

WT

## Marginal Blastozone A (tip) vs Marginal Blastzone in C (base) 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)
```

```
##          tf2ambr1 tf2ambr3 tf2ambr4 tf2ambr6 tf2aother1  
## Solyc00g005040.2.1      12      NA      3      12      NA  
## Solyc00g005050.2.1      33       1     14     17      NA  
## Solyc00g005060.1.1       1       5      1      1      NA  
## Solyc00g005070.1.1      14      22     23      5       3  
## Solyc00g005080.1.1      19       2     25     32      NA  
## Solyc00g005150.1.1       3      NA     NA      4      NA  
##          tf2aother2 tf2aother4 tf2aother7 tf2bmbr2 tf2bmbr5  
## Solyc00g005040.2.1       1      NA       2      NA      NA  
## Solyc00g005050.2.1       2       3      NA      NA       3  
## Solyc00g005060.1.1      NA      NA      NA     19      NA  
## Solyc00g005070.1.1       6      33       2    230       7  
## Solyc00g005080.1.1      12      10       3      6      12  
## Solyc00g005150.1.1      NA       2       1     NA       1  
##          tf2bmbr6 tf2bother1 tf2bother3 tf2bother4 tf2bother6  
## Solyc00g005040.2.1       1       6       3       5       2  
## Solyc00g005050.2.1      16      46       9      23      22  
## Solyc00g005060.1.1      NA      NA       1       1       1  
## Solyc00g005070.1.1      12      25       4      11      11  
## Solyc00g005080.1.1      33      52      12      15      12  
## Solyc00g005150.1.1       1      11      NA      NA       1  
##          tf2cmbr1.4 tf2cmbr3 tf2cmbr6 tf2cmbr7 tf2cother2  
## Solyc00g005040.2.1      NA       6       8       4       3  
## Solyc00g005050.2.1       1      34      17      12       4  
## Solyc00g005060.1.1      NA       1      NA      NA       1  
## Solyc00g005070.1.1      23      11       8       9       4  
## Solyc00g005080.1.1      22       7       8      12       9  
## Solyc00g005150.1.1       1       3      NA      NA       1  
##          tf2cother5 tf2cother6 tf2cother7 wtambr2 wtambr4  
## Solyc00g005040.2.1       8       4       1     NA       2  
## Solyc00g005050.2.1      10      16      12     NA       6
```

## Solyc00g005060.1.1	1	2	1	NA	NA	
## Solyc00g005070.1.1	11	5	5	24	3	
## Solyc00g005080.1.1	21	14	3	9	15	
## Solyc00g005150.1.1	6	1	NA	NA	1	
##	wtambr5	wtaoth1	wtaoth5	wtaoth6	wtaoth7	
## Solyc00g005040.2.1	8	1	1	1	NA	
## Solyc00g005050.2.1	6	17	16	9	2	
## Solyc00g005060.1.1	1	NA	NA	NA	NA	
## Solyc00g005070.1.1	9	8	6	5	5	
## Solyc00g005080.1.1	19	18	37	6	10	
## Solyc00g005150.1.1	2	2	5	NA	NA	
##	wtaoth8	wtbmr2	wtbmr3	wtbmr6	wtbmr8	wtbth1.4
## Solyc00g005040.2.1	2	2	4	3	NA	NA
## Solyc00g005050.2.1	3	20	5	18	NA	NA
## Solyc00g005060.1.1	2	1	2	1	1	NA
## Solyc00g005070.1.1	6	14	6	12	14	NA
## Solyc00g005080.1.1	7	25	15	27	NA	NA
## Solyc00g005150.1.1	2	NA	NA	3	NA	NA
##	wtbth3	wtbth5	wtbth8	wtcbr10	wtcbr1.4.6	
## Solyc00g005040.2.1	8	NA	3	NA	9	
## Solyc00g005050.2.1	25	NA	14	5	38	
## Solyc00g005060.1.1	NA	NA	NA	1	3	
## Solyc00g005070.1.1	6	2	4	5	12	
## Solyc00g005080.1.1	29	NA	11	NA	7	
## Solyc00g005150.1.1	2	NA	2	NA	1	
##	wtcbr2	wtcbr3	wtcbr7	wtcbr9	wtcot1.3.4	wtcot2
## Solyc00g005040.2.1	3	1	NA	NA	NA	NA
## Solyc00g005050.2.1	21	11	4	7	2	6
## Solyc00g005060.1.1	NA	NA	1	NA	13	NA
## Solyc00g005070.1.1	7	4	6	1	169	6
## Solyc00g005080.1.1	19	45	4	7	11	26
## Solyc00g005150.1.1	3	3	2	1	2	1
##	wtcot6					
## Solyc00g005040.2.1	12					
## Solyc00g005050.2.1	37					
## Solyc00g005060.1.1	NA					
## Solyc00g005070.1.1	24					
## Solyc00g005080.1.1	35					
## Solyc00g005150.1.1	5					

#### summary(counts)

##	tf2ambr1	tf2ambr3	tf2ambr4	tf2ambr6
##	Min. : 1	Min. : 1	Min. : 1	Min. : 1
##	1st Qu.: 5	1st Qu.: 1	1st Qu.: 4	1st Qu.: 4
##	Median : 18	Median : 2	Median : 17	Median : 15
##	Mean : 69	Mean : 7	Mean : 69	Mean : 56
##	3rd Qu.: 54	3rd Qu.: 6	3rd Qu.: 53	3rd Qu.: 45
##	Max. :10289	Max. :620	Max. :10818	Max. :8858
##	NA's :11302	NA's :17949	NA's :9602	NA's :10812
##	tf2aoth1	tf2aoth2	tf2aoth4	tf2aoth7
##	Min. : 1	Min. : 1	Min. : 1	Min. : 1
##	1st Qu.: 2	1st Qu.: 3	1st Qu.: 3	1st Qu.: 2
##	Median : 5	Median : 10	Median : 10	Median : 3

## Mean : 24	Mean : 41	Mean : 45	Mean : 10
## 3rd Qu.: 15	3rd Qu.: 30	3rd Qu.: 34	3rd Qu.: 9
## Max. :8136	Max. :5044	Max. :9013	Max. :1087
## NA's :19457	NA's :13064	NA's :12662	NA's :16587
## tf2bmbr2	tf2bmbr5	tf2bmbr6	tf2bother1
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 1	1st Qu.: 3	1st Qu.: 4	1st Qu.: 7
## Median : 3	Median : 11	Median : 17	Median : 31
## Mean : 29	Mean : 41	Mean : 63	Mean : 110
## 3rd Qu.: 11	3rd Qu.: 32	3rd Qu.: 50	3rd Qu.: 95
## Max. :5894	Max. :7353	Max. :6879	Max. :13995
## NA's :23681	NA's :12434	NA's :10534	NA's :8421
## tf2bother3	tf2bother4	tf2bother6	tf2cmbr1.4
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 3	1st Qu.: 4	1st Qu.: 4	1st Qu.: 2
## Median : 10	Median : 15	Median : 13	Median : 8
## Mean : 33	Mean : 49	Mean : 42	Mean : 41
## 3rd Qu.: 28	3rd Qu.: 44	3rd Qu.: 38	3rd Qu.: 35
## Max. :7837	Max. :6695	Max. :5456	Max. :3024
## NA's :11340	NA's :9938	NA's :9895	NA's :19540
## tf2cmbr3	tf2cmbr6	tf2cmbr7	tf2cother2
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 5	1st Qu.: 4	1st Qu.: 4	1st Qu.: 3
## Median : 19	Median : 14	Median : 14	Median : 11
## Mean : 65	Mean : 43	Mean : 44	Mean : 43
## 3rd Qu.: 56	3rd Qu.: 38	3rd Qu.: 39	3rd Qu.: 33
## Max. :8623	Max. :5767	Max. :6601	Max. :5476
## NA's :9762	NA's :12143	NA's :11416	NA's :13555
## tf2cother5	tf2cother6	tf2cother7	wtambr2
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 5	1st Qu.: 4	1st Qu.: 4	1st Qu.: 2
## Median : 18	Median : 13	Median : 12	Median : 5
## Mean : 59	Mean : 42	Mean : 40	Mean : 37
## 3rd Qu.: 52	3rd Qu.: 37	3rd Qu.: 35	3rd Qu.: 17
## Max. :9251	Max. :5670	Max. :5035	Max. :25469
## NA's :9545	NA's :10431	NA's :13298	NA's :19734
## wtambr4	wtambr5	wtaoother1	wtaoother5
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 3	1st Qu.: 3	1st Qu.: 4	1st Qu.: 5
## Median : 10	Median : 11	Median : 14	Median : 19
## Mean : 45	Mean : 35	Mean : 48	Mean : 75
## 3rd Qu.: 31	3rd Qu.: 30	3rd Qu.: 39	3rd Qu.: 58
## Max. :19492	Max. :3743	Max. :16475	Max. :28791
## NA's :12584	NA's :12089	NA's :10884	NA's :9646
## wtaoother6	wtaoother7	wtaoother8	wtbmr2
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 3	1st Qu.: 3	1st Qu.: 3	1st Qu.: 5
## Median : 9	Median : 8	Median : 10	Median : 18
## Mean : 27	Mean : 26	Mean : 29	Mean : 68
## 3rd Qu.: 25	3rd Qu.: 22	3rd Qu.: 26	3rd Qu.: 55
## Max. :2303	Max. :4900	Max. :2875	Max. :8049
## NA's :11518	NA's :11894	NA's :12597	NA's :10339
## wtbmr3	wtbmr6	wtbmr8	wtbother1.4
## Min. : 1	Min. : 1	Min. : 1	Min. : 1

```
## 1st Qu.: 4 1st Qu.: 5 1st Qu.: 1 1st Qu.: 1
## Median : 16 Median : 20 Median : 2 Median : 1
## Mean : 62 Mean : 80 Mean : 6 Mean : 4
## 3rd Qu.: 51 3rd Qu.: 63 3rd Qu.: 4 3rd Qu.: 1
## Max. :9833 Max. :12959 Max. :567 Max. :634
## NA's :10663 NA's :10203 NA's :22000 NA's :29943
## wtbother3 wtbother5 wtbother8 wtcnbr10
## Min. : 1 Min. : 1 Min. : 1 Min. : 1
## 1st Qu.: 4 1st Qu.: 1 1st Qu.: 3 1st Qu.: 3
## Median : 16 Median : 4 Median : 10 Median : 9
## Mean : 52 Mean : 22 Mean : 26 Mean : 26
## 3rd Qu.: 46 3rd Qu.: 13 3rd Qu.: 25 3rd Qu.: 23
## Max. :12673 Max. :4164 Max. :3054 Max. :3176
## NA's :9678 NA's :21339 NA's :11247 NA's :12664
## wtcnbr1.4.6 wtcnbr2 wtcnbr3 wtcnbr7
## Min. : 1 Min. : 1 Min. : 1 Min. : 1
## 1st Qu.: 5 1st Qu.: 5 1st Qu.: 5 1st Qu.: 3
## Median : 17 Median : 16 Median : 19 Median : 7
## Mean : 56 Mean : 60 Mean : 80 Mean : 23
## 3rd Qu.: 50 3rd Qu.: 48 3rd Qu.: 59 3rd Qu.: 20
## Max. :5921 Max. :10238 Max. :45264 Max. :3087
## NA's :9760 NA's :11531 NA's :10802 NA's :13907
## wtcnbr9 wtcother1.3.4 wtcother2 wtcother6
## Min. : 1 Min. : 1 Min. : 1 Min. : 1
## 1st Qu.: 3 1st Qu.: 1 1st Qu.: 2 1st Qu.: 5
## Median : 7 Median : 4 Median : 6 Median : 20
## Mean : 23 Mean : 13 Mean : 19 Mean : 73
## 3rd Qu.: 20 3rd Qu.: 10 3rd Qu.: 15 3rd Qu.: 60
## Max. :3538 Max. :4073 Max. :3330 Max. :9016
## NA's :13611 NA's :15020 NA's :13069 NA's :9330
```

```
colnames(counts)
```

```
## [1] "tf2ambr1" "tf2ambr3" "tf2ambr4" "tf2ambr6"
## [5] "tf2aother1" "tf2aother2" "tf2aother4" "tf2aother7"
## [9] "tf2bmbr2" "tf2bmbr5" "tf2bmbr6" "tf2bother1"
## [13] "tf2bother3" "tf2bother4" "tf2bother6" "tf2cnbr1.4"
## [17] "tf2cnbr3" "tf2cnbr6" "tf2cnbr7" "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" "wtcnbr10"
## [41] "wtcnbr1.4.6" "wtcnbr2" "wtcnbr3" "wtcnbr7"
## [45] "wtcnbr9" "wtcother1.3.4" "wtcother2" "wtcother6"
```

```
#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] "tf2aoth1"      "tf2aoth2"      "tf2aoth4"      "tf2aoth7"
## [9] "tf2bmbr2"      "tf2bmbr5"      "tf2bmbr6"      "tf2both1"
## [13] "tf2both3"      "tf2both4"      "tf2both6"      "tf2cmbr1.4"
## [17] "tf2cmbr3"      "tf2cmbr6"      "tf2cmbr7"      "tf2coth2"
## [21] "tf2coth5"      "tf2coth6"      "tf2coth7"      "wtambr2"
## [25] "wtambr4"      "wtambr5"      "wtaoth1"      "wtaoth5"
## [29] "wtaoth6"      "wtaoth7"      "wtaoth8"      "wtbmbr2"
## [33] "wtbmbr3"      "wtbmbr6"      "wtbmbr8"      "wtboth1.4"
## [37] "wtboth3"      "wtboth5"      "wtboth8"      "wtcmbr10"
## [41] "wtcmbr1.4.6"  "wtcmbr2"      "wtcmbr3"      "wtcmbr7"
## [45] "wtcmbr9"      "wtcoth1.3.4" "wtcoth2"      "wtcoth6"
```

```
WTambrVScmbr <- counts[,c(24:26,42:45)]
head(WTambrVScmbr)
```

```
##          wtambr2 wtambr4 wtambr5 wtcnbr2 wtcnbr3 wtcnbr7 wtcnbr9
## Solyc00g005040.2.1      0      2      8      3      1      0      0
## Solyc00g005050.2.1      0      6      6     21     11      4      7
## Solyc00g005060.1.1      0      0      1      0      0      1      0
## Solyc00g005070.1.1     24      3      9      7      4      6      1
## Solyc00g005080.1.1      9     15     19     19     45      4      7
## Solyc00g005150.1.1      0      1      2      3      3      2      1
```

```
colnames(WTambrVScmbr)
```

```
## [1] "wtambr2" "wtambr4" "wtambr5" "wtcmbr2" "wtcmbr3" "wtcmbr7" "wtcmbr9"
```

```
group <- c(rep("wtambr", 3), rep("wtcmbr", 4))
d <- DGEList(counts=WTambrVScmbr,group=group)
```

```
d$samples
```

```
##          group lib.size norm.factors
## wtambr2 wtambr   395165           1
## wtambr4 wtambr   792542           1
## wtambr5 wtambr   632686           1
## wtcnbr2 wtcnbr  1130695           1
## wtcnbr3 wtcnbr  1560130           1
## wtcnbr7 wtcnbr   374882           1
## wtcnbr9 wtcnbr   386974           1
```

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

```
## Disp = 0.3527 , BCV = 0.5939
```

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

```
##               logFC logCPM    PValue
## Solyc00g005050.2.1  0.8130  3.648 2.374e-01
## Solyc00g005070.1.1 -2.8673  4.409 2.045e-04
## Solyc00g005080.1.1 -0.7528  4.633 2.250e-01
## Solyc00g005440.1.1  0.1915  4.502 8.108e-01
## Solyc00g005840.2.1 -0.4983  5.033 4.268e-01
## Solyc00g006470.1.1 -3.0931 11.734 5.860e-07
```

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

```
## Comparison of groups: wtcnbr-wtambr
##               logFC logCPM    PValue    FDR
## Solyc12g098590.1.1 -9.336  8.137 7.723e-24 1.105e-19
## Solyc00g071180.2.1 -9.136  6.687 1.440e-20 7.985e-17
## Solyc12g010020.1.1 -9.317  6.950 1.674e-20 7.985e-17
## Solyc00g187050.2.1 -9.093  6.614 2.801e-20 9.370e-17
## Solyc09g059170.1.1 -10.439 6.614 3.273e-20 9.370e-17
## Solyc03g098780.1.1 -6.814  9.703 4.490e-20 1.071e-16
```

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

```
## [1] 14312      4
```

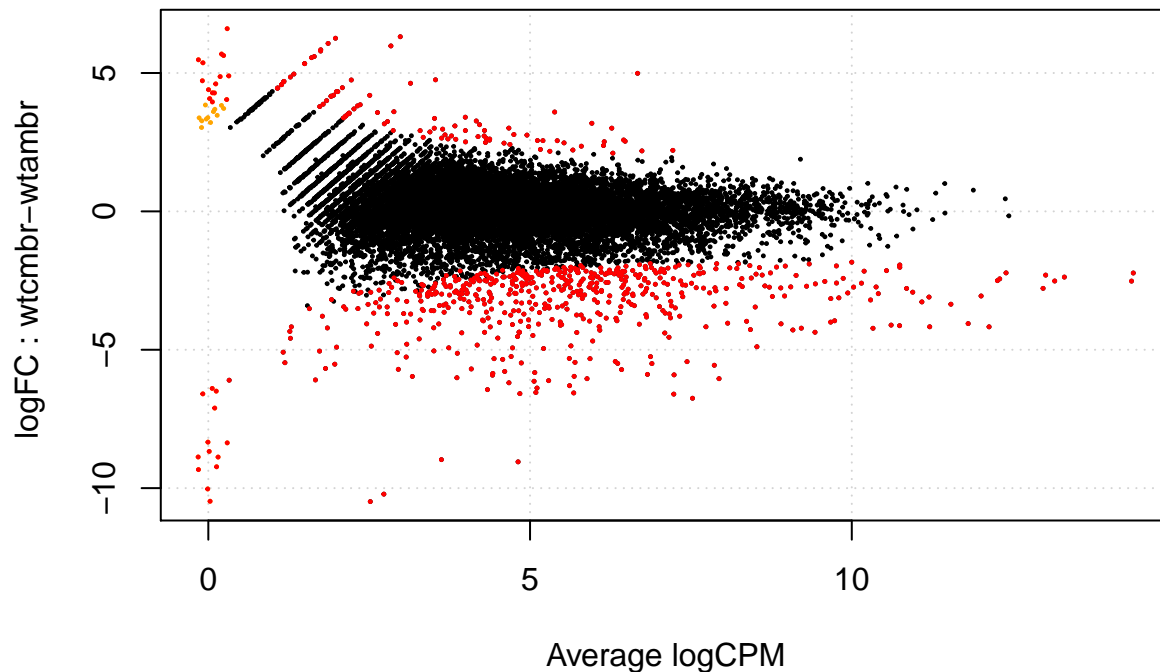
```
sum(results$table$FDR<.05) # How many are DE genes?
```

```
## [1] 666
```

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

```
##      [,1]
## -1    558
##  0   13646
##  1     108
```

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