1. This assignment uses a subset of the Singh prostate cancer data that was acquired for
   Homework 1. Download the "subcel.txt" and "subsamples.txt" files for homework 3
   from the course web site. These files list the 20 CEL files to be used in this assignment.
   (You will have to edit "subcel.txt" so the path points to the correct location on your
   computer.)

   (a) Use the supplied "subsamples.txt" file to create a `phenoData` object. Use short
       sample names to identify the arrays.

   (b) Load in all 20 CEL files to create an `AffyBatch` object. Create a boxplot and a
       histogram of the raw data. Do the arrays need to be normalized?

   (c) Look at images of several of the arrays. Do you see anything unusual?

2. Use BioConductor to decide if the RNA used in any of these experiments was excep-
   tionally degraded.

   Bonus: What happens if you plot the degradation slopes as a function of the factors
   in the `pData` object in the AffyBatch?

3. In Homework 1, we saw that clustering the samples after standard processing with
   dChip, there were two dominant clusters that did not match the biological division
   into normal and cancer. The 20 samples used in this homework are evenly balanced
   between normal/cancer and between the two dChip clusters. Using the `qc` function
   in the `simpleaffy` package, explore the various quality control metrics. Do any of
   the quality control metrics suggest an explanation for the differences between the two
   clusters?

4. Perform background correction on all the arrays using two different methods: "mas"
   and "rma". Prepare histograms and boxplots of the AffyBatch objects for both back-
   ground correction methods. What can you say about the differences? Which back-
   ground correction method do you think is better? Why?

5. This problem is a continuation of the previous problem. This problem, like several of
   the later problems, uses the `simpleCluster` function that is available from the course
   web site. This function produces cluster dendrograms based on the 25% of genes with
   the highest average expression.

   (a) After processing the arrays using "mas" background correction, use `expresso`
       to perform "quantiles" normalization, select the "pmonly" features, and quantify
       using the "medianpolish" summarization method. Use `simpleCluster` to produce
       a dendrogram.

   (b) Repeat part (a) using "rma" background correction.

   (c) Does the background correction method have any effect on clustering based on
       the highest expressing genes?