MULTIVARIATE ANALYSIS METHODS IN SYSTEMS BIOLOGY

E. Kurali[†], L. Zhu, A. Anderson, D. Rajagopalan,
A. Bhattacharyya, K. Lee

GlaxoSmithKline King of Prussia, PA, USA

[†] E-mail: *Edit.2.Kurali@gsk.com*

Systems biology presents an alternative paradigm to the reductionist approach of traditional molecular biology. Instead of focusing on the components of the system 'one protein at a time', it seeks to understand how biological systems function through the integration of multiple levels of biological information. Advances in high-throughput platform technologies have made this 'global' view accessible for the first time. Analysing and interpreting such complex datasets integrated across mulitple molecular platforms presents many challenges and requires novel analysis methods. The challenges include data pre-processing, normalization, integration, highdimensionality, multi-collinearity, multiple testing, predictive modelling,and model validation. Multivariate analysis methods such as principal component analysis and partial least squares discriminant analysis were applied to integrate multiple platform data (transcriptomics, proteomics, lipidomics, mass spectometry and NMR spectroscopy) from a clinical diabetes study. Issues and solutions related to proper data normalization, data cleaning, handling of nuisance factors and covariates, predictive modeling, internal and external model validation, and biomarker discovery are presented. In conclusion, multivariate analysis approaches provide a natural paradigm for analyzing and interpreting systems biology data.