Analysis of Smchd1 data set

Matt Ritchie

17 December 2014

Introduction

This data set was generated by Natasha Jansz from the Blewitt lab at the WEHI. FASTQ files from this experiment are available from GEO under accession number GSE64099. The matrix of counts analysed below was obtained by running Subread and featureCounts from the Rsubread package. The R code below generated the results presented in Liu et al. (2014).

Importing the data

```
library(edgeR)
load("x.rda")
fullanno = x$genes
sel = rowSums(cpm(x$counts)>0.5)>=3
x = x[sel,]
selanno = x$genes
x$genes = x$genes[,c(1,3)]
des = model.matrix(~x$samples$group)
colnames(des)[2] = "Smchd1nullvsWt"
x = calcNormFactors(x, method="TMM")

genotype = x$samples$group

# Make an MDS plot: Figure 1B
plotMDS(x, labels=1:7, col=as.numeric(genotype), main="MDS plot for Smchd1 experiment")
legend("topright", legend=c("WT", "Smchd1 null"), pch=15, col=1:2)
```

MDS plot for Smchd1 experiment
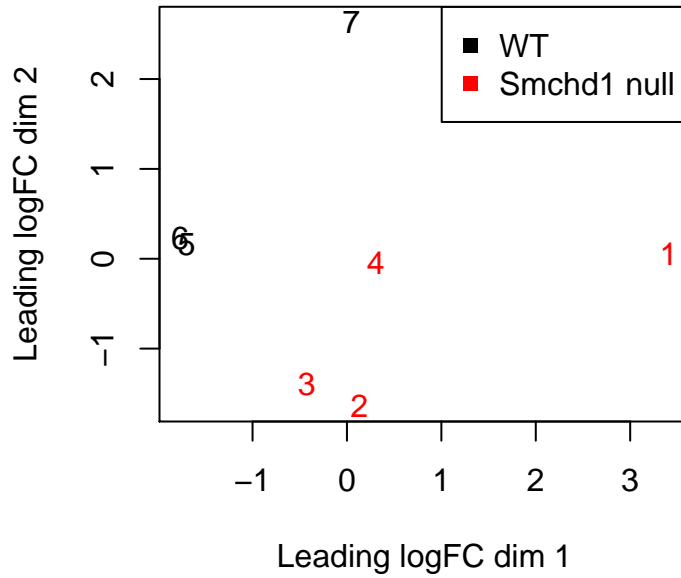
Voom analysis

```
# Analysis with voom only on full data set
v = voom(x, design=des)
vfit = lmFit(v)
vfit = eBayes(vfit)
options(digits=3)
topTable(vfit,coef=2, sort.by="P")

##            GeneID    Symbols logFC AveExpr      t  P.Value adj.P.Val    B
## 74355      74355       Smchd1 -3.12   6.067 -23.35 2.16e-08  0.000266 9.97
## 18028      18028         Nfib  8.98   1.714  12.60 2.17e-06  0.013355 3.15
## 75605      75605        Kdm5b -3.55   3.618 -11.75 3.62e-06  0.014857 5.06
## 667435    667435  Igkv17-121 -5.35  -1.435 -10.22 9.95e-06  0.025513 2.57
## 381126    381126        Garem  6.17   0.113  10.08 1.10e-05  0.025513 2.35
## 381413    381413       Gpr176 -4.02   1.328  -9.90 1.25e-05  0.025513 3.39
## 75033      75033         Mei4  6.44   0.259   9.69 1.45e-05  0.025513 2.23
## 69136      69136        Tusc1  5.67  -0.184   8.90 2.67e-05  0.040995 1.87
## 233552    233552        Gdpd5 -2.82   1.948  -8.56 3.49e-05  0.042754 2.81
## 80890      80890        Trim2 -1.43   4.491  -8.40 4.00e-05  0.042754 2.72

top = topTable(vfit,coef=2,number=Inf,sort.by="P")
sum(top$adj.P.Val<0.05)

## [1] 12
```
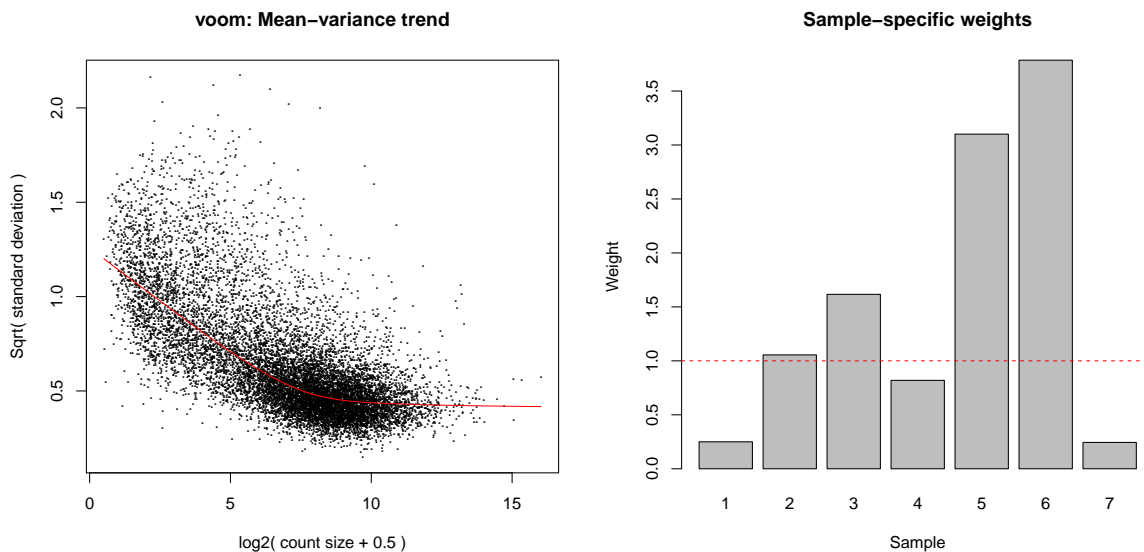
```
sum(top$adj.P.Val<0.01)
```

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

Voom with sample weights

```
# Analysis with combined voom and sample quality weights
# Figure 2A & 2B
vwts = voomWithQualityWeights(x, design=des, normalization="none", plot=TRUE)
```



```
vfit2 = lmFit(vwts)
vfit2 = eBayes(vfit2)
topTable(vfit2,coef=2,sort.by="P")
```

```
##          GeneID Symbols  logFC  AveExpr      t  P.Value adj.P.Val     B
## 74355     74355  Smchd1  -3.17    6.067  -28.5 1.61e-09  1.98e-05 12.57
## 18028     18028    Nfib   9.23    1.714   19.0 4.44e-08  2.73e-04  6.91
## 381126   381126   Garem   6.45    0.113   15.9 1.85e-07  7.58e-04  6.02
## 75033     75033    Mei4   6.56    0.259   15.0 2.84e-07  8.73e-04  5.83
## 69136     69136   Tusc1   5.88   -0.184   13.6 6.16e-07  1.11e-03  5.31
## 54354     54354  Rassf5   5.74    4.554   13.6 6.26e-07  1.11e-03  6.63
## 75605     75605   Kdm5b  -3.80    3.618  -13.5 6.53e-07  1.11e-03  6.67
## 58998     58998   Pvrl3   7.69    0.961   13.1 8.46e-07  1.11e-03  5.33
## 320398   320398   Lrig3   7.39    1.584   13.1 8.49e-07  1.11e-03  5.32
## 17069     17069    Ly6e   2.63    7.605   13.0 9.01e-07  1.11e-03  6.26
```

```
top2 = topTable(vfit2,coef=2,number=Inf,sort.by="P")
sum(top2$adj.P.Val<0.05)
```

```
## [1] 1478
```

```
sum(top2$adj.P.Val<0.01)

## [1] 94
```

Voom after removing outlier samples

```
# Analysis where outlier samples have been removed
v3 = voom(x[,-c(1,7)], design=des[-c(1,7),])
vfit3 = lmFit(v3)
vfit3 = eBayes(vfit3)
topTable(vfit3,coef=2, sort.by="P")

##              GeneID Symbols logFC AveExpr     t  P.Value adj.P.Val
## 74355         74355  Smchd1 -3.13    5.95 -23.7 3.48e-07   0.00428
## 54354         54354  Rassf5  5.89    3.83  14.6 6.32e-06   0.01380
## 17069         17069    Ly6e  2.64    7.40  14.5 6.57e-06   0.01380
## 14958         14958    H1f0 -1.88    5.50 -14.4 6.80e-06   0.01380
## 70012         70012   Cep85  1.70    5.82  14.1 7.51e-06   0.01380
## 80879         80879 Slc16a3 -4.93    5.29 -14.0 8.07e-06   0.01380
## 100043324 100043324 Gm10480 -1.60    5.61 -13.8 8.80e-06   0.01380
## 15511         15511  Hspa1b -3.46    4.03 -13.7 8.97e-06   0.01380
## 66848         66848   Fuca2 -1.76    4.39 -12.6 1.45e-05   0.01902
## 106869       106869 Tnfaip8  1.98    5.76  12.4 1.62e-05   0.01902
##                 B
## 74355        7.38
## 54354        2.84
## 17069        4.76
## 14958        4.72
## 70012        4.62
## 80879        4.50
## 100043324 4.47
## 15511        4.28
## 66848        3.96
## 106869       3.87

top3 = topTable(vfit3,coef=2,number=Inf,sort.by="P")
sum(top3$adj.P.Val<0.05)

## [1] 492

sum(top3$adj.P.Val<0.01)

## [1] 1
```
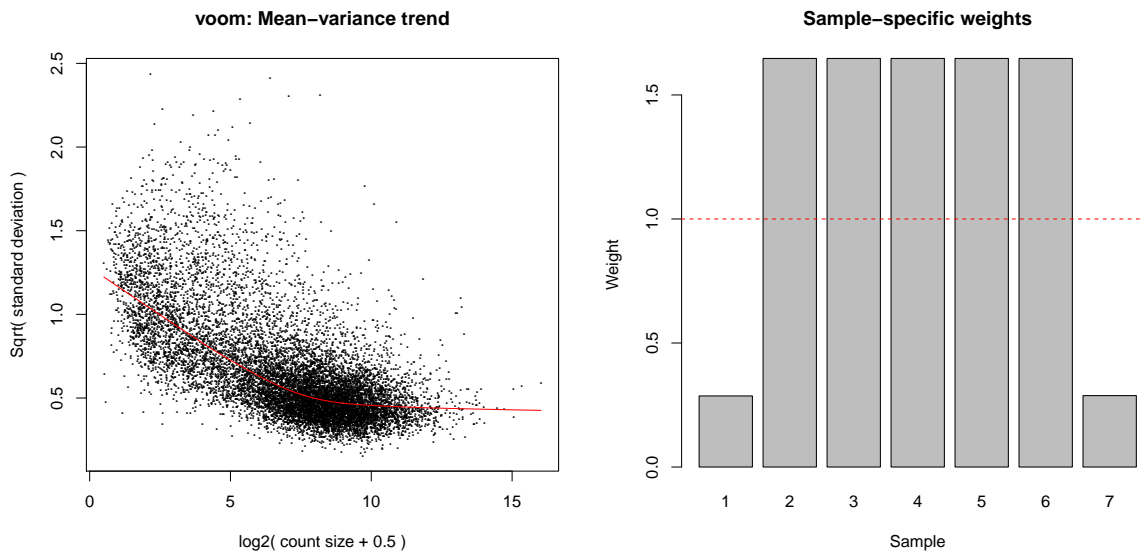
Voom with block weights

```
# Analysis with voom and block weights
Z = cbind(c(1,0,0,0,0,0,-1),c(0,1,1,1,1,1,-5))
vwts2 = voomWithQualityWeights(x, design=des, normalization="none", var.design=Z, plot=TRUE)
```



```
vfit4 = lmFit(vwts2)
vfit4 = eBayes(vfit4)
topTable(vfit4,coef=2,sort.by="P")

##          GeneID    Symbols logFC AveExpr     t  P.Value adj.P.Val     B
## 74355     74355      Smchd1 -3.13   6.067 -27.1 4.18e-09  5.14e-05 11.67
## 75605     75605       Kdm5b -3.67   3.618 -13.5 9.59e-07  5.90e-03  6.39
## 667435   667435  Igkv17-121 -5.18  -1.435 -11.8 2.66e-06  7.48e-03  3.83
## 18028     18028        Nfib  9.16   1.714  11.6 2.90e-06  7.48e-03  3.54
## 381126   381126       Garem  6.44   0.113  11.2 3.98e-06  7.48e-03  3.31
## 80890     80890       Trim2 -1.33   4.491 -10.7 5.64e-06  7.48e-03  4.62
## 18787     18787    Serpine1 -7.17  -0.618 -10.6 5.84e-06  7.48e-03  3.62
## 17069     17069        Ly6e  2.54   7.605  10.6 5.97e-06  7.48e-03  4.38
## 320435   320435        Rinl  1.77   5.075  10.6 6.01e-06  7.48e-03  4.56
## 381413   381413      Gpr176 -3.88   1.328 -10.5 6.08e-06  7.48e-03  4.28

top4 <- topTable(vfit4,coef=2,number=Inf,sort.by="P")
sum(top4$adj.P.Val<0.05)

## [1] 488

sum(top4$adj.P.Val<0.01)

## [1] 12
```

Roast analysis of Imprinted genes

```r
# Gene set testing below using ROAST (Wu et al. 2010)
# Imprinted genes downloaded from http://www.mousebook.org/imprinting-gene-list
imprints = read.table("imprinted_genelist.txt", sep="\t", header=TRUE)
nrow(imprints)
```

```
## [1] 150
```

```r
impind = match(alias2SymbolTable(imprints[,1], species="Mm"),
               alias2SymbolTable(x$genes$Symbols, species="Mm"))
impind = unique(impind[!is.na(impind)])
length(impind)
```

```
## [1] 46
```

```r
set.seed(2102014)
# Voom only
roast(v, index=impind, contrast=2, nrot=99999)
```

```
##           Active.Prop P.Value
## Down            0.196  0.9406
## Up              0.196  0.0594
## UpOrDown        0.196  0.1188
## Mixed           0.391  0.0202
```

```r
# Voom with sample weights
roast(vwts, index=impind, contrast=2, nrot=99999)
```

```
##           Active.Prop P.Value
## Down            0.326  0.9802
## Up              0.283  0.0198
## UpOrDown        0.326  0.0396
## Mixed           0.609  0.0001
```

```r
# Voom with outlier removal
roast(v3, index=impind, contrast=2, nrot=99999)
```

```
##           Active.Prop P.Value
## Down            0.174 0.66312
## Up              0.217 0.33688
## UpOrDown        0.217 0.67376
## Mixed           0.391 0.00208
```

```r
# Voom with block weights
roast(vwts2, index=impind, contrast=2, nrot=99999)
```

```
##           Active.Prop P.Value
## Down            0.217  0.9393
## Up              0.261  0.0607
## UpOrDown        0.261  0.1213
## Mixed           0.478  0.0002
```

Roast analysis of Protocadherins

```r
# Repeat for Protocadherins
proind = grep("protocadherin",as.character(fullanno$GeneName))
length(proind)
```

```
## [1] 71
```

```r
proind = grep("protocadherin",as.character(selanno$GeneName))
proind = unique(proind)
length(proind)
```

```
## [1] 8
```

```r
set.seed(2102014)
# Voom only
roast(v, index=proind, contrast=2, nrot=99999)
```

```
##          Active.Prop P.Value
## Down           0.125 0.94190
## Up             0.375 0.05810
## UpOrDown       0.375 0.11620
## Mixed          0.500 0.00655
```

```r
# Voom with sample weights
roast(vwts, index=proind, contrast=2, nrot=99999)
```

```
##          Active.Prop P.Value
## Down           0.125 0.99387
## Up             0.625 0.00614
## UpOrDown       0.625 0.01227
## Mixed          0.750 0.00050
```

```r
# Voom with outlier removal
roast(v3, index=proind, contrast=2, nrot=99999)
```

```
##          Active.Prop P.Value
## Down           0.125  0.9293
## Up             0.250  0.0707
## UpOrDown       0.250  0.1414
## Mixed          0.375  0.0410
```

```r
# Voom with block weights
roast(vwts2, index=proind, contrast=2, nrot=99999)
```

```
##          Active.Prop P.Value
## Down           0.125 0.97649
## Up             0.625 0.02352
## UpOrDown       0.625 0.04703
## Mixed          0.750 0.00269
```

```
sessionInfo()

## R version 3.1.1 (2014-07-10)
## Platform: x86_64-apple-darwin10.8.0 (64-bit)
##
## locale:
## [1] en_AU.UTF-8/en_AU.UTF-8/en_AU.UTF-8/C/en_AU.UTF-8/en_AU.UTF-8
##
## attached base packages:
## [1] parallel  stats     graphics  grDevices utils     datasets
## [7] methods   base
##
## other attached packages:
##  [1] org.Mm.eg.db_2.14.0  RSQLite_0.11.4       DBI_0.2-7
##  [4] AnnotationDbi_1.26.0 GenomeInfoDb_1.0.2   Biobase_2.24.0
##  [7] BiocGenerics_0.10.0  edgeR_3.6.8          limma_3.23.1
## [10] knitr_1.6
##
## loaded via a namespace (and not attached):
## [1] evaluate_0.5.5  formatR_1.0     highr_0.3        IRanges_1.22.10
## [5] stats4_3.1.1    stringr_0.6.2   tools_3.1.1
```