

limma_debug

Maciek Sykulski

5/12/2014

```
%default_mode r
%r R.version.string
%r print("email: 'Maciek Sykulski' <macieksk@gmail.com>")
[1] "R version 3.0.2 (2013-09-25)"
[1] "email: 'Maciek Sykulski' <macieksk@gmail.com>"

library(limma)
library(statmod)
sessionInfo()
packageDescription('limma')$Maintainer
R version 3.0.2 (2013-09-25)
Platform: x86_64-unknown-linux-gnu (64-bit)

locale:
 [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C              LC_TIME=en_US.UTF-8
 [4] LC_COLLATE=en_US.UTF-8   LC_MONETARY=en_US.UTF-8  LC_MESSAGES=en_US.UTF-8
 [7] LC_PAPER=en_US.UTF-8     LC_NAME=C                 LC_ADDRESS=C
[10] LC_TELEPHONE=C           LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C

attached base packages:
[1] stats      graphics  grDevices  utils      datasets  methods    base

other attached packages:
[1] statmod_1.4.19 limma_3.18.13
[1] "Gordon Smyth <smyth@wehi.edu.au>"

#system('wget http://bioinf.wehi.edu.au/marray/IlluminaCaseStudy/Targets.\
txt')
#system('wget http://bioinf.wehi.edu.au/marray/IlluminaCaseStudy/probe%20\
profile.txt.gz')
#system('wget http://bioinf.wehi.edu.au/marray/IlluminaCaseStudy/control\
%20probe%20profile.txt.gz')
dir()
[1] "control probe profile.txt.gz" "limma_debug.sagews"      "probe profile.txt.gz"
[4] "R"                          "Targets.txt"             "Targets.txt.1"
[7] "Targets.txt.2"              "Targets.txt.3"          "terminal.term"

#This is not a proper limma analysis workflow, this is only to demnstrate\
the problem
```

```

targets <- readTargets('Targets.txt')
x <- read.ilmn(files='probe profile.txt.gz', ctrlfiles='control probe \
  profile.txt.gz',other.columns='Detection')
y <- neqc(x)
dim(y)
expressed <- rowSums(y$other$Detection < 0.05) >= 3
y <- y[expressed,]
dim(y)
ct <- factor(targets$Type)
design <- model.matrix(~ 0 + ct)
colnames(design) <- levels(ct)
#dupcor <- duplicateCorrelation(y,design,block=targets$Donor)
#dupcor$consensus.correlation

fit <- lmFit(y,design,robust=True)

fit[,2:3]$design # this subsets on samples in design
head(fit[,2:3]$coef)
Reading file probe profile.txt.gz ... ...
Reading file control probe profile.txt.gz ... ...

```

```
[1] 48803 12
```

```
[1] 24691 12
```

```

mL MS pL stroma
2 0 0 0 1
3 1 0 0 0
MS pL
ILMN_2055271 5.238395 5.426670
ILMN_1653355 5.760232 5.842400
ILMN_1787689 5.302525 5.047953
ILMN_1745607 10.348355 5.217474
ILMN_1735045 7.273259 7.537589
ILMN_1659452 5.036892 5.022157

```

```

fit$design
mL MS pL stroma
1 0 1 0 0
2 0 0 0 1
3 1 0 0 0
4 0 0 1 0
5 0 1 0 0
6 0 0 0 1
7 1 0 0 0
8 0 0 1 0
9 0 1 0 0

```

```

10 0 0 0 1
11 1 0 0 0
12 0 0 1 0
attr("assign")
[1] 1 1 1 1
attr("contrasts")
attr("contrasts")$ct
[1] "contr.treatment"

```

```

####
#### Problem with subsetting of fit$design
####

```

```

is.fullrank(fit$design)
is.fullrank(fit$design[,2:3])
is.fullrank(fit[,2:3]$design)
[1] TRUE
[1] TRUE
[1] FALSE

```

```

c()
topTable(eBayes(fit,robust=TRUE), coef=2:3,sort.by="F")
topTable(eBayes(fit,robust=FALSE), coef=2:3,sort.by="F")
NULL

```

	SYMBOL	MS	pL	AveExpr	F	P.Value	adj.P.Val
ILMN_1746485	LOC645899	13.70583	13.73622	13.71490	40312.97	1.257293e-20	2.779984e-17
ILMN_1656807	RPL27	14.05692	13.98869	14.00480	38770.95	1.534313e-20	2.779984e-17
ILMN_2038774	EEF1A1	13.90857	13.98534	13.95823	38752.81	1.537984e-20	2.779984e-17
ILMN_2412521	KIAA0101	13.70286	13.67485	13.70604	38260.49	1.641725e-20	2.779984e-17
ILMN_1752285	RPL4	13.84533	13.95270	13.80513	37650.23	1.782178e-20	2.779984e-17
ILMN_2218277	RPS2	13.82234	13.75690	13.77192	37236.16	1.885693e-20	2.779984e-17
ILMN_1715926	LOC642210	14.13195	14.16180	14.15469	37095.87	1.922385e-20	2.779984e-17
ILMN_2041368	RN7SL1	13.77140	13.72141	13.78550	35971.57	2.249470e-20	2.779984e-17
ILMN_2215061	ORC6L	13.74861	13.64835	13.72340	35840.50	2.291785e-20	2.779984e-17
ILMN_2219134	RPS15	13.81265	13.80921	13.78184	35705.42	2.336393e-20	2.779984e-17
	SYMBOL	MS	pL	AveExpr	F	P.Value	adj.P.Val
ILMN_1746485	LOC645899	13.70583	13.73622	13.71490	41852.05	1.038390e-20	2.418938e-17
ILMN_1656807	RPL27	14.05692	13.98869	14.00480	40129.82	1.286865e-20	2.418938e-17
ILMN_2038774	EEF1A1	13.90857	13.98534	13.95823	40125.71	1.287538e-20	2.418938e-17
ILMN_2412521	KIAA0101	13.70286	13.67485	13.70604	39650.91	1.368211e-20	2.418938e-17
ILMN_1752285	RPL4	13.84533	13.95270	13.80513	38954.06	1.497845e-20	2.418938e-17
ILMN_2218277	RPS2	13.82234	13.75690	13.77192	38532.01	1.583511e-20	2.418938e-17
ILMN_1715926	LOC642210	14.13195	14.16180	14.15469	38315.58	1.629711e-20	2.418938e-17
ILMN_2041368	RN7SL1	13.77140	13.72141	13.78550	37187.33	1.898368e-20	2.418938e-17
ILMN_2215061	ORC6L	13.74861	13.64835	13.72340	37056.00	1.932968e-20	2.418938e-17
ILMN_2219134	RPS15	13.81265	13.80921	13.78184	36891.43	1.977396e-20	2.418938e-17

```

topTable
function (fit, coef = NULL, number = 10, genelist = fit$genes,
  adjust.method = "BH", sort.by = "B", resort.by = NULL, p.value = 1,
  lfc = 0, confint = FALSE)
{

```

```

if (!is(fit, "MArrayLM"))
  stop("fit must be an MArrayLM object")
if (is.null(fit$coefficients))
  stop("coefficients not found in fit object")
if (confint && is.null(fit$stdev.unscaled))
  stop("stdev.unscaled not found in fit object")
if (is.null(coef))
  coef <- 1:ncol(fit)
if (length(coef) > 1) {
  coef <- unique(coef)
  if (length(fit$coef[1, coef]) < ncol(fit))
    fit <- eBayes(fit[, coef])
  if (sort.by == "B")
    sort.by <- "F"
  return(topTableF(fit, number = number, genelist = genelist,
    adjust.method = adjust.method, sort.by = sort.by,
    p.value = p.value, lfc = lfc))
}
fit <- unclass(fit)
ebcols <- c("t", "p.value", "lods")
if (confint)
  ebcols <- c("s2.post", "df.total", ebcols)
topTable(fit = fit[c("coefficients", "stdev.unscaled")],
  coef = coef, number = number, genelist = genelist, A = fit$Amean,
  eb = fit[ebcols], adjust.method = adjust.method, sort.by = sort.by,
  resort.by = resort.by, p.value = p.value, lfc = lfc,
  confint = confint)
}
<environment: namespace:limma>

```

```

topTable.fixed<-function (fit, coef = NULL, number = 10, genelist = \
fit$genes,
  adjust.method = "BH", sort.by = "B", resort.by = NULL, p.value = 1,
  lfc = 0, confint = FALSE, robust=FALSE)
{
  if (!is(fit, "MArrayLM"))
    stop("fit must be an MArrayLM object")
  if (is.null(fit$coefficients))
    stop("coefficients not found in fit object")
  if (confint && is.null(fit$stdev.unscaled))
    stop("stdev.unscaled not found in fit object")
  if (is.null(coef))
    coef <- 1:ncol(fit)
  if (length(coef) > 1) {
    coef <- unique(coef)
    if (length(fit$coef[1, coef]) < ncol(fit)) {
      design <- fit$design[, coef]
      fit<-fit[,coef]
      fit$design<-design
      fit <- eBayes.fixed(fit, robust=robust)
    }
    if (sort.by == "B")

```

```

        sort.by <- "F"
        return(topTableF(fit, number = number, genelist = genelist,
            adjust.method = adjust.method, sort.by = sort.by,
            p.value = p.value, lfc = lfc))
    }
    fit <- unclass(fit)
    ebcols <- c("t", "p.value", "lods")
    if (confint)
        ebcols <- c("s2.post", "df.total", ebcols)
    toptable(fit = fit[c("coefficients", "stdev.unscaled")],
        coef = coef, number = number, genelist = genelist, A = fit$Amean,
        eb = fit[ebcols], adjust.method = adjust.method, sort.by = sort.\
        by,
        resort.by = resort.by, p.value = p.value, lfc = lfc,
        confint = confint)
}

```

```

eBayes
function (fit, proportion = 0.01, stdev.coef.lim = c(0.1, 4),
    trend = FALSE, robust = FALSE, winsor.tail.p = c(0.05, 0.1))
{
    if (trend)
        if (is.null(fit$Amean))
            stop("Need Amean component in fit to estimate trend")
    eb <- ebayes(fit = fit, proportion = proportion, stdev.coef.lim = stdev.coef.lim,
        trend = trend, robust = robust, winsor.tail.p = winsor.tail.p)
    fit$df.prior <- eb$df.prior
    fit$s2.prior <- eb$s2.prior
    fit$var.prior <- eb$var.prior
    fit$proportion <- proportion
    fit$s2.post <- eb$s2.post
    fit$t <- eb$t
    fit$df.total <- eb$df.total
    fit$p.value <- eb$p.value
    fit$lods <- eb$lods
    if (!is.null(fit$design) && is.fullrank(fit$design)) {
        F.stat <- classifyTestsF(fit, fstat.only = TRUE)
        fit$F <- as.vector(F.stat)
        df1 <- attr(F.stat, "df1")
        df2 <- attr(F.stat, "df2")
        if (df2[1] > 1e+06)
            fit$F.p.value <- pchisq(df1 * fit$F, df1, lower.tail = FALSE)
        else fit$F.p.value <- pf(fit$F, df1, df2, lower.tail = FALSE)
    }
    fit
}
<environment: namespace:limma>

```

```

eBayes.fixed<-function (fit, proportion = 0.01, stdev.coef.lim = c(0.1, \
    4),
    trend = FALSE, robust = FALSE, winsor.tail.p = c(0.05, 0.1))
{

```

```

print(paste("eBayes(robust=",robust))
if (trend)
  if (is.null(fit$Amean))
    stop("Need Amean component in fit to estimate trend")
eb <- ebayes(fit = fit, proportion = proportion, stdev.coef.lim = \
  stdev.coef.lim,
  trend = trend, robust = robust, winsor.tail.p = winsor.tail.p)
fit$df.prior <- eb$df.prior
fit$s2.prior <- eb$s2.prior
fit$var.prior <- eb$var.prior
fit$proportion <- proportion
fit$s2.post <- eb$s2.post
fit$t <- eb$t
fit$df.total <- eb$df.total
fit$p.value <- eb$p.value
fit$lods <- eb$lods
if (!is.null(fit$design) && is.fullrank(fit$design)) {
  print("Computing Fstat")
  F.stat <- classifyTestsF(fit, fstat.only = TRUE)
  fit$F <- as.vector(F.stat)
  df1 <- attr(F.stat, "df1")
  df2 <- attr(F.stat, "df2")
  if (df2[1] > 1e+06)
    fit$F.p.value <- pchisq(df1 * fit$F, df1, lower.tail = FALSE)
  else fit$F.p.value <- pf(fit$F, df1, df2, lower.tail = FALSE)
}
fit
}

```

```

N<-10
topTable(eBayes(fit,robust=TRUE), coef=2:3,sort.by="F",number=N)
topTable.fixed(eBayes.fixed(fit,robust=TRUE), coef=2:3,sort.by="F",robust\
=TRUE,number=N)
topTable.fixed(eBayes.fixed(fit,robust=TRUE), coef=2:3,sort.by="F",robust\
=FALSE,number=N)
topTable(eBayes(fit,robust=FALSE), coef=2:3,sort.by="F",number=N)

```

SYMBOL	MS	pL	AveExpr	F	P.Value	adj.P.Val
ILMN_1746485	LOC645899	13.70583	13.73622	13.71490	40312.97	1.257293e-20
ILMN_1656807	RPL27	14.05692	13.98869	14.00480	38770.95	1.534313e-20
ILMN_2038774	EEF1A1	13.90857	13.98534	13.95823	38752.81	1.537984e-20
ILMN_2412521	KIAA0101	13.70286	13.67485	13.70604	38260.49	1.641725e-20
ILMN_1752285	RPL4	13.84533	13.95270	13.80513	37650.23	1.782178e-20
ILMN_2218277	RPS2	13.82234	13.75690	13.77192	37236.16	1.885693e-20
ILMN_1715926	LOC642210	14.13195	14.16180	14.15469	37095.87	1.922385e-20
ILMN_2041368	RN7SL1	13.77140	13.72141	13.78550	35971.57	2.249470e-20
ILMN_2215061	ORC6L	13.74861	13.64835	13.72340	35840.50	2.291785e-20
ILMN_2219134	RPS15	13.81265	13.80921	13.78184	35705.42	2.336393e-20

```

[1] "eBayes(robust= TRUE"
[1] "Computing Fstat"
[1] "eBayes(robust= TRUE"
[1] "Computing Fstat"

```

	SYMBOL	MS	pL	AveExpr	F	P.Value	adj.P.Val
ILMN_1746485	LOC645899	13.70583	13.73622	13.71490	40312.97	1.257293e-20	2.779984e-17
ILMN_1656807	RPL27	14.05692	13.98869	14.00480	38770.95	1.534313e-20	2.779984e-17
ILMN_2038774	EEF1A1	13.90857	13.98534	13.95823	38752.81	1.537984e-20	2.779984e-17
ILMN_2412521	KIAA0101	13.70286	13.67485	13.70604	38260.49	1.641725e-20	2.779984e-17
ILMN_1752285	RPL4	13.84533	13.95270	13.80513	37650.23	1.782178e-20	2.779984e-17
ILMN_2218277	RPS2	13.82234	13.75690	13.77192	37236.16	1.885693e-20	2.779984e-17
ILMN_1715926	LOC642210	14.13195	14.16180	14.15469	37095.87	1.922385e-20	2.779984e-17
ILMN_2041368	RN7SL1	13.77140	13.72141	13.78550	35971.57	2.249470e-20	2.779984e-17
ILMN_2215061	ORC6L	13.74861	13.64835	13.72340	35840.50	2.291785e-20	2.779984e-17
ILMN_2219134	RPS15	13.81265	13.80921	13.78184	35705.42	2.336393e-20	2.779984e-17

[1] "eBayes(robust= TRUE"

[1] "Computing Fstat"

[1] "eBayes(robust= FALSE"

[1] "Computing Fstat"

	SYMBOL	MS	pL	AveExpr	F	P.Value	adj.P.Val
ILMN_1746485	LOC645899	13.70583	13.73622	13.71490	41852.05	1.038390e-20	2.418938e-17
ILMN_1656807	RPL27	14.05692	13.98869	14.00480	40129.82	1.286865e-20	2.418938e-17
ILMN_2038774	EEF1A1	13.90857	13.98534	13.95823	40125.71	1.287538e-20	2.418938e-17
ILMN_2412521	KIAA0101	13.70286	13.67485	13.70604	39650.91	1.368211e-20	2.418938e-17
ILMN_1752285	RPL4	13.84533	13.95270	13.80513	38954.06	1.497845e-20	2.418938e-17
ILMN_2218277	RPS2	13.82234	13.75690	13.77192	38532.01	1.583511e-20	2.418938e-17
ILMN_1715926	LOC642210	14.13195	14.16180	14.15469	38315.58	1.629711e-20	2.418938e-17
ILMN_2041368	RN7SL1	13.77140	13.72141	13.78550	37187.33	1.898368e-20	2.418938e-17
ILMN_2215061	ORC6L	13.74861	13.64835	13.72340	37056.00	1.932968e-20	2.418938e-17
ILMN_2219134	RPS15	13.81265	13.80921	13.78184	36891.43	1.977396e-20	2.418938e-17

	SYMBOL	MS	pL	AveExpr	F	P.Value	adj.P.Val
ILMN_1746485	LOC645899	13.70583	13.73622	13.71490	41852.05	1.038390e-20	2.418938e-17
ILMN_1656807	RPL27	14.05692	13.98869	14.00480	40129.82	1.286865e-20	2.418938e-17
ILMN_2038774	EEF1A1	13.90857	13.98534	13.95823	40125.71	1.287538e-20	2.418938e-17
ILMN_2412521	KIAA0101	13.70286	13.67485	13.70604	39650.91	1.368211e-20	2.418938e-17
ILMN_1752285	RPL4	13.84533	13.95270	13.80513	38954.06	1.497845e-20	2.418938e-17
ILMN_2218277	RPS2	13.82234	13.75690	13.77192	38532.01	1.583511e-20	2.418938e-17
ILMN_1715926	LOC642210	14.13195	14.16180	14.15469	38315.58	1.629711e-20	2.418938e-17
ILMN_2041368	RN7SL1	13.77140	13.72141	13.78550	37187.33	1.898368e-20	2.418938e-17
ILMN_2215061	ORC6L	13.74861	13.64835	13.72340	37056.00	1.932968e-20	2.418938e-17
ILMN_2219134	RPS15	13.81265	13.80921	13.78184	36891.43	1.977396e-20	2.418938e-17