Enrichment analysis for breast cancer

Loading required libraries

```
library(DESeq2)
library(clusterProfiler)
library(org.Hs.eg.db)
library(pathview)
library(biomaRt)
library(ReactomePA)
library(ggplot2)
library(RColorBrewer)
library(ggsci)
```

Read count matrix into CSV

```
read_count_data <- function(file_path){
  counts_data <- read.csv(file_path, row.names = 1)
  expression_data <- round(counts_data)
  return(expression_data)
}

expression_matrix <- read_count_data("../Enrichment analysis for breast cancer/Data/GSE183947_fpkm.csv")

head(expression_matrix,2)
```

```
##        tumor.rep1 tumor.rep2 tumor.rep3 tumor.rep4 tumor.rep5 tumor.rep6
## TSPAN6          1          2          0          5          5          5
## TNMD            0          0          0          0          0          0
##        tumor.rep7 tumor.rep8 tumor.rep9 tumor.rep10 tumor.rep11 tumor.rep12
## TSPAN6          4          4          6          12           6           4
## TNMD            0          0          0           0           0           0
##        tumor.rep13 tumor.rep14 tumor.rep15 tumor.rep16 tumor.rep17 tumor.rep18
## TSPAN6           8          11           4          14          10           7
## TNMD             0           1           0           0           0           0
##        tumor.rep19 tumor.rep20 tumor.rep21 tumor.rep22 tumor.rep23 tumor.rep24
## TSPAN6           6           7           9           7          10           7
## TNMD             0           0           0           0           0           0
##        tumor.rep25 tumor.rep26 tumor.rep27 tumor.rep28 tumor.rep29 tumor.rep30
## TSPAN6           5           2           5           5           9           2
## TNMD             0           0           0           1           0           1
##        normal.rep1 normal.rep2 normal.rep3 normal.rep4 normal.rep5 normal.rep6
## TSPAN6          12           3          13          15           7           0
## TNMD             6           2           0           2           0           0
##        normal.rep7 normal.rep8 normal.rep9 normal.rep10 normal.rep11
## TSPAN6          10           7           5           6           6
## TNMD             0           0          11           0           3
##        normal.rep12 normal.rep13 normal.rep14 normal.rep15 normal.rep16
```

```
## TSPAN6            7           11           16           12           12
## TNMD              0            0            0            0            1
##        normal.rep17 normal.rep18 normal.rep19 normal.rep20 normal.rep21
## TSPAN6           10            9            7            9            7
## TNMD              1            0            0            0            0
##        normal.rep22 normal.rep23 normal.rep24 normal.rep25 normal.rep26
## TSPAN6            8            9            6            6            6
## TNMD              0            1            0            0            0
##        normal.rep27 normal.rep28 normal.rep29 normal.rep30
## TSPAN6            4            5           10            5
## TNMD              9            1            0            0
```

Read metadata into CSV

```
read_metadata <- function(file_path){
  coldata <- read.csv(file_path, row.names = 1)
  return (coldata)
}

meta_data <- read_metadata("../Enrichment analysis for breast cancer/Data/metadata.csv")
head(meta_data)
```

```
##            condition description
## tumor rep1     tumor   CA.102548
## tumor rep2     tumor   CA.104338
## tumor rep3     tumor   CA.105094
## tumor rep4     tumor   CA.109745
## tumor rep5     tumor  CA.1906415
## tumor rep6     tumor  CA.1912627
```

Convert condition column in metadata to factor

```
meta_data$condition <- as.factor(meta_data$condition)
meta_data$description <- as.factor(meta_data$description)
```

Make sure the row names in metadata matches to the column names in expression matrix

```
all(rownames(meta_data) %in% colnames(expression_matrix))
```

```
## [1] FALSE
```

Match the row names in metadata to the column names in expression matrix

```
rownames(meta_data) = colnames(expression_matrix)
```

Pre-filtering: removing rows with low gene counts

keep rows that have at least 10 reads total

```
pre_filter <- function(){
  # Only keep rows that have total counts above the cutoff
  keep <- expression_matrix %>% rowSums(.) >= 10
  filtered_counts <- expression_matrix[keep,]
  return (filtered_counts)
}
filtered_expression_counts <- pre_filter()
head(filtered_expression_counts,2)
```

```
##        tumor.rep1 tumor.rep2 tumor.rep3 tumor.rep4 tumor.rep5 tumor.rep6
## TSPAN6          1          2          0          5          5          5
## TNMD            0          0          0          0          0          0
##        tumor.rep7 tumor.rep8 tumor.rep9 tumor.rep10 tumor.rep11 tumor.rep12
## TSPAN6          4          4          6          12           6           4
## TNMD            0          0          0           0           0           0
##        tumor.rep13 tumor.rep14 tumor.rep15 tumor.rep16 tumor.rep17 tumor.rep18
## TSPAN6           8          11           4          14          10           7
## TNMD             0           1           0           0           0           0
##        tumor.rep19 tumor.rep20 tumor.rep21 tumor.rep22 tumor.rep23 tumor.rep24
## TSPAN6           6           7           9           7          10           7
## TNMD             0           0           0           0           0           0
##        tumor.rep25 tumor.rep26 tumor.rep27 tumor.rep28 tumor.rep29 tumor.rep30
## TSPAN6           5           2           5           5           9           2
## TNMD             0           0           0           1           0           1
##        normal.rep1 normal.rep2 normal.rep3 normal.rep4 normal.rep5 normal.rep6
## TSPAN6          12           3          13          15           7           0
## TNMD             6           2           0           2           0           0
##        normal.rep7 normal.rep8 normal.rep9 normal.rep10 normal.rep11
## TSPAN6          10           7           5           6           6
## TNMD             0           0          11           0           3
##        normal.rep12 normal.rep13 normal.rep14 normal.rep15 normal.rep16
## TSPAN6            7           11           16           12           12
## TNMD              0            0            0            0            1
##        normal.rep17 normal.rep18 normal.rep19 normal.rep20 normal.rep21
## TSPAN6           10            9            7            9            7
## TNMD              1            0            0            0            0
##        normal.rep22 normal.rep23 normal.rep24 normal.rep25 normal.rep26
## TSPAN6            8            9            6            6            6
## TNMD              0            1            0            0            0
##        normal.rep27 normal.rep28 normal.rep29 normal.rep30
## TSPAN6            4            5           10            5
## TNMD              9            1            0            0
```

Construct a DESeqDataSet.

```
deseqdataset <- function(){
  deseqdataset <- DESeqDataSetFromMatrix(countData = filtered_expression_counts,
                                         colData = meta_data,
                                         design = ~ condition)

  return(deseqdataset)
}

deseqdataset_object <- deseqdataset()
```

```
## converting counts to integer mode
```

```
deseqdataset_object
```

```
## class: DESeqDataSet
## dim: 19687 60
## metadata(1): version
## assays(1): counts
## rownames(19687): TSPAN6 TNMD ... RP11-474G23.1 AC005358.1
## rowData names(0):
## colnames(60): tumor.rep1 tumor.rep2 ... normal.rep29 normal.rep30
## colData names(2): condition description
```

Differential expression analysis

```
diff_expr_analysis <- function(){
  deseq_analysis <- DESeq(deseqdataset_object)
  result <- results(deseq_analysis)
  return (result)
}

deseq_result <- diff_expr_analysis()
```

```
## estimating size factors

## estimating dispersions

## gene-wise dispersion estimates

## mean-dispersion relationship

## -- note: fitType='parametric', but the dispersion trend was not well captured by the
##    function: y = a/x + b, and a local regression fit was automatically substituted.
##    specify fitType='local' or 'mean' to avoid this message next time.
```

```
## final dispersion estimates

## fitting model and testing

## -- replacing outliers and refitting for 944 genes
## -- DESeq argument 'minReplicatesForReplace' = 7
## -- original counts are preserved in counts(dds)

## estimating dispersions

## fitting model and testing
```

```
deseq_result
```

```
## log2 fold change (MLE): condition tumor vs normal
## Wald test p-value: condition tumor vs normal
## DataFrame with 19687 rows and 6 columns
##                baseMean log2FoldChange     lfcSE       stat     pvalue
##               <numeric>      <numeric> <numeric>  <numeric>  <numeric>
## TSPAN6         7.021979      -0.393064  0.195072   -2.01497 0.04390819
## TNMD           0.741177      -2.611175  1.063323   -2.45568 0.01406202
## DPM1           6.355119       0.403925  0.277570    1.45522 0.14560966
## SCYL3          6.478091       0.533370  0.177740    3.00084 0.00269238
## C1orf112       8.698340       0.274895  0.204442    1.34461 0.17875208
## ...                 ...            ...       ...        ...        ...
## ZBTB8B         8.161848     -0.1272060  0.305882 -0.4158666 0.67750761
## RP11-1084J3.4  0.184986      0.0967778  1.681015  0.0575711 0.95409030
## FLJ00388       0.214750     -0.4802992  2.414921 -0.1988882 0.84235023
## RP11-474G23.1  0.203916     -0.1917594  1.356854 -0.1413265 0.88761204
## AC005358.1     0.642168     -2.8026881  1.087189 -2.5779211 0.00993967
##                    padj
##               <numeric>
## TSPAN6        0.1023107
## TNMD          0.0415335
## DPM1          0.2530477
## SCYL3         0.0110673
## C1orf112      0.2939838
## ...                 ...
## ZBTB8B        0.7660112
## RP11-1084J3.4        NA
## FLJ00388             NA
## RP11-474G23.1        NA
## AC005358.1    0.0314222
```

Convert DESeq result into DataFrame

```
df_deseq_result <- as.data.frame(deseq_result)
```

Extract differentially expressed genes that have padj <= 0.01

```
sig_genes <- function(){
significant_genes <- df_deseq_result[df_deseq_result$padj <= 0.01,] %>% na.omit(significant_genes)
ordered_sig_genes <- significant_genes[order(significant_genes$padj, decreasing = FALSE), ]
return(ordered_sig_genes)
}

sign_genes <- sig_genes()
head(sign_genes)
```

```
##         baseMean log2FoldChange     lfcSE      stat       pvalue         padj
## DEFB130 13.34575      -4.633728 0.2701234 -17.15412 5.855508e-66 1.084557e-61
## LCN6    35.01702      -3.524980 0.2097024 -16.80944 2.081287e-63 1.927480e-59
## CCDC177 12.20217      -4.314425 0.2680149 -16.09770 2.647458e-58 1.634541e-54
## MDGA2   17.26324      -2.859002 0.2012085 -14.20915 8.039282e-46 3.722590e-42
## SOX7    85.13424      -2.776340 0.1980979 -14.01499 1.262172e-44 4.675591e-41
## KLK9    15.08985      -3.979799 0.2906030 -13.69497 1.088078e-42 3.358896e-39
```

Write significant genes into CSV file

```
write_sig_genes <- function(out_path){
  write.csv(sign_genes, file = out_path )
}

write_sig_genes("../Enrichment analysis for breast cancer/outputs/significant_genes.csv")
```

Convert Gene SYMBOLs to ENTREZ IDs

```
entrez_ids <- function(){
# copy the rownames of significant genes and store it in gene_names
gene_names <- rownames(sign_genes)
# convert gene names into ENTREZID
entrez_ids <- mapIds(org.Hs.eg.db,
                     keys = gene_names,
                     column = "ENTREZID",
                     keytype = "SYMBOL",
                     multiVals = "first")

# create column named ENTREZID in sign_genes that contain ENTREZID
# of significant genes
sign_genes$ENTREZID <- entrez_ids

# remove ENTREZID that contain NA
sign_genes <- sign_genes[!is.na(sign_genes$ENTREZID), ]
return(sign_genes)
```

```
}

signficant <- entrez_ids()
```

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

```
head(signficant)
```

```
##         baseMean log2FoldChange     lfcSE      stat       pvalue         padj
## LCN6    35.01702      -3.524980 0.2097024 -16.80944 2.081287e-63 1.927480e-59
## CCDC177 12.20217      -4.314425 0.2680149 -16.09770 2.647458e-58 1.634541e-54
## MDGA2   17.26324      -2.859002 0.2012085 -14.20915 8.039282e-46 3.722590e-42
## SOX7    85.13424      -2.776340 0.1980979 -14.01499 1.262172e-44 4.675591e-41
## KLK9    15.08985      -3.979799 0.2906030 -13.69497 1.088078e-42 3.358896e-39
## UGT2A1  10.74854      -2.658262 0.2027555 -13.11068 2.860340e-39 7.568459e-36
##         ENTREZID
## LCN6      158062
## CCDC177    56936
## MDGA2     161357
## SOX7       83595
## KLK9      284366
## UGT2A1     10941
```

Up-regulated significant genes

```
up_sig_genes <- function(){
  upregulated_genes <- signficant[signficant$log2FoldChange>0 , ]
  return(upregulated_genes)
}

upregulated_sig_genes <- up_sig_genes()
head(upregulated_sig_genes,5)
```

```
##         baseMean log2FoldChange     lfcSE      stat       pvalue         padj
## MYBL2  12.758600       3.205769 0.3136118 10.222092 1.578957e-24 1.169817e-21
## E2F1    8.894549       2.548282 0.2552266  9.984390 1.783952e-23 1.180084e-20
## CXCL10 11.179072       3.760349 0.3999615  9.401776 5.364966e-21 2.614998e-18
## MMP11  96.083267       3.434542 0.3687198  9.314776 1.222105e-20 5.520934e-18
## NUSAP1 16.450466       2.393583 0.2583756  9.263967 1.969752e-20 8.291763e-18
##        ENTREZID
## MYBL2      4605
## E2F1       1869
## CXCL10     3627
## MMP11      4320
## NUSAP1    51203
```

ENTREZIDs of up-regulated significant genes

```
up_entrez_ids <- upregulated_sig_genes$ENTREZID
head(up_entrez_ids)
```

```
## [1] "4605"   "1869"   "3627"   "4320"   "51203"  "124222"
```

Write up-regulated significant genes into CSV file

```
write_up_sig_genes <- function(out_path){
  write.csv(upregulated_sig_genes, file = out_path )
}

write_up_sig_genes("../Enrichment analysis for breast cancer/outputs/up-regulated_significant_genes.csv
```

Down-regulated significant genes

```
down_sig_genes <- function(){
  downregulated_genes <- signficant[signficant$log2FoldChange < 0 , ]
  return(downregulated_genes)
}

downregulated_sig_genes <- down_sig_genes()
head(downregulated_sig_genes,5)
```

```
##         baseMean log2FoldChange     lfcSE      stat       pvalue         padj
## LCN6    35.01702      -3.524980 0.2097024 -16.80944 2.081287e-63 1.927480e-59
## CCDC177 12.20217      -4.314425 0.2680149 -16.09770 2.647458e-58 1.634541e-54
## MDGA2   17.26324      -2.859002 0.2012085 -14.20915 8.039282e-46 3.722590e-42
## SOX7    85.13424      -2.776340 0.1980979 -14.01499 1.262172e-44 4.675591e-41
## KLK9    15.08985      -3.979799 0.2906030 -13.69497 1.088078e-42 3.358896e-39
##         ENTREZID
## LCN6      158062
## CCDC177    56936
## MDGA2     161357
## SOX7       83595
## KLK9      284366
```

ENTREZIDs of down-regulated significant genes

```
down_entrez_ids <- downregulated_sig_genes$ENTREZID
head(down_entrez_ids)
```

```
## [1] "158062" "56936"  "161357" "83595"  "284366" "10941"
```

Write down-regulated significant genes into CSV file

```
write_down_sig_genes <- function(out_path){
  write.csv(downregulated_sig_genes, file = out_path )
}

write_down_sig_genes("../Enrichment analysis for breast cancer/outputs/down-regulated_significant_genes
```

Gene Ontology

Group up-regulated significant genes that have similar BP GO terms

```
go_up <- function(){
    go <- groupGO( gene = up_entrez_ids,
                   OrgDb = org.Hs.eg.db,
                   ont = "BP",  # Biological Process
                   readable = TRUE)
    return(go)

}

go_terms_up <- go_up()
head(go_terms_up)
```

```
##                    ID           Description Count GeneRatio
## GO:0000003 GO:0000003          reproduction   135  135/1650
## GO:0002376 GO:0002376 immune system process   272  272/1650
## GO:0008152 GO:0008152     metabolic process  1088 1088/1650
## GO:0009987 GO:0009987      cellular process  1464 1464/1650
## GO:0016032 GO:0016032         viral process    75   75/1650
## GO:0022414 GO:0022414  reproductive process   134  134/1650
##
## GO:0000003
## GO:0002376
## GO:0008152
## GO:0009987 MYBL2/E2F1/CXCL10/MMP11/NUSAP1/MTHFD2/TK1/CCNB1/COL10A1/IDH2/PYCR1/PITX1/TROAP/LMNB1/CBX2/
## GO:0016032
## GO:0022414
```

Convert GO terms of up-regulated significant genes to DataFrame

```
dataframe_go_up <- function(){
  df_go_terms_up <- as.data.frame(go_terms_up)
  return(df_go_terms_up)
}

df_go_group_up <- dataframe_go_up()
```

```r
df_go_group_up_top7 <- head(df_go_group_up,7)
```

Barplot of up-regulated BP GO terms

```r
bar_plot_up <- function(){
p <- ggplot(df_go_group_up_top7, aes(x = reorder(Description, - Count),
                                     y = Count, fill = Description)) +
                                     geom_bar(stat = "identity") +
                                     ggtitle("BP of up-regulated GO terms") +
                                     coord_flip() +
                                     theme_bw() +
                                     scale_fill_jama()+
                                     theme(plot.title = element_text(size = 12,
                                                                     face = "bold",
                                                                     hjust = 0.5))+
                                     xlab("Description")


  jpeg("../Enrichment analysis for breast cancer/outputs/BPgroup_up_barplot.jpeg")
  print(p)
  dev.off()

}

bar_plot_up()
```

```
## pdf
##   2
```

Group down-regulated significant genes that have similar BP GO terms

```r
go_down <- function(){
  go <- groupGO(gene = down_entrez_ids,
                OrgDb = org.Hs.eg.db,
                ont = "BP",
                readable = TRUE)
  return(go)
}

go_terms_down <- go_down()
head(go_terms_down)
```

```
##                    ID           Description Count GeneRatio
## GO:0000003 GO:0000003          reproduction   217  217/2336
## GO:0002376 GO:0002376 immune system process   269  269/2336
## GO:0008152 GO:0008152     metabolic process  1234 1234/2336
## GO:0009987 GO:0009987      cellular process  1947 1947/2336
## GO:0016032 GO:0016032         viral process    29   29/2336
```

```
## GO:0022414 GO:0022414  reproductive process   216  216/2336
##
## GO:0000003
## GO:0002376
## GO:0008152
## GO:0009987 MDGA2/SOX7/UGT2A1/OC90/GOLGA7B/SYNPO2/SIGLEC5/ZNF709/DES/MYH11/LTB4R2/ABCA8/FGF10/NOVA1/MI
## GO:0016032
## GO:0022414
```

Convert GO terms of down-regulated significant genes to DataFrame

```
dataframe_go_down <- function(){
  df_go_terms_down <- as.data.frame(go_terms_down)
  return(df_go_terms_down)
}

df_go_group_down <- dataframe_go_down()
```

```
df_go_group_down_top7 <- head(df_go_group_down,7)
```

Barplot of down-regulated BP GO terms

```
bar_plot_down <- function(){

p <- ggplot(df_go_group_down_top7, aes(x = reorder(Description, - Count),
                                       y = Count, fill = Description)) +
                                       geom_bar(stat = "identity") +
                                       ggtitle("BP of down-regulated GO terms")+
                                       coord_flip() +
                                       theme_bw() +
                                       scale_fill_jama()+
                                      theme(plot.title = element_text(size = 12,
                                                                      face = "bold",
                                                                      hjust = 0.5))+
                                       xlab("Description")


  jpeg("../Enrichment analysis for breast cancer/outputs/BPgroup_down_barplot.jpeg")
  print(p)
  dev.off()

}
bar_plot_down()
```

```
## pdf
##   2
```

Over-representation analysis

Go enrichment analysis

Enriched GO terms among up-regulated significant genes

```
enrich_go_up <- function(){
        ego_up <- enrichGO( gene          = up_entrez_ids,
                            universe      = signficant$ENTREZID,
                            OrgDb         = org.Hs.eg.db,
                            keyType       = "ENTREZID",
                            ont           = "BP",
                            pvalueCutoff  = 0.05,
                            qvalueCutoff  = 0.01,
                            pAdjustMethod = "BH",
                            readable      = TRUE)
        return(ego_up)

}

enrichment_go_up <- enrich_go_up()
head(enrichment_go_up)
```

```
##                    ID                          Description GeneRatio  BgRatio
## GO:0051276 GO:0051276              chromosome organization  126/1544 151/3609
## GO:0006259 GO:0006259                DNA metabolic process  172/1544 243/3609
## GO:0007059 GO:0007059               chromosome segregation   89/1544 107/3609
## GO:0098813 GO:0098813       nuclear chromosome segregation   75/1544  88/3609
## GO:0007049 GO:0007049                           cell cycle  263/1544 430/3609
## GO:0000070 GO:0000070 mitotic sister chromatid segregation   54/1544  58/3609
##                  pvalue     p.adjust       qvalue
## GO:0051276 8.948182e-26 2.780200e-22 2.372681e-22
## GO:0006259 6.294809e-20 9.778986e-17 8.345591e-17
## GO:0007059 3.033977e-18 3.142189e-15 2.681610e-15
## GO:0098813 9.602902e-17 7.459054e-14 6.365713e-14
## GO:0007049 2.470995e-16 1.535476e-13 1.310408e-13
## GO:0000070 3.634273e-16 1.881948e-13 1.606094e-13
##
## GO:0051276
## GO:0006259
## GO:0007059
## GO:0098813
## GO:0007049 MYBL2/E2F1/NUSAP1/TK1/CCNB1/AURKB/PLK1/CDCA8/MCM2/FANCA/KIFC1/KIF2C/STMN1/FOXM1/MCM4/ZWIN
## GO:0000070
##            Count
## GO:0051276   126
## GO:0006259   172
## GO:0007059    89
## GO:0098813    75
## GO:0007049   263
## GO:0000070    54
```

Dataframe of enriched GO terms among up-regulated significant genes

```
df_go_terms_up <- function(){
  df_go_term_up <- as.data.frame(enrichment_go_up)
  return(df_go_term_up)
}

df_go_up <- df_go_terms_up()
```

```
df_go_up_top7 <- head(df_go_up, 7)
```

BarPlot of up-regulated BP enriched GO terms

```
bar_plot_enriched_up <- function() {

  jama_colors <- pal_jama("default")(7)
  p <- ggplot(df_go_up_top7, aes(x = reorder(Description, -Count),
                                 y = Count,
                                 fill = p.adjust)) +
       geom_bar(stat = "identity") +
       coord_flip() +
       scale_fill_gradientn(name = "p.adjust", colors = jama_colors) +
       labs(title = "Bar plot of BP of Enriched Up-regulated GO terms",
            x = "GO Term",
            y = "Gene Count") +
       theme_bw() +
       theme(plot.title = element_text(size = 10, face = "bold", hjust = 0.5))

  jpeg("../Enrichment analysis for breast cancer/outputs/BP_enriched_up_barplot.jpeg")
  print(p)
  dev.off()
}

bar_plot_enriched_up()
```

```
## pdf
##   2
```

Dotplot of up-regulated BP enriched GO terms

```
dot_plot_enriched_up <- function(){
  jama_colors <- pal_jama("default")(7)
  p <- ggplot(df_go_up_top7, aes(x = reorder(Description, -Count),
                                 y = Count, size = Count, color = p.adjust)) +
    geom_point(alpha = 0.6) +
    coord_flip() +
    scale_size_continuous(range = c(3, 8), name = "Gene Count") +
    scale_color_gradientn(name = "p.adjust", colors = jama_colors) +
```

```
    labs(title = "Dot plot of BP of enriched up-regulated genes",
         x = "GO Term",
         y = "Gene Count") +
    theme_bw() +
    theme(plot.title = element_text(size = 10, face = "bold", hjust = 0.5))


  jpeg("../Enrichment analysis for breast cancer/outputs/BP_enriched_up_dotplot.jpeg")
  print(p)
  dev.off()
}

dot_plot_enriched_up()
```

```
## pdf
##   2
```

```
jpeg("../Enrichment analysis for breast cancer/outputs/Network_plot_up.jpeg")
cnet_plot_up <- cnetplot(enrichment_go_up, showCategory = 2, vertex.label.cex = 1.2)
print(cnet_plot_up)
```

```
## Warning: ggrepel: 130 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

```
dev.off()
```

```
## pdf
##   2
```

```
jpeg("../Enrichment analysis for breast cancer/outputs/GO_graph_up.jpeg")
go_graph_up <- plotGOgraph(enrichment_go_up)
```

```
##
## groupGOTerms:    GOBPTerm, GOMFTerm, GOCCTerm environments built.

##
## Building most specific GOs .....

##  ( 8320 GO terms found. )

##
## Build GO DAG topology ..........

##  ( 8320 GO terms and 18464 relations. )

##
## Attaching package: 'SparseM'

## The following object is masked from 'package:base':
##
##     backsolve
```

```
##
## Annotating nodes ...............

##  ( 3609 genes annotated to the GO terms. )

## Loading required package: Rgraphviz

## Loading required package: graph

## Loading required package: grid

##
## Attaching package: 'Rgraphviz'

## The following objects are masked from 'package:IRanges':
##
##     from, to

## The following objects are masked from 'package:S4Vectors':
##
##     from, to
```

```
print(go_graph_up)
```

```
## $dag
## A graphNEL graph with directed edges
## Number of Nodes = 30
## Number of Edges = 42
##
## $complete.dag
## [1] "A graph with 30 nodes."
```

```
dev.off()
```

```
## pdf
##   2
```

Enriched GO terms among down-regulated significant genes

```
enrich_go_down <- function(){
        ego_down <- enrichGO( gene          = down_entrez_ids,
                              universe      = signficant$ENTREZID,
                              OrgDb         = org.Hs.eg.db,
                              keyType       = "ENTREZID",
                              ont           = "BP",
                              pvalueCutoff  = 0.05,
                              qvalueCutoff  = 0.01,
                              pAdjustMethod = "BH",
```

```
                              readable      = TRUE)
        return(ego_down)

}

enrichment_go_down <- enrich_go_down()
head(enrichment_go_down)
```

```
##                    ID                   Description GeneRatio  BgRatio
## GO:0003008 GO:0003008                system process  363/2065 472/3609
## GO:0007267 GO:0007267            cell-cell signaling  319/2065 431/3609
## GO:0050877 GO:0050877         nervous system process  197/2065 251/3609
## GO:0044057 GO:0044057   regulation of system process  130/2065 155/3609
## GO:0099537 GO:0099537       trans-synaptic signaling  151/2065 187/3609
## GO:0007268 GO:0007268 chemical synaptic transmission  148/2065 183/3609
##                  pvalue     p.adjust       qvalue
## GO:0003008 9.829182e-22 3.069654e-18 2.408667e-18
## GO:0007267 1.097491e-14 1.713732e-11 1.344715e-11
## GO:0050877 2.040669e-13 2.124336e-10 1.666904e-10
## GO:0044057 4.259348e-13 3.325486e-10 2.609411e-10
## GO:0099537 2.659578e-12 1.521318e-09 1.193733e-09
## GO:0007268 3.525065e-12 1.521318e-09 1.193733e-09
##
## GO:0003008 UGT2A1/DES/MYH11/FGF10/MFRP/ANK2/CLN3/CNN1/GSTM2/NTRK2/TLR9/AKAP12/CACNA1G/CRYBG3/PDE2A/LI
## GO:0007267
## GO:0050877
## GO:0044057
## GO:0099537
## GO:0007268
##            Count
## GO:0003008   363
## GO:0007267   319
## GO:0050877   197
## GO:0044057   130
## GO:0099537   151
## GO:0007268   148
```

Dataframe of enriched GO terms among down-regulated significant genes

```
df_go_terms_down <- function(){
  df_go_term_down <- as.data.frame(enrichment_go_down)
  return(df_go_term_down)
}

df_go_down <- df_go_terms_down()
```

```
df_go_down_top7 <- head(df_go_down, 7)
```

BarPlot of down-regulated BP enriched GO terms

```
bar_plot_enriched_down<- function(){
  jama_colors <- pal_jama("default")(7)
  p <- ggplot(df_go_down_top7, aes(x = reorder(Description, - Count) ,
                                   y = Count, fill = p.adjust)) +
    geom_bar(stat = "identity") +
    coord_flip() +
    scale_fill_gradientn(name = "p.adjust", colors = jama_colors) +
    labs(title = "Bar plot of BP of Enriched down-regulated Genes",
         x = "GO Term",
         y = "Gene Count") +
    theme_bw() +
    theme(plot.title = element_text(size = 10, face = "bold", hjust = 0.5))

  jpeg("../Enrichment analysis for breast cancer/outputs/BP_enriched_down_barplot.jpeg")
  print(p)
  dev.off()
}

bar_plot_enriched_down()
```

```
## pdf
##   2
```

Dotplot of down-regulated BP enriched GO terms

```
dot_plot_enriched_down<- function(){

  jama_colors <- pal_jama("default")(7)
  p <- ggplot(df_go_down_top7, aes(x = reorder(Description, -Count),
                                   y = Count, size = Count,
                                   color = p.adjust)) +
        geom_point(alpha = 0.6) +
        coord_flip() +
        scale_size_continuous(range = c(3, 8), name = "Gene Count") +
        scale_color_gradientn(name = "p.adjust", colors = jama_colors) +
        labs(title = "Dot plot of BP of enriched down-regulated genes",
             x = "GO Term",
             y = "Gene Count") +
        theme_bw() +
        theme(plot.title = element_text(size = 10, face = "bold", hjust = 0.5))

  jpeg("../Enrichment analysis for breast cancer/outputs/BP_enriched_down_dotplot.jpeg",
       width = 700, height = 800)
  print(p)
  dev.off()
}

dot_plot_enriched_down()
```

```
## pdf
##   2
```

```
jpeg("../Enrichment analysis for breast cancer/outputs/GO_graph_down.jpeg")
go_graph_down <- plotGOgraph(enrichment_go_down)
```

```
##
## groupGOTerms:    GOBPTerm, GOMFTerm, GOCCTerm environments built.

##
## Building most specific GOs .....

##  ( 9373 GO terms found. )

##
## Build GO DAG topology ..........

##  ( 9373 GO terms and 20799 relations. )

##
## Annotating nodes ...............

##  ( 3609 genes annotated to the GO terms. )
```

```
print(go_graph_down)
```

```
## $dag
## A graphNEL graph with directed edges
## Number of Nodes = 18
## Number of Edges = 21
##
## $complete.dag
## [1] "A graph with 18 nodes."
```

```
dev.off()
```

```
## pdf
##   2
```

Pathway Enrichment Analysis

KEGG pathway enrichment analysis among up-regulated significant genes

```
kegg_enrichment_up <- function(){

kegg_up <- enrichKEGG(gene = up_entrez_ids,
                      universe = signficant$ENTREZID,
                      organism = "hsa",
```

```
                      pvalueCutoff = 0.05,
                      qvalueCutoff = 0.01,
                      pAdjustMethod = "BH")
  return(kegg_up)

}

kegg_enrich_up <- kegg_enrichment_up()
```

```
## Reading KEGG annotation online: "https://rest.kegg.jp/link/hsa/pathway"...
```

```
## Reading KEGG annotation online: "https://rest.kegg.jp/list/pathway/hsa"...
```

```
head(kegg_enrich_up)
```

```
##                                category                    subcategory
## hsa04110             Cellular Processes          Cell growth and death
## hsa05169                 Human Diseases       Infectious disease: viral
## hsa04141 Genetic Information Processing Folding, sorting and degradation
## hsa04612             Organismal Systems                  Immune system
## hsa03013 Genetic Information Processing                    Translation
## hsa05014                 Human Diseases       Neurodegenerative disease
##                ID                                  Description GeneRatio BgRatio
## hsa04110 hsa04110                                   Cell cycle    48/832 57/1867
## hsa05169 hsa05169                 Epstein-Barr virus infection    47/832 58/1867
## hsa04141 hsa04141 Protein processing in endoplasmic reticulum    37/832 45/1867
## hsa04612 hsa04612         Antigen processing and presentation    22/832 24/1867
## hsa03013 hsa03013                 Nucleocytoplasmic transport    24/832 27/1867
## hsa05014 hsa05014               Amyotrophic lateral sclerosis    57/832 86/1867
##                pvalue     p.adjust       qvalue
## hsa04110 4.197373e-10 1.099712e-07 8.969124e-08
## hsa05169 8.187553e-09 1.072569e-06 8.747754e-07
## hsa04141 1.732074e-07 1.512678e-05 1.233723e-05
## hsa04612 1.506544e-06 9.309637e-05 7.592833e-05
## hsa03013 1.776648e-06 9.309637e-05 7.592833e-05
## hsa05014 2.715432e-05 1.048458e-03 8.551107e-04
##
## hsa04110                                                                  1869/891/9212/5347/4171/4173/
## hsa05169                                                                       1869/3627/6890/365
## hsa04141
## hsa04612
## hsa03013
## hsa05014 56893/23225/581/7186/203068/637/10376/4728/79139/4708/5710/9631/5690/7388/4704/5688/10762/8
##          Count
## hsa04110    48
## hsa05169    47
## hsa04141    37
## hsa04612    22
## hsa03013    24
## hsa05014    57
```

DataFrame of KEGG enriched among up-regulated significant genes

```
dataframe_kegg_up <- function(){
  df_kegg_up_genes <- as.data.frame(kegg_enrich_up)
  return(df_kegg_up_genes)
}

df_kegg_up <- dataframe_kegg_up()
```

```
df_kegg_up_top7 <- head(df_kegg_up, 7)
```

Barplot for KEGG enriched up-regulated significant gene

```
bar_plot_kegg_enriched_up<- function(){

  jama_colors <- pal_jama("default")(7)
  p <- ggplot(df_kegg_up_top7, aes(x = reorder(Description, - Count),
                                   y = Count, fill = p.adjust)) +
         geom_bar(stat = "identity", width = 0.8) +
         coord_flip() +
         scale_fill_gradientn(name = "p.adjust", colors = jama_colors) +
         labs(title = "Bar plot of Enriched KEGG pathway among up-regulated Genes",
             x = "Enriched pathway",
             y = "Gene Count") +
         theme_bw() +
         theme(plot.title = element_text(size = 10, face = "bold", hjust = 0.5))

  jpeg("../Enrichment analysis for breast cancer/outputs/kegg_enriched_up_barplot.jpeg",
  width = 700, height = 800)
  print(p)
  dev.off()

}

bar_plot_kegg_enriched_up()
```

```
## pdf
##   2
```

Dotplot for KEGG enriched up-regulated significant gene

```
dot_plot_kegg_enriched_up<- function(){

  jama_colors <- pal_jama("default")(7)
  p  <- ggplot(df_kegg_up_top7, aes(x = reorder(Description, -Count),
                                   y = Count, size = Count, color = p.adjust))+
        geom_point(alpha = 0.6) +
        coord_flip() +
```

```
        scale_size_continuous(range = c(3, 8), name = "Gene Count") +
        scale_color_gradientn(name = "p.adjust", colors = jama_colors) +
        labs(title = "Dot plot of Enriched KEGG pathway among up-regulated genes",
             x = "GO Term",
             y = "Gene Count") +
        theme_bw() +
        theme(plot.title = element_text(size = 10, face = "bold", hjust = 0.5))


  jpeg("../Enrichment analysis for breast cancer/outputs/kegg_enriched_up_dotplot.jpeg",
  width = 700, height = 800)
  print(p)
  dev.off()
}

dot_plot_kegg_enriched_up()
```

```
## pdf
##   2
```

Visualize the top enriched pathway that have smallest qval

```
visualize_top_path_up <- function(){
  pathview(gene.data = up_entrez_ids,
           pathway.id = "hsa04110",
           species = "hsa",
           kegg.dir = "../Enrichment analysis for breast cancer/outputs/")
}
visualize_top_path_up()
```

Browse the top enriched pathway for up-regulated sig genes

```
browseKEGG(kegg_enrich_up, 'hsa04110')
```

KEGG pathway enrichment analysis among down-regulated significant genes

```
kegg_enrichment_down <- function(){
kegg_down <- enrichKEGG(gene = down_entrez_ids,
                        universe = signficant$ENTREZID,
                        organism = "hsa",
                        pvalueCutoff = 0.05,
                        qvalueCutoff = 0.01,
                        pAdjustMethod = "BH")
return(kegg_down)
}
```

```
kegg_enrich_down <- kegg_enrichment_down()
head(kegg_enrich_down)
```

```
##                                     category
## hsa04080 Environmental Information Processing
## hsa04740                    Organismal Systems
## hsa04020 Environmental Information Processing
## hsa04024 Environmental Information Processing
## hsa04014 Environmental Information Processing
## hsa00590                           Metabolism
##                                  subcategory       ID
## hsa04080 Signaling molecules and interaction hsa04080
## hsa04740                       Sensory system hsa04740
## hsa04020                  Signal transduction hsa04020
## hsa04024                  Signal transduction hsa04024
## hsa04014                  Signal transduction hsa04014
## hsa00590                     Lipid metabolism hsa00590
##                                        Description GeneRatio BgRatio       pvalue
## hsa04080 Neuroactive ligand-receptor interaction   72/1035 84/1867 1.497664e-09
## hsa04740                   Olfactory transduction   29/1035 29/1867 3.112181e-08
## hsa04020                Calcium signaling pathway   64/1035 76/1867 5.753330e-08
## hsa04024                   cAMP signaling pathway   52/1035 65/1867 2.275312e-05
## hsa04014                    Ras signaling pathway   50/1035 63/1867 4.944153e-05
## hsa00590              Arachidonic acid metabolism   19/1035 20/1867 1.202375e-04
##              p.adjust       qvalue
## hsa04080 3.938858e-07 3.263332e-07
## hsa04740 4.092518e-06 3.390639e-06
## hsa04020 5.043753e-06 4.178735e-06
## hsa04024 1.496018e-03 1.239446e-03
## hsa04014 2.600624e-03 2.154610e-03
## hsa00590 5.270411e-03 4.366520e-03
##
## hsa04080 56413/3953/2899/7068/1511/6863/117/2901/185/10800/5745/5179/1910/6865/5733/2893/130576/154/1
## hsa04740
## hsa04020                                                    56413/2255/4915/8913/2252/845/185/10800/1910/19
## hsa04024
## hsa04014
## hsa00590
##          Count
## hsa04080    72
## hsa04740    29
## hsa04020    64
## hsa04024    52
## hsa04014    50
## hsa00590    19
```

DataFrame of KEGG enriched among down-regulated significant genes

```
dataframe_kegg_down <- function(){
  df_kegg_down_genes <- as.data.frame(kegg_enrich_down)
  return(df_kegg_down_genes)
```

```
}

df_kegg_down <- dataframe_kegg_down()
```

```
df_kegg_down_top7 <- head(df_kegg_down, 7)
```

Barplot for KEGG enriched down-regulated significant gene

```
bar_plot_kegg_enriched_down<- function(){

  jama_colors <- pal_jama("default")(7)
  p <- ggplot(df_kegg_down_top7, aes(x = reorder(Description, - Count),
                                     y = Count, fill = p.adjust)) +
       geom_bar(stat = "identity", width = 0.8) +
       coord_flip() +
       scale_fill_gradientn(name = "p.adjust", colors = jama_colors) +
       labs(title = "Bar plot of Enriched KEGG pathway of down-regulated Genes",
           x = "Enriched pathway",
           y = "Gene Count") +
       theme_bw() +
       theme(plot.title = element_text(size = 10, face = "bold", hjust = 0.5))

  jpeg("../Enrichment analysis for breast cancer/outputs/kegg_enriched_down_barplot.jpeg",
  width = 700, height = 800)
  print(p)
  dev.off()

}

bar_plot_kegg_enriched_down()
```

```
## pdf
##   2
```

Dotplot for KEGG enriched down-regulated significant gene

```
dot_plot_kegg_enriched_down <- function(){
  jama_colors <- pal_jama("default")(7)
  p <- ggplot(df_kegg_down_top7, aes(x = reorder(Description, -Count),
                                     y = Count, size = Count,
                                     color = p.adjust)) +
       geom_point(alpha = 0.6) +
       coord_flip() +
       scale_size_continuous(range = c(3, 8), name = "Gene Count") +
       scale_color_gradientn(name = "p.adjust", colors = jama_colors) +
       labs(title = "Dot plot of BP of enriched down-regulated genes",
           x = "GO Term",
           y = "Gene Count") +
```

```
      theme_bw() +
      theme(plot.title = element_text(size = 10, face = "bold", hjust = 0.5))


  jpeg("../Enrichment analysis for breast cancer/outputs/kegg_enriched_down_dotplot.jpeg",
  width = 700, height = 800)
  print(p)
  dev.off()
}
dot_plot_kegg_enriched_down()
```

```
## pdf
##   2
```

Visualize the top enriched pathway that have smallest qval

```
visualize_top_path_down <- function(){
  pathview(gene.data = down_entrez_ids,
           pathway.id = "hsa04080",
           species = "hsa",
           kegg.dir = "../Enrichment analysis for breast cancer/outputs/")
}

visualize_top_path_down()
```

Browse the top enriched pathway for down-regulated sig genes

```
browseKEGG(kegg_enrich_down, 'hsa04080')
```

Reactome pathway enrichment analysis among up-regulated significant genes

```
reactome_up_genes <- function(){
  reactome_enrichment <- enrichPathway(gene          = up_entrez_ids,
                                       universe      = signficant$ENTREZID,
                                       organism      = "human",
                                       pvalueCutoff  = 0.05,
                                       qvalueCutoff  = 0.01,
                                       pAdjustMethod = "BH")

  return(reactome_enrichment)
}

reactome_enriched_path_up <- reactome_up_genes()
head(reactome_enriched_path_up)
```

```
##                            ID              Description GeneRatio  BgRatio
## R-HSA-1640170 R-HSA-1640170               Cell Cycle  148/1123 180/2487
## R-HSA-69278     R-HSA-69278      Cell Cycle, Mitotic  131/1123 156/2487
## R-HSA-68886     R-HSA-68886                  M Phase   83/1123  95/2487
## R-HSA-69620     R-HSA-69620   Cell Cycle Checkpoints   71/1123  80/2487
## R-HSA-5663205 R-HSA-5663205       Infectious disease  156/1123 221/2487
## R-HSA-9824446 R-HSA-9824446 Viral Infection Pathways  127/1123 171/2487
##                      pvalue     p.adjust       qvalue
## R-HSA-1640170 2.615800e-26 1.687191e-23 9.251672e-24
## R-HSA-69278   4.972379e-25 1.603592e-22 8.793260e-23
## R-HSA-68886   2.611283e-18 5.614258e-16 3.078565e-16
## R-HSA-69620   1.287509e-16 2.076108e-14 1.138429e-14
## R-HSA-5663205 1.109163e-15 1.247911e-13 6.842892e-14
## R-HSA-9824446 1.160848e-15 1.247911e-13 6.842892e-14
##
## R-HSA-1640170                                        4605/1869/7083/891/4001/9212/5347/55143/4171/1100
## R-HSA-69278
## R-HSA-68886
## R-HSA-69620
## R-HSA-5663205 142/3654/9636/23225/3159/3838/6772/2214/4939/6184/1104/7428/5230/1174/10095/59345/2030
## R-HSA-9824446
##               Count
## R-HSA-1640170   148
## R-HSA-69278     131
## R-HSA-68886      83
## R-HSA-69620      71
## R-HSA-5663205   156
## R-HSA-9824446   127
```

DataFrame of Reactome enriched pathways of up-regulated significant genes

```
dataframe_reactome_up <- function(){
  df_reactome_up_genes <- as.data.frame(reactome_enriched_path_up)
  return(df_reactome_up_genes)
}

df_reactome_up <- dataframe_reactome_up()
```

Visualize Reactome Pathway Enrichment Results

```
df_reactome_up_top7 <- head(df_reactome_up, 7)
```

Barplot for Reactome enriched up-regulated significant gene

```
bar_plot_reactome_enriched_up<- function(){
  jama_colors <- pal_jama("default")(7)
  p <- ggplot(df_reactome_up_top7, aes(x = reorder(Description, - Count),
```

```
                                      y = Count, fill = p.adjust)) +
       geom_bar(stat = "identity", width = 0.8) +
       coord_flip() +
       scale_fill_gradientn(name = "p.adjust", colors = jama_colors) +
       labs(title = "Bar plot of Enriched Reactome pathway of up-regulated Genes",
            x = "Enriched pathway",
            y = "Gene Count") +
       theme_bw() +
       theme(plot.title = element_text(size = 10, face = "bold", hjust = 0.5))

  jpeg("../Enrichment analysis for breast cancer/outputs/reactome_enriched_up_barplot.jpeg",
  width = 700, height = 800)
  print(p)
  dev.off()

}

bar_plot_reactome_enriched_up()
```

```
## pdf
##   2
```

Dotplot for Reactome enriched up-regulated significant gene

```
dot_plot_reactome_enriched_up<- function(){
 jama_colors <- pal_jama("default")(7)
 p <- ggplot(df_reactome_up_top7, aes(x = reorder(Description, -Count),
                                      y = Count, size = Count,
                                      color = p.adjust)) +
      geom_point(alpha = 0.6) +
      coord_flip() +
      scale_size_continuous(range = c(3, 8), name = "Gene Count") +
      scale_color_gradientn(name = "p.adjust", colors = jama_colors) +
      labs(title = "Dot plot of BP of enriched up-regulated genes",
           x = "GO Term",
           y = "Gene Count") +
      theme_bw() +
      theme(plot.title = element_text(size = 10, face = "bold", hjust = 0.5))


jpeg("../Enrichment analysis for breast cancer/outputs/reactome_enriched_up_dotplot.jpeg",
width = 700, height = 800)
print(p)
dev.off()
}

dot_plot_reactome_enriched_up()
```

```
## pdf
##   2
```

visualize the top enriched Reactome pathway for up-regulated sig genes

Take the generated URL and browse it

```
visualize_reactome_path_up <- function(){
  # Convert the Reactome ID to a URL for visualization
  reactome_url <- paste0("https://reactome.org/PathwayBrowser/#/", "R-HSA-1640170")
  # Print the URL for manual review
  print(reactome_url)

}

visualize_reactome_path_up()
```

```
## [1] "https://reactome.org/PathwayBrowser/#/R-HSA-1640170"
```

Reactome pathway enrichment analysis among down-regulated significant genes

```
reactome_down_genes <- function(){
  reactome_enrichment <- enrichPathway(gene = down_entrez_ids,
                                       universe = signficant$ENTREZID,
                                       organism = "human",
                                       pvalueCutoff = 0.05,
                                       qvalueCutoff = 0.01,
                                       pAdjustMethod = "BH")

  return(reactome_enrichment)
}

reactome_enriched_path_down <- reactome_down_genes()
head(reactome_enriched_path_down)
```

```
##                          ID                                Description GeneRatio
## R-HSA-397014   R-HSA-397014                        Muscle contraction   56/1364
## R-HSA-9709957 R-HSA-9709957                        Sensory Perception   56/1364
## R-HSA-211945   R-HSA-211945 Phase I - Functionalization of compounds   26/1364
## R-HSA-5576891 R-HSA-5576891                        Cardiac conduction   40/1364
## R-HSA-372790   R-HSA-372790                         Signaling by GPCR  117/1364
## R-HSA-500792   R-HSA-500792                       GPCR ligand binding   76/1364
##                BgRatio       pvalue     p.adjust       qvalue
## R-HSA-397014   61/2487 2.026560e-10 1.305104e-07 1.190337e-07
## R-HSA-9709957  65/2487 5.375500e-08 1.730911e-05 1.578700e-05
## R-HSA-211945   26/2487 1.480544e-07 2.882265e-05 2.628807e-05
## R-HSA-5576891  44/2487 1.912268e-07 2.882265e-05 2.628807e-05
## R-HSA-372790  158/2487 2.237783e-07 2.882265e-05 2.628807e-05
## R-HSA-500792   96/2487 3.660541e-07 3.928981e-05 3.583477e-05
##
## R-HSA-397014
## R-HSA-9709957
## R-HSA-211945
```

```
## R-HSA-5576891
## R-HSA-372790  56413/5138/111/5296/6863/115557/117/2840/185/10800/5745/5179/1910/10850/1956/7225/6865/
## R-HSA-500792
##               Count
## R-HSA-397014     56
## R-HSA-9709957    56
## R-HSA-211945     26
## R-HSA-5576891    40
## R-HSA-372790    117
## R-HSA-500792     76
```

DataFrame of Reactome enriched pathways among down-regulated significant genes

```
dataframe_reactome_down <- function(){
  df_reactome_down_genes <- as.data.frame(reactome_enriched_path_down)
  return(df_reactome_down_genes)
}

df_reactome_down <- dataframe_reactome_down()
```

Visualize Reactome Pathway Enrichment Results

Select the top 7 enriched reactome pathways

```
df_reactome_down_top7 <- head(df_reactome_down, 7)
```

Barplot for Reactome enriched down-regulated significant gene

```
bar_plot_reactome_enriched_down<- function(){
  jama_colors <- pal_jama("default")(7)
  p <- ggplot(df_reactome_down_top7, aes(x = reorder(Description, - Count) ,
                                         y = Count, fill = p.adjust)) +
         geom_bar(stat = "identity", width = 0.8) +
         coord_flip() +
         scale_fill_gradientn(name = "p.adjust", colors = jama_colors) +
         labs(title = "Bar plot of Enriched Reactome pathway of down-regulated Genes",
              x = "Enriched pathway",
              y = "Gene Count") +
         theme_bw() +
         theme(plot.title = element_text(size = 10, face = "bold", hjust = 0.5))

  jpeg("../Enrichment analysis for breast cancer/outputs/reactome_enriched_down_barplot.jpeg",
  width = 700, height = 800)
  print(p)
  dev.off()
```

```
}
bar_plot_reactome_enriched_down()
```

```
## pdf
##   2
```

Dotplot for Reactome enriched downregulated significant gene

```
dot_plot_reactome_enriched_down<- function(){
  jama_colors <- pal_jama("default")(7)
  p <- ggplot(df_reactome_down_top7, aes(x = reorder(Description, -Count),
                                         y = Count, size = Count,
                                         color = p.adjust)) +
       geom_point(alpha = 0.6) +
       coord_flip() +
       scale_size_continuous(range = c(3, 8), name = "Gene Count") +
       scale_color_gradientn(name = "p.adjust", colors = jama_colors) +
       labs(title = "Dot plot of BP of enriched down-regulated genes",
            x = "GO Term",
            y = "Gene Count") +
       theme_bw() +
       theme(plot.title = element_text(size = 10, face = "bold", hjust = 0.5))


jpeg("../Enrichment analysis for breast cancer/outputs/reactome_enriched_down_dotplot.jpeg",
width = 700, height = 800)
print(p)
dev.off()
}
dot_plot_reactome_enriched_down()
```

```
## pdf
##   2
```

visualize the top enriched Reactome pathway for down-regulated sig genes

Take the generated URL and browse it

```
visualize_reactome_path_down <- function(){
  # Convert the Reactome ID to a URL for visualization
  reactome_url <- paste0("https://reactome.org/PathwayBrowser/#/", "R-HSA-397014")
  # Print the URL for manual review
  print(reactome_url)
}

visualize_reactome_path_down()
```

```
## [1] "https://reactome.org/PathwayBrowser/#/R-HSA-397014"
```