

Software needs (I)

- Software for computational biology and bioinformatics, especially the analysis of microarray data had to meet the needs of both biologists, statisticians and computer scientists.
- The analysis of microarray data was not simply the domain of statisticians, an entry point for biologist was required.
- New conceptual needs and challenges, requirement for biologists to access high quality statistical methods

1

Software needs (II)

- Data acquisition
 - Data management
 - Data transformation
 - Background correction
 - Combining data sources
 - Various normalisation steps
 - Finding differentially expressed genes
 - Interpreting and visualising high dimensional data
- Transparency
Reproducibility
Efficiency

2

R

- Free software environment for statistical computing and graphics
- High level interpreted language
- Packaging protocols
- Interfaces with Perl, Python, Java, C
- Active user community

Get the latest R version
(precompiled binaries for Windows and Mac OS X)
<http://cran.r-project.org/>



3

Bioconductor

- Open source software for bioinformatics
- Bioconductor is a series of R packages
- Core emphasis on reproducible research, good documentation and training, re-usable data structures, designed to work with different variations of data
- Questions about the analysis of array data using Bioconductor can be posted on their mailing list. This is a very informative mailing list for the analysis of data from a wide variety of high throughput genomic technologies.

<http://www.bioconductor.org/docs/mailList.html>
<http://www.bioconductor.org/>

For a quick install of a subset of the most common packages:
> source("http://www.bioconductor.org/biocLite.R")
> biocLite()

4

The screenshot shows the Bioconductor website with a navigation menu on the left and a main content area. The main content area features a 'project news' section with a list of recent updates, including the release of Bioconductor 2.1 and the 6th edition of the Computational and Statistical Aspects of Microarray Analysis (CSAMAA) conference. The website has a clean, professional layout with a blue and white color scheme.

5

Limma

- Especially for the application of linear models for analysing designed experiments and the assessment of differential expression
- Includes processing capabilities for two-colour spotted arrays and affymetrix chip data. The differential expression methods treat two-colour, affymetrix and single-colour experiments in a unified way.

<http://bioinf.wehi.edu.au/limma> (home page)
<http://www.bioconductor.org/packages/2.2/bioc/html/limma.html>
(link to userguide)

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, pages 397- 420

6

References

Good collection of references

<http://www.statsci.org/micrarra/refs/index.html>

This is a good overview (old, but good!)

Smyth, G. K., Yang, Y.-H., Speed, T. P. (2003). Statistical issues in microarray data analysis. In: *Functional Genomics: Methods and Protocols*, M. J. Brownstein and A. B. Khodursky (eds.), Methods in Molecular Biology Volume 224, Humana Press, Totowa, NJ, pages 111-136. (PDF)

Bioconductor Software packages list

<http://www.bioconductor.org/packages/2.2/Software.html>

links to packages according to whether they are for: Microarray, Annotation, Visualisation, Statistics, GraphsAndNetworks, Technology, Infrastructure

There are also Annotation packages and ExperimentData packages

<http://www.bioconductor.org/packages/2.2/BiocViews.html>

7

Other key packages

- Microarray data
 - limma, marray, aroma.light....
- Affymetrix data
 - affy, oligo, simpleaffy, affyPLM, gcrma, affyio....
- aCGH data
 - snapCGH, aCGH, DNACopy, GLAD....
- High density arrays (Nimblegen)
 - tilingArray, oligo,...
- Illumina beadarray data
 - beadarray, lumi, beadarraySNP, BeadExplorer...

8

snapCGH

- Designed to be compatible with the popular R package *limma*
- This allows smooth progression between pre-processing, normalization and the segmentation steps
- Potentially reduces learning curve for anyone already familiar with *limma* e.g. biologists

9