ex5: Differential expression analysis with edgeR

arab2データの遺伝子発現の2群間比較を、edgeRで行う。

edgeRは複雑なパッケージである。開発者が詳細のユーザーガイドやマニュアルを提供しているので、これらを活用して欲しい（リンクは下記参照）。

Import library

```
> library(edgeR)
```

Import data

```
> dat <- read.delim("arab2.txt", row.names=1)

# ... dat中身の確認作業 ...
```

2グループ、各3繰り返し実験、という実験デザインを定義する。

```
> grp <- c("M", "M", "M", "H", "H", "H")
> grp
[1] "M" "M" "M" "H" "H" "H"
```

edgeRのDGEList関数でカウントデータを読み込む。

```
> D <- DGEList(dat, group=grp)
> head(D)
...
```

Normalization

TMM法で、ノーマライズする。calcNormFactorsを使う。

```
> D <-calcNormFactors(D, method="TMM")
```

計算結果の確認

```
> D$samples
   group lib.size norm.factors
m1     M  1902032    1.0399197
m2     M  1934029    1.0611305
m3     M  3259705    0.8841923
h1     H  2129854    1.0266944
h2     H  1295304    1.1412144
h3     H  3526579    0.8747345
```

DE testing

estimate dispersion

```
> D <- estimateCommonDisp(D)
> D$common.dispersion
```

```
[1] 0.342609

> D <- estimateTagwiseDisp(D)
> summary(D$tagwise.dispersion)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.1173  0.1834  0.4728  1.0540  1.7400  3.7390
```

DE test

```
> de.tagwise <- exactTest(D, pair=c("M", "H"))
```

Multiple comparison correction and View results

```
> topTags(de.tagwise)
Comparison of groups:  H-M
             logFC    logCPM       PValue          FDR
AT5G48430 6.233066  6.706315 3.281461e-21 8.604319e-17
AT3G46280 5.078716  8.120404 1.110955e-19 1.456517e-15
AT2G19190 4.620707  7.381817 1.710816e-19 1.495310e-15
AT4G12500 4.334870 10.435847 4.689616e-19 3.074161e-15
AT2G44370 5.514376  5.178263 9.902189e-18 5.192906e-14
AT2G39380 5.012163  5.765848 2.010501e-17 8.786223e-14
AT3G55150 5.809677  4.871425 3.065826e-17 1.148414e-13
AT4G12490 3.901996 10.198755 8.068822e-17 2.455369e-13
AT1G51820 4.476647  6.369685 8.490613e-17 2.455369e-13
AT2G39530 4.366709  6.710299 9.364131e-17 2.455369e-13
```

Dump the table into a text file

```
> write.table(de.tagwise$table, "de.tagwise.txt", sep="\t", quote=F)
```

もしくは、

```
> tmp <- topTags(de.tagwise, n=nrow(de.tagwise$table))
> write.table(tmp$table, "de.tagwise2.txt", sep="\t", quote=F)
```

後者はFDRの値も出力される。

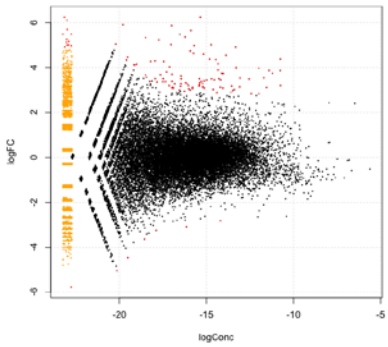
MA plot

edgeR提供のplotSmear関数を使うと便利。

```
plotSmear(D)
```

有意に発現差のある遺伝子を赤色でハイライトすることもできる。

```
> de.names <- row.names(dat[decideTestsDGE(de.tagwise, p.value=0.05) !=0, ])
> plotSmear(D, de.tags=de.names)
```



Inspect DE result

example: fold-change > 10を抽出・カウント

```
> detab <- tmp$table
# get fold-change > 10
> detab[detab$logFC > log2(10),]
> nrow(detab[detab$logFC > log2(10),])
```

example: FC > 5 AND FDR < 0.01

```
> detab[(detab$logFC > log2(2)  & detab$FDR < 0.05), ]
```

edgeRに組み込まれている''decideTestDGE''関数も便利。

```
> summary(decideTestsDGE(de.tagwise, p.value=0.05))
   [,1]
-1    49
0  25903
1    269
```

dumpしたタブ区切りテキストをMS Excelで読み込んで、フィルタ機能やソート機能を駆使してデータを探索するのも良いだろう。

Links

- http://www.bioconductor.org/packages/release/bioc/html/edgeR.html | edgeR
- http://www.bioconductor.org/packages/release/bioc/vignettes/edgeR/inst/doc/edgeRUsersGuide.pdf | edgeR User's Guide