Bioconductor Homework

March 28, 2011

1 Working with meta-data in Bioconductor

1. Install Bioconductor and load the packages Biobase, limma, and ALL.
2. Load the data ALL. how many arrays (samples) does the data set include? How many different genes (features) does the array include?
3. Now explore what is in ALL. What does the abstract tell you about the patient population for the arrays in ALL? What additional variables other than gene expression are included in the data set?
4. Plot the density for any 4 arrays from ALL. Use par(mfrow=c()) so they are all on the same plot.
5. Now subset the data into two datasets, one with males and one with females, using the "which" command. How many arrays are in each subset?
6. Use the "grep" command to generate a subset of data that includes patients whose bell type is B-cell. How many arrays are in this subset?
7. Although we didn't discuss it, identify groups of individuals from multiple categories. For instance, mol.biol for patients has 6 possible categories. What if we only want to consider those patients with mol.biol of either "NEG" or "BCR/ABL". We can do this by using creating a vector containing the factor names we are interested in and then using the %in% command as follows:

```
> vec<-c("NEG", "BCR/ABL")
> grp.ids<-which(ALL$mol.biol%in%vec)
> grpALL<-ALL[,grp.ids]
```

Create a subset of the data the includes only those individuals with mol.biol ALL/AF4 or E2A/PBX1.

2 An Example of Gene-Expression analysis

Work through the Swirl Zebrafish example in the limma package Vignette. Data are easily obtained by following the directions in the vignette- just be sure they the files from swirl.zip are in your R working directory.