

Volcano Plot Visualization of Differentially Expressed Genes in Triple Negative Breast Cancer

Code ▼

2023-10-19

Hide

```
library(tidyverse)
```

1. Load required file

Hide

```
# save DGE results of 114 PT vs 179 NT from XENA as a dataframe
deg_results = read.csv("DEG_results_TCGA TNBC Primary_Tumor (114) vs. GTEx Breast Normal_Tissue (179).csv",header = TRUE,
                      row.names = 1)
```

Hide

```
# check dimension of the DGE results
dim(deg_results)
```

```
[1] 20976      6
```

Hide

```
# view the first six rows of the DGE results
head(deg_results, n = 6)
```

	logFC	AveExpr	t	P.Value	adj.P.Val	B
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
CENPF	6.478870	3.4653718	56.01866	2.585348e-159	5.423025e-155	354.1512
RP11-64K7.1	4.888965	-4.0958638	55.81132	7.031004e-159	7.374117e-155	350.2475
HMGN1P37	5.419908	-3.7650094	53.74119	1.817965e-154	1.271121e-150	340.6081
CCNB1	4.331793	3.3783003	52.35235	1.987900e-151	1.042455e-147	336.0379
HNRNPCP2	3.312463	0.9255393	51.99882	1.208700e-150	5.070739e-147	333.7665
AURKA	4.699976	2.2348892	51.80356	3.289540e-150	1.150023e-146	333.0575
6 rows						

2. Create a Volcano Plot of Differentially Expressed Genes in Triple Negative Breast Cancer Primary Tumor vs Normal Breast Tissue

[Hide](#)

```
# create a column named diff.expressed with all values = "NO"  
deg_results$diff.expressed = "NO"
```

[Hide](#)

```
# label diffexpressed = "UP" for significant upregulated genes.  
# significant upregulated genes have log fold change > 1 and adjusted p value < 0.05  
# Check the counts of "UP"  
deg_results$diff.expressed[deg_results$logFC>1 &  
                           deg_results$adj.P.Val <0.05] = "UP"  
  
nrow(deg_results[deg_results$diff.expressed == "UP",])
```

```
[1] 3738
```

[Hide](#)

```
# label diffexpressed = "DOWN" for significant downregulated genes.  
# significant downregulated genes have log fold change < -1 and adjusted p value < 0.05  
# Check the counts of DOWN"  
deg_results$diff.expressed[deg_results$logFC < (-1) &  
                           deg_results$adj.P.Val <0.05] = "DOWN"  
  
nrow(deg_results[deg_results$diff.expressed == "DOWN",])
```

```
[1] 5551
```

[Hide](#)

```
# create a volcano plot
ggplot(data=deg_results,
      aes(x=logFC,y=-log10(adj.P.Val),
        col = diff.expressed)) +
geom_point(size = 1.25)+
geom_vline(xintercept = c(-1,1),col = 'black', linetype = 'dashed')+
geom_hline(yintercept = -log10(0.05),col = 'black',linetype = 'dashed')+
scale_color_manual(values = c("#00AFBB","grey","#E6676B"),
      labels = c('downregulated genes',
        'not significant genes',
        'upregulated genes'))+
coord_cartesian(ylim = c(0,175), xlim = c(-10,10))+
scale_x_continuous(breaks = seq(-10,10,2)) +
scale_y_continuous(breaks = seq(0,175,25)) +
labs(color = '', x = expression("log"[2]*"fold_change"),
      y = expression("-log"[10]*"adjusted_p-value" ))+
ggtitle(label = "TNBC Primary Tumor vs Normal Tissue",) +
theme_bw()+
theme(plot.title = element_text(hjust = 0.5, vjust = + 3, face = "bold",
      margin = margin(t=20,r=0,b=0,l=0)),
      axis.title.x = element_text(size=12,hjust = 0.5, vjust = -3,
      margin = margin(t=0,r=0,b=20,l=0)),
      axis.title.y = element_text(size=12, vjust = +3, hjust = 0.5,
      margin = margin(t=0,r=0,b=0,l=20)),
      legend.text = element_text(size=11, color = 'black'),
      axis.text.x = element_text(size = 10, color = 'black'),
      axis.text.y = element_text(size = 10,color = 'black'))
```

