A novel algorithm to identification of breast cancer biomarkers

Masoome Oveisi [a], Sama Goliaei [a*], Zahra Razaghi-Moghadam [a],

[a] Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran

*sgoliaei@ut.ac.ir

Abstract: The breast cancer, caused by abnormal growth of abnormal cells in the breast tissue, is the most common type of cancer in women and the second most common cancer in the world [1]. With recent advances in genetic studies, genomic technologies provide a comprehensive understanding of cancer and tumors at the molecular level. Several achievements have shown that these information and technologies can be used in early diagnosis of cancer and therapeutic goals. One of the main factors in early detection of cancer is biomarker that can be useful in predicting an effective treatment [2]. The cancer is a complex disease and its mechanism remains largely uncertain, thus, biomarkers play an important role in detecting, evaluating, and directing cancer treatment. In recent years, several researches have been conducted to discover biomarkers of breast cancer. In general, identified biomarkers can be divided into four categories of single gene biomarkers, gene set biomarkers, pathway biomarkers, and network biomarkers [3]. In this paper, we presented a novel algorithm to identify subnetwork markers to building a prediction model for breast cancer. The idea of this algorithm is to obtain subnetwork markers by identifying differentially expressed genes (DE genes) in the gene interaction network corresponding to the protein-protein interaction network, obtaining linear paths between these genes, connecting the best paths inspired by the Steiner tree algorithm and then clustering the tree. The proposed algorithm is applied on a gene expression profile and two separate protein interactions networks. The gene expression data is collected from 12 different data sets. All of these datasets are published by GEO. The data contains 616 samples, including 220 healthy samples and 792 samples of breast cancer patients. Also, two protein-protein interaction networks called BioGRID and TissueNet are used. The results of the evaluation indicate the high accuracy of the markers obtained from both networks in predicting new samples. Also, the gene ontology analysis of the markers shows that these subnetworks play an important role in the biological processes associated with breast cancer.

Keywords: Cancer; Breast cancer; Breast cancer biomarker; Subnetwork marker; Differentially expressed gene; Cancer prediction model

References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2015," *CA: a cancer journal for clinicians,* vol. 65, no. 1, pp. 5-29, 2015.

[2] W. Awada, T. M. Khoshgoftaar, D. Dittman, R. Wald, and A. Napolitano, "A review of the stability of feature selection techniques for bioinformatics data," in *Information Reuse and Integration (IRI), 2012 IEEE 13th International Conference on*, 2012, pp. 356-363: IEEE.

[3] T. Zeng, S. y. Sun, Y. Wang, H. Zhu, and L. Chen, "Network biomarkers reveal dysfunctional gene regulations during disease progression," *The FEBS journal,* vol. 280, no. 22, pp. 5682-5695, 2013.