



BioC2005 Conference *Where Software and Biology Connect*

Differential Expression Analysis of Microarray Experiments

Gordon Smyth
August 16, 2005

Walter and Eliza Hall Institute of Medical Research



Aims

This laboratory explores some of the features of the limma package for assessing differential expression in microarray experiments. Examples are included of cDNA two-color microarrays and Affymetrix one-channel microarrays. Some pre-processing issues are also discussed for two-color arrays.

Lab exercises

Exercise	Platform	Design	Topics covered
apoAI data	cDNA	Two group comparison with common reference	Introduction to linear models. Obtaining empirical Bayes statistics. Getting lists of differentially expressed genes.
integrin beta7 data	cDNA	Direct comparisons with dye-swaps	Data entry for two color data. Highlighting control probes. Exploring different background correction methods. Allowing for genewise dye effects.
Estrogen data	Affymetrix	2x2 Factorial	More on linear models. Use of contrasts. Venn diagrams. Linking gene lists to annotation information on the internet. Gene set tests.
Drosophila embryogenesis dataset	Affymetrix	Time course with series-level replication	Time course analysis using linear models and moderated F-statistics.

Datasets used in the exercises

Please check whether you already have the "Drosophila Embryo" and "Estrogen" packages from the Required Software. (These data sets are stored within R packages.)

- [ApoAI Knockout Data](#) (1.2 MB)
- [Integrin beta7 data](#) (14.6 MB)
- [Estrogen Data](#) (23.6 MB)
- Drosophila Embryo Data (as an R package): [Windows](#) - [Source \(Mac or Linux\)](#) (3.5 MB)

Required R packages

Please install these packages before attempting to repeat the lab exercises. Note: if you don't have write permission to the system library directory of your R installation, you can use the `.libPaths()` function with something like `.libPaths("C:/mylibdir")` before you run `install.packages()` (or equivalent) to install the packages in a customized directory location.

Package	Windows	MacOS X	Source
limma_2.0.4	limma_2.0.4.zip	limma_2.0.4.tar.gz	limma_2.0.4.tar.gz
statmod_1.2.0	statmod_1.2.0.zip	statmod_1.2.0.tar.gz	statmod_1.2.0.tar.gz
affy_1.6.7	affy_1.6.7.zip	affy_1.6.7.tgz	affy_1.6.7.tar.gz
Biobase_1.5.12	Biobase_1.5.12.zip	Biobase_1.5.12.tgz	Biobase_1.5.12.tar.gz
hgu95av2_1.8.4	hgu95av2_1.8.4.zip	hgu95av2_1.8.4.tar.gz	hgu95av2_1.8.4.tar.gz
hgu95av2cdf_1.5.1	hgu95av2cdf_1.5.1.zip	hgu95av2cdf_1.5.1.tar.gz	hgu95av2cdf_1.5.1.tar.gz
xtable_1.2-5	xtable_1.2-5.zip	xtable_1.2-5.tgz	xtable_1.2-5.tar.gz

Getting started

You should be running R 2.1.0 or 2.1.1 and limma 2.0.4. A good way to get started is to open up the Limma User's Guide:

```
library(limma)
```

If you're using Windows, just use the drop-down menu "Vignettes". Otherwise, type

```
limmaUsersGuide()
```

References

1. Smyth, G. K., Thorne, N. P. and Wettenhall J. (2005) limma: Linear Models for Microarray Data User's Guide. <http://bioinf.wehi.edu.au/limma> (Included as part of the limma package.)

2. Smyth, G. K. (2005). Limma: linear models for microarray data. In: *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds.), Springer, New York.
3. Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* **3**, No. 1, Article 3. <http://www.bepress.com/sagmb/vol3/iss1/art3/>