# Chapter 83
# Cancer Pathway Network Analysis Using Cellular Automata

**Kalyan Mahata**
*Government College of Engineering and Leather Technology, India*

**Anasua Sarkar**
*Government College of Engineering and Leather Technology, India*

## ABSTRACT

*Identification of cancer pathways is the central goal in the cancer gene expression data analysis. Data mining refers to the process analyzing huge data in order to find useful pattern. Data classification is the process of identifying common properties among a set of objects and grouping them into different classes. A cellular automaton is a discrete, dynamical system with simple uniformly interconnected cells. Cellular automata are used in data mining for reasons such as all decisions are made locally depend on the state of the cell and the states of neighboring cells. A high-speed, low-cost pattern-classifier, built around a sparse network referred to as cellular automata (ca) is implemented. Lif-stimulated gene regulatory network involved in breast cancer has been simulated using cellular automata to obtain biomarker genes. Our model outputs the desired genes among inputs with highest priority, which are analysed for their functional involvement in relevant oncological functional enrichment analysis. This approach is a novel one to discover cancer biomarkers in cellular spaces.*

## INTRODUCTION

For a given type of cancer, there are often stimulating factors that produces different patterns of gene expression in patient data. Analyzing them, it reveals the discovery of gene networks and regulatory pathways involved in those tumor formations (Sarkar, 2013). In this respect, the gene expression profiles in breast cancer has been analysed extensively recently. This reveals the identification of breast cancer molecular subtypes and the development of prognostic and predictive gene signatures, resulting in an improved knowledge on the heterogeneity of breast cancer and its biomarkers (Perou et al., 2000; Sorlie et al., 2001; Sorlie, 2003; Sotiriou, 2009).

DOI: 10.4018/978-1-5225-8903-7.ch083

In pioneering work, (Perou et al., 2000) classified the expression of approximately 8,000 genes in samples from 42 breast cancer patients into specific 'intrinsic subtypes' of primary breast carcinomas. Later, Sorlie et al. (Sorlie et al., 2001) experiments for correlations between gene expression patterns and clinically relevant parameters to detect the breast cancer subtypes. They reveal the prognostic markers with respect to overall and relapse-free survival among a subset of patients who had received uniform therapy (Sorlie et al., 2001; Sorlie, 2003).

Sotiriou and Pusztai (Sotiriou, 2009) define that there are three approaches for analyzing the global gene expression profiling to obtain genomic signatures. In 'top-down' approach, tumor or cell line gene expressions are correlated with the clinical outcome of patients to identify prognostic gene signatures. In 'bottom-up' approach, a gene expression signature is defined to be a prognostic predictor which is related to a particular biological pathway or process. Third, among the list of candidate genes, some biomarkers are prospectively selected based on previous biological knowledge like recurrence score signature (Sotiriou, 2009).

In complex disease states such as breast cancer, genes may overlap in their involvement of relevant pathways and networks and may influence the disease significantly. Therefore, the shared neighbourhood in gene expression networks exhibits the significance of one gene in the network. In a rank-based network analysis (Corban et al., 2010), they construct the networks of similarity among gene expression profiles based on shared neighborhood. The hierarchical agglomerative clustering is used to group nodes in clusters by computing a rank on the log odds ratio based on Poisson parameter over shared neighbors. The nodes with highest ranking on degrees and neighbourhoods with cross-connectivity, which belong to some dense bipartite networks, are selected (Corban et al., 2010). Implicit notion of applying such analysis to biologically important tumor gene expression profiles, is to rank genes according to their correlated expression patterns in gene expression networks of cancer cell line.

We present an integrated approach to detect cancer biomarkers using cellular automata, which we experiment over the gene expression profiles in a data set (Icardi et al., 2012) of 6 samples of leukemia inhibitory factor effect on breast cancer cell line. We analyse the LIF-responsive genes in MCF7 cells. We develop the correlation expression network over these genes using the shared neighborhood ranking score method (Corban et al., 2010). Therefore, we choose significant genes based on their ranking values to simulate the cellular automata approach over their ranking score matrices in both control and LIF-stimulated expression stages. The 2-dimensional analysis in cellular automata over these two stages, chooses finally the most significant 4 biomarkers based on our newly developed algorithm to traverse gene networks in cellular space. We further analyse the extended biological KEGG pathways associated with those selected biomarkers using DAVID (http://david.abcc.ncifcrf.gov/) (Huang, 2009). The detailed analysis of the GO annotations of those identified biomarkers also exhibits how the priority in shared neighbourhood is reflected in the priority in biological functions to detect biomarker oncogenes.

## LITERATURE REVIEW

Cellular Automata is a powerful tool to model the complex dynamic system. To bridge between bioinformatics with artificial intelligence, CA is played a vital role.

Protein synthesis is done by CA nowadays which shows extraordinary results. Previously, cellular automata has been applied to predict protein attributes using pseudo amino acid composition (Xiao et al., 2011).To implement it, the pseudo amino acid composition (PseAAC) was developed and stimulated