Computational Biology and Chemistry 68 (2017) 260–265

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Research Article

# Identifying dynamic pathway interactions based on clinical information

Shinuk Kim

*Department of Civil Engineering, Sangmyung University, Cheonan Chungnam 31066, Republic of Korea*

ARTICLE INFO

*Article history:*
Received 24 November 2016
Received in revised form 16 April 2017
Accepted 17 April 2017
Available online 24 April 2017

*Keywords:*
Dynamic pathway
Pathway interaction
Clinical information

ABSTRACT

In this paper, we introduce approaches for inferring dynamic pathway interactions by converting static datasets into dynamic datasets using patients' clinical information. One approach uses survival time–based dynamic datasets, and the other uses grade- and stage-based dynamic datasets. Based on cancer grades and stages, we generated six dynamic levels and obtained two pairs of significant pathways out of twelve enriched pathways. One pair of the pathways included CELL ADHESION MOLECULES CAMS and SYSTEMIC LUPUS ERYTHEMATOSUS (correlation coefficient = 1.00), in which CD28, CD86, HLA-DOA, and HLA-DOB were identified as common genes in the pathways. The other pair of the pathways included SPLICEOSOME and PRIMARY IMMUNODEFICIENCY (correlation coefficient = 0.94) with no common genes identified.



## 1. Introduction

Pathway-based analysis in cancer research has advanced rapidly in recent years. Pathway analysis can be used to discover more stable biomarkers and to calculate more accurate classifications than analysis of individual gene markers (Kim et al., 2012; Guo et al., 2005; Lee et al., 2008; Liu et al., 2015; Su et al., 2009). More importantly, pathway analyses yield functional insights into a wide range of biological phenomena (Visakh and Abdul Nazeer, 2016). Many gene set analysis tools have been developed for identification of meaningful pathways (Hung et al., 2010; Li et al., 2015). Gene set enrichment analysis (GSEA) is one of the most well-known and stable algorithms for pathway analysis (Subramanian et al., 2007). GSEA is a method to identify a group of genes that are over-represented in a large set of genes that may be associated with disease phenotypes. The method consists of three steps to identify significantly enriched groups of genes: (i) calculation of an enrichment scores by calculating ranked list genes, (ii) estimation of significance level of enrichment score by using an empirical nominal $p$ value, (iii) adjustment for multiple hypotheses to evaluate an entire database of gene sets.

The availability of many different types of gene expression datasets, including cancer patients' clinical information, in conjunction with advances in high-throughput biotechnology has resulted in the generation of unprecedented amounts of data. Using these data, it became possible to analyze systemic changes in cancer development. Despite the existence of numerous datasets, however, cost and ethical concerns have largely prevented dynamic measurements of gene expression levels for dynamic analysis. In this regard, mathematical modeling is a plausible alternative to dynamic analysis. Many mathematical models have been proposed to address problems in systems biology and to aid in deciphering the variations of heterogeneous diseases such as a cancer (Kim et al., 2007; Ay and Arnosti, 2011; Chen et al., 1999). For this purpose Ay (Ay and Arnosti, 2011) described three major types of models: thermodynamic models, differential-based models, and Boolean networks.

Tumors are classified according to numeric grades and stages, allowing doctors to quickly grasp the condition of a patient and assign an appropriate treatment. Caner grades are stratified from 1 (G1) to 4 (G4) based on the level of differentiation of tumor cell or tissue. On the other hand, cancer stages are described based on dissemination of disease as primary tumor (T), regional lymph nodes (N), or distant metastasis (M). The three stages combined with other information yield five stages: 0, I, II, III, and IV. In the stage description, higher numbers indicate larger tumors and amounts of affected tissue.

In this study, we used grade and stage information to stratify data as a substitute for information regarding time-dependent cancer growth. Then, we inferred dynamic pathway interaction based on clinical information, including cancer survival times and

*E-mail address:* kshinuk@smu.ac.kr (S. Kim).

http://dx.doi.org/10.1016/j.compbiolchem.2017.04.009

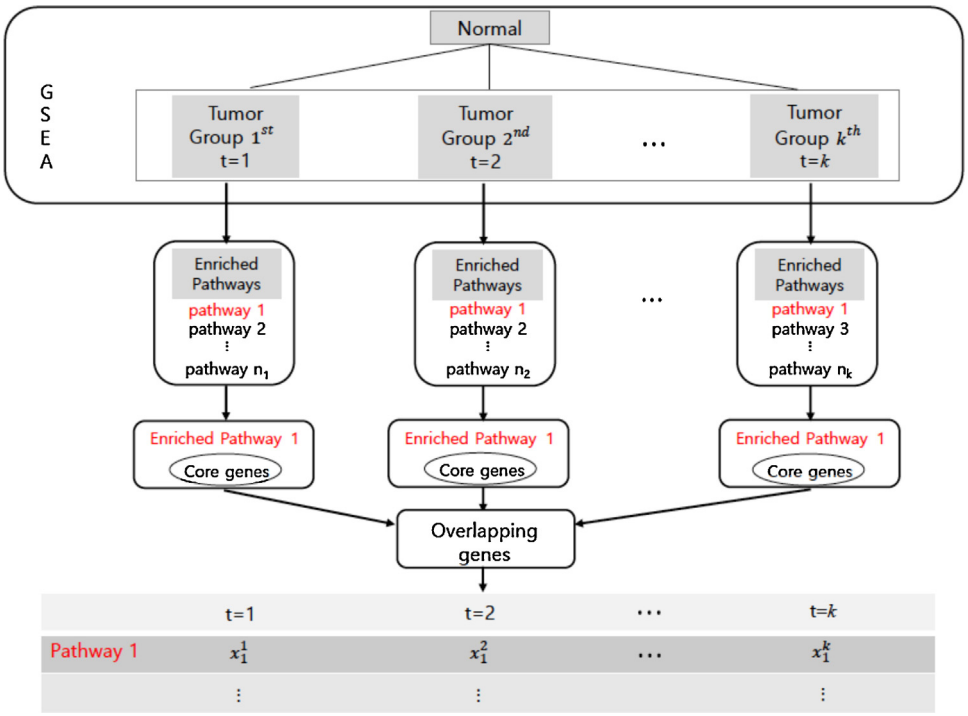


**Fig. 1.** Diagram for generating pathway module $x_i^k$, pathway 1 in red character appears in each implementation is called overlapping pathway.

**Table 1**
Dynamic pathway interactions of GBM with correlation coefficients greater than 0.7 with less than 0.05 *p*-value.

| Pair of pathways (OLG/GSP/mean NOM *p*-val/mean FWER *p*-val) | | Coef/*p*-val | NCG* |
|---|---|---|---|
| ASTHMA (6/30/0.0085/0.1811) | LEISHMANIA INFECTION (15/72/0.0007/0.0433) | 0.8776/1.74e−7 | 4 |

OLG: size of overlapping genes in the pathway; GSP: original gene size of the pathway; mean NOM *p*-val: mean of nominal *p*-value from GSEA performance; mean FWER *p*-val: mean of family wise *p*-value; Coef.: correlation coefficient; *p*-val: *p*-value; NCG*: number of common genes of a pair of pathways.

stages and grades. To obtain dynamic gene expression datasets, we implemented an alternative approach in which patients were grouped based on survival time and/or stages and grades.

Previous studies of dynamic interactions considered static gene–gene or pathway–pathway interactions based on patient information (Kim et al., 2007; Brophy and Voigt, 2014; Dahiya and Chaudhuri, 2013). Recently, Brophy and Voigt (2014) developed hypothetical tools for identifying cellular circuits involved in pathway regulation. Those studies inferred gene–gene regulatory networks within a pathway. In this study, by contrast, we inferred interactions among gene sets (especially pathways) to observe global changes in dynamic pathways. Furthermore, a suitable computational method for this purpose is urgently needed because pathway network analysis can predict the functions of gene sets associated with cancer/disease progression.

Despite the demand for such a technique, however, no tools are currently available for the analysis of either static or dynamic pathway interactions. Here, we present novel approaches for the determination of dynamic pathway interactions using static cancer datasets, which we believe will lead to a better understanding of cancer progression and the factors that determine survival time.

## 2. Materials and methods

### 2.1. Materials

Two sets of gene expression profiles were downloaded from The Cancer Genome Analysis (TCGA; available at http://cancergenome.nih.gov/) in 2016: one for ovarian (OV) cancer, and the other for glioblastoma multiforme (GBM). Both datasets were generated by the University of North Carolina Cancer Genome Characterization Center using an Agilent G4502A microarray platform containing 17814 genes and uploaded in Dec. 2014. The TCGA also provides various aspects of patient clinical information, including pathological classifications such as disease stage and grade, degree of

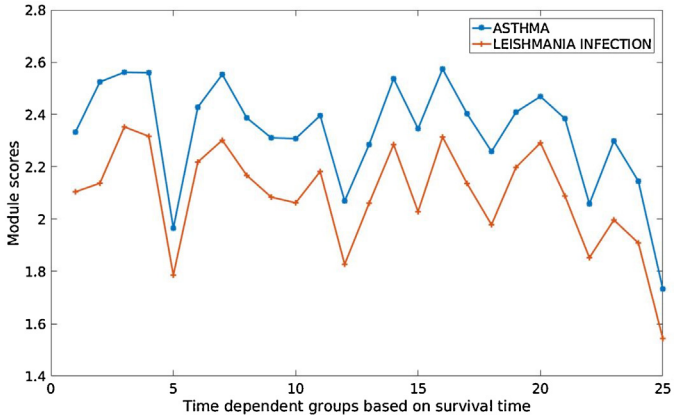


**Fig. 2.** Dynamic changes of pair of pathways. *X*-axis presents time dependent groups based on survival time in incremental order (from short to long) and *Y*-axis presents the module of pathways at given time.