Differential Gene Expression Analysis using EdgeR

Tissue datasets are processed separately

```
library(baySeq)
library(edgeR)
seqdata = read.delim("/Users/jzs6897/Downloads/Supplementary\ File\ 29.txt")
row.names(seqdata) = seqdata[,1]
tmp = seqdata[2:4]
tmp2 = seqdata[47:49]
leaf = cbind(tmp,tmp2)

tmp = seqdata[5:7]
tmp2 = seqdata[50:52]
root = cbind(tmp,tmp2)

tmp = seqdata[8:10]
tmp2 = seqdata[53:55]
stem = cbind(tmp,tmp2)
L_groups = c("LCtrl","LCtrl","LCtrl","LDro","LDro","LDro")
R_groups = c("RCtrl","RCtrl","RCtrl","RDro","RDro","RDro")
S_groups = c("SCtrl","SCtrl","SCtrl","SDro","SDro","SDro")
dL <- DGEList(counts=leaf,group=factor(L_groups))
dS <- DGEList(counts=stem,group=factor(S_groups))
dR <- DGEList(counts=root,group=factor(R_groups))
```

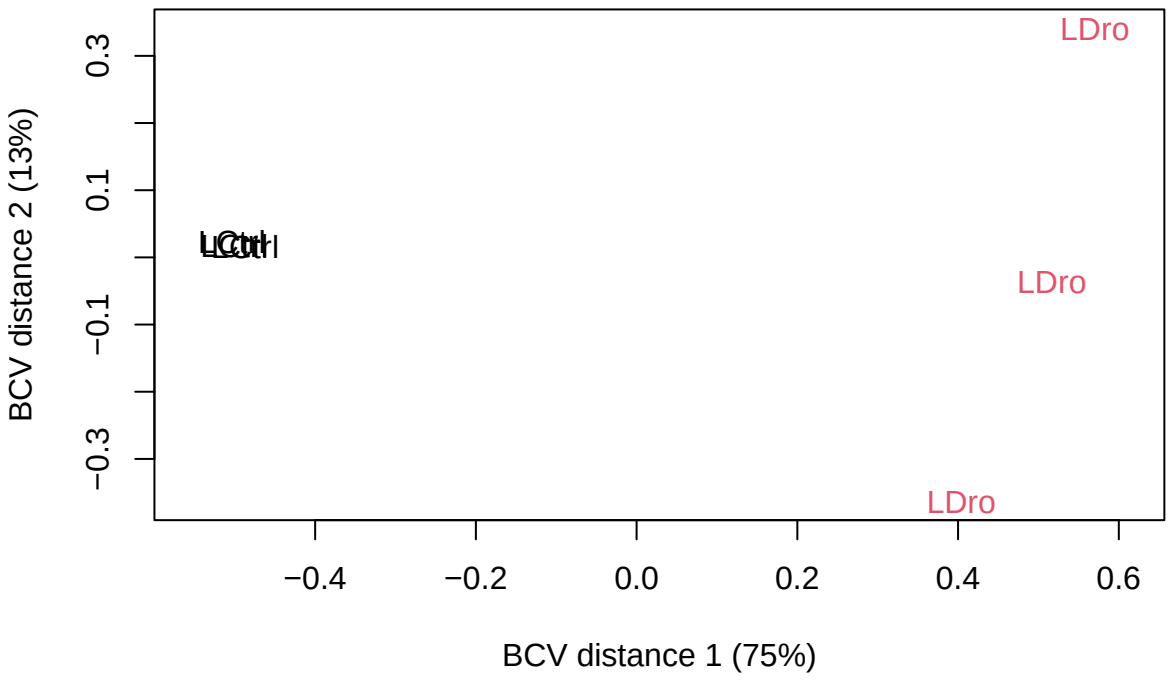
Filtering out genes with very low expression reduces the number of genes to be considered by approximately half

```
keep <- rowSums(cpm(dL)>30) >= 3
nonExp_L = dL[!keep,]
dL <- dL[keep,]
keep <- rowSums(cpm(dS)>30) >= 3
nonExp_S = dS[!keep,]
dS <- dS[keep,]
keep <- rowSums(cpm(dR)>30) >= 3
nonExp_R = dR[!keep,]
dR <- dR[keep,]
```

##Data exploration

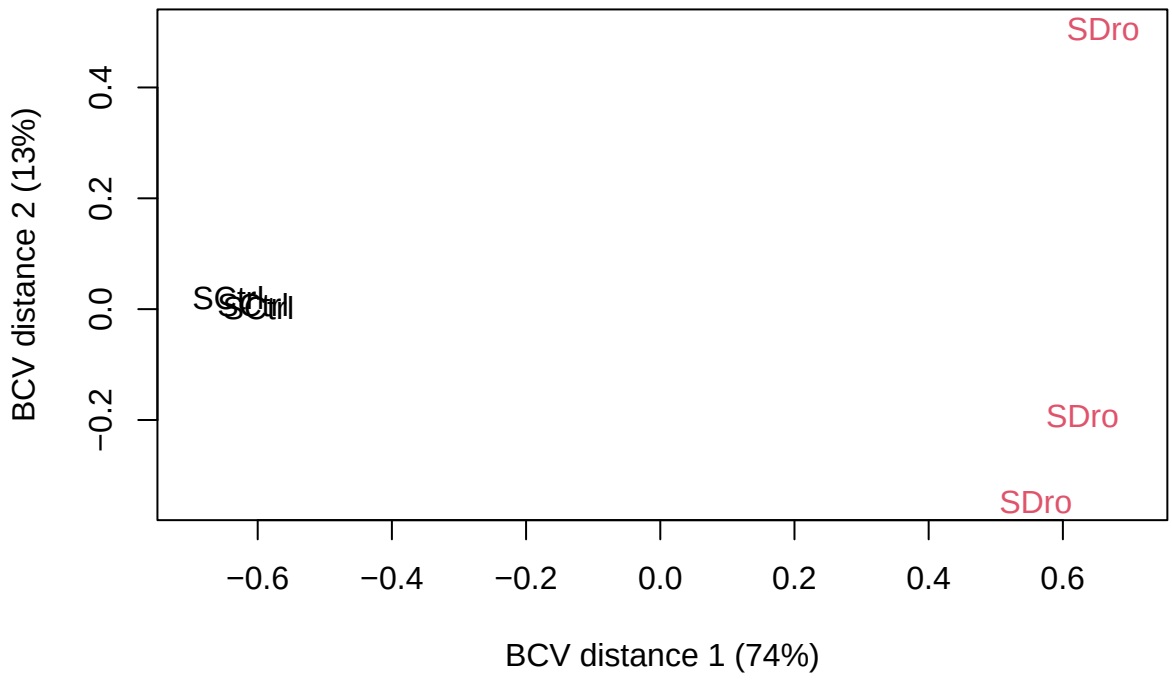
Make a plot indicating similarities between the datasets

```
dL$samples$lib.size <- colSums(dL$counts)
dL <- calcNormFactors(dL)
plotMDS(dL, method="bcv", col=as.numeric(dL$samples$group), labels = dL$samples$group)
```



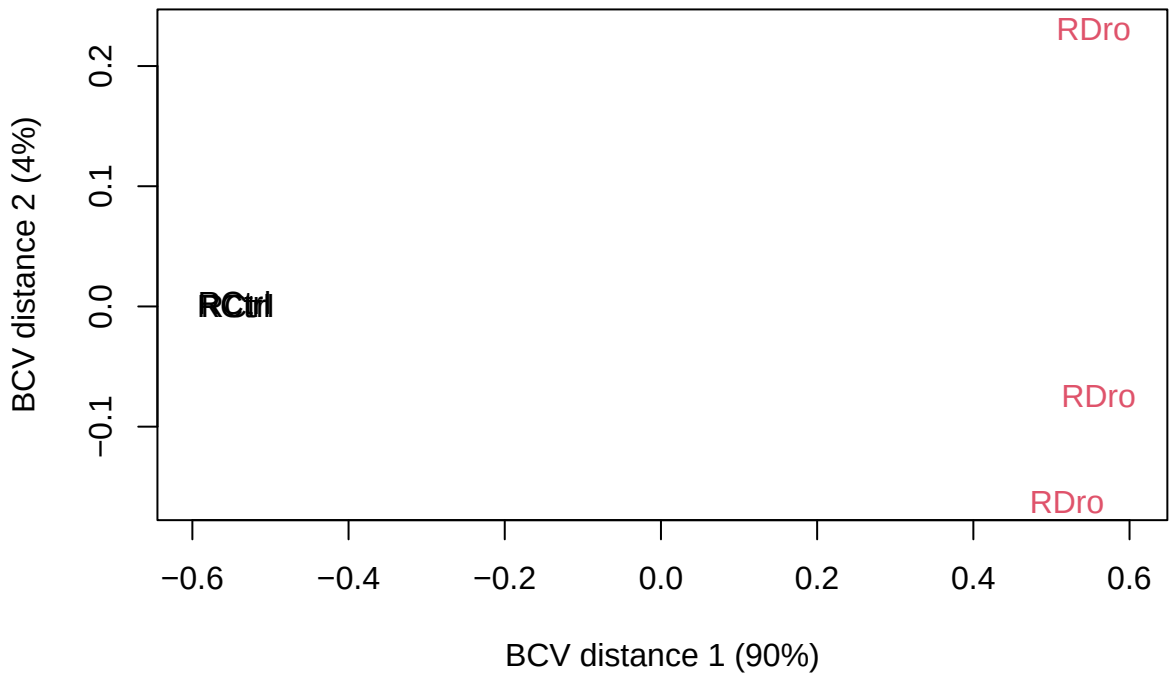

```
#legend("bottomleft", as.character(unique(d$samples$group)), col=1:3, pch=20)
```

```
dS$samples$lib.size <- colSums(dS$counts)
dS <- calcNormFactors(dS)
plotMDS(dS, method="bcv", col=as.numeric(dS$samples$group), labels = dS$samples$group)
```




```
#legend("bottomleft", as.character(unique(d$samples$group)), col=1:3, pch=20)
```

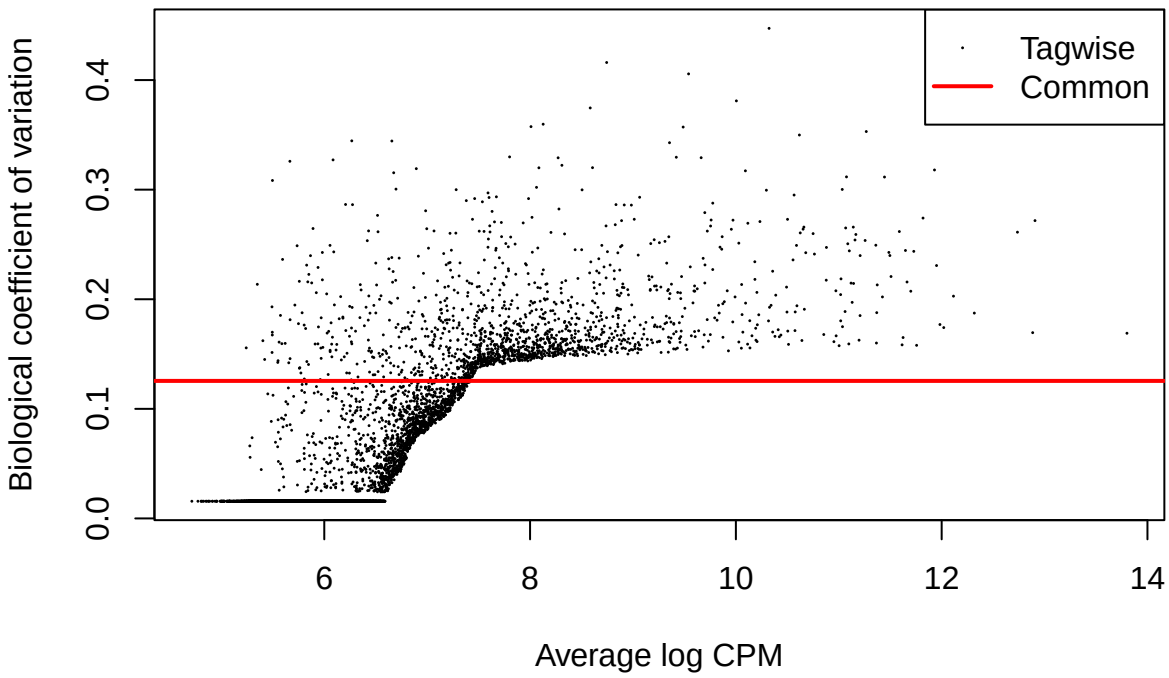
```
dR$samples$lib.size <- colSums(dR$counts)
dR <- calcNormFactors(dR)
plotMDS(dR, method="bcv", col=as.numeric(dR$samples$group), labels = dR$samples$group)
```




```
dL1 <- estimateCommonDisp(dL, verbose=T)
```

```
## Disp = 0.01575 , BCV = 0.1255
```

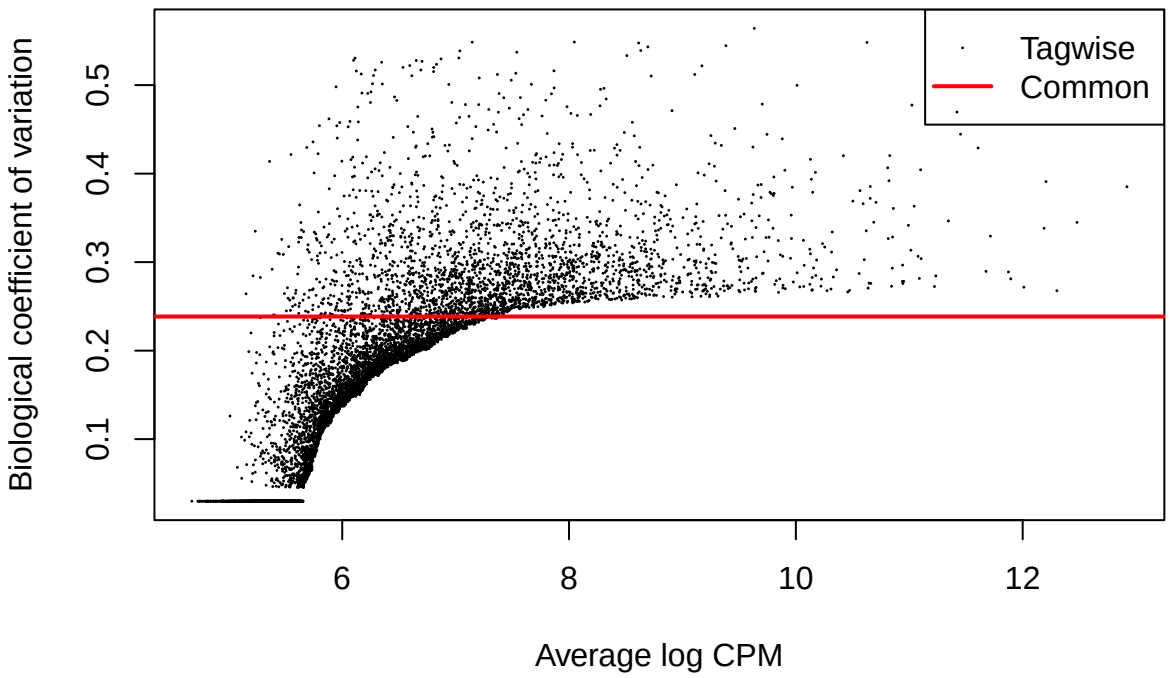
```
dL1 <- estimateTagwiseDisp(dL1)
plotBCV(dL1)
```




```
dS1 <- estimateCommonDisp(dS, verbose=T)
```

```
## Disp = 0.05689 , BCV = 0.2385
```

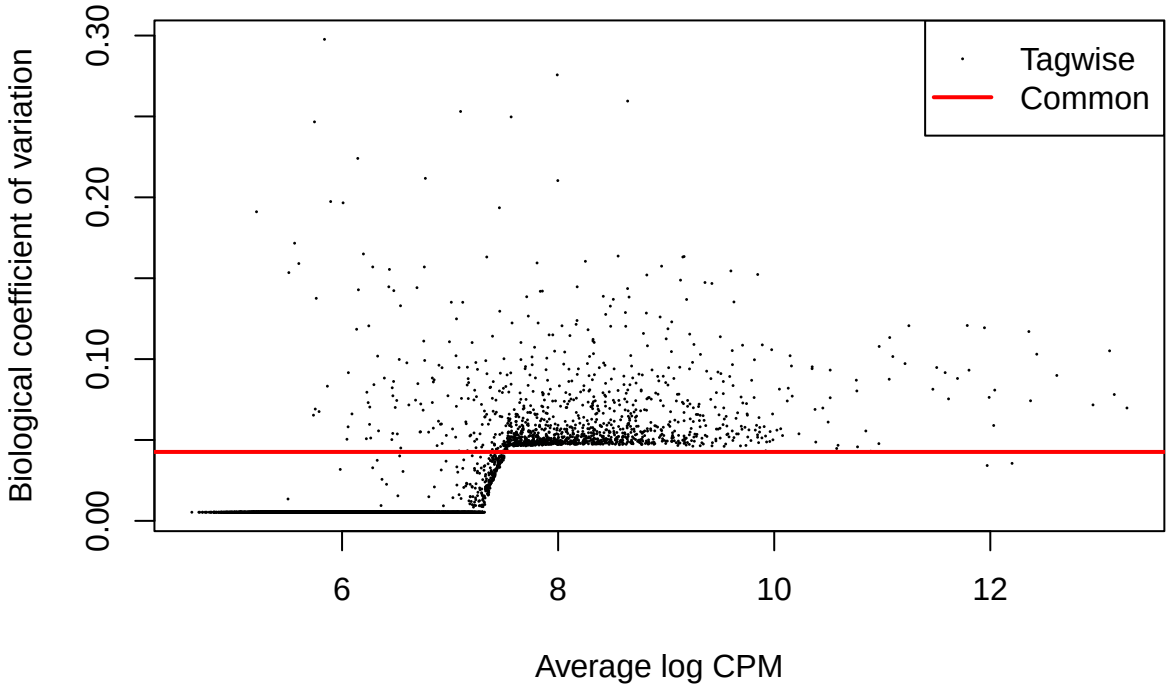
```
dS1 <- estimateTagwiseDisp(dS1)
plotBCV(dS1)
```




```
dR1 <- estimateCommonDisp(dR, verbose=T)
```

```
## Disp = 0.00181 , BCV = 0.0426
```

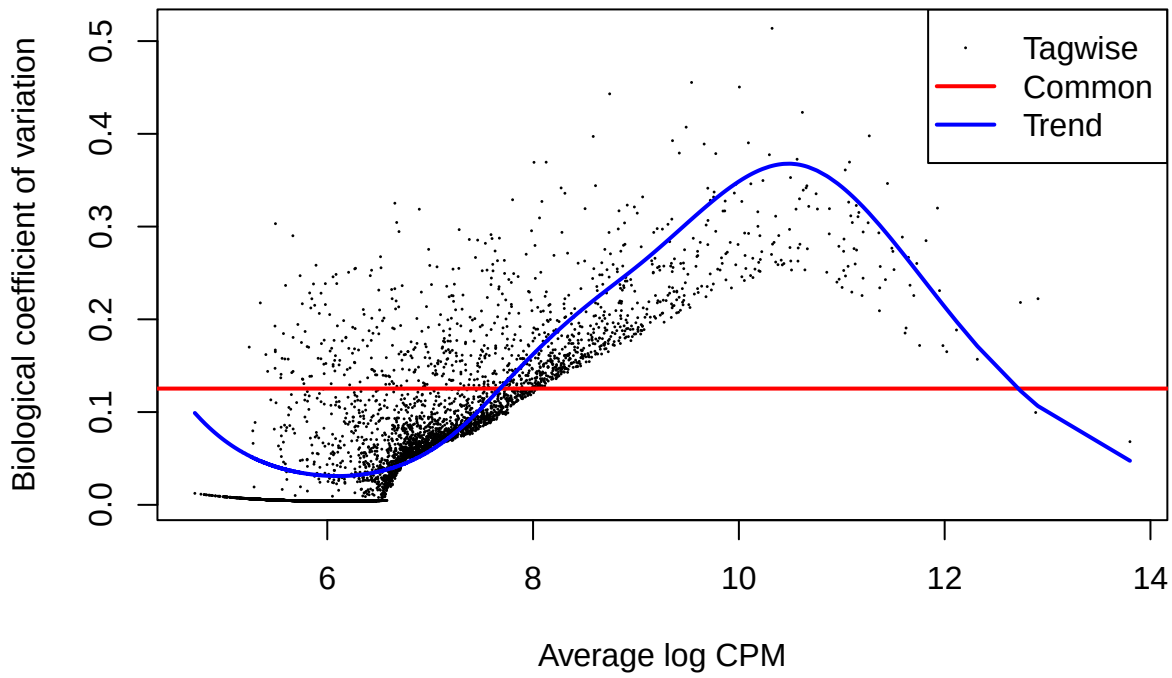
```
dR1 <- estimateTagwiseDisp(dR1)
plotBCV(dR1)
```



Generalized Linear Model of Dispersion

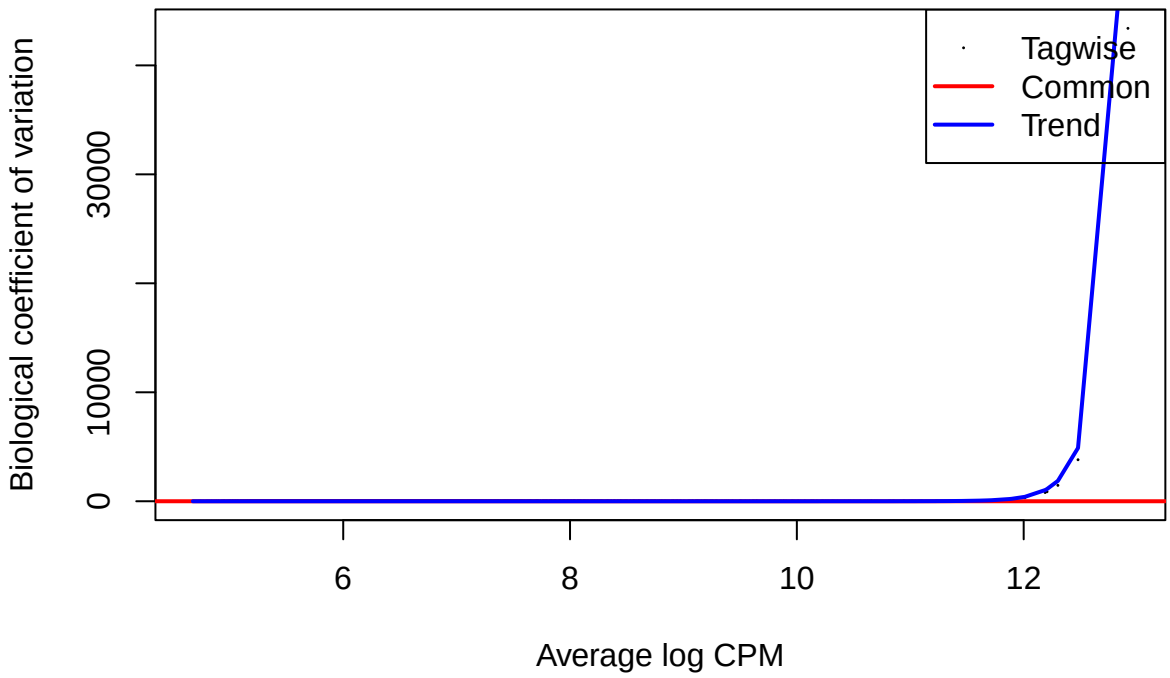
Leaf

```
design.matL <- model.matrix(~ 0 + dL$samples$group)
colnames(design.matL) <- levels(dL$samples$group)
dL2 <- estimateGLMCommonDisp(dL,design.matL)
dL2 <- estimateGLMTrendedDisp(dL2,design.matL, method="spline")
dL2 <- estimateGLMTagwiseDisp(dL2,design.matL)
plotBCV(dL2)
```



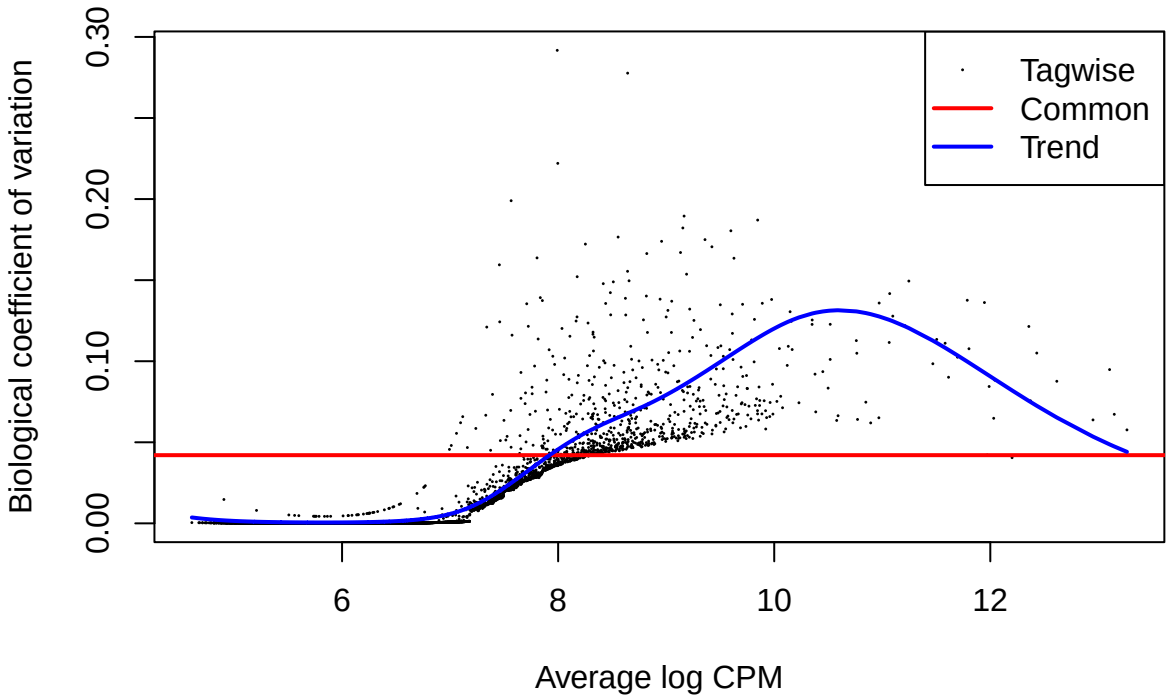
Stem

```
design.matS <- model.matrix(~ 0 + dS$samples$group)
colnames(design.matS) <- levels(dS$samples$group)
dS2 <- estimateGLMCommonDisp(dS,design.matS)
dS2 <- estimateGLMTrendedDisp(dS2,design.matS, method="spline")
dS2 <- estimateGLMTagwiseDisp(dS2,design.matS)
plotBCV(dS2)
```



Root

```
design.matR <- model.matrix(~ 0 + dR$samples$group)
colnames(design.matR) <- levels(dR$samples$group)
dR2 <- estimateGLMCommonDisp(dR,design.matR)
dR2 <- estimateGLMTrendedDisp(dR2,design.matR, method="spline")
dR2 <- estimateGLMTagwiseDisp(dR2,design.matR)
plotBCV(dR2)
```



DGE Analysis

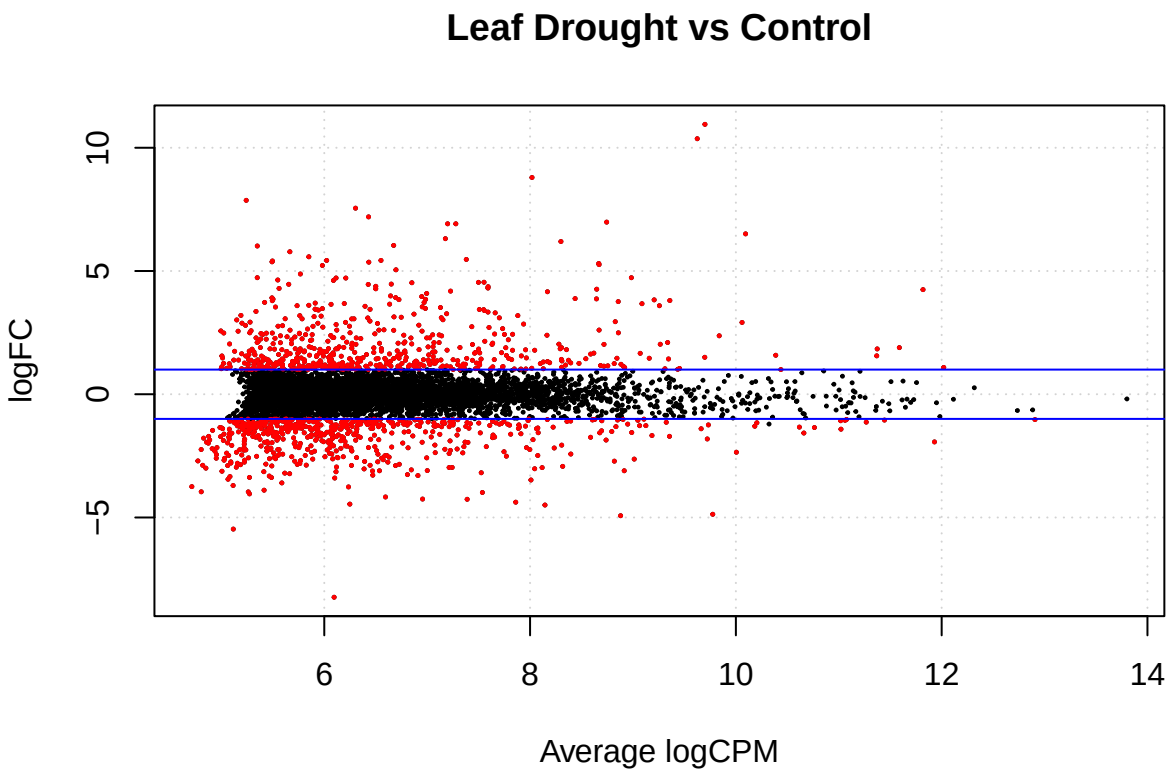
Root

```
et_L <- exactTest(dL1, pair=c(1,2)) # compare groups 1 and 2
et_R <- exactTest(dR1, pair=c(1,2)) # compare groups 1 and 3
et_S <- exactTest(dS1, pair=c(1,2)) # compare groups 2 and 3
```

```
de_L <- decideTestsDGE(et_L, adjust.method="BH", p.value=0.05, lfc=1)
summary(de_L)
```

```
##        LDro-LCtrl
## Down          567
## NotSig       6039
## Up            554
```

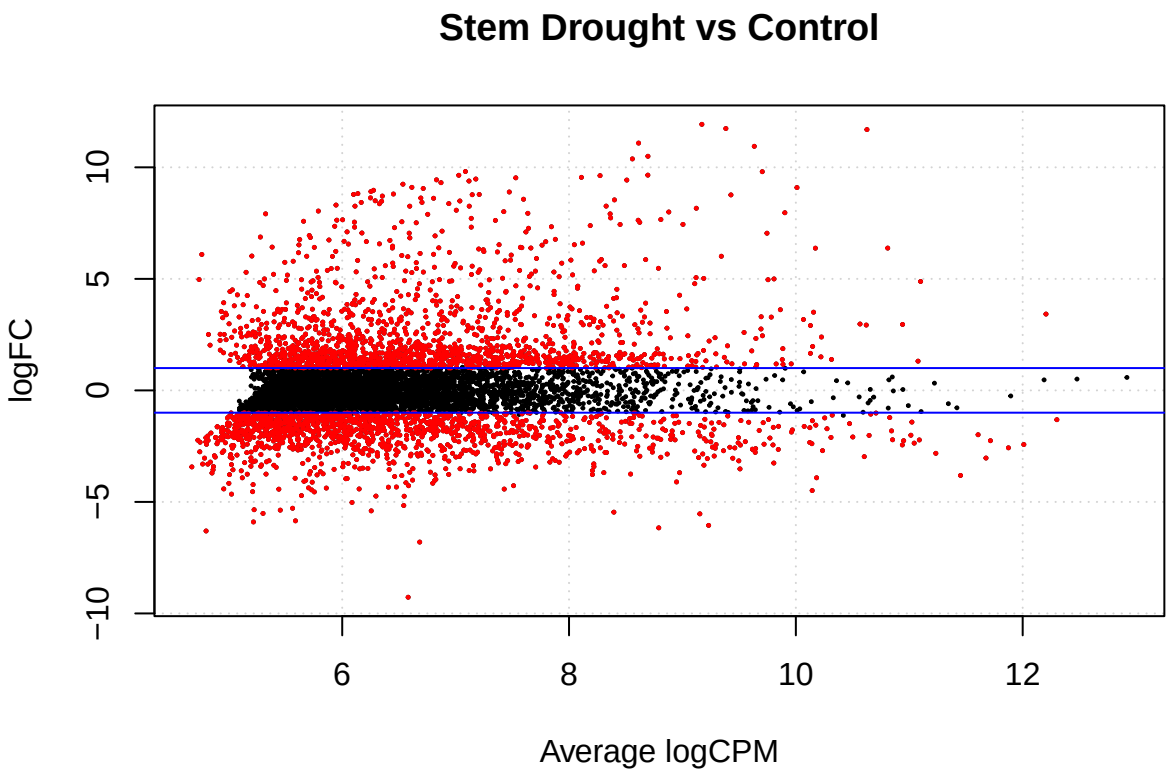
```
DEtags_L <- rownames(dL)[as.logical(de_L)]
plotSmear(et_L, de.tags=DEtags_L, main = "Leaf Drought vs Control")
abline(h = c(-1, 1), col = "blue")
```




```
de_S <- decideTestsDGE(et_S, adjust.method="BH", p.value=0.05, lfc=1)
summary(de_S)
```

```
##        SDro-SCtrl
## Down         1402
## NotSig       4744
## Up           1594
```

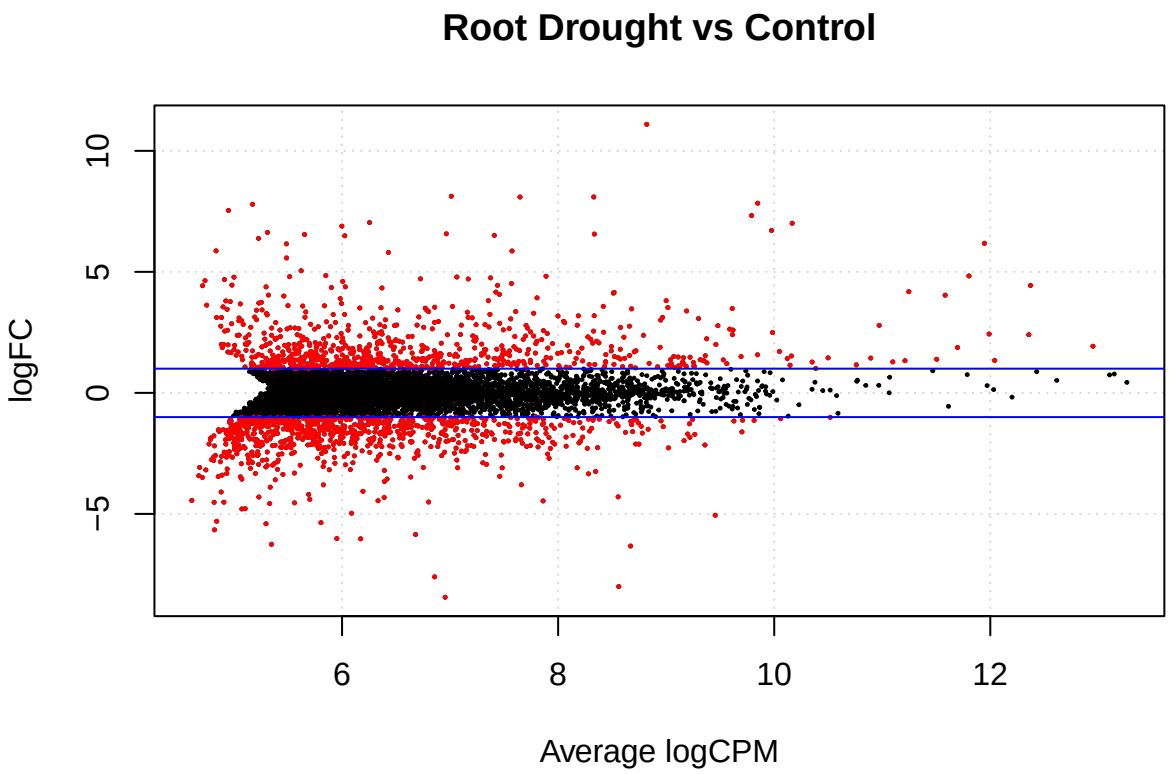
```
DEtags_S <- rownames(dS)[as.logical(de_S)]
plotSmear(et_S, de.tags=DEtags_S, main = "Stem Drought vs Control")
abline(h = c(-1, 1), col = "blue")
```




```
de_R <- decideTestsDGE(et_R, adjust.method="BH", p.value=0.05, lfc=1)
summary(de_R)
```

```
##        RDro-RCtrl
## Down          839
## NotSig       6199
## Up            762
```

```
DEtags_R <- rownames(dR)[as.logical(de_R)]
plotSmear(et_R, de.tags=DEtags_R, main = "Root Drought vs Control")
abline(h = c(-1, 1), col = "blue")
```




```
outR <- topTags(et_R, adjust.method="BH", n=Inf)
write.csv(outR, file="Root_DEGs.csv")
outS <- topTags(et_S, adjust.method="BH", n=Inf)
write.csv(outS, file="Stem_DEGs.csv")
outL <- topTags(et_L,adjust.method="BH",  n=Inf)
write.csv(outL, file="Leaf_DEGs.csv")
```