

1. Download the data set for homework 3 from the course web site. Use the supplied “SampleInfo.txt” file to create a `phenoData` object, then load in all 10 CEL files to create an `AffyBatch` object. Create a boxplot and a histogram of the raw data. Look at images of several of the arrays.
2. Use BioConductor to decide if the RNA used in any of these experiments was exceptionally degraded.
3. Perform background correction on all the arrays using two different methods: “mas” and “rma”. Prepare boxplots of the Affy batches for both background correction methods. What can you say about the differences? Next, prepare a histogram of the difference in intensity on one or two arrays. Explain why the histograms look the way they do.
4. For this problem, we will work with the data that has already been background corrected using “mas”. Normalize the background-corrected data using four different methods: “constant”, “quantile”, “invariantset” and “loess”. For each method, prepare boxplots and histograms (density plots) of the normalized data. Based on these plots, which method would you choose?
5. This is a continuation of the previous problem. Select one array, and construct a data frame with four columns. Each column should contain the intensity measurements from a different normalization method for the same array. Prepare M-versus-A plots (using `mva.pairs`) of this data frame. Explain any interesting qualitative features of these plots. Do these plots change your conclusion to the previous problem?
6. In this problem, we will work with the data that has already been background-corrected using “mas” and normalized using “constant” scaling. Use the `expresso` function to select the “pmonly” features and summarize them using 4 different methods: the “avgdiff” method of MAS4.0, the “liwong” method of dChip, the “mas” method of MAS5.0, and the “medianpolish” method of RMA. Prepare boxplots of the processed data for each method. (Note: because normalization produces an `exprSet`, you will have to manually extract the `exprs`, log transform them (in most cases: read about RMA!), and convert them to a data frame before creating the boxplot.) Do these plots lead you to prefer one method?
7. This is a continuation of the previous problem. The sample information tells us that there are two kinds of samples. Select one array of each type. For each method, prepare an M-versus-A plot of the summarized data on the two arrays. (Watch out for log transforms and the special case of RMA.) Do these plots give you any reason to prefer one method over another?