A Novel Cancer Classifier based on Differentially Expressed Gene Network

Jaegyoon Ahn
Department of Computer Science
Yonsei University
3rd Engineering Bldg. 533-1, Shinchon-dong, Seodaemun-gu, Seoul, Korea
00822-2123-7757
ajk@cs.yonsei.ac.kr

Youngmi Yoon
Information Technology Department
Gachon Univ. of Medicine & Science
1108 Gachon-Kwan, Yonsu-dong, Yonsu-gu, Incheon, Korea
008232-820-4393
ymyoon@gachon.ac.kr

Sanghyun Park
Department of Computer Science
Yonsei University
3rd Engineering Bldg. 520, Shinchon-dong, Seodaemun-gu, Seoul, Korea
00822-2123-5714
sanghyun@cs.yonsei.ac.kr

ABSTRACT
It is fundamental and essential to elucidate how cancer-related genes interact with each other. In this study, we build two undirected graphs: one is a graph consisting of edges only observed in tumor samples, and the other is a graph consisting of edges only observed in normal samples. We apply a genetic algorithm for searching sub-networks of these genetic networks. Those gene sub-networks identify new cancer-related genes that might be related with previously known cancer-related genes, and also show a higher accuracy in classifying tumor and normal samples than the current methods.

Categories and Subject Descriptors
H.2.8 [**Database Management**]: Data mining; J.3 [**Life and Medical Sciences**]: Biology and Genetics

General Terms
Algorithms

Keywords
Cancer Classification, Microarray, Genetic algorithm

1. INTRODUCTION
It is important to identify cancer-related genes, and to develop cancer classification methods using microarray experiments. Especially, elucidating how cancer-related genes interact with each other is more fundamental and essential. Using the microarray data, implementing a cancer classifier in the form of gene sub-network gives a hint to understand the interaction of cancer-related genes.

Large numbers of cancer classification methods based on microarray data use various machine learning techniques which extract cancer-related genes by examining gene expression profiles which are differentially expressed in cancer tissues, classify a new sample with these genes. These showed that the machine learning methods are effectively applied to cancer classification [1-6]. These machine learning classification methods generally filter individual marker genes out, and use them collectively as a classifier without considering gene-gene interactions. We propose a novel method by adopting genetic network which is recently recognized as a model to describe a complex biological occurrences and diseases such as cancers. In this study we consider two kinds of undirected graphs: one is a graph consisting of edges only observed in tumor samples, and the other is a graph consisting of edges only observed in normal samples.

The search space for sub-networks that can differentiate tumor versus normal in the complicated and massive gene network is extremely large. This study applies a genetic algorithm for efficient search. Consequently, this study identifies a classifier with minimally 18 genes, and also exhibits a high accuracy rate when is applied to prostate cancer microarray data. Moreover, the resulting classifier includes new cancer-related genes that might be related with previously known cancer-related genes.

ACM-BCB 2010, Niagara Falls, NY, USA

2. ALGORITHM

2.1 Constructing the gene network
We construct the differentially expressed gene network which is built up with genes whose expression values show significant difference between tumor and normal samples. The definitions for the differentially expressed gene network are as follows.

Definition 1 ***Tumor and Normal edge*****:** Let $exp(A)$ be expression value of gene A. For genes A and B, edge (A, B) is defined as tumor edge, if $exp(A) > exp(B)$ on all the samples in tumor sample set T and on less than p% samples in normal sample set N. On the contrary, edge (A, B) is defined as normal edge, if $exp(A) > exp(B)$ on all the samples in N and on less than p% samples in T.

Definition 2 ***Differentially expressed gene network*****:** Differentially expressed gene network is defined as a