Research: Statistical methods for the analysis of microarray data

Kellie J. Archer

Department of Biostatistics

Sanger Hall B1-069, 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

The technology for hybridizing archived tissue specimens and the use of laser capture microdissection for selecting cell populations for RNA extraction has increased over the past few years. Both of these methods contribute to RNA degradation. Therefore, quality assessments of RNA hybridized to microarrays is becoming increasingly more important. Existing methods for estimating the quality of RNA hybridized to a GeneChip, from resulting microarray data, suffer from subjectivity and lack of estimates of variability.

Specifically, 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 GCOS 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.

We have proposed a method for assessing RNA quality for a hybridized array which overcomes this drawback. Since there is an inherent hierarchical structure to GeneChip data, where pixels are nested within probes and probes are nested within probe sets, and multiple probe sets may interrogate the same transcript, the proposed method is to assess RNA degradation by fitting mixed effects ANOVA models to estimate the 3':5' ratio considering end (3' or 5') as a fixed effect, while treating probe set and PM level data as random effects, and pixel level data as a subsample nested within PM probe. The effectiveness of the proposed method has been demonstrated by application of the methods (proposed and existing) to two microarray datasets for which external verification of RNA quality is known. 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.

Dr. Archer's research pertains to the development of innovative statistical methods and software for the analysis of microarray data. Other specific research interests include (i) comparison of intensity extraction methods for oligonucleotide microarray data; (ii) enhancement of the random forest supervised learning methods for microarray data analysis; and (iii) development of software libraries for preprocessing steps and for microarray data analysis.