

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

9 December 2004

Lecture 26: Open Questions

- Tissue Arrays
- CGH Arrays
- SNP chips
- protein arrays
- data integration
- etc

General Themes

How is the data measured?

Processed?

Normalized?

Stored? (Have we kept track of things in an organized fashion?)

General Themes

Multiple Testing

Cross-Validation

Clustering

Experimental Design

Tissue Arrays

Not really high-throughput in the same sense as the others; instead of making hundreds of measurements on a single sample, we make a single measurement on hundreds of samples.

How do we define this measurement?

Integrated Optical Density

Is this the right measurement to make?

CGH Arrays

DNA, not RNA

Still worried about normalization

Scale of changes often much smaller on a log scale than with expression arrays

New feature: sequence/position information

How do we check for contiguous blocks?

What type of distance should we use?

SNP Chips

Again, DNA

Affy here

How do we define a probeset?

How do we combine probes?

Splice Variant Chips

Affy again

How do we define a gene?

Is this the natural measure a biological unit? This wording is vague. How should we sharpen the definition to make clear what we're looking for?

Protein Arrays

Smaller number of proteins being measured

Can we do absolute quantification?

Multiple spottings and dilution series – how can this be exploited?

Mass Spectrometry

Hundreds of proteins, not thousands of genes

How do we calibrate the data?

How do we quantify the data?

How do we design the studies?

Data Integration

How can we combine expression data with DNA data?

How can we combine expression data with protein data?

Lung cancer and FUS1

A Case Study

Comparing cancers and tissues

Different subtypes of ovarian cancer

Thought to be associated with different surrounding normal tissues

How can we test this?

What's hard about this problem?

Some Sample Sizes

How many samples do you need?

What type of an effect are you looking for?

How big do you expect it to be?