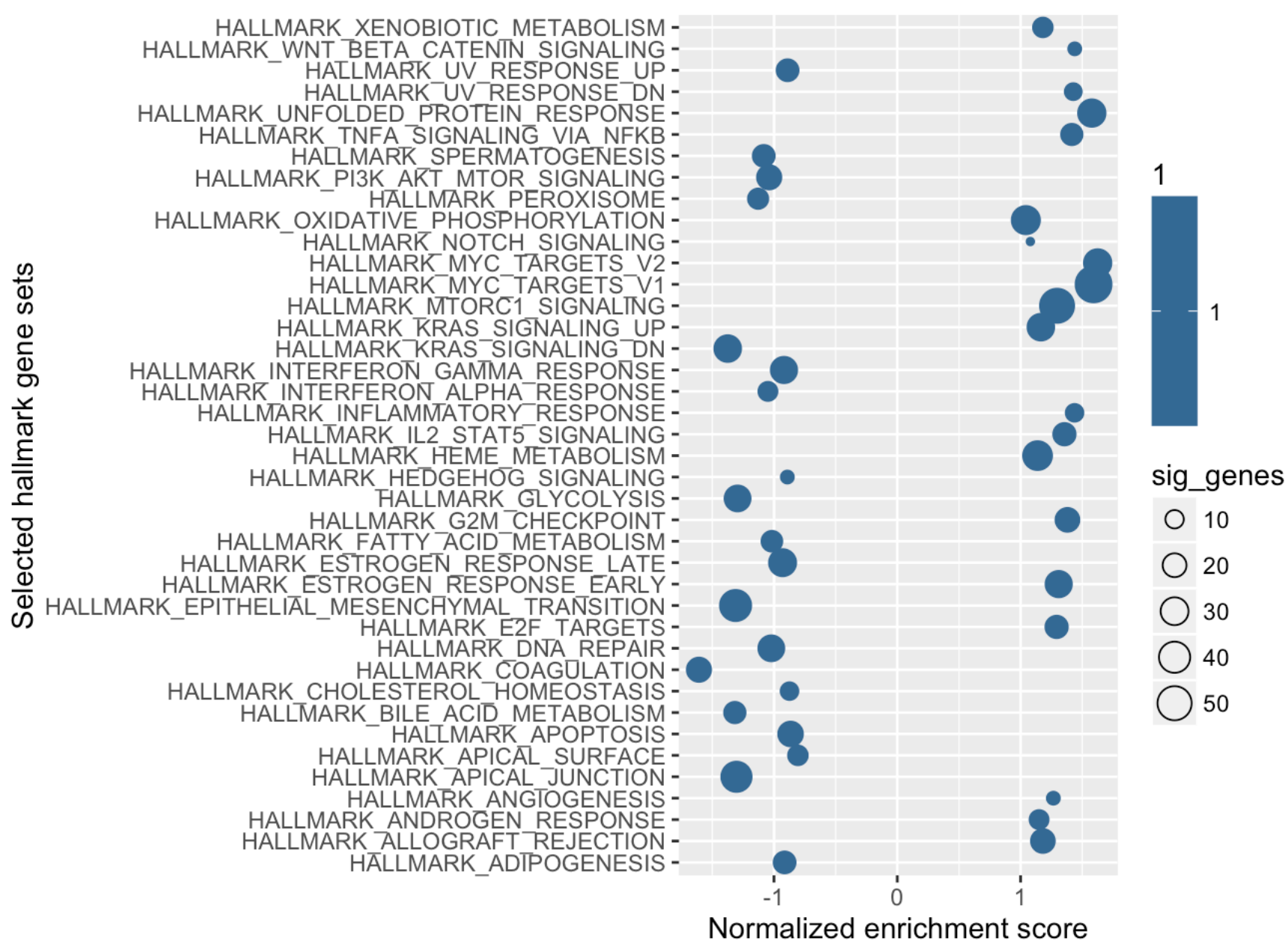
```
##              NAME      NES sig_genes
## 1  HALLMARK_MYC_TARGETS_V2 1.624757      29
## 2  HALLMARK_MYC_TARGETS_V1 1.591507      57
## 3  HALLMARK_UNFOLDED_PROTEIN_RESPONSE 1.577254      29
## 4  HALLMARK_WNT_BETA_CATENIN_SIGNALING 1.438713       4
## 5  HALLMARK_INFLAMMATORY_RESPONSE 1.437345       9
## 6  HALLMARK_UV_RESPONSE_DN 1.426889       8
```

Plot to have a look

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.3
```

```
p <- ggplot(dt, aes(x=NES, y=NAME, size=sig_genes,
                    colour=1, fill=1)) +
  geom_point(shape=21) +
  xlab("Normalized enrichment score") + ylab("Selected hallmark gene sets")
p
```



Color balloons according to p values Mark red if p value < 0.1

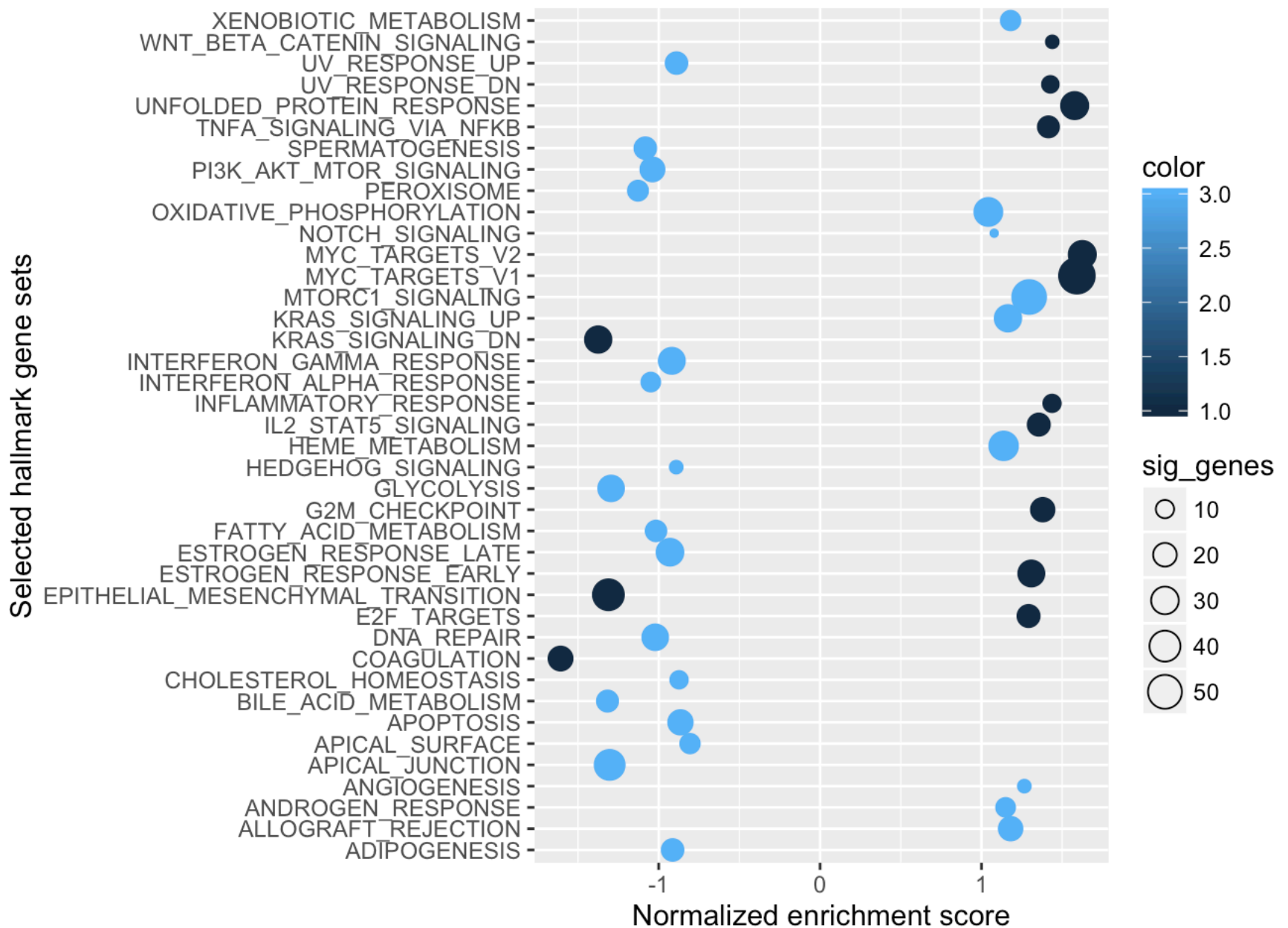
```

dt2 <- read.csv("data_for_balloon_plot_p_val.csv")
# Add color info
col <- vector()
for (i in 1:length(dt2[,1])) {
  if (dt2$NOM.p.val[i] <= 0.1) {
    col[i] <- 1
  }
  else {col[i] <- 3}
}
# Add col column to dt2
dt2$color <- col
# remove "HALLMARK_" in names
dt2$NAME <- gsub("^.*?_", "", dt2[,1])

# plot again
p <- ggplot(dt2, aes(x=NES, y=NAME, size=sig_genes, colour=color, fill=color)) +
  geom_point(shape=21) +
  xlab("Normalized enrichment score") + ylab("Selected hallmark gene sets")

```

p



Details to improve next. Color and color legends.