KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model

Junli Du[a], Zhifa Yuan[*,a], Ziwei Ma[a,b]

[a] College of Science, Northwest A&F University, Yangling 712100, P. R. China
[b] Department of Mathematical Sciences, New Mexico State University, Las Cruces,
New Mexico,88001, U.S.A.
E-mail: liml75@126.com (J. Du) , ziweima@nmsu.edu (Z. Ma)

Abstract: Overrepresentation analysis (ORA) is a widely used and commonly acceptable method to identity the most revelant pathways for functional analysis of genomic data in last few decases. The dynamic impact approach (DIA) represents an alternative to overrepresentation analysis (ORA) for functional analysis of time-course experiments or those involving multiple treatments since it can deal with time-course experiments or multiple treatments under different experiments conditions which ORA fails to analysis. Though DIA can estimate the biological impact of the differentially expressed genes (DEGs) associated with particular biological functions, but it does not take into account the correlated dependence structure of these DEGs. So path analysis model (KEGG-PATH) is develpoed to explor the correlated dependence structure of the KEGG pathway hierarchy in this paper. In this model, based on the impact estimated by DIA, we susdivide the total effect of each KEGG pathway into the direct effect and indirect effect by taking account not only each KEGG pathway itself, but also its related pathways. Also, the impact direction of each KEGG pathway is estimated by a gradient analysis from principle componet analysis (PCA). At last, the advantage of this model is demonstrated through the functional analysis of the bovine mammary transcriptome during lactation.