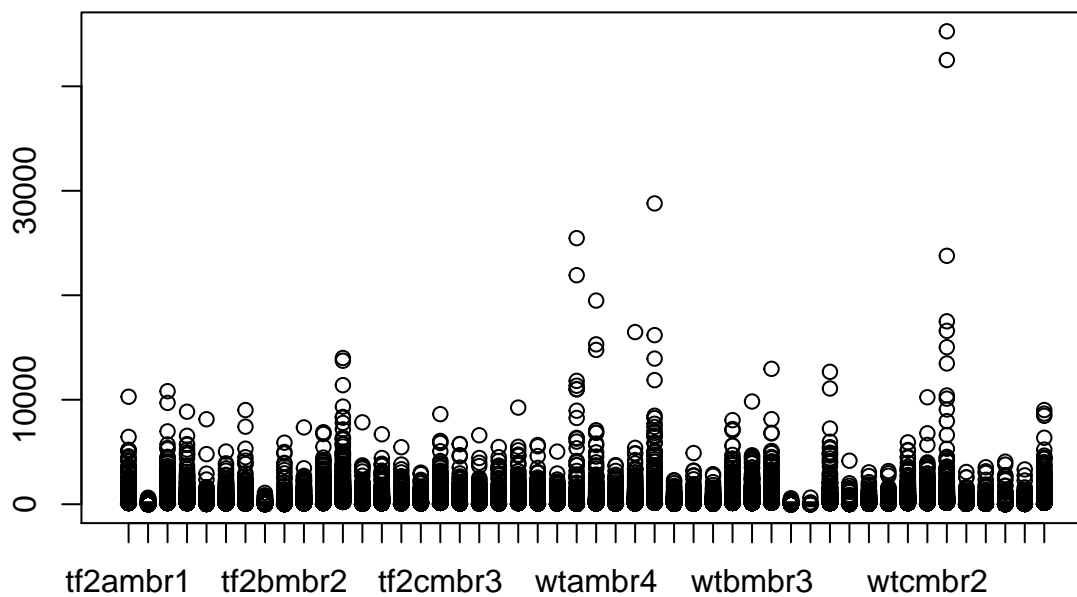


## Analyzing RNAseq for differential expression

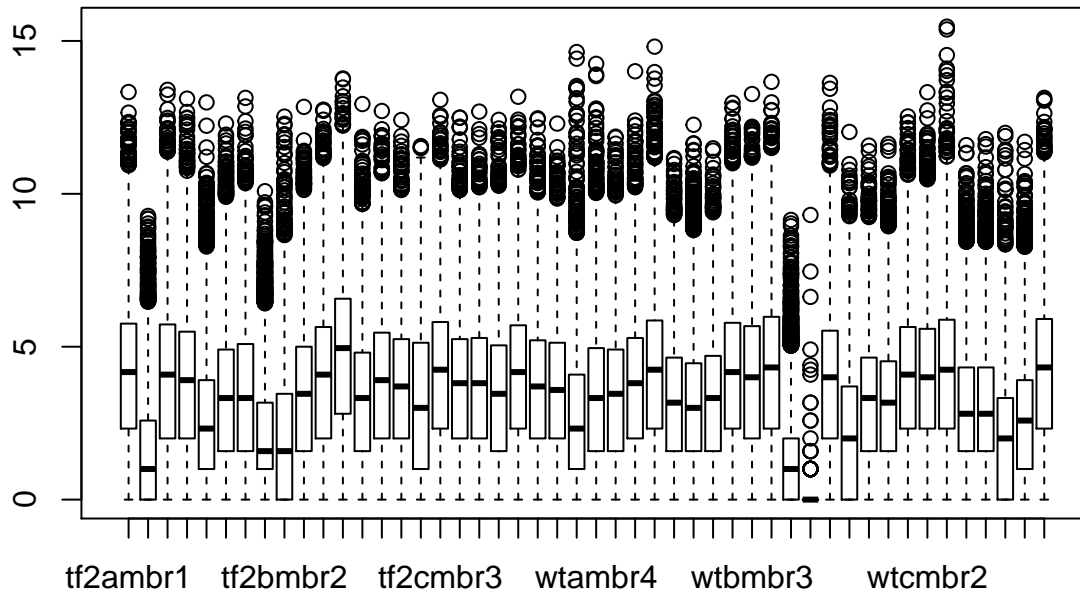
```
library(edgeR)
```

read in raw count data per gene

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



```
#Look at boxplot of log2  
normexplog <- log(counts,2)  
boxplot(normexplog)
```



```
#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 tf2bibr2 tf2bibr5
## 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
##
##          tf2bibr6 tf2bither1 tf2bither3 tf2bither4 tf2bither6
## 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 wtaother1 wtaother5 wtaother6 wtaother7
## 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
##
##      wtaother8 wtbmbr2 wtbmbr3 wtbmbr6 wtbmbr8 wtbother1.4
## 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
##
##      wtbother3 wtbother5 wtbother8 wtcnbr10 wtcnbr1.4.6
## 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
##
##      wtcnbr2 wtcnbr3 wtcnbr7 wtcnbr9 wtcother1.3.4 wtcother2
## 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
##
##      wtcother6
## 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)
```

Library wtbother1.4 is disturbingly small, I am going to remove it from analysis.

```
counts <- subset(counts, select=-c(wtbother1.4))
```

Need to convert NA to 0 counts

```
counts[is.na(counts)] <- 0
```

We are adding a new column “group”, that identifies which library belongs to which group sample type.

```
group <- c(
  rep("tf2ambr", 4),
  rep("tf2aother", 4),
  rep("tf2bibr", 3),
  rep("tf2bother", 5),
  rep("tf2cmbr", 4),
  rep("tf2cother", 4),
  rep("wtambr", 3),
  rep("wtaother", 4),
  rep("wtbibr", 4),
  rep("wtbother", 3),
  rep("wtcmbr", 6),
  rep("wtcother", 3)
)
```

Now we have to put the data into a DGEList object. [Here is a nice summary](#) of what the DGEList object is all about.

```
d <- DGEList(counts=counts, group=group)
```

Check to make sure each group matches the correct sample.

```
head(d$counts)
```

```
##          tf2ambr1 tf2ambr3 tf2ambr4 tf2ambr6 tf2aother1
## Solyc00g005040.2.1      12      0      3      12      0
## Solyc00g005050.2.1      33      1     14     17      0
## Solyc00g005060.1.1       1      5      1      1      0
## Solyc00g005070.1.1      14     22     23      5      3
## Solyc00g005080.1.1      19      2     25     32      0
## Solyc00g005150.1.1       3      0      0      4      0
##          tf2aother2 tf2aother4 tf2aother7 tf2bibr2 tf2bibr5
## Solyc00g005040.2.1       1      0      2      0      0
## Solyc00g005050.2.1       2      3      0      0      3
## Solyc00g005060.1.1       0      0      0     19      0
## Solyc00g005070.1.1       6     33      2    230      7
## Solyc00g005080.1.1      12     10      3      6     12
## Solyc00g005150.1.1       0      2      1      0      1
##          tf2bibr6 tf2bother1 tf2bother3 tf2bother4 tf2bother6
## Solyc00g005040.2.1       1      6      3      5      2
## Solyc00g005050.2.1      16     46      9     23     22
## Solyc00g005060.1.1       0      0      1      1      1
## Solyc00g005070.1.1      12     25      4     11     11
## Solyc00g005080.1.1      33     52     12     15     12
## Solyc00g005150.1.1       1     11      0      0      1
##          tf2cmbr1.4 tf2cmbr3 tf2cmbr6 tf2cmbr7 tf2cother2
## Solyc00g005040.2.1       0      6      8      4      3
## Solyc00g005050.2.1       1     34     17     12      4
## Solyc00g005060.1.1       0      1      0      0      1
## Solyc00g005070.1.1      23     11      8      9      4
## Solyc00g005080.1.1      22      7      8     12      9
## Solyc00g005150.1.1       1      3      0      0      1
```

```

##          tf2cother5 tf2cother6 tf2cother7 wtambr2 wtambr4
## Solyc00g005040.2.1      8      4      1      0      2
## Solyc00g005050.2.1     10     16     12      0      6
## Solyc00g005060.1.1      1      2      1      0      0
## Solyc00g005070.1.1     11      5      5     24      3
## Solyc00g005080.1.1     21     14      3      9     15
## Solyc00g005150.1.1      6      1      0      0      1
##          wtambr5 wtaother1 wtaother5 wtaother6 wtaother7
## Solyc00g005040.2.1      8      1      1      1      0
## Solyc00g005050.2.1      6     17     16      9      2
## Solyc00g005060.1.1      1      0      0      0      0
## Solyc00g005070.1.1      9      8      6      5      5
## Solyc00g005080.1.1     19     18     37      6     10
## Solyc00g005150.1.1      2      2      5      0      0
##          wtaother8 wtbmbr2 wtbmbr3 wtbmbr6 wtbmbr8 wtbother3
## Solyc00g005040.2.1      2      2      4      3      0      8
## Solyc00g005050.2.1      3     20      5     18      0     25
## Solyc00g005060.1.1      2      1      2      1      1      0
## Solyc00g005070.1.1      6     14      6     12     14      6
## Solyc00g005080.1.1      7     25     15     27      0     29
## Solyc00g005150.1.1      2      0      0      3      0      2
##          wtbother5 wtbother8 wtcnbr10 wtcnbr1.4.6 wtcnbr2
## Solyc00g005040.2.1      0      3      0      9      3
## Solyc00g005050.2.1      0     14      5     38     21
## Solyc00g005060.1.1      0      0      1      3      0
## Solyc00g005070.1.1      2      4      5     12      7
## Solyc00g005080.1.1      0     11      0      7     19
## Solyc00g005150.1.1      0      2      0      1      3
##          wtcnbr3 wtcnbr7 wtcnbr9 wtcnother1.3.4 wtcnother2
## Solyc00g005040.2.1      1      0      0      0      0
## Solyc00g005050.2.1     11      4      7      2      6
## Solyc00g005060.1.1      0      1      0     13      0
## Solyc00g005070.1.1      4      6      1    169      6
## Solyc00g005080.1.1     45      4      7     11     26
## Solyc00g005150.1.1      3      2      1      2      1
##          wtcnother6
## Solyc00g005040.2.1     12
## Solyc00g005050.2.1     37
## Solyc00g005060.1.1      0
## Solyc00g005070.1.1     24
## Solyc00g005080.1.1     35
## Solyc00g005150.1.1      5

```

d\$samples

```

##          group lib.size norm.factors
## tf2ambr1      tf2ambr 1313540      1
## tf2ambr3      tf2ambr   91726      1
## tf2ambr4      tf2ambr 1438416      1
## tf2ambr6      tf2ambr 1088653      1
## tf2aother1     tf2aother 263117      1
## tf2aother2     tf2aother 698710      1
## tf2aother4     tf2aother 792325      1
## tf2aother7     tf2aother 142504      1

```

```

## tf2bmbr2      tf2bmbr  189160      1
## tf2bmbr5      tf2bmbr  727355      1
## tf2bmbr6      tf2bmbr 1244342      1
## tf2bother1    tf2bother 2415227      1
## tf2bother3    tf2bother  626786      1
## tf2bother4    tf2bother 1003586      1
## tf2bother6    tf2bother  854903      1
## tf2cmbr1.4    tf2bother  443572      1
## tf2cmbr3      tf2cmbr 1337575      1
## tf2cmbr6      tf2cmbr  790129      1
## tf2cmbr7      tf2cmbr  832907      1
## tf2cother2    tf2cmbr  723602      1
## tf2cother5    tf2cother 1216379      1
## tf2cother6    tf2cother  838942      1
## tf2cother7    tf2cother  676969      1
## wtambr2       tf2cother  395165      1
## wtambr4       wtambr   792542      1
## wtambr5       wtambr   632686      1
## wtaother1     wtambr   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
## wtcnbr10      wtcnbr    459717      1
## wtcnbr1.4.6   wtcnbr   1158809      1
## wtcnbr2       wtcnbr   1130695      1
## wtcnbr3       wtcnbr   1560130      1
## wtcnbr7       wtcnbr    374882      1
## wtcnbr9       wtcnbr    386974      1
## wtcnbr1.3.4   wtcnbr    197345      1
## wtcnbr2       wtcnbr    319043      1
## wtcnbr6       wtcnbr   1525172      1

```

```

#Filter to exclude genes that have <2 counts in (N Rep)-1
cpm.d<- cpm(d)
d <- d[rowSums(cpm.d>1)>=3,]

#normalize library
d <- calcNormFactors(d)
d$pseudo<- equalizeLibSizes(d)
write.csv (d$pseudo$d$pseudo,"normalized_read_count.csv")

# read in norm data
data_norm <- read.csv("normalized_read_count.csv", row.names=1)
dim(data_norm)

```

```
## [1] 24570    47
```

## MDS using dist and cmdscale

library

```
library(ggplot2)
```

Prepare for MDS plot

```
m <- cmdscale(dist(scale(t(data_norm))))  
colnames(m) <- c("x", "y")  
m <- as.data.frame(m)
```

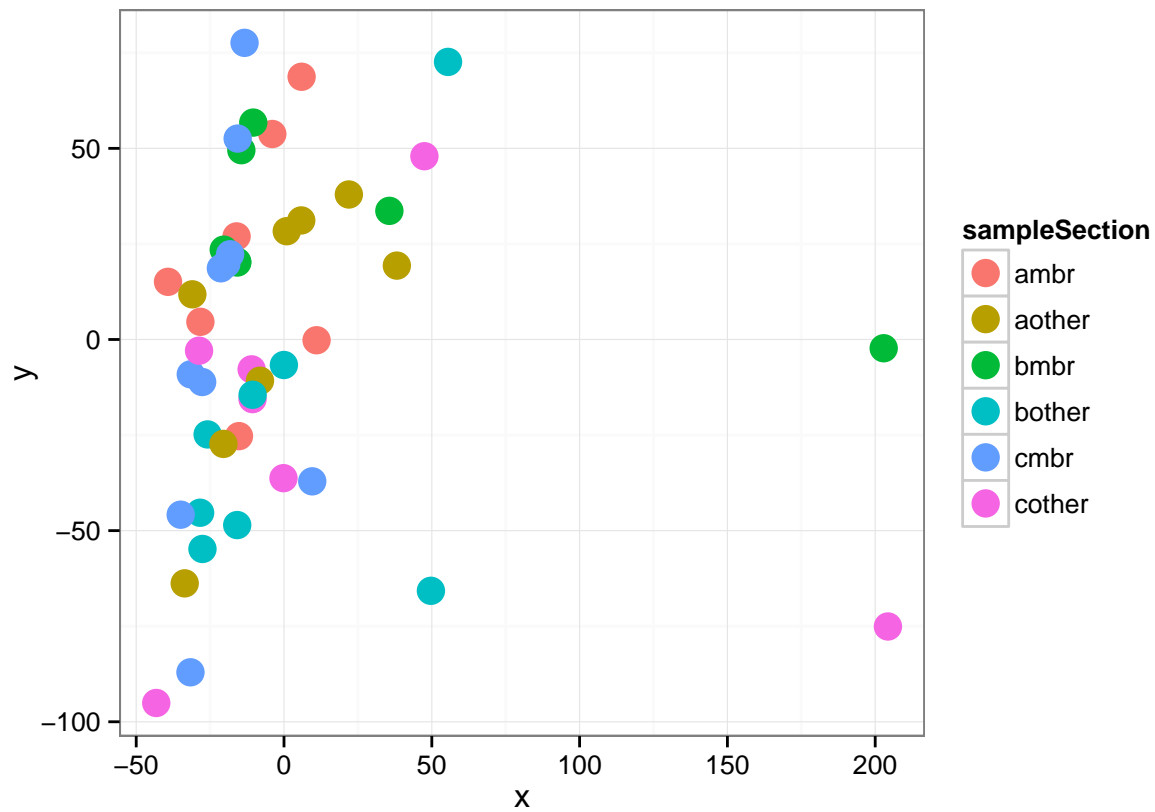
Label for visualization

```
m$sampleSection <- c(  
  rep("ambr", 4),  
  rep("aother", 4),  
  rep("bmbr", 3),  
  rep("bother", 5),  
  rep("cmbr", 4),  
  rep("cother", 4),  
  rep("ambr", 3),  
  rep("aother", 4),  
  rep("bmbr", 4),  
  rep("bother", 3),  
  rep("cmbr", 6),  
  rep("cother", 3)  
)  
  
m$genotype <- c(  
  rep("tf2", 24),  
  rep("wt", 23)  
)  
  
m$type <- c(  
  rep("mbr", 4),  
  rep("other", 4),  
  rep("mbr", 3),  
  rep("other", 5),  
  rep("mbr", 4),  
  rep("other", 4),  
  rep("mbr", 3),  
  rep("other", 4),  
  rep("mbr", 4),  
  rep("other", 3),  
  rep("mbr", 6),  
  rep("other", 3)  
)
```

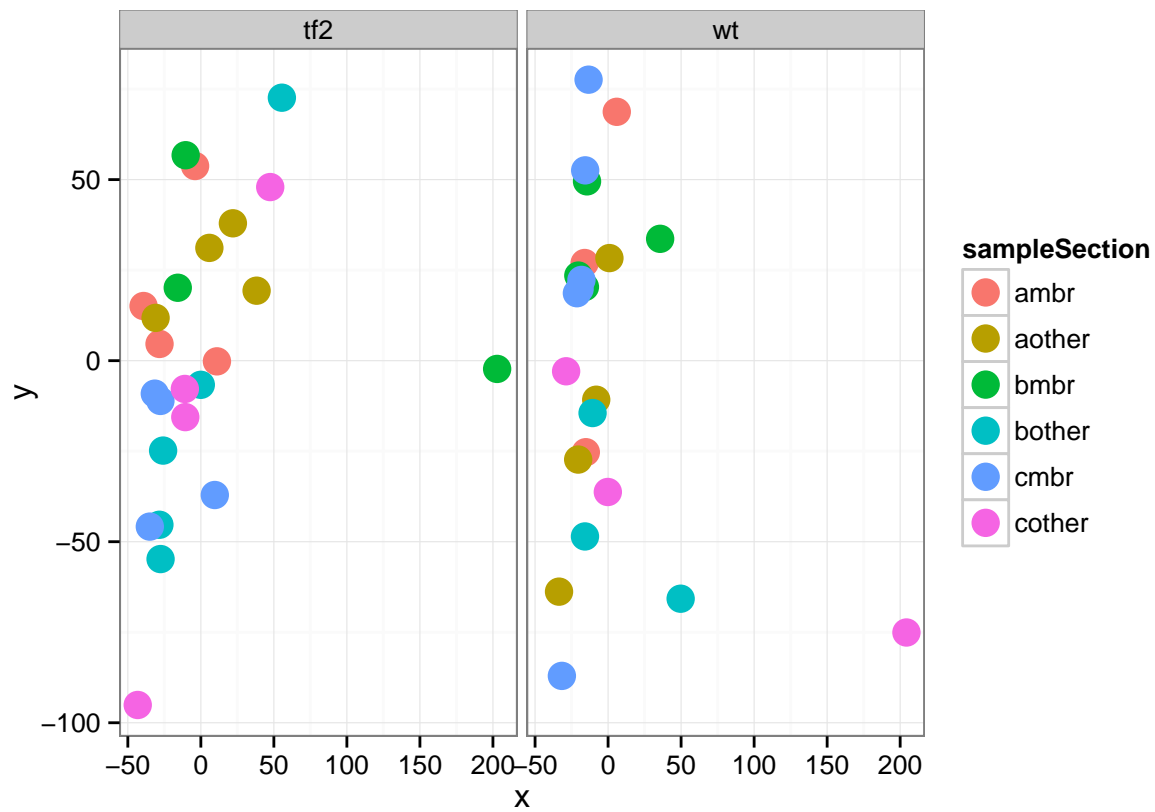
### All the samples

There is no visible trends, even when subsetting by sample.

```
ggplot(m, aes(x,y,colour=sampleSection))+geom_point(size=5)+theme_bw()
```



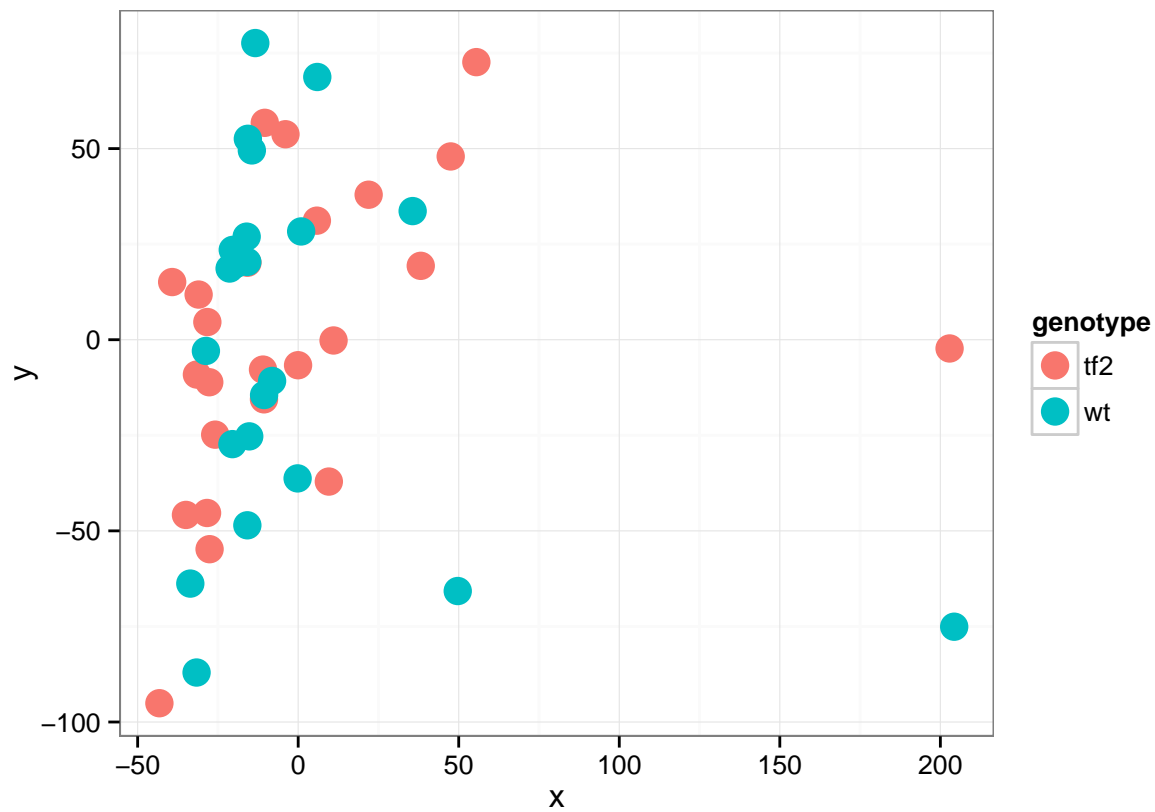




### By Genotype

There is no visible trend

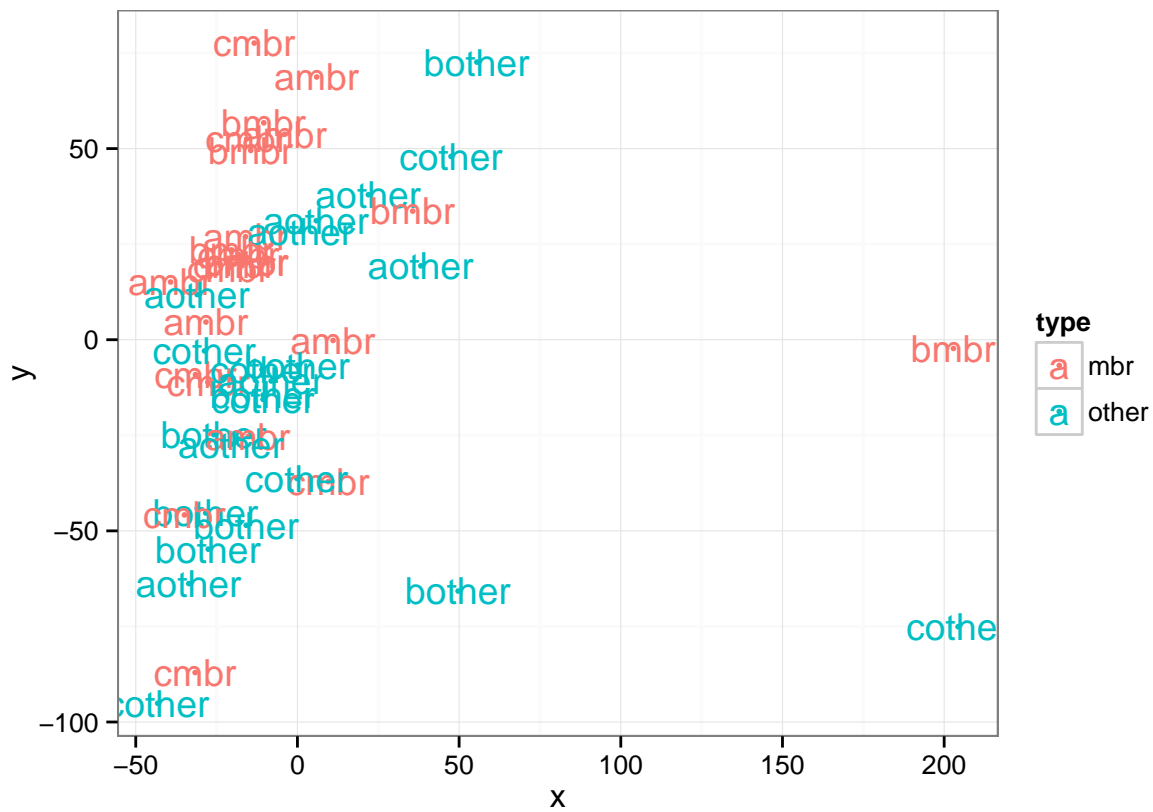
```
ggplot(m, aes(x,y,colour=genotype))+geom_point(size=5)+theme_bw()
```



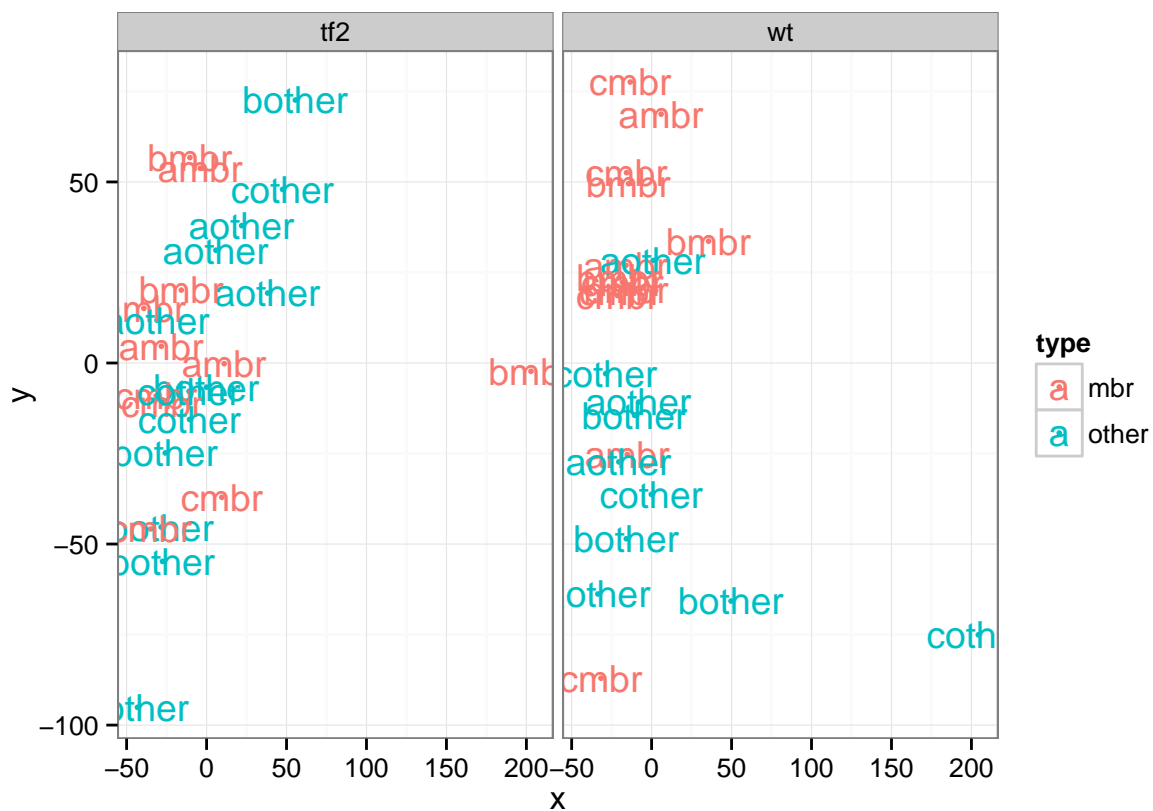
### All together

The only time a slight trend appears is between Marginal Blastozone Regions (MBR) and rachis regions (other).

```
ggplot(m, aes(x,y,colour=type, label=sampleSection))+geom_point(size=1)+theme_bw() + geom_text()
```



```
ggplot(m, aes(x,y,colour=type, label=sampleSection))+geom_point(size=1)+theme_bw() + geom_text() + facet
```



```
library(rmarkdown) render("MDS.Rmd", "pdf_document")
```