Example Analysis of a Two-Color Array Experiment Using LIMMA

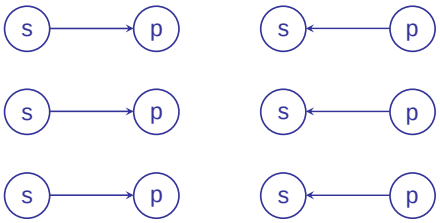
3/30/2011

Example Dataset

- Two-color microarray experiment described in

Brooks, L., Strable, J., Zhang, X., Ohtsu, K., Zhou, R., Sarkar, A., Hargreaves, S., Eudy, D., Pawlowska, T., Ware, D., Janick-Buckner, D., Buckner, B., Timmermans, M.C.P., Schnable, P.S., Nettleton, D., Scanlon, M.J. (2009). Microdissection of shoot meristem functional domains. *PLoS Genetics*. **5**(5): e1000476.

Comparison of Two Maize Cell Types (p vs. s)




```
######
#
#Load Limma package.  (Must first be installed.)
#
######

library(limma)

######
#
#Examine the limma users guide.
#
######

limmaUsersGuide()
```



```
######
#
#Set the working directory to the directory containing
#the raw data files.  The files are available at
#http://www.public.iastate.edu/~dnett/microarray/sam3.0/
#You will need to download them to your hard drive
#and change the path accordingly.
#
######

setwd("C:\\z\\sam3.0")
```



```
######
#
#Read the Targets.txt file that contains information
#about the slides and what was hybridized to each
#channel.
#
######

targets=readTargets("Targets.txt")
targets

  SlideNumber  FileName Cy3 Cy5
1           1 sam31.gpr   s   p
2           2 sam32.gpr   s   p
3           3 sam33.gpr   s   p
4           4 sam34.gpr   p   s
5           5 sam35.gpr   p   s
6           6 sam36.gpr   p   s
```

```
######
#
#Read the data files into R and examine content.
#
######

RG=read.maimages(targets,source="genepix")

attributes(RG)

$class
[1] "RGList"
attr(,"package")
[1] "limma"

$names
[1] "R"  "G"  "Rb"  "Gb"  "targets"  "genes"  "source"
[8] "printer"
```

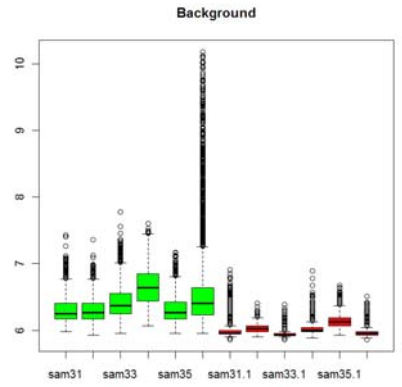
7

```
head(RG$R)
     sam31 sam32 sam33 sam34 sam35 sam36
[1,]   292   231   113   141   134    78
[2,]   406   231   143   148   179   116
[3,]   128   153   104    92   114    81
[4,]   168   134   120    86   260    80
[5,]   200   175   129   157   220    79
[6,]   112   186   107   107   197   364

head(RG$Rb)
     sam31 sam32 sam33 sam34 sam35 sam36
[1,]    61    65    61    64    71    65
[2,]    63    65    61    62    70    63
[3,]    62    63    60    65    82    62
[4,]    63    64    61    62    85    63
[5,]    63    65    62    65    74    62
[6,]    62    65    60    62    74    63
```

8

```
######
#
#Examine boxplots of red and green backgrounds.
#
######

boxplot(data.frame(cbind(log2(RG$Gb),log2(RG$Rb))),
        main="Background",
        col=rep(c("green","red"),each=6))
```

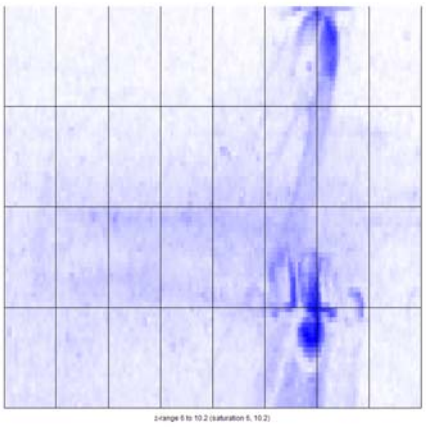
9



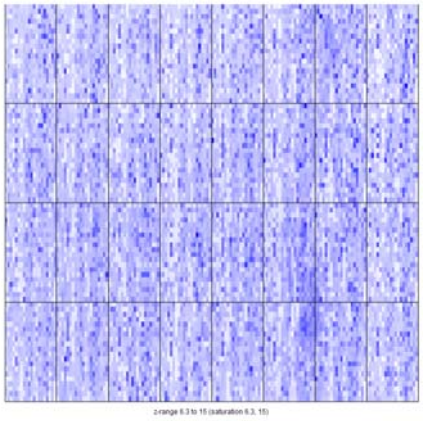
10

```
######
#
#For slide 6, examine an image of the green background
#and green signal corresponding to the slide layout.
#
######

imageplot(log2(RG$Gb[,6]),RG$printer)

x11()

imageplot(log2(RG$G[,6]),RG$printer)
```

11



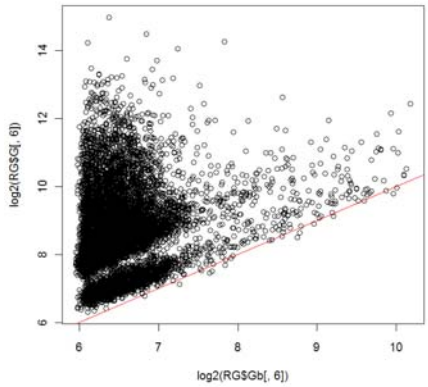
12



13

```
######
#
#Plot signal vs. background for the green channel of
#slide 6.
#
######

plot(log2(RG$Gb[,6]),log2(RG$G[,6]))
lines(c(-9,99),c(-9,99),col=2)
```

14



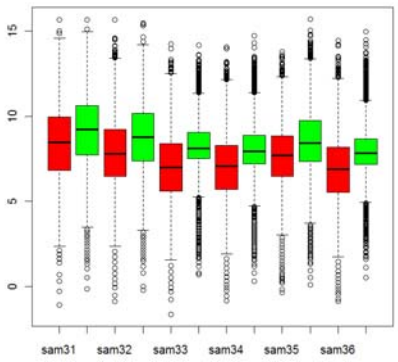
15

```
######
#
#Perform background correction for all slides.
#
######

RG=backgroundCorrect(RG,method="normexp")

attributes(RG)

$names
[1] "R"  "G"  "targets" "genes"   "source"  "printer"

$class
[1] "RGList"
attr(,"package")
[1] "limma"
```

16

```
head(RG$R)

        sam31     sam32    sam33    sam34     sam35      sam36
[1,] 228.83502 158.03662 46.60193 68.38565  53.91669   6.720112
[2,] 340.83502 158.03662 76.60193 77.38565  99.91669  46.692513
[3,]  63.83502  82.03662 38.60193 18.38565  22.91669  12.692516
[4,] 102.83502  62.03662 53.60193 15.38565 165.91669  10.692618
[5,] 134.83502 102.03662 61.60193 83.38565 136.91669  10.692618
[6,]  47.83502 113.03662 41.60193 36.38565 113.91669 294.692513

######
#
#Examine side-by-side boxplots of
#background-corrected data.
#
######

boxplot(log2(data.frame(
             RG$R,RG$G))[,as.vector(rbind(1:6,7:12))],
             col=rep(c("red","green"),6))
```

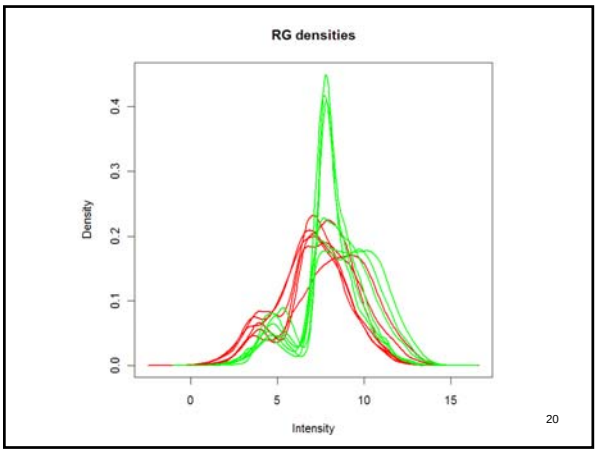
17



18

```
######
#
#Examine background corrected signal densities
#for each channel.
#
######

plotDensities(RG)
```

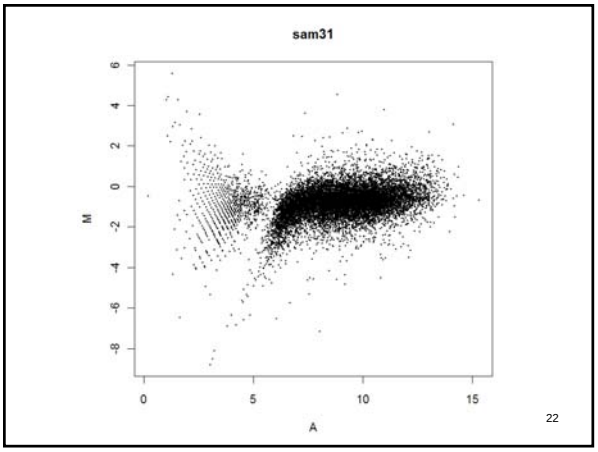
19



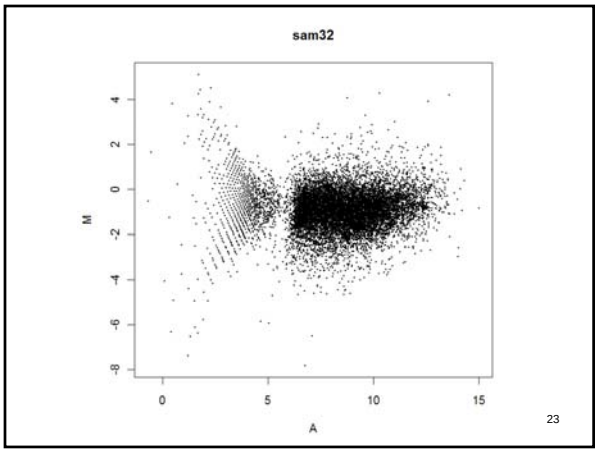
20

```
######
#
#Examine plots of the difference vs. the average
#log background corrected signal.
#
######

plotMA(RG)
plotMA(RG[,2])
plotMA(RG[,3])
plotMA(RG[,4])
plotMA(RG[,5])
plotMA(RG[,6])
```

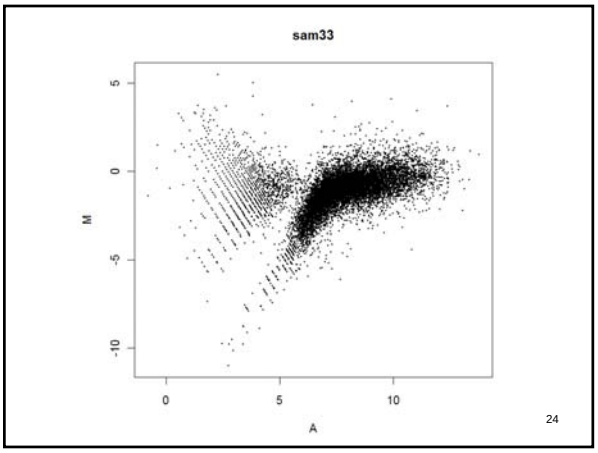
21



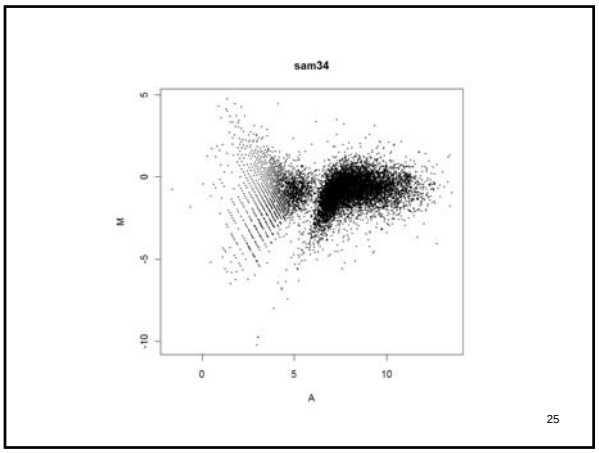
22

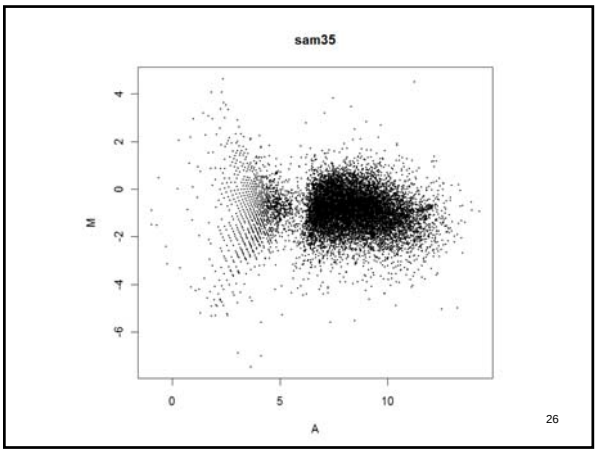


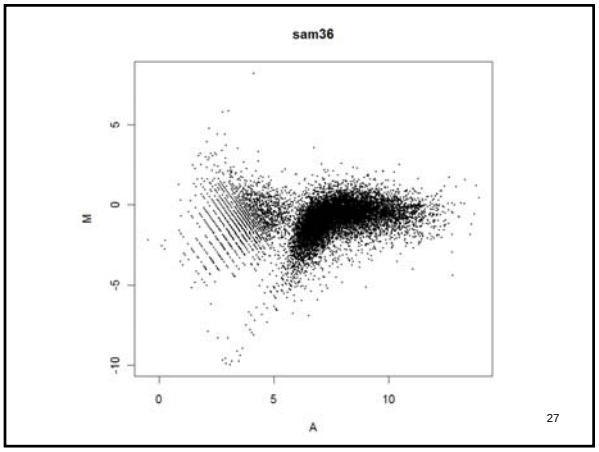
23

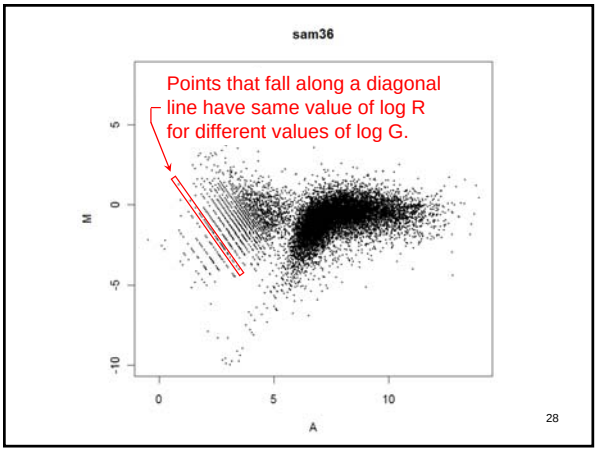


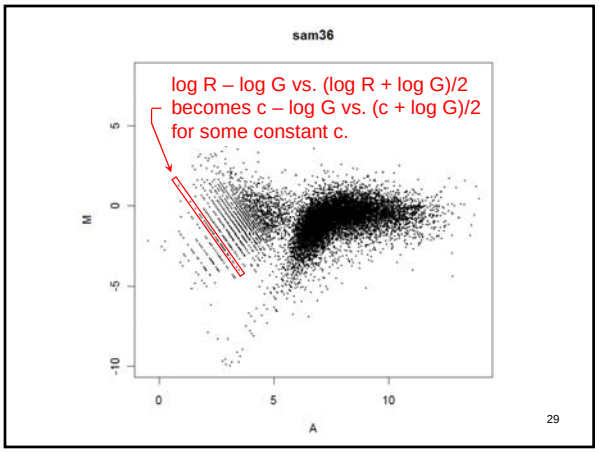
24

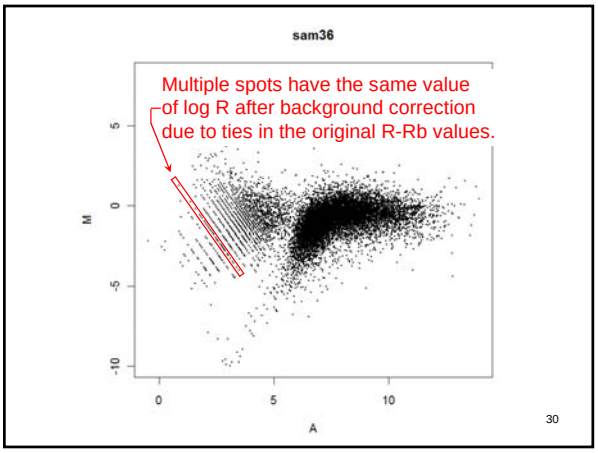

sam34
M
A
25


sam35
M
A
26


sam36
M
A
27


sam36
Points that fall along a diagonal line have same value of log R for different values of log G.
M
A
28


sam36
log R – log G vs. (log R + log G)/2 becomes c – log G vs. (c + log G)/2 for some constant c.
M
A
29


sam36
Multiple spots have the same value of log R after background correction due to ties in the original R-Rb values.
M
A
30

```
######
#
#Loess normalize and median center the data.
#Examine the resulting Red-Green differences.
#
######

MA=normalizeWithinArrays(RG,method="loess")
MA=normalizeWithinArrays(MA,method="median")

attributes(MA)

$class
[1] "MAList"
attr(,"package")
[1] "limma"

$names
[1] "targets" "genes" "source" "printer" "M" "A"
```

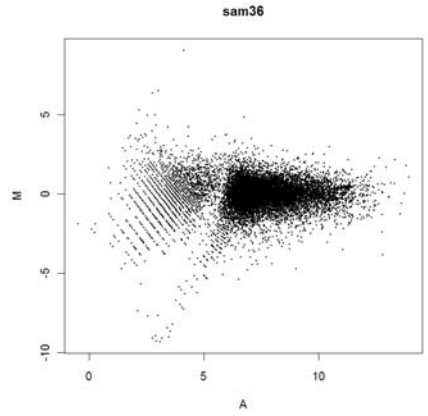
31

```
head(MA$M)
          sam31      sam32      sam33      sam34      sam35      sam36
[1,] -0.1880557  0.28739007 -0.4896094  0.45571496 -1.1221094 -0.105308479
[2,]  0.3088745  0.20587708  1.0333266 -0.09791323 -1.1995438 -0.126305139
[3,]  1.1172339  0.35352269  1.0495153  0.67218498 -0.3438164 -2.700574497
[4,]  0.8310556 -0.42577257  0.7128648 -0.80643398 -2.7923563  0.002659864
[5,]  0.6358650  0.06359337  0.6134892  0.81796018  0.2055294 -0.296545845
[6,] -0.1323416  1.67347491  0.6485572 -0.79897047  0.6615731 -0.691363072

 head(MA$A)
         sam31    sam32    sam33    sam34    sam35    sam36
[1,] 8.290617 7.591408 6.663639 6.606916 6.730848 2.801016
[2,] 8.619009 7.632863 6.644862 6.923444 7.586340 6.404087
[3,] 6.072136 6.652326 5.969590 4.228227 5.011339 5.603872
[4,] 6.812037 6.635686 6.435603 4.726025 9.200278 3.606411
[5,] 7.202041 7.074841 6.581947 6.688394 7.357936 3.774414
[6,] 6.272107 6.476244 6.228192 6.278806 6.910750 8.749031
```

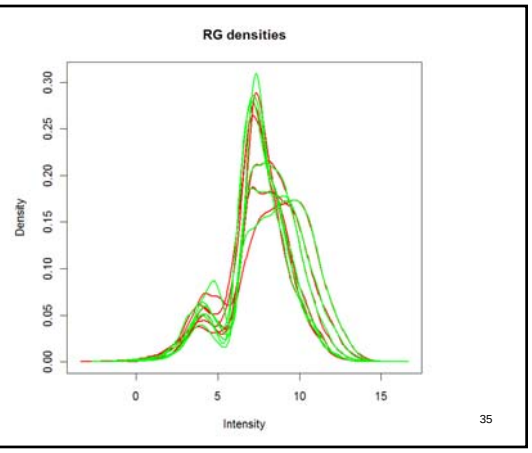
32

```
plotMA(MA[,6])

plotDensities(MA)

boxplot(data.frame(MA$M))
```

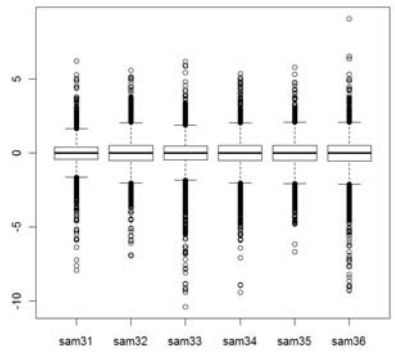
33



34



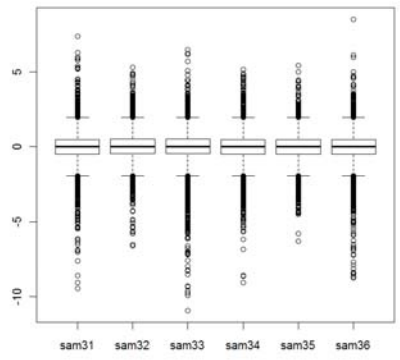
35



36

```
######
#
#Scale normalize the differences.
#
######

MA=normalizeBetweenArrays(MA,method="scale")
boxplot(data.frame(MA$M))
```

37



38

```
#Create a design matrix for a gene-specific vector of
#differences as the response.

targets
 SlideNumber  FileName Cy3 Cy5
1           1 sam31.gpr   s   p
2           2 sam32.gpr   s   p
3           3 sam33.gpr   s   p
4           4 sam34.gpr   p   s
5           5 sam35.gpr   p   s
6           6 sam36.gpr   p   s

design=cbind(rep(1,6),rep(c(1,-1),each=3))
colnames(design)=c("Cy5minusCy3","PminusS")
design
     Cy5minusCy3 PminusS
[1,]           1       1
[2,]           1       1
[3,]           1       1
[4,]           1      -1
[5,]           1      -1
[6,]           1      -1
```

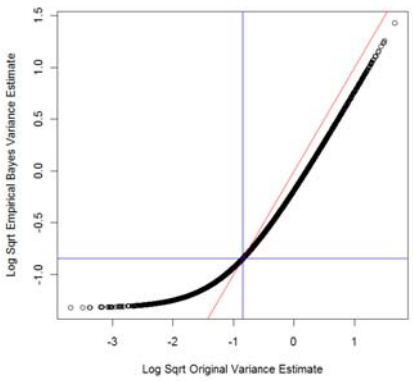
39

```
#Use the limma function lmFit to fit the model.

fit=lmFit(MA, design)

#Set up contrasts and compute p-values
#using the limma approach.

contrast.matrix=makeContrasts(PminusS, levels=design)
fit2=contrasts.fit(fit, contrast.matrix)
efit=eBayes(fit2)
attributes(efit)
$names
 [1] "coefficients"  "rank"      "assign"           "qr"
 [5] "df.residual"   "sigma"     "cov.coefficients" "stdev.unscaled"
 [9] "genes"         "Amean"     "method"           "design"
[13] "contrasts"     "df.prior"  "s2.prior"         "var.prior"
[17] "proportion"    "s2.post"   "t"                "p.value"
[21] "lods"          "F"         "F.p.value"

$class
[1] "MArrayLM"
attr(,"package")
[1] "limma"
```

40

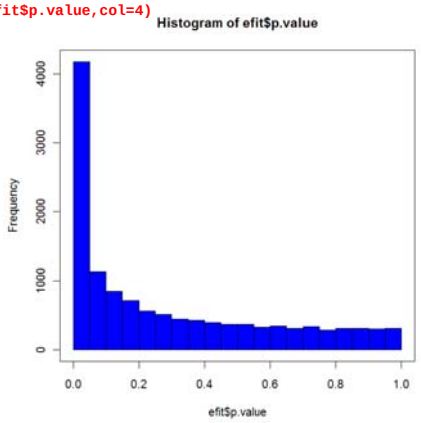
```
#Examine estimates of the prior df and variance.
efit$df.prior
[1] 2.508525

efit$s2.prior
[1] 0.1831196

#Compare variance estimates before and after shrinkage.

plot(log(efit$sigma),log(sqrt(efit$s2.post)),
     xlab="Log Sqrt Original Variance Estimate",
     ylab="Log Sqrt Empirical Bayes Variance Estimate")
lines(c(-99,99),c(-99,99),col=2)
lsrs20=log(sqrt(efit$s2.prior))
abline(h=lsrs20,col=4)
abline(v=lsrs20,col=4)
```

41



42

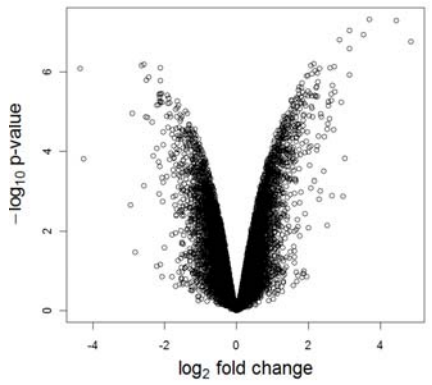
```
hist(efit$p.value,col=4)
box()
```



43

```
#Get log base 2 fold change estimates

log2fc=efit$coefficients

#Create a volcano plot.
p=efit$p.value[,1]
plot(log2fc,-log(p,base=10),
     xlab=expression(paste(log[2]," fold change")),
     ylab=expression(paste(-log[10]," p-value")),
     cex.lab=1.7)
```

44



45

```
#Load some functions for multiple testing.

source("http://www.public.iastate.edu/
        ~dnett/microarray/multtest.txt")

#Estimate the number of EE (m0) and DE (m1) genes.

m=length(p)
m0=estimate.m0(p)
m1=m-m0

m
[1] 12800

m0
[1] 6084

m1
[1] 6716
```

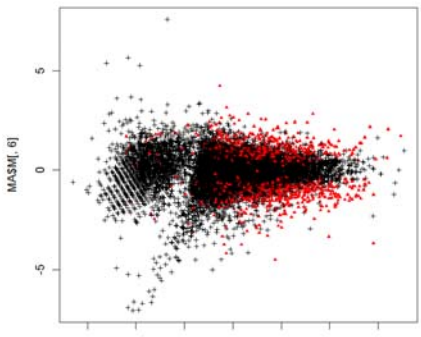
46

```
#Convert p-values to q-values.

qvals=jabes.q(p)

#Find number of genes declared to be DE
#for various FDR thresholds.

fdrcuts=c(0.001,0.01,0.05)
cbind(fdrcuts,NumberOfGenes=
              apply(outer(qvals,fdrcuts,"<="),2,sum))

     fdrcuts NumberOfGenes
[1,]   0.001           145
[2,]   0.010          1210
[3,]   0.050          3352
```

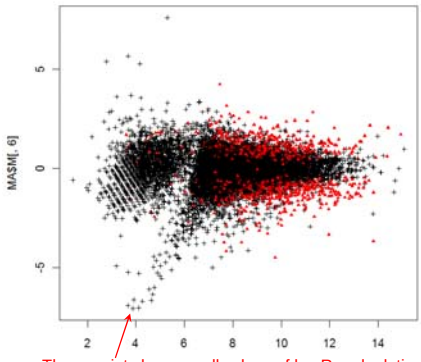
47

```
#Examine position of genes declared to be DE at FDR=0.01
#in MA plot for slide 6.

plot(MA$A[,6],MA$M[,6],
     col=1+(qvals<=0.01),
     pch=3+14*(qvals<=0.01),cex=.6)
```

48



49

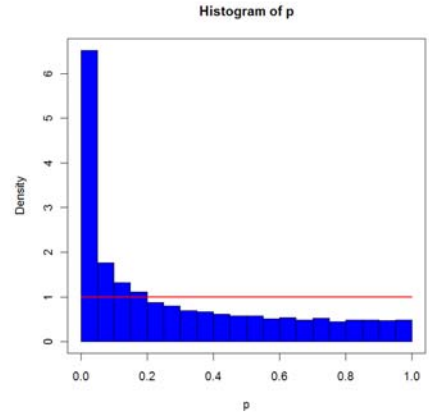


These points have small values of log R and relatively big values of log G, but these differences are not replicated on other slides.

50

```
#Fit uniform-beta mixture model.

out=ub.mix(p)
out

[1]  1.000000  0.414665 35.335280


plot.ub.mix(p,out)
```

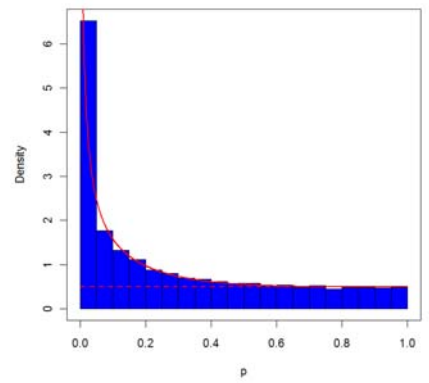
51



52

```
#Change starting values and refit
#uniform-beta mixture model.

out=ub.mix(p,c(.5,1,10))
out

[1] 0.4934128 0.3863018 4.4009335


plot.ub.mix(p,out)
```

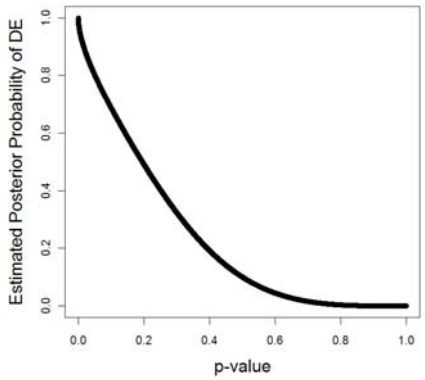
53



54

```
#Compute estimated ppde for each gene.

eppde=ppde(p,out)

plot(p,eppde,xlab="p-value",
             ylab="Estimated Posterior Probability of DE",
             cex.lab=1.5)
```

55



56

```
ppdecuts=seq(.6,.95,by=0.05)
cbind(ppdecuts,NumberOfGenes=
               apply(outer(eppde,ppdecuts,">="),2,sum))

     ppdecuts NumberOfGenes
[1,]     0.60          6048
[2,]     0.65          5650
[3,]     0.70          5208
[4,]     0.75          4737
[5,]     0.80          4200
[6,]     0.85          3546
[7,]     0.90          2781
[8,]     0.95          1830
```

57