Observed log fold changes depend on the mean for simulated counts

Michael Love

October 14, 2014

```r
library("DESeq2")
library("DESeq2paper")
makeSimScript <- system.file("script/makeSim.R", package = "DESeq2paper", mustWork = TRUE)
source(makeSimScript)
data("meanDispPairs")
data("pickrell_sumexp")
```

```r
m <- 8
pickrell.sub <- DESeqDataSet(pickrell[, seq_len(m)], ~1)
pickrell.sub <- estimateSizeFactors(pickrell.sub)
K <- counts(pickrell.sub, normalized = TRUE)
library("ggplot2")
library("gridExtra")
simpleMA <- function(K) {
    baseMean <- rowMeans(K)
    lfc <- log2(rowMeans(K[, 1:(m/2)])) - log2(rowMeans(K[, (m/2 + 1):m]))
    data <- data.frame(log10baseMean = log10(baseMean), lfc = lfc)
    data <- subset(data, is.finite(log10baseMean) & is.finite(lfc))
    ggplot(data, aes(log10baseMean, lfc)) + stat_binhex() + ylim(-6, 6) + xlab("observed log10 mean of cou
        ylab("observed log2 fold change") + scale_fill_gradient(trans = "log10")
}
plt.real <- simpleMA(K)
```

```r
plt.real
```

```r
set.seed(1)
n <- 10 * round(nrow(pickrell)/10)
es <- 0
condition <- factor(rep(c("A", "B"), each = m/2))
x <- model.matrix(~condition)
beta <- rep(0, n)
sim <- makeSim(n, m, x, beta, meanDispPairs)
mat <- sim$mat
ylim <- c(-6, 6)
xlab <- "true log10 base mean"
ylab <- "true log2 fold change"
d <- data.frame(log10mu = log10(sim$mu0), beta = beta)
plt.null1 <- ggplot(d, aes(log10mu, beta)) + geom_point() + xlab(xlab) + ylab(ylab) +
    ylim(ylim[1], ylim[2])
plt.null2 <- simpleMA(mat)
```
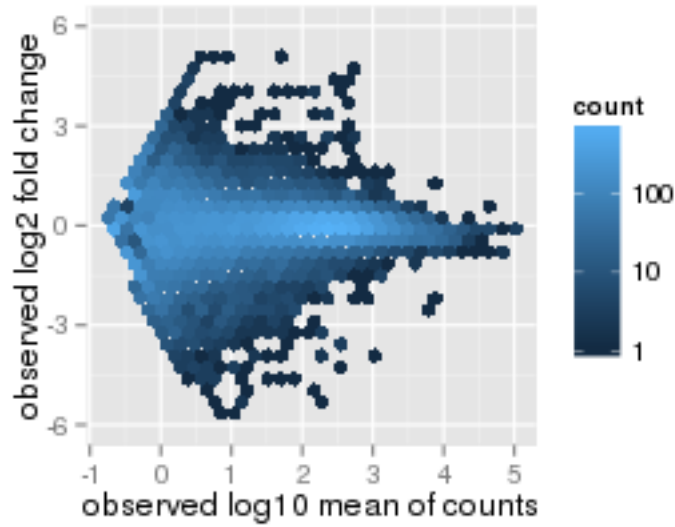
Figure 1: MA plot from a 4 vs 4 comparison of the Pickrell et al. dataset, wherein there is no known phenotypic difference dividing the groups.

```
n <- 10 * round(nrow(pickrell)/10)
es <- 1
condition <- factor(rep(c("A", "B"), each = m/2))
x <- model.matrix(~condition)
beta <- c(rep(0, n * 8/10), sample(c(-es, es), n * 2/10, TRUE))
sim <- makeSim(n, m, x, beta, meanDispPairs)
mat <- sim$mat
d <- data.frame(log10mu = log10(sim$mu0), beta = beta)
plt.flat1 <- ggplot(d, aes(log10mu, beta)) + geom_point() + xlab(xlab) + ylab(ylab) +
    ylim(ylim[1], ylim[2])
plt.flat2 <- simpleMA(mat)
```

```
grid.arrange(plt.null1, plt.null2, plt.flat1, plt.flat2, ncol = 2)
```
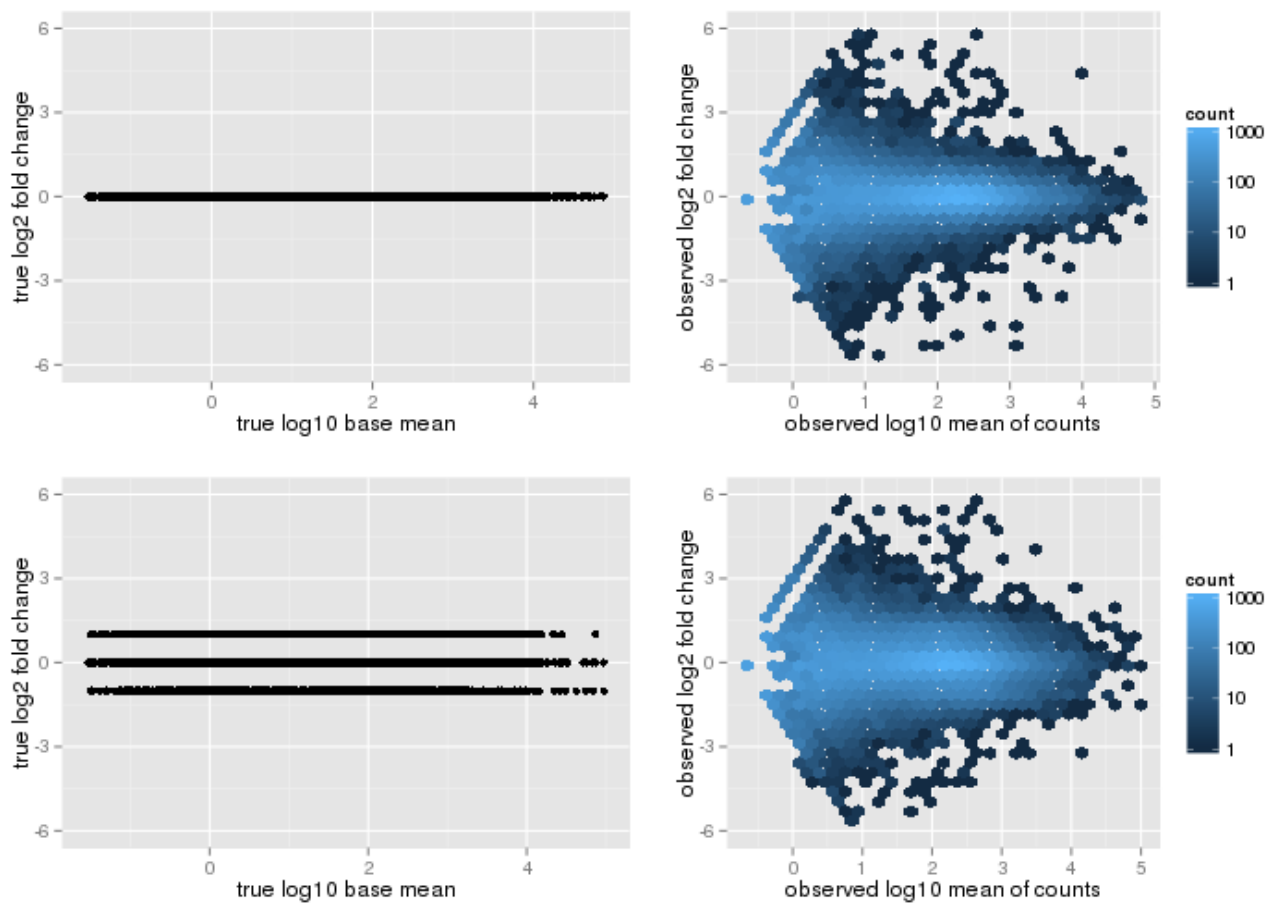
Figure 2: True log fold changes and the observed log fold changes from simulated counts. In the top row, all true log fold changes are equal to zero. On the bottom row, 20% of true log fold changes are set to a fixed value.