

A novel approach to identify highly connected and differentially expressed gene subnetworks in metastasizing endometrial cancer

K. Kusonmano^{1,2,3*}, M.K. Halle^{2,4}, E. Wik^{5,6}, E.A. Hoivik^{2,4}, C. Krakstad^{2,4}, K.K. Mauland^{2,4}, I.L. Tangen^{2,4}, A. Berg^{2,4}, H.M.J. Werner^{2,4}, J. Trovik^{2,4}, A.M. Øyan^{4,7}, K.-H. Kalland^{4,7}, I. Jonassen^{1,8}, H.B. Salvesen^{2,4}, K. Petersen¹

¹ Computational Biology Unit, Department of Informatics, University of Bergen, Bergen, Norway

² Department of Obstetrics and Gynecology, Haukeland University Hospital, Bergen, Norway

³ Bioinformatics and Systems Biology Program, School of Bioresources and Technology, King Mongkut's University of Technology Thonburi, Bangkok, Thailand

⁴ Center for Cancer Biomarkers, Department of Clinical Science, University of Bergen, Bergen, Norway

⁵ Center for Cancer Biomarkers, Department of Clinical Medicine, University of Bergen, Bergen, Norway

⁶ Department of Pathology, The Gade Institute, Haukeland University Hospital, Bergen, Norway

⁷ Department of Microbiology, Haukeland University Hospital, Bergen, Norway

⁸ Center for Cancer Biomarkers, Department of Informatics, University of Bergen, Bergen, Norway

* e-mail: kanthida.kus@mail.kmutt.ac.th

Key words: differential expression analysis, gene set analysis, biological network analysis, differentially expressed subnetworks, metastasis, endometrial cancer

Motivation and Aim: Differential expression analyses is a common approach to study molecular changes between two phenotypes of interest, and gene set and network analysis complementary identifying subset of associated genes for functional study and interpretation. The standard application of gene set analysis evaluates enrichment of differentially expressed genes in publicly available gene sets, while the network analysis identifies likely functional gene modules with their interactions.

Methods and Algorithms: We propose a novel approach combining gene set and network analyses to identify phenotypically relevant highly connected and differentially expressed gene subnetworks for better understanding biological mechanisms involved. The method allows data integration between available biologic knowledge of *a priori* gene-gene relations and assessment of internal correlations of expression data. The proposed pipeline includes permutation tests to assess the significance of the identified subnetworks being differentially expressed.

Results: The pipeline was applied to study genes differentially expressed between primary tumors and metastatic lesions in endometrial carcinomas, and made freely available as an R implementation. The identified top ranked subnetworks are associated significantly with disease aggressiveness as well as patient survival, and include genes related to cell proliferation and epithelial-mesenchymal transition.

Conclusions: We have developed a workflow for differential expression analysis providing gene-gene relations inside detected subgroup of related to the studied phenotype. We integrated of known protein-protein interactions and gene expression correlations of aggressive primary tumors and metastatic lesions. The identified subnetworks links to biological processes relevant for the metastasizing process, and also significantly associates with patient survival. We suggest further functional studies of their biological relevance for development of metastatic disease in endometrial cancer.