Bioinformatics
Stat 697K / CS 691K / Microbio 690K
Microarray Analysis
Erin M. Conlon
Department of Mathematics and Statistics

Microarray Technology
• Microarrays (genechips) are small glass slides about ½ inch square
• Hold information for 10,000 genes at one time
• Tell us which genes are active (expressed) in an experiment

Statistics and Microarrays
• Large data sets require large number of statistical hypothesis tests
• Require other statistical methods of analysis
• Data, i.e.
  – 10,000 genes on one slide
  – 20 slides
Microarrays and UMass

- Cost used to be limiting:
  - 2000: $1000 per slide
  - 2006: $200-300 per slide (or less)
- On campus, we have many groups producing microarray data
- More data than statisticians
- Goal of this course:
  - Prepare students to work with microarray studies

9 Microarray Classes

- 1 homework set to analyze microarray data
  - Use statistical language R (freeware)
- 2 classes will be tutorials using R
  - 1 in lecture, 1 in computer lab
- Sample programs will be given for homework
- Introductory statistics will be covered in lecture
  - Means, medians
  - Histograms, boxplots
  - Hypothesis testing

Lecture Schedule

- Introduction to microarray technology
  - How they are produced
- Pre-processing data, image quality
- Hypothesis testing
  - Determine genes different between genechips
- Clustering
  - Genes or samples (patients)
  - Common patterns of expression