# Bulk RNA Seq - Volcano Plots

Brandon Hancock

1/14/2021

This page shows how volcano plots were generated for the paper

The dataset can be downloaded from this link

Analysis derived from the DESeq2 tutorial

Load the required R packages:

```
library(readxl)
library(readr)
library(DESeq2)
library(apeglm)
library(EnhancedVolcano)
```

Load in the Bulk RNA seq STAR gene counts file:

```
STAR_gene_counts <- read_csv("C:/Users/bh719/Dropbox (Partners HealthCare)/Harvard Cytof/for Brandon/Sally/STAR_Gene_Counts.csv")
```

Remove genes with duplicate entries (1-Mar and 2-Mar)

```
STAR_gene_counts <- STAR_gene_counts[!duplicated(STAR_gene_counts$Gene_ID),]
```

Genes come all capitalized, change so that only the first letter is capitalized

```
Cap <- function(g){
  g <- paste(toupper(substring(g,1,1)), tolower(substring(g,2)), sep = '')
  return(g)
}

STAR_gene_count$Gene_ID <- sapply(STAR_gene_counts$Gene_ID, Cap)
```

Define Meta Data

```
colmetadat <- data.frame(Injury = c(rep("Uninjured",6),rep("7D after Injury",10)),CD44 = c(rep("CD44 High",4),rep("CD44 Low",5),rep("CD44 High",6),rep("CD44 Low",6)))
row.names(colmetadat) <- colnames(STAR_gene_counts)[2:length(colnames(STAR_gene_counts))]
```

Define DESeq2 matrix

```
gene_row <- STAR_gene_counts$Gene_ID
cmat <- STAR_gene_counts
cmat <- cmat[,!(names(cmat) %in% c('Gene_ID'))]
row.names(cmat) <- gene_row
```

Create DESeq2 object

```
dds <- DESeqDataSetFromMatrix(countData = cmat,colData = colmetadat,design = ~ CD44 + Injury)
dds <- dds[rowSums(counts(dds)) >= 10,]
dds$Group <- factor(paste0(dds$CD44,dds$Injury))
design(dds) <- ~ Group
```

Factor levels must be refined for different differential expression (DE) analysis group comparisons Only the first group/level will be compared against, so in this case, we can only extract comparisons against 'CD44 LowUninjured'

Run the DE analysis for CD44 High Uninjured versus CD44 Low Uninjured & CD44 Low 7D after Injury versus CD44 Low Uninjured

```
dds$Group = factor(dds$Group, levels = c('CD44 LowUninjured','CD44 HighUninjured','CD44 LowTD after Injury','CD44 HighTD after Injury'))
dds <- DESeq(dds)
high_low <- lfcShrink(dds, coef = "Group_CD44.HighUninjured_vs_CD44.LowUninjured", type = "apeglm")
low7D_low <- lfcShrink(dds, coef = "Group_CD44.LowTD.after.Injury_vs_CD44.LowUninjured", type = "apeglm")
```

Run the DE analysis for the other group comparisons by first changing the factor levels

```
dds$Group <- factor(dds$Group, levels = c('CD44 HighUninjured','CD44 HighTD after Injury','CD44 LowUninjured','CD44 LowTD after Injury'))
dds <- DESeq(dds)
high7D_high <- lfcShrink(dds,coef = "Group_CD44.HighTD.after.Injury_vs_CD44.HighUninjured", type = "apeglm")
dds$Group <- factor(dds$Group, levels = c('CD44 LowTD after Injury','CD44 HighTD after Injury','CD44 LowUninjured','CD44 HighUninjured'))

dds <- DESeq(dds)
high7D_low7D <- lfcShrink(dds,coef = "Group_CD44.HighTD.after.Injury_vs_CD44.LowTD.after.Injury", type = "apeglm"
)
```

Create the Volcano Plot for CD44 Low 7D after Injury vs CD44 Low Uninjured

Define the genes we would like labeled, correct capitalization

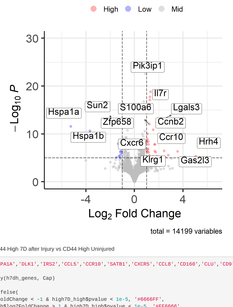
```
lfd1_genes <- c('HSPA1A','HSPA1B','SUN2','ZFP658','PIK3IP1','S100A6','IL7R','DDIT2','CCNB2','CXCR6','CCR10','KLRG1','HRH4','GAS2L3','LGALS3')
lfd1_genes <- sapply(lfd1_genes, Cap)
```

Define the colors for each gene, red for unregulated (High), grey for unaffected genes (Mid), blue for down regulated (Low)

```
keyvals.colour <- ifelse(
  low7D_low$log2FoldChange < -1 & low7D_low$pvalue < 1e-5, '#0666FF',
  ifelse(low7D_low$log2FoldChange > 1 & low7D_low$pvalue < 1e-5, '#FF6666',
         'grey')
)
keyvals.colour[is.na(keyvals.colour)] <- 'grey'
names(keyvals.colour)[keyvals.colour == '#0666FF'] <- 'Low'
names(keyvals.colour)[keyvals.colour == 'grey'] <- 'Mid'
names(keyvals.colour)[keyvals.colour == '#FF6666'] <- 'High'
```

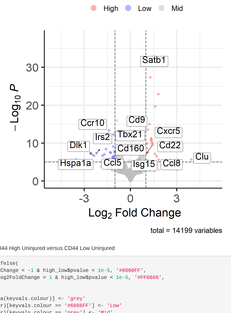
Plot

```
## 615, W 763
EnhancedVolcano(low7D_low,lab = rownames(low7D_low), title = "",subtitle = '',
                x = 'log2FoldChange',y = 'pvalue',
                labSize = 6.0, colCustom = keyvals.colour,
                titleLabSize = 1,subtitleLabSize = 1,axisLabSize = 22,
                boxedLabels = TRUE,drawConnectors = TRUE,
                selectLab = lfd1_genes,
                xlab = bquote(~Log[2]~ 'Fold Change'))
```



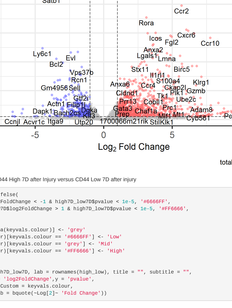

Now the same but for CD44 High 7D after Injury vs CD44 High Uninjured

```
h7dh_genes <- c('HSPA1A','DLK1','IRS2','CCL5','CCR10','SATB1','CXCR5','CCL8','CD160','Clu','CD9','ISG15','CD22',
  'TBX21')
h7dh_genes <- sapply(h7dh_genes, Cap)

keyvals.colour <- ifelse(
  high7D_high$log2FoldChange < -1 & high7D_high$pvalue < 1e-5, '#0666FF',
  ifelse(high7D_high$log2FoldChange > 1 & high7D_high$pvalue < 1e-5, '#FF6666',
         'grey')
)
keyvals.colour[is.na(keyvals.colour)] <- 'grey'
names(keyvals.colour)[keyvals.colour == '#0666FF'] <- 'Low'
names(keyvals.colour)[keyvals.colour == 'grey'] <- 'Mid'
names(keyvals.colour)[keyvals.colour == '#FF6666'] <- 'High'

EnhancedVolcano(high7D_high,lab = rownames(high7D_high), title = "",subtitle = '',
                x = 'log2FoldChange',y = 'pvalue',
                labSize = 6.0, colCustom = keyvals.colour,
                titleLabSize = 1,subtitleLabSize = 1,axisLabSize = 22,
                boxedLabels = TRUE,drawConnectors = TRUE,
                selectLab = h7dh_genes,
                xlab = bquote(~Log[2]~ 'Fold Change'))
```



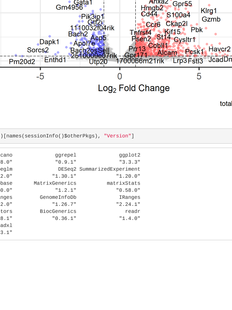

Supplemental Figure - CD44 High Uninjured versus CD44 Low Uninjured

```
keyvals.colour <- ifelse(
  high_low$log2FoldChange < -1 & high_low$pvalue < 1e-5, '#0666FF',
  ifelse(high_low$log2FoldChange > 1 & high_low$pvalue < 1e-5, '#FF6666',
         'grey')
)
keyvals.colour[is.na(keyvals.colour)] <- 'grey'
names(keyvals.colour)[keyvals.colour == '#0666FF'] <- 'Low'
names(keyvals.colour)[keyvals.colour == 'grey'] <- 'Mid'
names(keyvals.colour)[keyvals.colour == '#FF6666'] <- 'High'

EnhancedVolcano(high_low, lab = rownames(high_low), title = "", subtitle = "",
                x = 'log2FoldChange',y = 'pvalue',
                colCustom = keyvals.colour,
                xlab = bquote(~Log[2]~ 'Fold Change'))
```




Supplemental Figure - CD44 High 7D after Injury versus CD44 Low 7D after injury

```
keyvals.colour <- ifelse(
  high7D_low7D$log2FoldChange < -1 & high7D_low7D$pvalue < 1e-5, '#0666FF',
  ifelse(high7D_low7D$log2FoldChange > 1 & high7D_low7D$pvalue < 1e-5, '#FF6666',
         'grey')
)
keyvals.colour[is.na(keyvals.colour)] <- 'grey'
names(keyvals.colour)[keyvals.colour == '#0666FF'] <- 'Low'
names(keyvals.colour)[keyvals.colour == 'grey'] <- 'Mid'
names(keyvals.colour)[keyvals.colour == '#FF6666'] <- 'High'


EnhancedVolcano(high7D_low7D, lab = rownames(high7D_low7D), title = "", subtitle = "",
                x = 'log2FoldChange',y = 'pvalue',
                colCustom = keyvals.colour,
                xlab = bquote(~Log[2]~ 'Fold Change'))
```




Session Info

```
installed.packages()[names(sessionInfo()$otherPkgs), "Version"]
```

```
##    EnhancedVolcano           ggrepel           ggplot2 
##            "1.8.0"            "0.9.1"            "3.3.3" 
##             apeglm             DESeq2 SummarizedExperiment 
##            "1.12.0"           "1.30.1"           "1.20.0" 
##           Biobase       MatrixGenerics         matrixStats 
##           "2.50.0"           "1.2.1"            "0.58.0" 
##      GenomicRanges      GenomeInfoDb            IRanges 
##           "1.42.0"           "1.26.7"           "2.24.1" 
##         S4Vectors       BiocGenerics              readr 
##           "0.28.1"           "0.36.1"           "1.4.0" 
##             readxl 
##            "1.3.1"
```