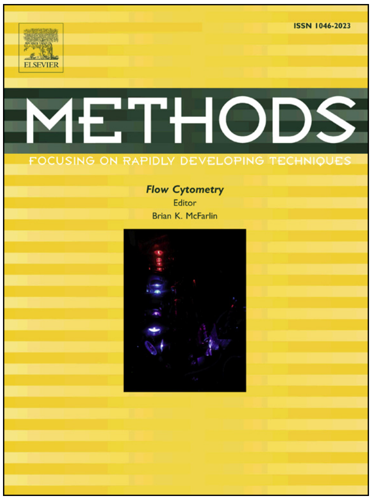

A Route-Based Pathway Analysis Framework Integrating Mutation Information and Gene Expression Data

Yue Zhao, Tham H. Hoang, Pujan Joshi, Seung-Hyun Hong, Charles Giardina, Dong-Guk Shin

PII: S1046-2023(17)30039-7
DOI: http://dx.doi.org/10.1016/j.ymeth.2017.06.016
Reference: YMETH 4246

To appear in: *Methods*

Please cite this article as: Y. Zhao, T.H. Hoang, P. Joshi, S-H. Hong, C. Giardina, D-G. Shin, A Route-Based Pathway Analysis Framework Integrating Mutation Information and Gene Expression Data, *Methods* (2017), doi: http://dx.doi.org/10.1016/j.ymeth.2017.06.016

A Route-Based Pathway Analysis Framework Integrating Mutation Information and Gene Expression Data

Yue Zhao[a,*], Tham H. Hoang[a], Pujan Joshi[a], Seung-Hyun Hong[a], Charles Giardina[b], Dong-Guk Shin[a]

[a] *Computer Science and Engineering Department, University of Connecticut, 371 Fairfield Way, Unit 4155, Storrs, CT 06269*
[b] *Department of Molecular and Cell Biology, University of Connecticut, 91 North Eagleville Road, Unit 3125, Storrs, CT 06269*

Abstract

We propose a new way of analyzing biological pathways in which the analysis combines both transcriptome data and mutation information and uses the outcome to identify routes of aberrant pathways potentially responsible for the etiology of disease. Each pathway route is encoded as a Bayesian Network which is initialized with a sequence of conditional probabilities which are designed to encode directionality of regulatory relationships encoded in the pathways, i.e. activation and inhibition relationships. First, we demonstrate the effectiveness of our model through simulation in which the model was able to easily separate Test samples from Control samples using fictitiously perturbed pathway routes. Second, we apply our model to analyze the Breast Cancer data set, available from TCGA, against many cancer pathways available from KEGG and rank the significance of identified pathways. The outcome is consistent with what have already been reported in the literature. Third, survival analysis has been carried out on the same data set by using pathway routes as features. Overall, we envision that our model of using pathway routes for analysis can further refine the conventional ways of subtyping cancer patients as it can discover additional characteristics specific to individuals tumor.

---

[*] Corresponding author
Email address: `yue.2.zhao@uconn.edu` (Yue Zhao)