International Conference on Molecular Systems Biology 2009

Statistical Analyses of High Dimensional Molecular Data (HDMD)

William R. Atchley

Department of Genetics
North Carolina State University
Raleigh, North Carolina, USA 27695-7614

High throughput technologies, such as microarrays and genome sequencing, are producing massive amounts of molecular data for analysis and interpretation. Unfortunately, this huge outpouring of data has quickly outpaced the ability of current statistical methods to derive biological insights from all these data. The multivariate statistical methods and models which have been used to summarize such data often do not take into consideration fundamental aspects of the processes that underlie the biological variation in these data.

This talk will explore some of the complications presented by HDMD and the resulting ramifications for meaningful statistical and biological analyses. A series of relevant biological questions will be posed that may help guide analyses of HDMD. Relevant mathematical, statistical, and biological models will be briefly explored for certain types of molecular data to facilitate selection of appropriate statistical methods. The utility of principal components and latent vector analyses for molecular data will be described and recommendations made with regard to biological analyses.