

Modeling and Analysis of SAGE Cerebellum Libraries

Shili Lin

Department of Statistics and the Mathematical Biosciences Institute, The Ohio State University, Columbus, USA.

A Serial Analysis of Gene Expression (SAGE) library is a collection of thousands of small DNA “tags”, each of which represents a distinct mRNA transcript. Existing methods have been proposed for analyzing single library data (i.e., one library per group) or one tag at a time. The practice of lumping all libraries together (in a multi-library setting) to form a “mega” library for each group is obviously unsatisfactory, but nonetheless performed frequently due to the lack of alternative methods. Since the tag counts within each library are inter-related as they are drawn from a multinomial distribution, analyzing thousands of tags one at a time is undoubtedly inadequate. Not only does such a practice ignore the dependency, but it also faces the multiple testing adjustment issue. This talk describes an attempt to address both of these issues so that all tags from multi-library groups can be analyzed together. The method proposed also gears toward multi-group data. Focusing on the problem of identifying genes that are differentially expressed, a Bayesian formulation is established. Under this formulation, the problem of separating the differentially expressed genes from the majority of similarly expressed ones is treated as a model selection problem, and the reversible jump Markov chain Monte Carlo method is adapted for this purpose. The method is applied to a set of mouse libraries to uncover genes that are associated with the process of aging in the cerebellum. The Gene Ontology (GO) analysis of the genes selected classifies them into several GO categories, which appear to be functionally relevant to aging.

[Shili Lin, Department of Statistics and the Mathematical Biosciences Institute, The Ohio State University, Columbus, OH 43210, USA.; shili@stat.ohio-state.edu]