WT

Marginal Blastozone A (distal) vs Marginal Blastzone in B (leaflet, mid) region

```
library(edgeR)
```

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"
```

```
WTambrVSbmbr <- counts[,c(24:26,32:34)]
head(WTambrVSbmbr)
```

```
##                   wtambr2 wtambr4 wtambr5 wtbmbr2 wtbmbr3 wtbmbr6
## Solyc00g005040.2.1       0       2       8       2       4       3
## Solyc00g005050.2.1       0       6       6      20       5      18
## Solyc00g005060.1.1       0       0       1       1       2       1
```

```
## Solyc00g005070.1.1      24       3       9      14       6      12
## Solyc00g005080.1.1       9      15      19      25      15      27
## Solyc00g005150.1.1       0       1       2       0       0       3
```

```
colnames(WTambrVSbmbr)
```

```
## [1] "wtambr2" "wtambr4" "wtambr5" "wtbmbr2" "wtbmbr3" "wtbmbr6"
```

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

```
d$samples
```

```
##          group lib.size norm.factors
## wtambr2 wtambr   395165            1
## wtambr4 wtambr   792542            1
## wtambr5 wtambr   632686            1
## wtbmbr2 wtbmbr  1355352            1
## wtbmbr3 wtbmbr  1213142            1
## wtbmbr6 wtbmbr  1598917            1
```

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

```
## Disp = 0.3611 , BCV = 0.6009
```

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

```
##                       logFC logCPM    PValue
## Solyc00g005050.2.1  0.3859  3.275 4.967e-01
## Solyc00g005070.1.1 -2.6718  3.844 4.260e-04
## Solyc00g005080.1.1 -1.1181  4.285 1.122e-01
## Solyc00g005440.1.1  0.4769  4.563 4.648e-01
## Solyc00g005840.2.1 -0.7871  4.646 2.347e-01
## Solyc00g006470.1.1 -5.5457 11.831 3.440e-13
```

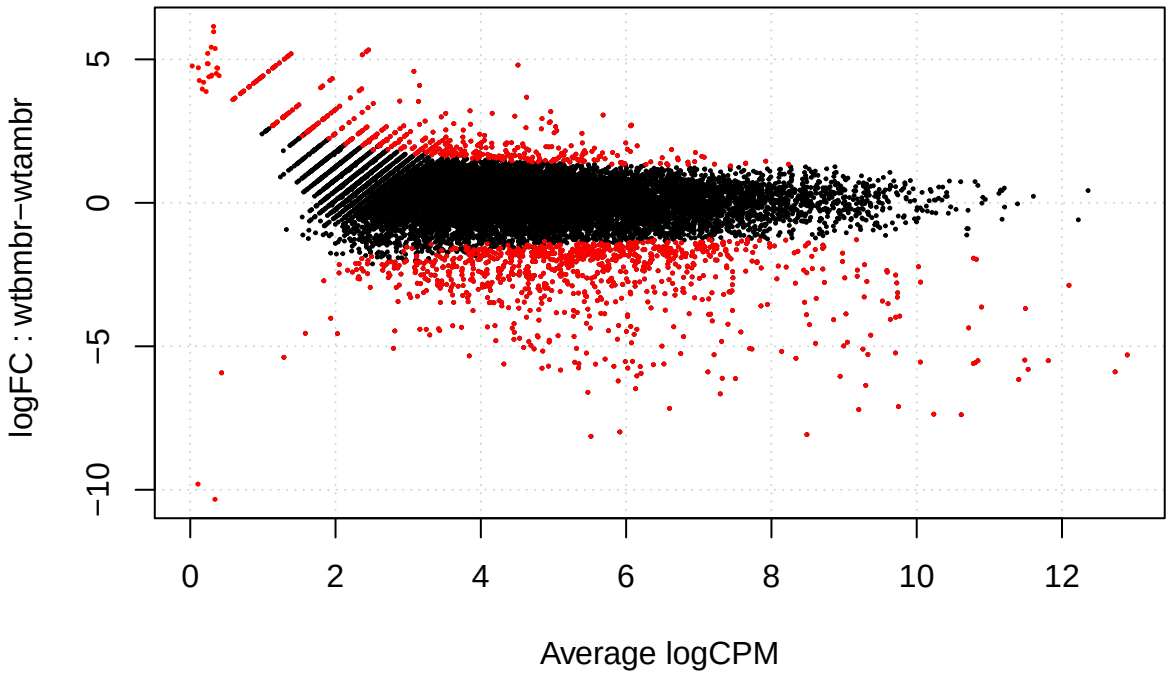
```
sum(DEtest$table$PValue<.05)
```

```
## [1] 1345
```

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

```
##    [,1]
## -1   381
## 0  13403
## 1     69
```

```
sig.genes <- rownames(DEtest$table[DEtest$table$PValue<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

```
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 =1,1, all.x=TRUE)
head(annotation)
```

```
##                 ITAG
## 1 Solyc00g005000.2.1
## 2 Solyc00g005020.1.1
## 3 Solyc00g005040.2.1
## 4 Solyc00g005050.2.1
## 5 Solyc00g005060.1.1
## 6 Solyc00g005070.1.1
##
## 1 Aspartic proteinase nepenthesin I (AHRD V1 **-- A9ZMF9_NEPAL); contains Interpro domain(s)  IPR001
## 2                                                                                          Unknown
## 3    Potassium channel (AHRD V1 ***- D0EM91_9ROSI); contains Interpro domain(s)  IPR000595  Cyclic nu
## 4                                                                  Arabinogalactan protein (AHRD V1 =
## 5                                                                                          Unknown
## 6                                                                                          Unknown
##         AGI symbol
## 1 AT3G20015   <NA>
```

```
## 2      <NA>   <NA>
## 3 AT5G46240   KAT1
## 4 AT5G11680   <NA>
## 5      <NA>   <NA>
## 6      <NA>   <NA>
##
## 1 pepsin A; similar to aspartyl protease family protein [Arabidopsis thaliana] (TAIR:AT3G18490.1); s
## 2
## 3
## 4
## 5
## 6
##   X..identity alignment.length e.value bit.score percent.query.align
## 1       63.76              447  7e-148       520               89.94
## 2          NA               NA      NA        NA                  NA
## 3       66.02              103   2e-37       150               85.71
## 4       76.96              204   1e-88       322               98.98
## 5          NA               NA      NA        NA                  NA
## 6          NA               NA      NA        NA                  NA
```

```
results.annotated <- merge(results.sig,annotation,by.x="row.names",by.y="ITAG",all.x=T,sort=F)
```

Write table with results

```
write.table(results.annotated,"wtambr_wtbmbr_DE1.txt",sep="\t",row.names=F)
```