Causal Inference of Gene Systems Network in Hepatocellular Carcinoma Progression by Graphical Chain Model

Masao Honda¹
mhonda@medf.m.kanazawa-u.ac.jp

Shu-ich Kaneko¹
skaneko@medf.m.kanazawa-u.ac.jp

Shigeru Saito²³
sh.saito@infocom.co.jp

Sachiyo Aburatani²
sachiyo@ims.u-tokyo.ac.jp

Katsuhis Horimoto²
khorimot@ims.u-tokyo.ac.jp

1 Department of Gastroenterology, Kanazawa University Graduate School of Medical Science, 13-1 Takara-machi, Kanazawa 920-8641, Japan
2 Laboratory of Biostatistics, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
3 Bioscience Department, INFOCOM CORPORATION, Mitsui Sumitomo Insurance Surugadai Annex Building, 3-11, Kanda-surugadai, Chiyoda-ku, Tokyo 101-0062, Japan

Keywords: causal graph, graphical chain model, hepatocellular carcinoma

1 Introduction

Hepatitis C virus (HCV) is the major etiologic agent of non-A non-B hepatitis and chronically infects about 170 million people worldwide. Many HCV carriers develop chronic hepatitis C (CH-C), finally complicated with hepatocellular carcinoma (HCC) in a liver with advanced stage CH-C.

Here, we analyzed the gene expression profiles of HCC and its background liver with advanced stage of CH-C [2-4], by a statistical method recently devised to infer the gene systems network in distinct biological stages from the gene expression profiles [1].

2 Materials and Methods

2.1 Graphical Chain Model (GCM)

The graphical chain model (GCM) infers a causal relationship between variables that can be naturally grouped (blocks) and ordered from prior knowledge [5]. In a GCM, any direct association between two variables in the same block is assumed to be non-causal, and any direct association between two variables from different blocks is assumed to be potentially causal. Thus, the GCM is one of the suitable models to infer the causal network between gene systems in distinct biological stages [1].

2.2 Procedure for Application of GCM to Expression Profiles

The procedure [1] is as follows: 1) The genes that express characteristically in distinct stages are selected from all genes.; 2) In each stage, the profiles of genes are subjected to a clustering analysis, and the gene groups (systems) are defined in terms of biological function.; 3) The gene systems are subjected to GCM, to infer the causal network between gene systems in different stages.; 4) The causal network between gene systems are evaluated by the biological knowledge.

2.3 Expression Profile Data

The expression profiles of 8516 genes were monitored in 27 samples in CH-C and 17 samples in HCC [2-4].
3 Results

Fig. 1 is a chain graph between CH-C and HCC. As seen in the figure, a causal graph for the gene systems in CH-C and HCC progression were inferred from the profiles.

![Figure 1: Orchestration of gene systems characterized by GOterm in hepatocellular carcinoma progression. Each circle with number corresponds to the clusters characterized by GOterm (the details will be described at the spot). Solid and bold lines and arrows indicate the association with 5% and 1% significance probabilities, respectively. The clusters connected with edges with 1% probability are enclosed by dashed lines.](image)

4 Discussion

The graphical chain model provides a snapshot for orchestrating the gene systems in hepatocellular carcinoma progression. The details of the model will be discussed at the spot.

References