WT

Marginal Blastozone vs Other, B region (middle) attempt 1

```
library(edgeR)
```

Read in Data

Read in raw count data per gene. Add checknames to FALSE because it was making the columns unique.

```
counts <- read.delim("../sam2countsResults.tsv",check.names=FALSE,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. In this case I am going to get rid of wtbmbr8 and wtbother1.4, because their count are very low and this could be the reason I am getting the errors from attempt 1.

```
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"
```

```
wtbregion <- counts[,c(32:34, 37:39)]
head(wtbregion)
```

```
##                    wtbmbr2 wtbmbr3 wtbmbr6 wtbother3 wtbother5 wtbother8
## Solyc00g005040.2.1       2       4       3         8         0         3
## Solyc00g005050.2.1      20       5      18        25         0        14
```

```
## Solyc00g005060.1.1       1       2       1         0         0         0
## Solyc00g005070.1.1      14       6      12         6         2         4
## Solyc00g005080.1.1      25      15      27        29         0        11
## Solyc00g005150.1.1       0       0       3         2         0         2
```

```
colnames(wtbregion)
```

```
## [1] "wtbmbr2"   "wtbmbr3"   "wtbmbr6"   "wtbother3" "wtbother5" "wtbother8"
```

```
group <- c(rep("wtbmbr", 3), rep("wtbother", 3))
d <- DGEList(counts=wtbregion,group=group)
```

```
d$samples
```

```
##              group lib.size norm.factors
## wtbmbr2     wtbmbr  1355352            1
## wtbmbr3     wtbmbr  1213142            1
## wtbmbr6     wtbmbr  1598917            1
## wtbother3 wtbother  1076939            1
## wtbother5 wtbother   200587            1
## wtbother8 wtbother   499487            1
```

```
cpm.d <- cpm(d)
d <- d[rowSums(cpm.d>5)>=3,]
d <- estimateCommonDisp(d,verbose=T) #No error this time.
```

```
## Disp = 0.3968 , BCV = 0.6299
```

```
d <- calcNormFactors(d)
d <- estimateCommonDisp(d)
DEtest <- exactTest(d,pair=c("wtbmbr","wtbother"))
head(DEtest$table)
```

```
##                       logFC logCPM    PValue
## Solyc00g005050.2.1  0.8282  4.123 4.221e-01
## Solyc00g005070.1.1 -0.0464  3.438 1.000e+00
## Solyc00g005080.1.1  0.1422  4.396 1.000e+00
## Solyc00g005440.1.1 -0.6585  4.759 4.205e-01
## Solyc00g005840.2.1  3.8416  7.469 2.742e-06
## Solyc00g006470.1.1  1.7425  8.647 1.898e-02
```

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

```
## Comparison of groups:  wtbother-wtbmbr
##                      logFC logCPM    PValue       FDR
## Solyc04g074380.2.1 13.062 12.151 9.223e-27 1.334e-22
## Solyc04g074390.2.1 10.654 10.512 2.340e-21 1.692e-17
## Solyc01g014280.2.1 10.224  9.312 1.193e-19 5.751e-16
## Solyc01g065610.1.1  8.188  8.033 3.638e-15 1.315e-11
## Solyc02g076780.2.1  7.164 10.740 2.879e-14 8.326e-11
## Solyc01g065620.1.1  6.941  8.375 4.528e-13 1.091e-09
```

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

```
## [1] 14460     4
```

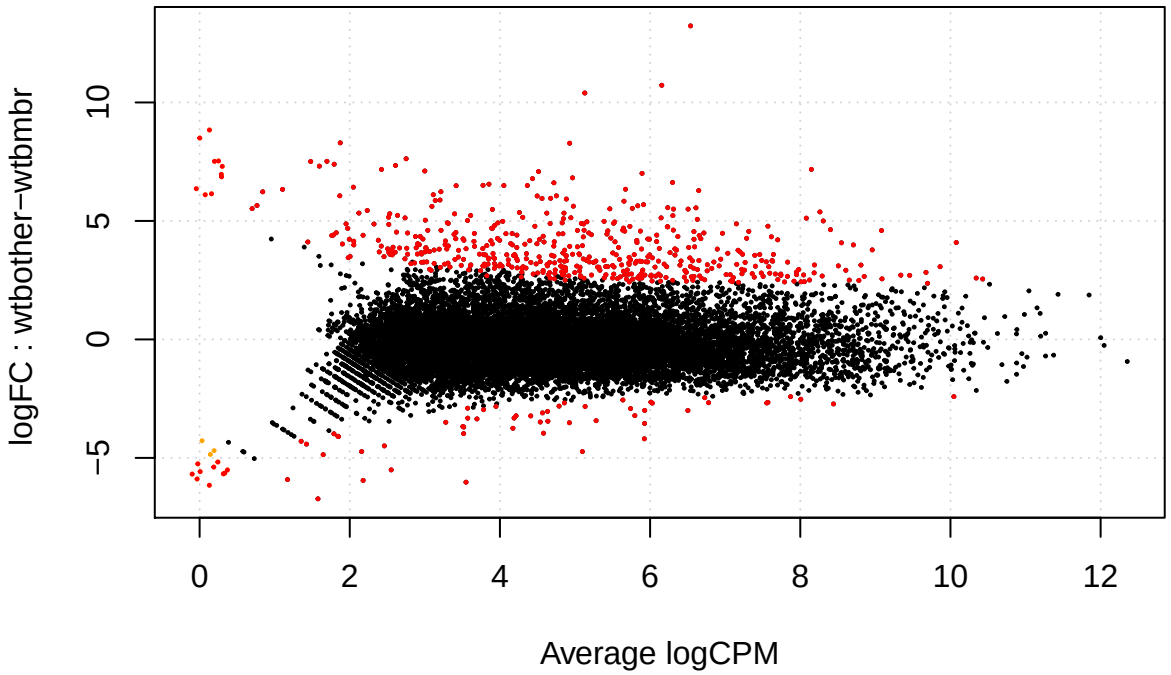
```
sum(results$table$FDR<.05) # How many are DE genes?
```

```
## [1] 566
```

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

```
##    [,1]
## -1   64
## 0 13894
## 1   502
```

```
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") #This s
```

Write table with results

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