Statistical methods for the analysis of microarray data

Kellie J. Archer
Sanger Hall B1-069, Department of Biostatistics
1101 East Marshall Street, Richmond, VA 23298
Tel: 804-827-2039, Fax: 804-828-8900, e-mail: kjarcher@vcu.edu
website: www.people.vcu.edu/~kjarcher

Dr. Archer's research pertains to the development of innovative statistical methods and software for the analysis of microarray data. Due to the high cost of microarray experiments, investigators typically select designs with biological rather than technical replicates. Therefore, it is essential that the quality of RNA hybridized to the microarray meets certain standards. The process of transcription begins with reverse transcriptase binding at the 3' end of a gene and continuing toward the 5' end. However, transcription generally does not continue to completion. That is, reverse transcription typically drops off before reaching the 5' end. Affymetrix GeneChips includes probe sets which interrogate both the 3' end and the 5' end for selected control genes to assess quality of transcription. The MAS 5.0 software estimates the 3':5' ratio after the PM and MM probes have been summarized into a probe set expression measure. Unfortunately, inherent to all probe set expression summary methods is that the 3' and 5' probe sets of interest are only represented on the GeneChip once. This leads to the unfortunate consequence of inadequate replications for variance estimation. The methodology proposed uses the pixel level intensities to increase the number of observations per probe set in order to obtain an estimate of the 3' to 5' ratio and an associated confidence interval. Since there is an inherent hierarchical structure to GeneChip data, where pixels are nested within probes and probes are nested within probe sets, the proposed method is to assess RNA degradation by fitting mixed effects ANOVA models to estimate the 3':5' ratio treating probe set as a fixed effect, while treating PM level data as random effects and treating pixel level data as a subsample nested within PM probe. In bioinformatics, algorithms are generally viewed as more important than models or statistical efficiency. Therefore, this BBSI project will focus on software development for making the proposed method available to the research community.

Other research interests: At VCU, gene expression is being quantified using the Affymetrix Genechip® technology, which is characterized by the ability to simultaneously study the expression of thousands of genes. Other areas of research include (i) development of novel methods for performing image analysis, normalization, and expression summaries for oligonucleotide microarray data; (ii) development of unsupervised and supervised learning methods for microarray data analysis; and (iii) development of software libraries for preprocessing steps and for microarray data analysis. In addition, possibilities exist to develop new methods for analyzing loss of heterozygosity data.