

Zemin_CRC_GSE108989_visualization

Hyun Yong Jin

August 8, 2019

```
knitr::opts_chunk$set(echo = TRUE)
```

Visualization of from stat_all table

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

```
library(ggrepel)
```

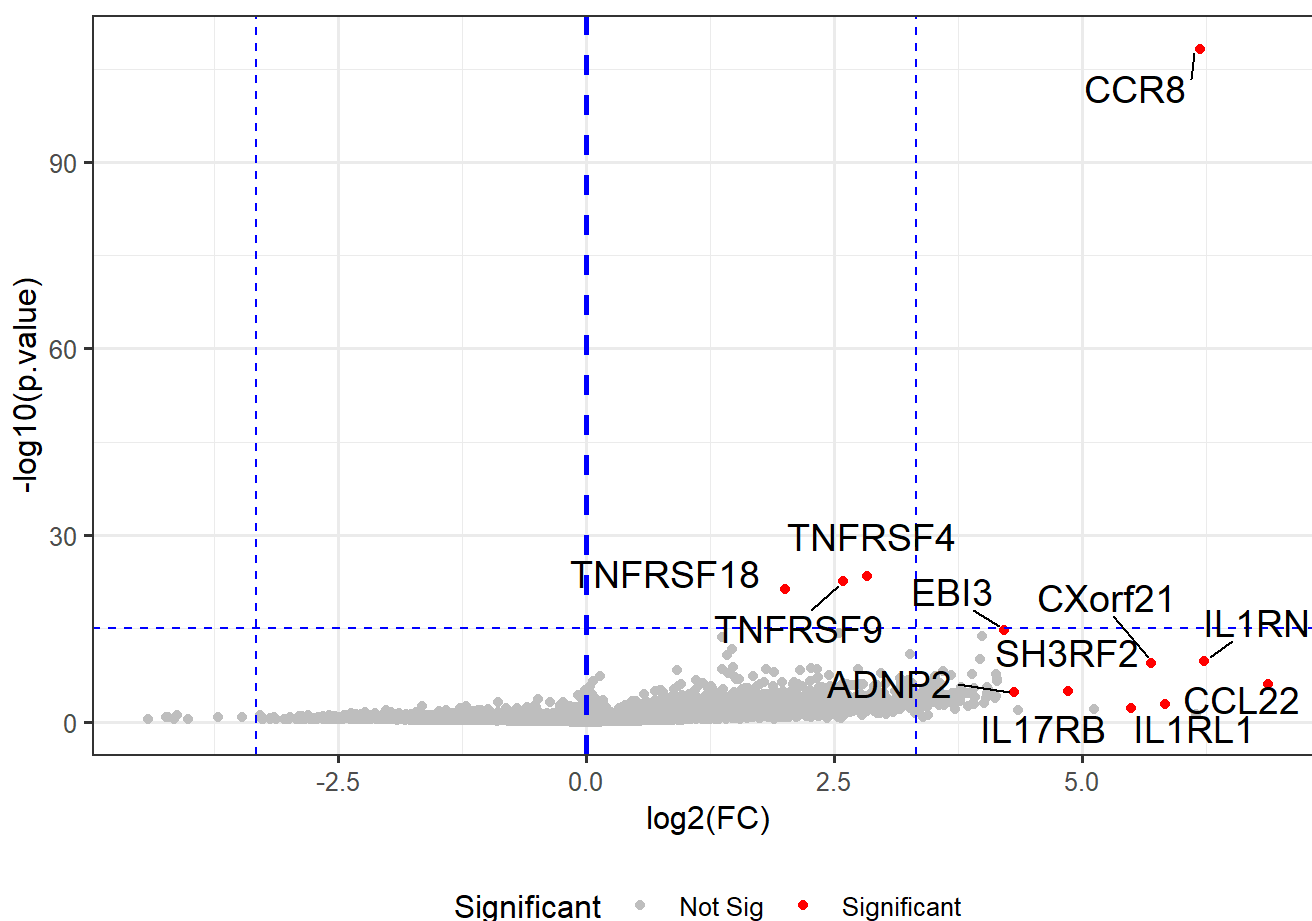
```
## Warning: package 'ggrepel' was built under R version 3.5.3
```

```
a <- read.csv("stat_all.csv", header=T, stringsAsFactors = FALSE)

a$Significant <- ifelse((a$FC>18&a$p.value<0.01)|a$p.value<1/10^15, "Significant", "Not Sig")

pl<-ggplot(a, aes(x = log2(FC), y = -log10(p.value))) +
  geom_point(aes(color = Significant)) +
  scale_color_manual(values = c("grey", "red")) +
  theme_bw(base_size = 12) + theme(legend.position = "bottom") +
  geom_vline(xintercept=0, linetype="dashed", color = "blue", size=1)+
  geom_vline(xintercept=log2(10), linetype="dashed", color = "blue")+
  geom_vline(xintercept=log2(0.1), linetype="dashed", color = "blue")+
  geom_hline(yintercept=-log10(1/10^15), linetype="dashed", color = "blue")+
  geom_text_repel(
    data = subset(a, (FC>18&p.value<0.01)|p.value<1/10^15),
    aes(label = gene_name),
    size = 5,
    box.padding = unit(0.35, "lines"),
    point.padding = unit(0.3, "lines")
  )

plot(pl)
```



```
#Want to add SD information on top of it.
#But P.value actually contains this SD informaiton, so may not be necessary.
```

```
#Next: Generate the one without CCR8
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.5.2
```

```
## -- Attaching packages ----- tidyverse 1.2.1 --
```

```
## v tibble 2.1.1      v purrr  0.3.2
## v tidyr  0.8.3      v dplyr  0.8.1
## v readr  1.3.1      v stringr 1.4.0
## v tibble 2.1.1      v forcats 0.4.0
```

```
## Warning: package 'tibble' was built under R version 3.5.3
```

```
## Warning: package 'tidyr' was built under R version 3.5.3
```

```
## Warning: package 'readr' was built under R version 3.5.2
```

```
## Warning: package 'purrr' was built under R version 3.5.3
```

```
## Warning: package 'dplyr' was built under R version 3.5.3
```

```
## Warning: package 'stringr' was built under R version 3.5.3
```

```
## Warning: package 'forcats' was built under R version 3.5.3
```

```
## -- Conflicts ----- tidyverse_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()
```

```
b <- read.csv("stat_all.csv", header=T, stringsAsFactors = FALSE)  
b<- b %>% filter(gene_name != "CCR8") #Remove row with CCR8  
dim(b)
```

```
## [1] 12545 17
```

```
dim(a)
```

```
## [1] 12546 18
```

```
b[c(1:10), c(1:10)]
```

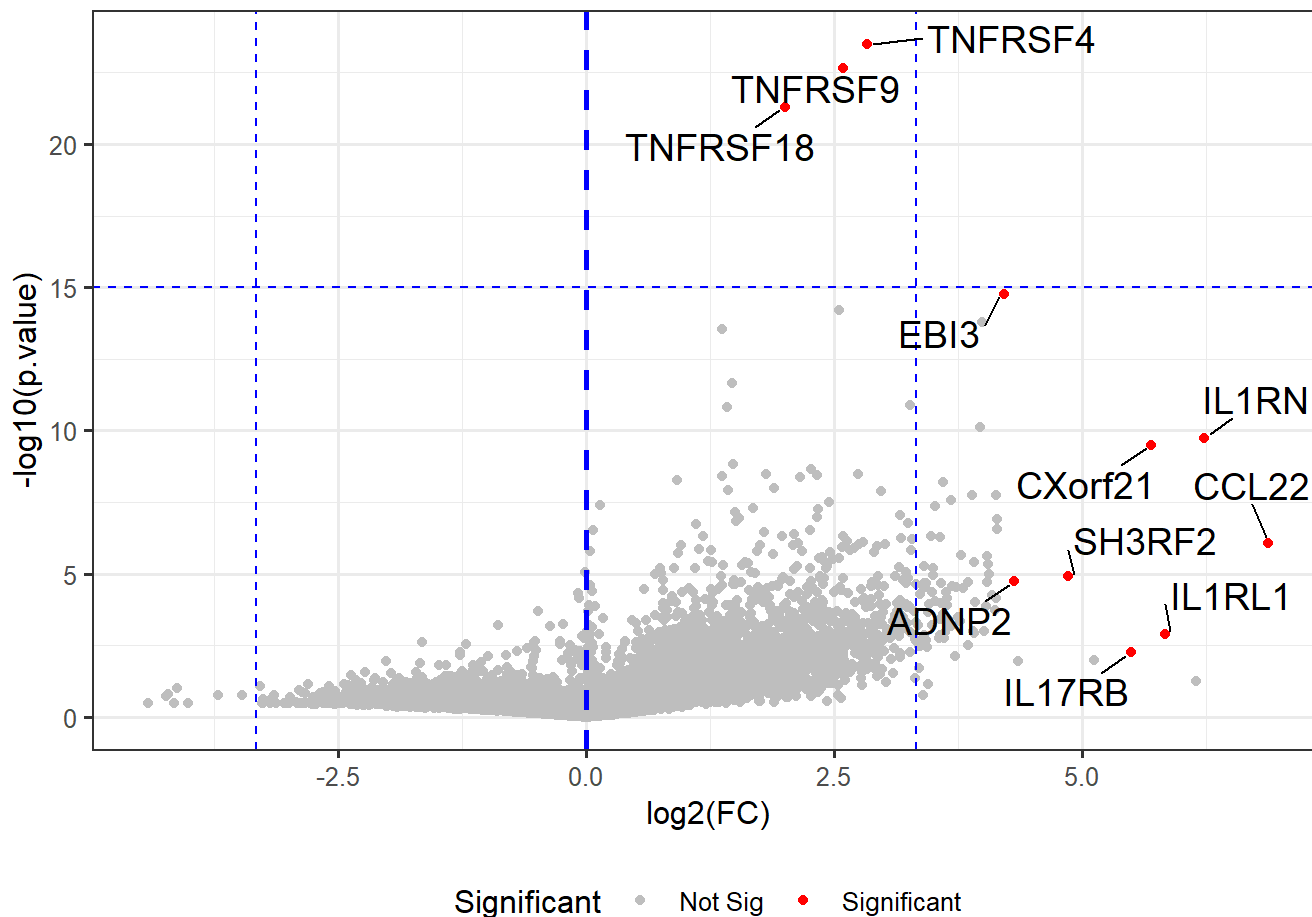
```
##      X  gene_name  hi_mean  lo_mean      FC  hi_SD  lo_SD  
## 1  1      A1BG  5.624701  4.0595055  1.3855631  18.886206  12.15253985  
## 2  2      ADA  9.637158  15.4005060  0.6257689  22.776546  48.72238384  
## 3  3      AKT3  3.094601  4.9781848  0.6216324  14.066178  30.02763392  
## 4  4  ZBTB11-AS1  3.659279  5.0819698  0.7200513  10.313471  24.73991295  
## 5  5      MED6  12.648734  12.0736096  1.0476348  25.797292  29.69652774  
## 6  6  NAALAD2  1.651442  0.9501306  1.7381207  6.311207  0.10058810  
## 7  7  SNORD116-26  1.723191  0.9174360  1.8782679  8.493994  0.01441694  
## 8  8      DDTL  4.317503  1.5256503  2.8299428  9.034162  2.82298572  
## 9  9  NAALADL1  9.981740  29.0250840  0.3439005  32.141850  71.76459543  
## 10 10  CDKN2B-AS1  1.318167  1.7236495  0.7647533  3.242690  5.61375752  
##      estimate estimate1 estimate2  
## 1  1.5651953  5.624701  4.0595055  
## 2 -5.7633477  9.637158  15.4005060  
## 3 -1.8835839  3.094601  4.9781848  
## 4 -1.4226909  3.659279  5.0819698  
## 5  0.5751246  12.648734  12.0736096  
## 6  0.7013110  1.651442  0.9501306  
## 7  0.8057546  1.723191  0.9174360  
## 8  2.7918529  4.317503  1.5256503  
## 9 -19.0433437  9.981740  29.0250840  
## 10 -0.4054828  1.318167  1.7236495
```

#This still contains CCR8 somehow.

```
b$Significant <- ifelse((b$FC>18&b$p.value<0.01)|b$p.value<1/10^15, "Significant", "Not Sig")
```

```
plb<-ggplot(b, aes(x = log2(FC), y = -log10(p.value))) +  
  geom_point(aes(color = Significant)) +  
  scale_color_manual(values = c("grey", "red")) +  
  theme_bw(base_size = 12) + theme(legend.position = "bottom") +  
  geom_vline(xintercept=0, linetype="dashed", color = "blue", size=1)+  
  geom_vline(xintercept=log2(10), linetype="dashed", color = "blue")+  
  geom_vline(xintercept=log2(0.1), linetype="dashed", color = "blue")+  
  geom_hline(yintercept=-log10(1/10^15), linetype="dashed", color = "blue")+  
  geom_text_repel(  
    data = subset(b, (FC>18&p.value<0.01)|p.value<1/10^15),  
    aes(label = gene_name),  
    size = 5,  
    box.padding = unit(0.35, "lines"),  
    point.padding = unit(0.3, "lines")  
  )  
)
```

```
plot(plb)
```



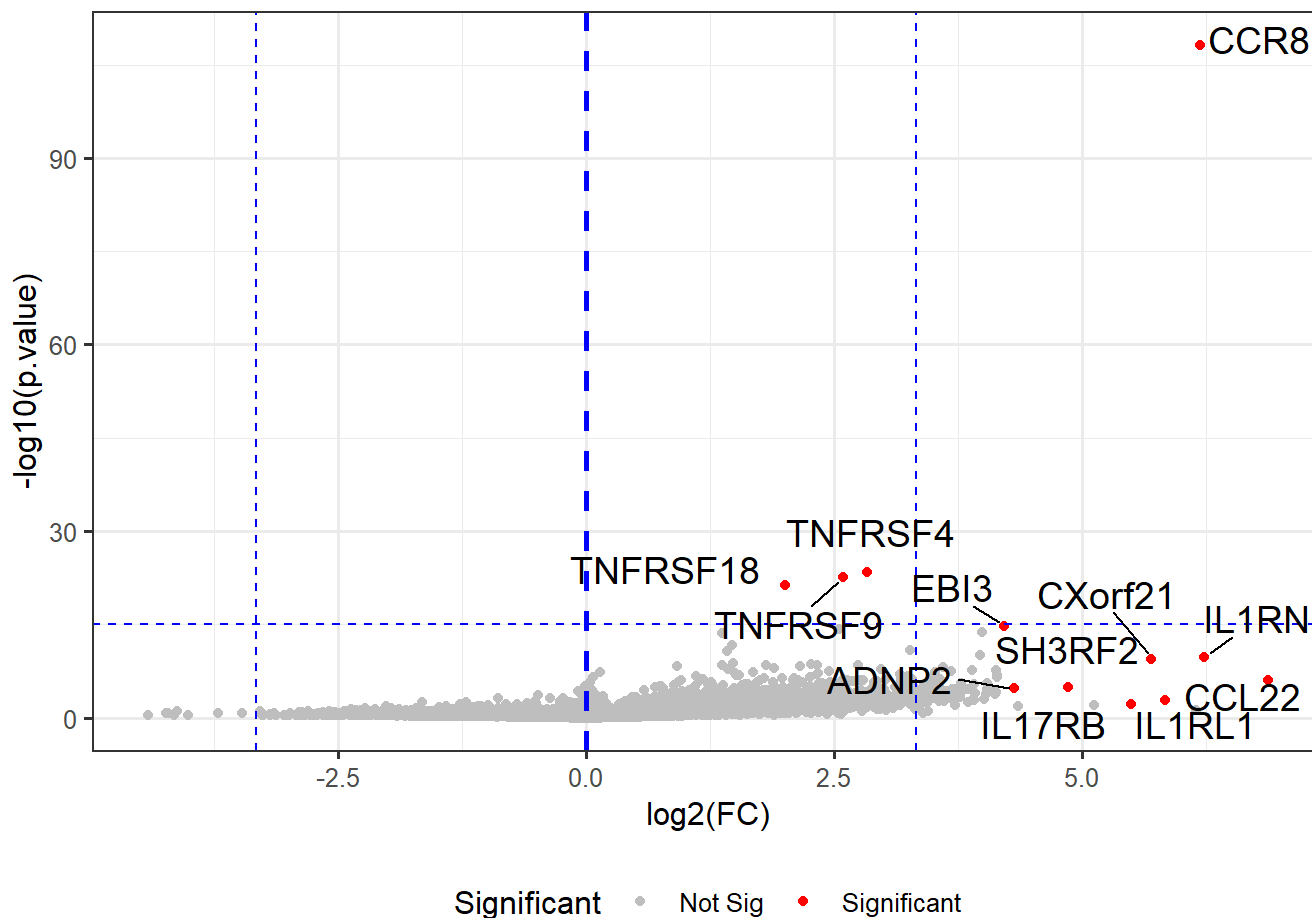
#Generate function to simplify this.

#This function called "volcano" takes filename as an argument.

```
library(ggplot2)
library(ggrepel)
library(tidyverse)

volcano.all <- function(filename){
  a <- read.csv(file = filename, header=T, stringsAsFactors = FALSE)
  a$Significant <- ifelse((a$FC>18&a$p.value<0.01)|a$p.value<1/10^15, "Significant", "Not Sig")
  pl<-ggplot(a, aes(x = log2(FC), y = -log10(p.value))) +
    geom_point(aes(color = Significant)) +
    scale_color_manual(values = c("grey", "red")) +
    theme_bw(base_size = 12) + theme(legend.position = "bottom") +
    geom_vline(xintercept=0, linetype="dashed", color = "blue", size=1)+
    geom_vline(xintercept=log2(10), linetype="dashed", color = "blue")+
    geom_vline(xintercept=log2(0.1), linetype="dashed", color = "blue")+
    geom_hline(yintercept=-log10(1/10^15), linetype="dashed", color = "blue")+
    geom_text_repel(
      data = subset(a, (FC>18&p.value<0.01)|p.value<1/10^15),
      aes(label = gene_name),
      size = 5,
      box.padding = unit(0.35, "lines"),
      point.padding = unit(0.3, "lines")
    )
  plot(pl)
}

volcano.all("stat_all.csv")
```



```

volcano.noccr8 <- function(filename){
a <- read.csv(file = filename, header=T, stringsAsFactors = FALSE)
a <- a %>% filter(gene_name != "CCR8") #Remove row with CCR8. Only addition
a$Significant <- ifelse((a$FC>18&a$p.value<0.01)|a$p.value<1/10^15, "Significant", "Not Sig")
p1<-ggplot(a, aes(x = log2(FC), y = -log10(p.value))) +
  geom_point(aes(color = Significant)) +
  scale_color_manual(values = c("grey", "red")) +
  theme_bw(base_size = 12) + theme(legend.position = "bottom") +
  geom_vline(xintercept=0, linetype="dashed", color = "blue", size=1)+
  geom_vline(xintercept=log2(10), linetype="dashed", color = "blue")+
  geom_vline(xintercept=log2(0.1), linetype="dashed", color = "blue")+
  geom_hline(yintercept=-log10(1/10^15), linetype="dashed", color = "blue")+
  geom_text_repel(
    data = subset(a, (FC>18&p.value<0.01)|p.value<1/10^15),
    aes(label = gene_name),
    size = 5,
    box.padding = unit(0.35, "lines"),
    point.padding = unit(0.3, "lines")
  )
plot(p1)
}

```

```
volcano.noccr8("stat_all.csv")
```

