GS01 0163
Analysis of Microarray Data

Keith Baggerly and Kevin Coombes
Section of Bioinformatics
Department of Biostatistics and Applied Mathematics
UT M. D. Anderson Cancer Center
kabagg@mdanderson.org
kcoombes@mdanderson.org

29 August 2006

Lecture 0: Outline and Details

Welcome!

- Why are we here?
- Who are we?
- What are we going to tell you about, and when?
- What do we assume you're familiar with?
- Where and when can you find ???
- What are we going to ask of you?

So, why are we here?

We want to learn about microarrays.

Microarrays have caught on as high-throughput assays for understanding molecular biology – they let us measure expression levels for thousands of genes in a single sample all at once.

This means new biology, and new data analysis.

We want to

- Understand how microarrays work and how they are analyzed.
- Perform some basic analyses of microarrays.

We're going to dive in here; we hope to have you analyzing data this week or next.

Who are we?

Keith Baggerly, HMB 13.336, kabagg@mdanderson.org, (713) 563-4290

Kevin Coombes, HMB 13.402, kcoombes@mdanderson.org (713) 794-4154

We're faculty in the section of Bioinformatics at MD Anderson, in the Dept of Biostatistics and Applied Mathematics. We've been working with microarray data for about 6 years, we've done pretty well with it, and we're going to trade off in telling you about it.

What are we going to tell you about, and when?

Week 1: How microarrays work, and Affymetrix file structures. The basics of dChip.

Week 2: Using dChip. Connecting numbers to biology – the basics of annotation and databases.

Week 3: Reviewing initial analyses. Introduction to R.

Week 4: Sample analysis using R/Bioconductor for Affymetrix.

Week 5: Reviewing R basics, quantification and normalization of Affymetrix data, Latin Square studies.

Week 6: Measures of differential expression (DE); R and DE.

What are we going to tell you about, and when?

Week 7: Reviewing DE assessments, introducing glass arrays.

Week 8: Case Study in R, start to finish

Weeks 9-10: Glass arrays. R and glass arrays. Putting it together for glass.

Week 11: Clustering microarray data, diagnostics, validation.

Week 12: Classification.

Week 13: Experimental design, proteomics.

Week 14: Meta-analysis

Week 15: Pathways, other arrays, open questions.

What do we assume you're familiar with?

Some biology.

Some programming.

Some statistics.

Computing

Microarray datasets are large. One of our initial data sets will involve 100 Affy gene chips; the files take up 1.5G of disk space when uncompressed. Another dataset takes 4G.

We will be using lots of freeware, including dChip (Windows-based), R, and the Bioconductor repository.

Textbooks

Required: Gentleman R, et al (eds). *Bioinformatics and Computational Biology Solutions Using R and Bioconductor.* Springer-Verlag, New York, 2005.

Optional: Simon RM, et al. *Design and Analysis of DNA Microarray Investigations.* Springer-Verlag, New York, 2003.

Optional: Dalgaard P. *Introductory Statistics with R.* Springer-Verlag, New York, 2002.

Optional: Speed T (ed). *Statistical Analysis of Gene Expression Microarray Data.* Chapman and Hall, New York, 2003.

Optional: Parmigiani G, et al (eds). *The Analysis of Gene Expression Data.* Springer-Verlag, New York, 2003.

Where can you find ???

- Course web site
  http://bioinformatics.mdanderson.org/MicroarrayCourse
- Office hours: T, Th, 4:30-5:30, HMB
  - Dr. Baggerly: HMB 13.336
  - Dr. Coombes: HMB 13.402

What are we going to ask of you?

Grading: Homeworks (roughly every two weeks)

1. Homeworks can be worked on jointly.
2. Assignments submitted electronically, in a single file (zipped is fine), by midnight on the assigned due date.
3. Submissions should include figures, R code, and text.