Seminar III: R/Bioconductor

August-December 2009

Leonardo Collado Torres
Bachelor in Genomic Sciences (LCG),
UNAM, Cuernavaca, Mexico
lcollado@lcg.unam.mx

http://www.lcg.unam.mx/~lcollado/

August 21, 2009

Assistants: Alejandro Reyes areyes@lcg.unam.mx, José Reyes jreyes@lcg.unam.mx and Víctor Moreno jmoreno@lcg.unam.mx

Note: Questions through the forum please. Those who are not from the sixth LCG generation send us an email so we can register you on the forum.

Abstract

With the following exercises you’ll take your first steps into using Sweave and explore the ALL dataset.

1 Sweave

1. Create your own template Sweave document.
   - Title: course name, homework number
   - Author: name, email, include a link to your personal academic webpage if you have one.\footnote{You will probably make one this semester on the PHP course.}
2 ALL DATASET

- Abstract: short description on the homework and any notes you might want to add
- A sample homework solution: meaning a short description and some code. For example, how to sum $2 + 3$.

2 ALL dataset

- You'll have to explore the ALL dataset and create your first homework as a vignette document.
- Install the ALL package and explore the ALL object.
  
  ```
  > library(ALL)
  > data(ALL)
  ```
- Select the samples from the B-cell tumors.
- Select those of molecular type BCR/ABL or NEG.
- Combine the previous two subsets and keep the intersection.
- Eliminate unused factor levels on your resulting subset.
- Use the `nsFilter` function from the `genefilter` package to keep those with entrez ID, GOBP, remove duplicate entrez and the following arguments:
  
  ```
  > nsFilter(var.fun = IQR, var.cutoff = 0.5, feature.exclude = "^AFFX")
  ```
- Meaning that we'll use the interquantile range with a variance cutoff of 0.5 to eliminate those with small variation and by excluding AFFX we'll take out the controls AFFY probes.
- How many:
  1. duplicates were removed?
  2. control features were excluded?
  3. had low variance (small variation)?
  4. had no GO?
  5. had no entrez ID?

---

2 John Quackenbush mentioned it on Monday as the most studied dataset.
3 What’s the name of the function to look for text in Unix? Well, a function with the same name is available in R. Use it.
4 A binary operator such as `%in%` is useful here.