

# Novel Multivariate Statistical Methods in Genomic Analysis

Nicholas Schork, Ph.D.  
University of California, San Diego

## **Abstract:**

The collection of large data sets resulting from, e.g, high-throughput genotyping studies, multi-phenotype clinical studies, and gene expression studies are now commonplace. Although standard genetic analysis tools can be used to identify important features and patterns in such data sets, new tools must be developed in order overcome shortcomings in traditional methodologies. We describe three methodological areas that show promise in this regard. The first involves simultaneous assessments and tests of differences in means and covariance matrices across groups as such genotypic categories or interventions. The second considers the use of Random Matrix Theory to test hypothesis about the structure of large data sets. The third takes advantage of smoothing techniques for time-dependent data to probe questions about the timing and significance of important biological events (such as the expression of a gene). Each of these techniques needs further refinement and development but represent extremely fertile and rich research areas for statistical geneticists.