# Chapter 40
# Hybrid High-Performance Computing Algorithm for Gene Regulatory Network

**Dina Elsayad**
*Faculty of Computer and Information Sciences, Ain Shams University, Egypt*

**Safawat Hamad**
*Ain Shams University, Egypt*

**Howida Abd-Alfatah Shedeed**
*Faculty of Computer and Information Sciences, Ain Shams University, Egypt*

**Mohamed Fahmy Tolba**
*Faculty of Computer and Information Sciences, Ain Shams University, Egypt*

## ABSTRACT

*This paper presents a parallel algorithm for gene regulatory network construction, hereby referred to as H2pcGRN. The construction of gene regulatory network is a vital methodology for investigating the genes interactions' topological order, annotating the genes functionality and demonstrating the regulatory process. One of the approaches for gene regulatory network construction techniques is based on the component analysis method. The main drawbacks of component analysis-based algorithms are its intensive computations that consume time. Despite these drawbacks, this approach is widely applied to infer the regulatory network. Therefore, introducing parallel techniques is indispensable for gene regulatory network inference algorithms. H2pcGRN is a hybrid high performance-computing algorithm for gene regulatory network inference. The proposed algorithm is based on both the hybrid parallelism architecture and the generalized cannon's algorithm. A variety of gene datasets is used for H2pcGRN assessment and evaluation. The experimental results indicated that H2pcGRN achieved super-linear speedup, where its computational speedup reached 570 on 256 processing nodes.*

DOI: 10.4018/979-8-3693-3026-5.ch040

## INTRODUCTION

One of the bioinformatics related research areas is the gene data analysis. Bioinformatics (Isea, 2015) can be defined as the integration of different sciences, that includes information engineering, mathematics, computer science, statistics, and biology. In fact, bioinformatics can be considered as a marriage of sciences. Providing comprehensive insights into the different biological processes is the main goal of Bioinformatics (Nair, 2007). Furthermore, bioinformatics focuses on the management, processing, analysis, and interpretation of the biological data (Kumar et al., 2019; Mahmoud & Muhi,2020; Guffroy et al., 2017; Trauth & Browning, 2018).

The analysis of genetic data is one of the bioinformatics research areas that focuses on studying the functions and behavior of the genes (Velculescu et al., 1995). One of the techniques for collecting gene data is microarrays (Muller & Nicolau, 2005). Microarrays are a microscope surface with pre-determined allocated tiny spots where each spot contains a well-known biological substance. Where the type of used biological material determines the type of microarrays. There are many types of microarrays such as DNA microarrays (Marzancola et al., 2016), MMChips (Chen, et al., 2011), Protein microarrays (Sutandy et al., 2013), Peptide microarrays (Zandian, et al., 2017), Tissue microarrays (Jawhar, 2009), Cellular microarrays (Bhatia & Flaim, 2014), Chemical compound microarrays (Ma & Horiuchi, 2006), Antibody microarrays (Kusnezow et al., 2003), Carbohydrate arrays (Horlacher & Seeberger, 2008), Phenotype microarrays (Bochner et al., 2001), and IRIS (Avci et al., 2015). Moreover, DNA microarrays (Marzancola et al., 2016) are furtherly classified into four types that are cDNA microarrays (Xiang & Chen, 2000; Nishimura et al., 2015; Churchill, 2002), oligo DNA microarrays (Mrowka et al., 2002), BAC microarrays (Cai et al., 2002), and SNP microarrays (Meaburn et al., 2006).

DNA microarrays (Marzancola et al., 2016) is a useful technique to measure the expression level of a huge number of genes simultaneously, where the expression level of genes is directly related to the amount of gene mRNA and where more mRNA amount means a high gene expression level (Schena et al., 1995). DNA microarrays play a vital role in disease-associated genes identification and analysis of gene expression. The analysis of gene data is a challenging process because of the huge data size and the many computational tasks of the analysis process (Yang, et al., 2008).

The gene data analysis computational tasks include subset extraction, search, clustering and gene regulatory network construction (Aluru, 2006), where the gene regulatory network construction techniques focus is demonstrating the genes' interactions topological sequence, as well as annotating the gene functionality (Segal, et al., 2003). Furthermore, Gene regulatory network inference techniques contribute to a comprehensive understanding of the cell signaling and regulatory process. The gene regulatory network is visualized as a graph. In other words, the gene regulatory network is represented by a set of nodes and a set of edges, where nodes represent genes while interactions among these genes are represented by the edges. The ideal gene regulatory network topology contains a few intensively connected genes (named hub genes) in addition to many lowly connected genes (Barabasi & Oltvai, 2004). The identification of these hub genes plays a vital role in clinical applications (Tang, et al., 2013). Despite the development of different computational techniques for the gene regulatory inference, the common limitations of these techniques are the intensive computation and time-consumption. As a result, parallel techniques are required to overcome these limitations. This paper presents a parallel algorithm for gene regulatory inference. This proposed algorithm will be named H2pcGRN (hybrid high performance-computing algorithm for gene regulatory network). This algorithm is a step towards completing our proposed parallel framework for gene regulatory network construction PAGeneRN (Parallel