Question 1

1. **Are there genes that are differentially expressed between MBR and Rachis in all three groups across the longitudinal axis?** Genotype specific. Basically asking what are the genes that define the Marginal Blastozone. What are the genes that define the Rachis. I am most interested in what is occuring in WT. This can be performed in each genotype and compared. **Hypothesis**: There will be genes that overlap between the analysis done between the two genotypes because the Marginal Blastozone.

Key to Samples

genotype: either wildtype of *tf2*

region: A. tip B. early emmerging leaflet C. base

type: MBR = Marginal Blastozone Region, other = the rachis or midvein region

libraries

```
library(edgeR)
```

```
## Loading required package: limma
```

```
library(locfit)
```

```
## locfit 1.5-9.1    2013-03-22
```

```
library(statmod)
```

```
## Warning: package 'statmod' was built under R version 3.1.3
```

Read in Data

Read in raw count data per gene.

```
counts <- read.delim("../../data/sam2countsResults.tsv")

colnames(counts)

#Remove tf2ambr.3, wtcmbr.1.4.6 and tf2aother7, ?wtbother1.4,
#because read count is so low.  See readsReps.Rmd for full report.

counts <- counts[,-c(3,9,37,42)]

#check the file
#head(counts)
#colnames(counts)

#need to convert NA to 0 counts
counts[is.na(counts)] <- 0
```

Normalization

```
#make the groups
colnames(counts)
```

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

```
sample <- gsub("[0-9]", "", names(counts))
sample <- gsub("\\.", "", sample)
sample <- sample[-1]

#set genotype
designTable <- as.data.frame(sample)

designTable$genotype <- ifelse(
  grepl("wt", designTable$sample, ignore.case = T), "wt",
    ifelse(
      grepl("tf", designTable$sample, ignore.case = T), "tf2", "unknown")
  )

#set type

designTable$tissue <- ifelse(
  grepl("other", designTable$sample, ignore.case = T), "rachis",
    ifelse(
      grepl("mbr", designTable$sample, ignore.case = T), "mbr",
         "unknown")
  )

#Set Region
designTable$region <- ifelse(
  grepl("a", designTable$sample, ignore.case = T), "A",
    ifelse(
      grepl("c", designTable$sample, ignore.case = T), "C", "B")
  )

head(designTable)
```

```
##     sample genotype tissue region
## 1   tfambr      tf2    mbr      A
## 2   tfambr      tf2    mbr      A
```

```
## 3   tfambr      tf2    mbr      A
## 4 tfaother      tf2 rachis      A
## 5 tfaother      tf2 rachis      A
## 6 tfaother      tf2 rachis      A
```

```
genotype <- designTable$genotype
sample <- designTable$sample
tissue <- designTable$region
region <- designTable$r
```

```
#put into DGE List
dim(counts)
```

```
## [1] 30308    45
```

```
y <- DGEList(counts=counts[,2:45], genes=counts[,1], group = sample)

cpm.y <- cpm(y) #counts per million
y <- y[rowSums(cpm.y > 5) >= 3,] # get rid of genes with low counts

y <- estimateCommonDisp(y,verbose=T) #Estimates common negative binomial dispersion by conditional maxi
```

```
## Disp = 0.4675 , BCV = 0.6837
```

```
y$samples
```

```
##                 group lib.size norm.factors
## tf2ambr1        tfambr  1313540            1
## tf2ambr4        tfambr  1438416            1
## tf2ambr6        tfambr  1088653            1
## tf2aother1    tfaother   263117            1
## tf2aother2    tfaother   698710            1
## tf2aother4    tfaother   792325            1
## tf2bmbr2        tfbmbr   189160            1
## tf2bmbr5        tfbmbr   727355            1
## tf2bmbr6        tfbmbr  1244342            1
## tf2bother1    tfbother  2415227            1
## tf2bother3    tfbother   626786            1
## tf2bother4    tfbother  1003586            1
## tf2bother6    tfbother   854903            1
## tf2cmbr1.4      tfcmbr   443572            1
## tf2cmbr3        tfcmbr  1337575            1
## tf2cmbr6        tfcmbr   790129            1
## tf2cmbr7        tfcmbr   832907            1
## tf2cother2    tfcother   723602            1
## tf2cother5    tfcother  1216379            1
## tf2cother6    tfcother   838942            1
## tf2cother7    tfcother   676969            1
## wtambr2         wtambr   395165            1
## wtambr4         wtambr   792542            1
## wtambr5         wtambr   632686            1
```

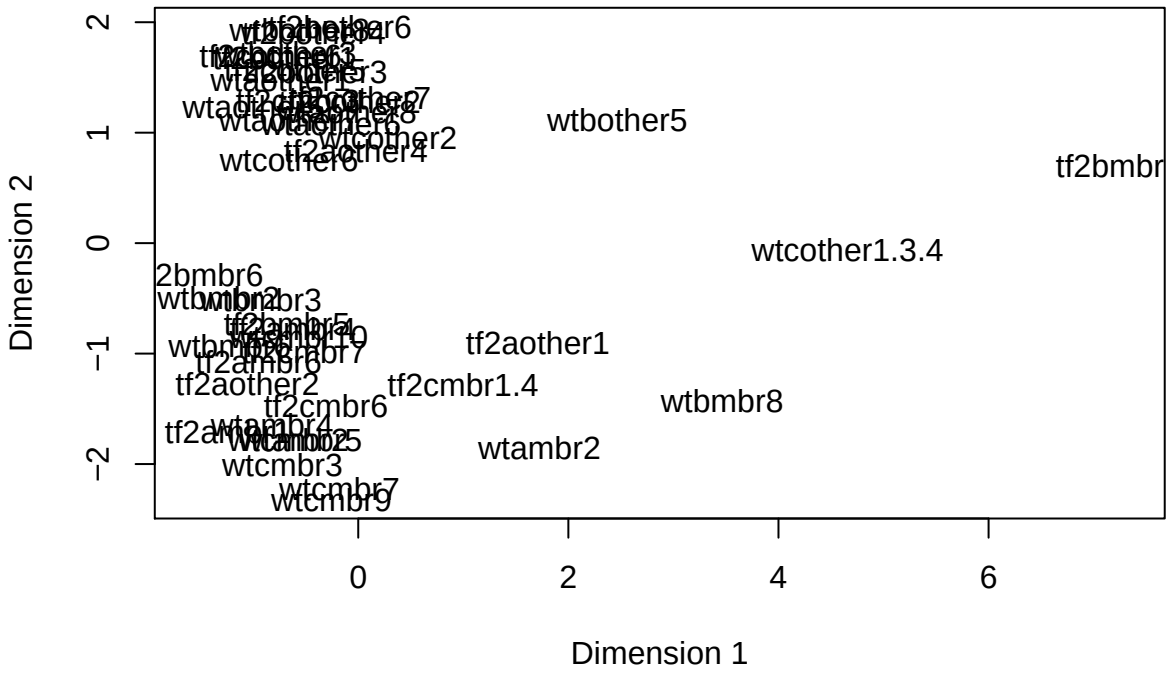
```
## wtaother1     wtaother   929017           1
## wtaother5     wtaother  1555921           1
## wtaother6     wtaother   498294           1
## wtaother7     wtaother   479003           1
## wtaother8     wtaother   510148           1
## wtbmbr2         wtbmbr  1355352           1
## wtbmbr3         wtbmbr  1213142           1
## wtbmbr6         wtbmbr  1598917           1
## wtbmbr8         wtbmbr    48352           1
## wtbother3     wtbother  1076939           1
## wtbother5     wtbother   200587           1
## wtbother8     wtbother   499487           1
## wtcmbr10        wtcmbr   459717           1
## wtcmbr2         wtcmbr  1130695           1
## wtcmbr3         wtcmbr  1560130           1
## wtcmbr7         wtcmbr   374882           1
## wtcmbr9         wtcmbr   386974           1
## wtcother1.3.4 wtcother   197345           1
## wtcother2     wtcother   319043           1
## wtcother6     wtcother  1525172           1
```

```
y <- calcNormFactors(y)
y <- estimateCommonDisp(y, verbose = T) #Disp = 0.46228 , BCV = 0.6799
```

```
## Disp = 0.4424 , BCV = 0.6651
```

```
plotMDS(cpm(y, log=TRUE), column=1) #Disp = 0.44804 , BCV = 0.6694
```

```
## Warning: "column" is not a graphical parameter
## Warning: "column" is not a graphical parameter
## Warning: "column" is not a graphical parameter
## Warning: "column" is not a graphical parameter
## Warning: "column" is not a graphical parameter
## Warning: "column" is not a graphical parameter
## Warning: "column" is not a graphical parameter
```



There are a few outliers, but these are kept in as they may reflect how homogeneous these tissues are. They are only in the tf2 mutant.

Analysis - WT

In order to answer the question: Are there genes that are differentially expressed between MBR and Rachis tissue in all three groups across the longitudinal axis? I am going to use an additive linear model, with the longitudinal axis as the blocking factor. This should be genotype specific.

```
#First I need to subset based on genotype

colnames(y)
```

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

```
wtY<- y[,22:44]

designTableWT <- designTable[22:44,]

designTableWT
```

```
##      sample genotype tissue region
## 22   wtambr       wt    mbr      A
## 23   wtambr       wt    mbr      A
## 24   wtambr       wt    mbr      A
## 25 wtaother       wt rachis      A
## 26 wtaother       wt rachis      A
## 27 wtaother       wt rachis      A
## 28 wtaother       wt rachis      A
## 29 wtaother       wt rachis      A
## 30   wtbmbr       wt    mbr      B
## 31   wtbmbr       wt    mbr      B
## 32   wtbmbr       wt    mbr      B
## 33   wtbmbr       wt    mbr      B
## 34 wtbother       wt rachis      B
## 35 wtbother       wt rachis      B
## 36 wtbother       wt rachis      B
## 37   wtcmbr       wt    mbr      C
## 38   wtcmbr       wt    mbr      C
## 39   wtcmbr       wt    mbr      C
## 40   wtcmbr       wt    mbr      C
## 41   wtcmbr       wt    mbr      C
## 42 wtcother       wt rachis      C
## 43 wtcother       wt rachis      C
## 44 wtcother       wt rachis      C
```

```
wtRegion <- designTableWT$region
wtTissue <- designTableWT$tissue

design <- model.matrix(~wtRegion + wtTissue)

rownames(design) <- colnames(wtY)

design
```

```
##               (Intercept) wtRegionB wtRegionC wtTissuerachis
## wtambr2                 1         0         0              0
## wtambr4                 1         0         0              0
## wtambr5                 1         0         0              0
## wtaother1               1         0         0              1
## wtaother5               1         0         0              1
## wtaother6               1         0         0              1
## wtaother7               1         0         0              1
## wtaother8               1         0         0              1
## wtbmbr2                 1         1         0              0
## wtbmbr3                 1         1         0              0
## wtbmbr6                 1         1         0              0
## wtbmbr8                 1         1         0              0
## wtbother3               1         1         0              1
## wtbother5               1         1         0              1
## wtbother8               1         1         0              1
## wtcmbr10                1         0         1              0
## wtcmbr2                 1         0         1              0
## wtcmbr3                 1         0         1              0
## wtcmbr7                 1         0         1              0
```

```
## wtcmbr9                 1         0         1              0
## wtcother1.3.4           1         0         1              1
## wtcother2               1         0         1              1
## wtcother6               1         0         1              1
## attr(,"assign")
## [1] 0 1 1 2
## attr(,"contrasts")
## attr(,"contrasts")$wtRegion
## [1] "contr.treatment"
##
## attr(,"contrasts")$wtTissue
## [1] "contr.treatment"
```

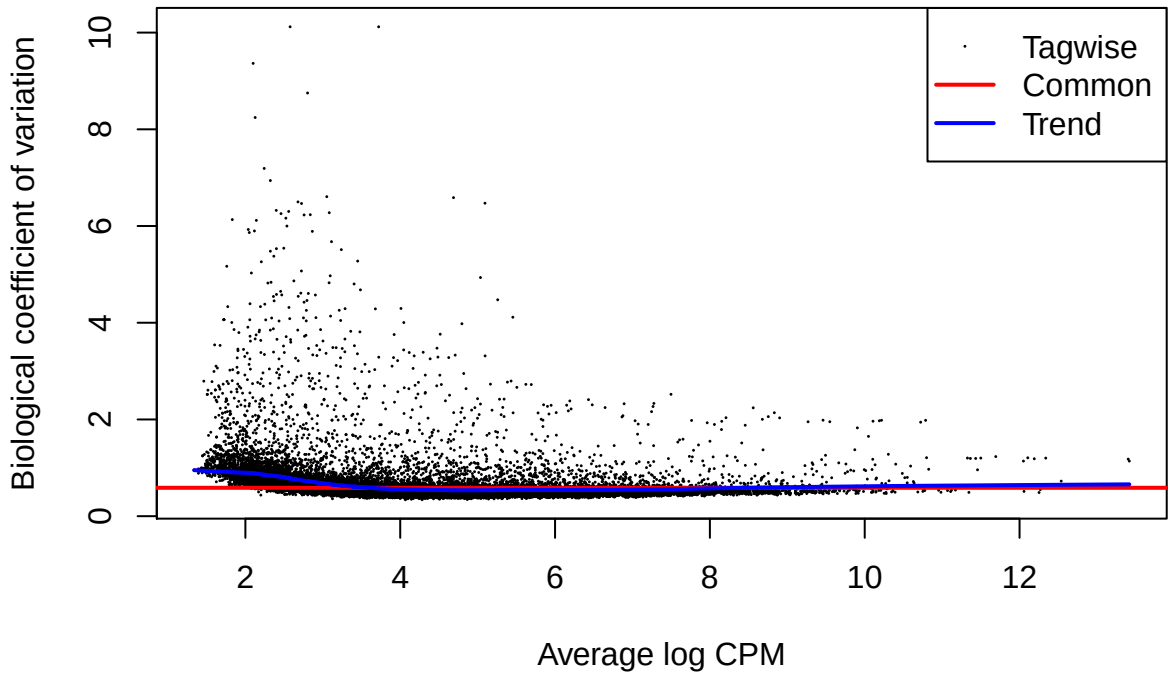
```
wtY <- estimateDisp(wtY, design, robust=TRUE) #Estimate Dispersion
```

```
## Loading required package: splines
```

```
wtY$common.dispersion # 0.3489676
```

```
## [1] 0.3448
```

```
plotBCV(wtY)
```



WT- Differential Gene Expression

```
fit <- glmFit(wtY, design)
lrt <- glmLRT(fit)
```

Here we see the top tags for MBR vs rachis tissue differential. It is adjusted for baseline differences between top, middle, and base.

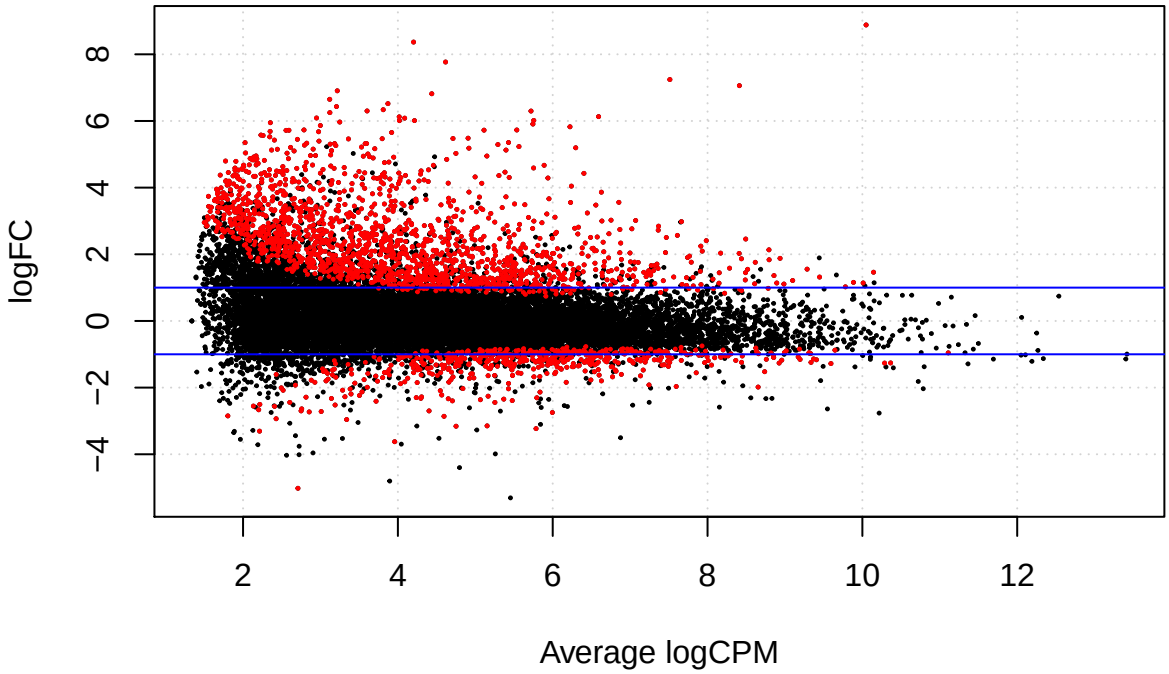
```
o <- order(lrt$table$PValue)

#cpm(wtY)[o[1:10],] #check to see differences
```

```
summary(de <- decideTestsDGE(lrt))
```

```
##    [,1]
## -1   529
## 0  16821
## 1   1707
```

```
detags <- rownames(y)[as.logical(de)]
plotSmear(lrt, de.tags=detags)
abline(h=c(-1, 1), col="blue")
```



All gene output

```
results <- topTags(lrt, n=Inf)

dim(results$table)
```

```
## [1] 19057     6
```

```
summary(de <- decideTestsDGE(lrt))
```

```
##    [,1]
## -1   529
## 0  16821
## 1   1707
```

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

```
## [1] 2236
```

```
#Subset only significant
results.sig <- subset(results$table, results$table$FDR < 0.05)

sig.genes <- results.sig$genes #only gene names
```

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

#Change headers for merging
colnames(results.sig)[1]<- "itag"
colnames(annotation)[1] <- "itag"
colnames(results$table)[1] <- "itag"

results.sig.annotated <- merge(results.sig, annotation, by = "itag", all.x=TRUE) #This is merging only

#Making all table

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

#Write out table to file
write.table(results.all.annotated, file = "WT.allresults.question1.txt", sep = "\t",row.names=F)

write.table(results.sig.annotated, file = "WT.onlysigresults.question1.txt", sep = "\t",row.names=F)
```

Set-Up

```
sessionInfo()
```

```
## R version 3.1.2 (2014-10-31)
## Platform: x86_64-apple-darwin13.4.0 (64-bit)
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] splines   stats     graphics  grDevices utils     datasets  methods
## [8] base
##
## other attached packages:
```

```
## [1] statmod_1.4.21 locfit_1.5-9.1 edgeR_3.6.7    limma_3.20.8
##
## loaded via a namespace (and not attached):
##  [1] digest_0.6.4    evaluate_0.5.5  formatR_0.10    grid_3.1.2
##  [5] htmltools_0.2.4 knitr_1.6       lattice_0.20-29 rmarkdown_0.6.1
##  [9] stringr_0.6.2   tools_3.1.2     yaml_2.1.13
```