WT vs *tf2*

tf2 Marginal Blastozone C (base) vs Marginal Blastzone in C (base) region

Read in Data

Read in raw count data per gene.

```
counts <- read.delim("../sam2countsResults.tsv",row.names=1)

#check the file
head(counts)
summary(counts)
colnames(counts)
#need to convert NA to 0 counts
counts[is.na(counts)] <- 0
```

Subset per DE expirement

I am going to start by subsetting the particular treatments I am looking at.

```
colnames(counts)
```

```
##  [1] "tf2ambr1"      "tf2ambr3"      "tf2ambr4"      "tf2ambr6"
##  [5] "tf2aother1"    "tf2aother2"    "tf2aother4"    "tf2aother7"
##  [9] "tf2bmbr2"      "tf2bmbr5"      "tf2bmbr6"      "tf2bother1"
## [13] "tf2bother3"    "tf2bother4"    "tf2bother6"    "tf2cmbr1.4"
## [17] "tf2cmbr3"      "tf2cmbr6"      "tf2cmbr7"      "tf2cother2"
## [21] "tf2cother5"    "tf2cother6"    "tf2cother7"    "wtambr2"
## [25] "wtambr4"       "wtambr5"       "wtaother1"     "wtaother5"
## [29] "wtaother6"     "wtaother7"     "wtaother8"     "wtbmbr2"
## [33] "wtbmbr3"       "wtbmbr6"       "wtbmbr8"       "wtbother1.4"
## [37] "wtbother3"     "wtbother5"     "wtbother8"     "wtcmbr10"
## [41] "wtcmbr1.4.6"   "wtcmbr2"       "wtcmbr3"       "wtcmbr7"
## [45] "wtcmbr9"       "wtcother1.3.4" "wtcother2"     "wtcother6"
```

```
tf2cmbrVSwtcmbr <- counts[,c(16:19,42:45)]
head(tf2cmbrVSwtcmbr)
```

```
##                   tf2cmbr1.4 tf2cmbr3 tf2cmbr6 tf2cmbr7 wtcmbr2 wtcmbr3
## Solyc00g005040.2.1          0        6        8        4       3       1
## Solyc00g005050.2.1          1       34       17       12      21      11
## Solyc00g005060.1.1          0        1        0        0       0       0
## Solyc00g005070.1.1         23       11        8        9       7       4
## Solyc00g005080.1.1         22        7        8       12      19      45
## Solyc00g005150.1.1          1        3        0        0       3       3
##                   wtcmbr7 wtcmbr9
```

```
## Solyc00g005040.2.1       0       0
## Solyc00g005050.2.1       4       7
## Solyc00g005060.1.1       1       0
## Solyc00g005070.1.1       6       1
## Solyc00g005080.1.1       4       7
## Solyc00g005150.1.1       2       1
```

```
colnames(tf2cmbrVSwtcmbr)
```

```
## [1] "tf2cmbr1.4" "tf2cmbr3"   "tf2cmbr6"   "tf2cmbr7"   "wtcmbr2"
## [6] "wtcmbr3"    "wtcmbr7"    "wtcmbr9"
```

```
group <- c(rep("tf2cmbr", 4), rep("wtcmbr", 4))
d <- DGEList(counts=tf2cmbrVSwtcmbr,group=group)
```

```
d$samples
```

```
##              group lib.size norm.factors
## tf2cmbr1.4 tf2cmbr   443572            1
## tf2cmbr3   tf2cmbr  1337575            1
## tf2cmbr6   tf2cmbr   790129            1
## tf2cmbr7   tf2cmbr   832907            1
## wtcmbr2     wtcmbr  1130695            1
## wtcmbr3     wtcmbr  1560130            1
## wtcmbr7     wtcmbr   374882            1
## wtcmbr9     wtcmbr   386974            1
```

```
cpm.d <- cpm(d)
d <- d[rowSums(cpm.d>5)>=3,] #change to 5
d <- estimateCommonDisp(d,verbose=T)
```

```
## Disp = 0.3738 , BCV = 0.6114
```

```
d <- calcNormFactors(d)
d <- estimateCommonDisp(d)

DEtest <- exactTest(d,pair=c("tf2cmbr","wtcmbr"))
head(DEtest$table)
```

```
##                      logFC logCPM  PValue
## Solyc00g005050.2.1 -0.2507  4.068 0.80984
## Solyc00g005070.1.1 -1.7894  4.127 0.02314
## Solyc00g005080.1.1 -0.1084  4.473 0.80647
## Solyc00g005440.1.1  0.1293  4.648 0.86712
## Solyc00g005840.2.1  0.3887  4.751 0.51008
## Solyc00g005880.1.1 -1.2454  3.443 0.13298
```

```
results <- topTags(DEtest, n=Inf)
head(results)
```

```
## Comparison of groups:  wtcmbr-tf2cmbr
##                       logFC logCPM    PValue       FDR
## Solyc02g023990.2.1 -6.785  6.703 1.811e-14 2.723e-10
## Solyc06g069460.1.1 -7.430  5.360 2.484e-12 1.867e-08
## Solyc07g044980.2.1 -5.249  7.993 1.184e-11 5.439e-08
## Solyc01g098190.2.1 -5.485  6.200 1.447e-11 5.439e-08
## Solyc02g071980.2.1 -5.012  7.469 2.462e-11 7.144e-08
## Solyc09g059170.1.1 -8.298  4.697 2.851e-11 7.144e-08
```

```
dim(results$table)
```

```
## [1] 15034     4
```

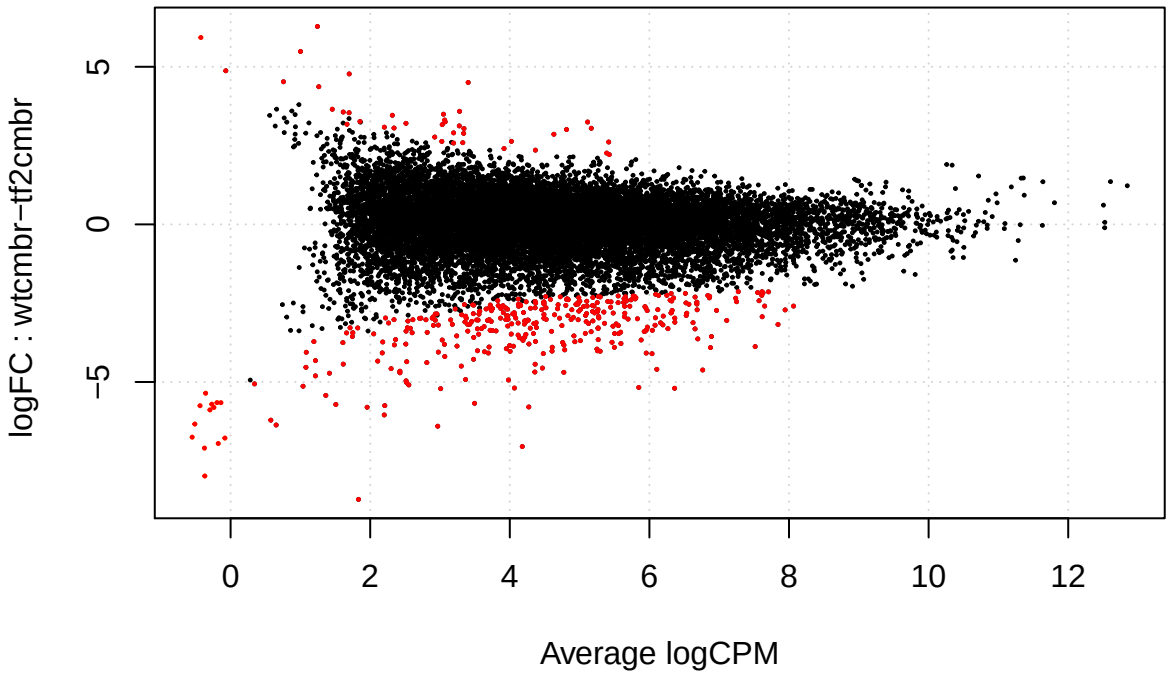
```
sum(results$table$FDR<.05) # How many are DE genes?
```

```
## [1] 396
```

```
summary(decideTestsDGE(DEtest,p.value=.05))
```

```
##    [,1]
## -1   356
## 0  14638
## 1     40
```

```
sig.genes <- rownames(results$table[results$table$FDR<0.05,])
plotSmear(d,de.tags=sig.genes)
```



Subset by all the ones with a significant score

```
results.sig <- subset(DEtest$table, DEtest$table$PValue < 0.05)
```

What are the genes that are misexpressed? For this we need to add some annotation

Essentially we are merging two annotations files to 1.) only sig genes 2.) all genes

```
annotation1<- read.delim("../ITAG2.3_all_Arabidopsis_ITAG_annotations.tsv", header=FALSE)  #Changed to
colnames(annotation1) <- c("ITAG", "SGN_annotation")
annotation2<- read.delim ("../ITAG2.3_all_Arabidopsis_annotated.tsv")
annotation <- merge(annotation1,annotation2, by = "ITAG")

#Making the only significant gene table
results.sig$ITAG <- rownames(results.sig)  #change row.names to ITAG for merging
results.sig.annotated <- merge(results.sig,annotation,by = "ITAG") #This is merging to only sig genes

#Making all table

results$table$ITAG <- rownames(results$table)
results.all.annotated <- merge(results$table, annotation,by = "ITAG")
```

Write table with results

```
write.table(results.all.annotated,"tf2cmbr_wtcmbr_DE_all.txt",sep="\t",row.names=F)
write.table(results.sig.annotated,"tf2cmbr_wtcmbr_DE.txt",sep="\t",row.names=F)
```