Solutions for chapter Differential Expression

Exercise 1

a There is a weak relationship but it is not dominant. We may safely proceed with the nonspecific filtering based on variablility.

b It switches between plotting the x-axis (means) on the original scale (`FALSE`) or on the rank scale (`TRUE`). The latter distributes the data more evenly along the x-axis and allows a better visual assessment of the standard deviation as a function of the mean.

Exercise 2

The number of probe sets with p-value less than 0.05 and mean $\log_2$ fold-change larger than 0.5 is

```
> sum(tt$p.value<0.05 & abs(tt$dm)>0.5)
[1] 224
```

This choice of thresholds is of course arbitrary.

Exercise 3

```
> mtyp = ALLset1$mol.biol
> sel = rep(1:2, each=rev(table(mtyp)))
> plot(exprs(ALLset1)[j, order(mtyp)], pch=c(1,15)[sel],
      col=c("black", "red")[sel],
      main=featureNames(ALLset1)[j],
      ylab=expression(log[2]~expression~level))
> legend("bottomleft", col=c("black", "red"),
      pch=c(1,15), levels(mtyp), bty="n")
```

Exercise 4

The curve for a bad discriminator would be close to the diagonal because the classification would be almost random. The curve for a perfect discriminator shows both high sensitivity and high specificity over the whole plot, that is, a rectangle from `[0,1]` to `[1,1]`.

Exercise 5

The identification of differentially expressed genes by area under the ROC curve is not so much affected by the sample size as the t-statistic is. For the t-test the number of differentially expressed genes increases constantly with the sample size. For the ROC curves this number stabilizes with a sufficient sample size.